THE RÔLE OF THE BANDWIDTH–DURATION PRODUCT $\mathcal{WT}$ IN THE DETECTABILITY OF DIOTIC SIGNALS

by

JUDI LAPSLEY MILLER

A thesis submitted
in fulfilment of the requirements for the degree of
Doctor of Philosophy
in Psychology

Victoria University of Wellington
New Zealand

August 13, 1999

I am in blood
Stept in so far that, should I wade no more,
Returning were as tedious as go o'er.

Shakespeare's Macbeth

Published by BFP Ink Publishers, New Zealand.

Dedicated
to the memory of
Carl Sagan

Abstract

The bandwidth–duration product, $\mathcal{WT}$, is a fundamental parameter in most theories of aural amplitude discrimination of Gaussian noise. These theories predict that detectability is dependent on $\mathcal{WT}$, but not on the individual values of bandwidth and duration. Due to the acoustical uncertainty principle, it is impossible to completely specify an acoustic waveform with both finite duration and finite bandwidth. An observer must decide how best to trade–off information in the time domain with information in the frequency domain. As Licklider (1963) states, "The nature of [the ear's] solution to the time–frequency problem is, in fact, one of the central problems in the psychology of hearing." This problem is still unresolved, primarily due to observer inconsistency in experiments, which degrades performance making it difficult to compare models.

The aim was to compare human observers' ability to trade bandwidth and duration, with simulated and theoretical observers. Human observers participated in a parametric study where the bandwidth and duration of 500 Hz noise waveforms was systematically varied for the same bandwidth–duration products ($\mathcal{WT}$=1, 2, and 4, where $\mathcal{W}$ varied over 2.5–160 Hz, and $\mathcal{T}$ varied over 400–6.25 ms, in octave steps). If observers can trade bandwidth and duration, detectability should be constant for the same $\mathcal{WT}$. The observers replicated the experiments six times so that group operating characteristic (GOC) analysis could be used to reduce the effects of their inconsistent decision making. Asymptotic errorless performance was estimated by extrapolating results from the GOC analysis, as a function of replications added.

Three simulated ideal observers: the energy, envelope, and full–linear (band–pass filter, full–wave rectifier, and true integrator) detectors were compared with each other, with mathematical theory and with human observers. Asymptotic detectability relative to the full–linear detector indicates that human observers best detect signals with a bandwidth of 40–80 Hz and a duration of 50–100 ms, and that other values are traded off in approximately concentric ellipses of equal detectability. Human detectability of Gaussian noise was best modelled by the full–linear detector using a non–optimal filter. Comparing psychometric functions for this detector with human data shows many striking similarities, indicating that human observers can sometimes perform as well as an ideal observer, once their inconsistency is minimised.

These results indicate that the human hearing system can trade bandwidth and duration of signals, but not optimally. This accounts for many of the disparate estimates of the critical band, rectifier, and temporal integrator, found in the literature, because (a) the critical band is adjustable, but has a minimum of 40–50 Hz, (b) the rectifier is linear, rather than square–law, and (c) the temporal integrator is either true or leaky with a very long time constant.

Keywords: ideal observers, energy detector, envelope detector, full–linear detector, bandwidth–duration product, bandwidth–duration reciprocity, acoustical uncertainty principle, narrow–band short–duration Gaussian noise, diotic, critical band, auditory filter, temporal integrator, auditory time constant, group operating characteristic (GOC) analysis, all–combinations analysis, function–of–replications–added (FORA), function of replications combined estimation (FORCE) analysis, asymptotic performance, non-parametric measures of detectability, information theory measures of detectability.

Acknowledgements

This research project was based on that of my supervisor, John Whitmore, and all his previous graduate students, from the Psychophysics Laboratory at Victoria University of Wellington: Kaye McAulay (GOC analysis), John Podd (GOC analysis), Rick Boven (GOC analysis), Alan Taylor (GOC analysis), Sue Galvin (Kaiser window, $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{2\mathrm{IFC}}$), Brian Scurfield (information theory measures of detectability, $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{2\mathrm{IFC}}$), and Vit Drga (GOC analysis, FORCE analysis, IFFT signal generation, $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{2\mathrm{IFC}}$, and envelope detection). The purpose of this research was to use these new methods and theories to study an old problem in psychophysics of how the human hearing system uses time and frequency information to detect acoustic waveforms.

I gratefully received financial support from a Helen Stewart Royle Scholarship, and from teaching and research employment opportunities in the School of Psychology. The use of the School of Mathematics and Computing Sciences computing resources was also appreciated. Thank you to the writers of the excellent freeware programs I used for graphics (GNUPLOT), text formatting (LaTeX 2_ε), and editing (Emacs)—they are all better than money can buy.

This is the fun bit—I finally get a chance to formally thank everyone who helped me get past the finishing post. Unfortunately, I've forgotten how to write in first person, present tense, with emotive words, so you will have to make do with the following burble!

I particularly want to thank the lab–rats: John Whitmore, for your friendship, guidance, and generosity with your time and resources; Vit Drga, for your constant support and encouragement; and Brian Scurfield, for believing I was capable.

Many thanks to all my friends, especially my fellow graduate students, for your empathy in the joys and sorrows of research: Vit Drga (again!), Emily Patterson–Kane (especially for your comments on the final draft), Jen Hay, Craig Gordon, Aaron Roydhouse, and David Tulloch.

Thanks too to my family, especially my immediate family: Eleanor Lapsley, Ian Lapsley, Edith Garland, and Laura Lapsley, for your love and support. Your letters and phone calls never failed to cheer me up and keep me going.

I gratefully thank Sue Galvin, for your cheery enthusiasm and good advice; Stuart Slater, for your helpful comments on an early draft; Jenny Miller and Derek Miller, for your support and also for helping us finance our home computer; Julie Liebrich and Wendy Parr, for writing nice things about me in scholarship applications; and Frank Walkey, for supporting the research at the Psychophysics Lab, and for supervising me after John's retirement.

I am eternally grateful to my constant companions: The Bunny, Fusty, and Pooh, who on more than one occassion stopped me going stark–raving mad!

Last, but not least, my love and thanks goes to Linton Miller—your input into this research project spanned every facet and made it all possible. Words cannot express my thanks and gratitude so…

:-XXXXXXXXXXXXXXXXXXXXX
{{{{{{{{{}}}}}}}}}
:-)))))))))))))))))))
IGYTTAASOMLAASOOR
/B/b

Contents

List of Figures

List of Tables

Notation

Table 1: Notation, acronyms, and abbreviations.

Symbol	Meaning
A	Amplitude of a tone
$\mathcal{A}$	Area under the ROC curve
$\mathcal{A}_{\mathrm{SIFC}}$	Area under the SIFC ROC curve
$\mathcal{A}_{\mathrm{2IFC}}$	Area under the 2IFC ROC curve
α^2	Proportion or percentage of energy constrained by duration $\mathcal{T}$
$\alpha(12\ldots n)$	Permutation index for the ordering of n events
ACA	All combinations analysis
AERB	Average equivalent rectangular bandwidth (Hz)
ATT	At The Time sampling strategy
$\lvert A(\omega)\rvert^2$	Energy spectrum (frequency is in radians)
β^2	Proportion or percentage of energy constrained by bandwidth $\mathcal{W}$
buffers	Mixed and scaled $\mathcal{SN}$ and $\mathcal{N}$ digital signals
c	Observer's criterion
$\mathcal{C}_{\mathrm{2IFC}}$	Channel capacity (bits) of an observer in a 2IFC task
CB	Critical band
c.d.f.	Cumulative distribution function
χ	Chi distribution
χ^2	Chi–square distribution
d'	Measure of sensitivity d–prime
d_z	Measure of sensitivity for normals with unequal variance
$\mathcal{D}$	Scurfield's discriminability measure (bits)
$\mathcal{D}_2$	Scurfield's discriminability measure for two events (bits)
$\mathcal{D}_6$	Scurfield's discriminability measure for six events (bits)
$\mathcal{D}_n$	Scurfield's discriminability measure for n events (bits)
$\mathcal{D}'_n$	Relative discriminability for n events
E	Energy
ERB	Equivalent rectangular bandwidth (Hz)
ERD	Equivalent rectangular duration (ms)
$\mathrm{Ess}\mathcal{T}_{\alpha^2\%}$	Essential duration constraining α^2% energy (ms)
$\mathrm{Ess}\mathcal{W}_{\beta^2\%}$	Essential bandwidth constraining β^2% energy (Hz)
$\mathrm{Ess}\mathcal{WT}_{\%}$	Essential $\mathcal{WT}$ based on $\mathrm{Ess}\mathcal{T}_{\%}$ and $\mathrm{Ess}\mathcal{W}_{\%}$
$\mathcal{F}$	$\mathcal{F}$ distribution
$\mathrm{FAR}(c)$	False–alarm rate at criterion c
FFT	Fast Fourier transform
FORA	Function of replications added
FORCE	Function of replications combined estimation
Γ	Gamma function
$Ga(x\lvert\lambda, n)$	Gamma distribution
GOC	Group operating characteristic

(continued...)

Table 1: Notation, acronyms, and abbreviations continued...

Symbol	Meaning
$\mathcal{H}_n$	Shannon entropy, average information, prior uncertainty of ordering of n events
$\mathrm{HR}(c)$	Hit rate at criterion c
IFFT	Inverse fast Fourier transform
i.i.d.	Independently, identically, distributed
κ	Parameter of Kaiser window
$\ell(x)$	likelihood ratio of x
$\mathcal{L}_a^2$	The signal space of finite energy signals over duration a
mROC	Mean receiver operating characteristic
$\mathcal{N}$	Noise–alone
"No"	Observer responds "No"
N_0	Noise power per unit cycle (in energy units)
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ, variance, σ^2
Ω	Bandwidth (in radians)
$P(\mathrm{C})_{\mathrm{SIFC}}$	Proportion correct in the SIFC task
$P(\mathrm{C})_{\mathrm{2IFC}}$	Proportion correct in the 2IFC task
p.d.f.	Probability distribution function
P	Probability
ROC	Receiver operating characteristic
ρ	Spearman's correlation coefficient: rho
$\mathcal{S}$	Signal–alone
SIFC	Single–interval forced–choice task
σ	Standard deviation
σ^2	Variance
$\sigma^2_{\mathcal{N}}$	Variance of $\mathcal{N}$ Gaussian waveform, also equal to V-rms and N_0
$\sigma^2_{\mathcal{S}}$	Variance of $\mathcal{S}$ Gaussian waveform, also equal to V-rms and S_0
SKE	Signal known exactly
SKS	Signal known statistically
$\mathcal{SN}$	Signal–plus–noise
SNR	Signal–to–noise ratio (decibels)
S_0	Signal power per unit cycle (in energy units)
SPL	Sound pressure level (decibels)
$\mathcal{T}$	Duration (seconds)
τ	Kendall's correlation coefficient: tau
TdB	Three–dB bandwidth (Hz)
2IFC	Two–interval forced–choice task
transients	Unmixed $\mathcal{S}$ and $\mathcal{N}$ digital signals, prior to conversion to buffers
TSD	Theory of signal detectability
$\mathcal{U}$	Mann–Whitney $\mathcal{U}$ statistic
$\mathcal{V}$	The hypervolume under an ROC hypersurface
V-pp	Peak–to–peak voltage
V-rms	RMS voltage
VLS	Very–large–sets of transient signals
$\mathcal{W}$	Bandwidth (Hz)
$\mathcal{WT}$	Bandwidth–duration product (dimensionless)
"Yes"	Observer responds "Yes"
$\mathcal{Z}$	Standard normal distribution, $\mathcal{N}(0,1)$

Chapter 1

Human hearing

> *"That's funny," said Pooh. "I dropped it on the other side," said Pooh, "and it came out on this side! I wonder if it would do it again?" And he went back for some more fir–cones.*
>
> *It did. It kept on doing it. Then he dropped two in at once, and leant over the bridge to see which of them would come out first; and one of them did; but as they were both the same size, he didn't know if it was the one which he wanted to win, or the other one. So the next time he dropped one big one and one little one, and the big one came out first, which was what he said it would do, and the little one came out last, which was what he had said it would do, so he had won twice... and when he went home for tea, he had won thirty–six and lost twenty–eight, which meant that he was—that he—well, you take twenty–eight from thirty–six, and* that's *what he was. Instead of the other way round.*
>
> *And that was the beginning of the game called Poohsticks ...*
>
> A. A. Milne.

. . . . . .

Psychophysics is concerned with describing how an organism uses its sensory systems to detect events in its environment. This description is functional, because the processes of the sensory systems are of interest, rather than their structure. One psychophysical theory, the Theory of Signal Detectability (TSD), uses a combination of statistical decision theory and the concept of the *ideal observer* to model an observer's sensitivity to events in its environment (Peterson, Birdsall, & Fox, 1954). TSD is *stimulus–oriented*, because properties of the stimuli are used to determine the theoretically best, or ideal, observer for a given detection task (Green & Swets, 1966; Jeffress, 1964). This observer may then be used to compare the performance of human, and other, observers. For instance, the ability of humans to detect simple acoustic waveforms can be modelled as a linear system consisting of a filter, rectifier, integrator, and sampler (e.g., Jeffress, 1964, 1967, 1968).

A number of psychophysicists have used the work of Peterson et al. (1954) to derive *energy* and *envelope* detectors for acoustic tones or noise, masked by noise (e.g., Green, 1960a; Green & Swets, 1966; Green & McGill, 1970; Jeffress, 1964, 1967, 1968; Marill, 1956; McGill, 1968a; Whitmore, 1969). These detectors are considered to be ideal observers for acoustic waveforms that are known only statistically, rather than exactly. This includes Gaussian noise, and deterministic signals

where the observer does not have access to all the properties of the waveform (e.g., sinusoidal phase is unknown). These energy and envelope detectors have been able to explain many of the human hearing system's abilities.

The energy and envelope detectors have two main parameters: signal–to–noise ratio and the bandwidth–duration product, $\mathcal{WT}$. Signal–to–noise ratio is the strength of a waveform, usually measured in decibels relative to some standard. $\mathcal{WT}$ is a measure of the spread of a waveform in time and frequency, where $\mathcal{W}$ is bandwidth in hertz, $\mathcal{T}$ is duration in seconds, and $\mathcal{WT}$ is dimensionless. Do these fundamental parameters play a rôle in determining the detectability of acoustic waveforms? One theory of tone–in–noise and noise–in–noise energy detection suggests that any combination of $\mathcal{W}$ and $\mathcal{T}$, for the same $\mathcal{WT}$, results in equal detectability, because the product of $\mathcal{W}$ and $\mathcal{T}$ is the parameter (Green & McGill, 1970). This is known as bandwidth–duration reciprocity. This theory also relates signal–to–noise ratio directly to detectability, dependent on $\mathcal{WT}$ (Green & McGill, 1970). It is not clear, however, how well this theory (and other similar theories) applies to human hearing.

The focus of the current study is to consider the detectability of narrow–band, short–duration, Gaussian noise (also called small–$\mathcal{WT}$ noise) by humans, when bandwidth and duration are systematically varied. Detectability of noise has not been studied in nearly as much detail as tones. Noise, however, resembles the types of sounds the human hearing system normally has to process (Miller, 1947; Moller, 1983, 1989); sounds that appear limited in time, bandwidth, and energy. Noise also has the advantage over tones, because its bandwidth and duration can be varied relatively independently of one another.

Acoustic waveforms can be represented as a *Fourier transform pair*. This relationship allows a waveform to be defined in the time domain and in the frequency domain. A direct result of the Fourier representation of waveforms is the acoustical uncertainty principle. This principle specifies that a waveform cannot have a finite bandwidth and a finite duration, and that the $\mathcal{WT}$ of a waveform has a lower bound. Because it is impossible to completely specify a band–limited acoustic waveform in finite time, and vice–versa, an observer (ideal, human, or electronic) must decide how best to trade–off information in the time domain with information in the frequency domain. As Licklider (1963, p. 993) states, "The nature of [the ear's] solution to the time–frequency problem is, in fact, one of the central problems in the psychology of hearing." This problem is still unresolved, primarily due to observer inconsistency in experiments, which degrades performance making it difficult to compare models.

1.1 The theory of signal detectability

In the 1950's, some psychophysicists realised that classical psychophysics was inherently flawed, and so modified their theories based on new ideas coming out of the engineering fields of information theory and signal analysis (e.g., Munson & Karlin, 1954; Smith & Wilson, 1953). The new engineering paradigm of TSD (Peterson et al., 1954), in particular, was suited to analogous detection problems in sensory psychology (Green, 1960b; Jeffress, 1964; Marill, 1956; Tanner & Swets, 1954; Tanner, 1960a, 1961).

The improvement over classical psychophysical methods was twofold. One, the decision making processes were separated from the sensory processes, exposing the classical *sensory* threshold to be a *response* threshold. Two, properties of the environment, other than the signal were included in the theories. In particular, Tanner and Swets (1954) postulated that noise was *always* present in

a human detection task. This view was backed up by Jeffress, who commented that 'the idea of an "absolute threshold" is meaningless—all thresholds are masked thresholds, all involve detecting the signal in a background of noise' (Jeffress, 1968, p. 187). In general, psychophysical theories derived from TSD have lead to a better understanding of human sensory processes.

1.1.1 The theory of ideal observers

An ideal observer may be derived using the *fundamental detection problem* paradigm. In modelling the task of an observer detecting a signal in noise, the fundamental detection problem is the simplest, nontrivial, detection task (Egan, 1975). In a trial, an observer is presented with a stimulus (evidence) in a finite observation interval, which could have come from either a noise–alone ($\mathcal{N}$) or a signal–plus–noise ($\mathcal{SN}$) event. Over a series of trials, evidence from both types of events is presented. This evidence randomly fluctuates from trial to trial, to the extent that there is confusability about which event the evidence came from. The observer must decide, at the end of each trial, which event occurred given the evidence. Because the evidence is confusable, the observer sometimes makes errors in judgement about which event occurred. The detectability of the signal is quantified by analysing how the errors are traded off as a function of the observer's criterion.

The theory of ideal observers relates the *detectability* of a signal to definite *physical* characteristics of the signal (Green & Swets, 1966). The evidence is modelled as a random variable, derived from transformations of the physical stimulus. No observer can make perfect detections of a signal masked by noise if there is overlap between the evidence distributions associated with the $\mathcal{N}$ and $\mathcal{SN}$ events. The aim of deriving an *ideal observer* is, therefore, to establish the best possible performance, given the events are confusable. It is not always known whether an ideal observer is the ideal for a given signal, because the ideal observer depends on a mathematical theory of the physical stimulus. Theories, by their very nature, never truly capture the properties of the real world, and, therefore, involve simplifications and assumptions.

The concept of the ideal observer (Lawson & Uhlenbeck, 1950; Peterson et al., 1954) originated from machine (especially radar) detection tasks. The ideal observer was a mathematical theory of a detection task, where the signal to be detected was noise degraded, and the observation of the signal was limited to a finite time. The aim of the theory was to determine to what extent noise limited the detection of the signal. Peterson et al. (1954) derived the theory of signal detectability and showed that the optimal detector invariably used the likelihood ratio decision axis, or a decision axis strictly monotonic with likelihood ratio, as a basis for decisions about the existence of the signal. If the signal is known to the observer exactly, and the observer can transform the evidence to a quantity monotonic with likelihood ratio, then the observer is considered to be an ideal signal–known–exactly (SKE) observer. If the observer does not have an exact representation of a signal (e.g., if the signal is random noise), or if the observer is unable to use information about some property of a deterministic signal (e.g., the phase of a tone), the observer is considered to be a signal–known–statistically (SKS) observer.

The early radar engineers considered the ideal observer as a mathematical theory that predicted the best possible performance for a particular class of signals, with particular restrictions on the information the observer had about the signals (Lawson & Uhlenbeck, 1950; Peterson et al., 1954). When the theory was extended to psychophysics, the emphasis changed, because unlike engineers, psychophysicists were not interested in *designing* detection systems; they were trying

to understand *existing* systems that did not necessarily perform ideally, and whose internal processes were usually inaccessible.

Marill (1956) and Green and Swets (1966) argued for the usefulness of the ideal observer in human psychophysical research, because (a) it puts an upper bound on performance measures, (b) differences between human and ideal observers give interesting insights, and (c) a difference between human and ideal observers maybe specifiable, and therefore lead to a powerful predictive model. As Green and Swets (1966) said, "The principle is parsimony—it is unnecessary to invent psychological mechanisms to explain a change that may be traced to the stimulus situation itself" (p. 152). The rôle of the ideal observer in psychophysics, however, has been controversial.

According to McGill and Teich (1991), the early psychophysical theories and experimental results based on TSD (Green, 1960a; Jeffress, 1964; Pfafflin & Mathews, 1962) showed that many auditory masking phenomena could be explained solely by the statistics of the stimulus. This suggested that detection tasks could be described by an ideal observer without leaving the stimulus domain. This was a *conservative* view of an ideal observer. Optimal detectability was described by the physical properties of the stimulus in the absence of further processing. The best example of a conservative ideal observer is the detection of a tone in narrow–band noise by a cross–correlation detector. Using the conservative view, the SKS observer is not an ideal observer, because it is outperformed by the SKE observer. McGill and Teich (1991) explained that these pure stimulus–oriented theories were not enough to describe human hearing. Some psychophysicists (e.g., Green & Swets, 1966), therefore, adopted a more *liberal* view of the ideal observer, where the aim was to derive the best observer given particular constraints on the signal, *and* on the observer. Therefore, there may be a number of possible ideal observers for a given task, depending on which signal properties the observer can use. The conservative psychophysicists would argue that the constraints on the observer mean the observer is not ideal, but for the purpose of psychophysics, the liberal view is the more useful, because the emphasis is on describing existing systems, and on developing meaningful and practical benchmarks for comparison.

Critics of the ideal observer concept, such as Martel and Mathews (1961), have come up with a variety of detection schemes to discount the ideal observer in general. These schemes result in perfect detection at any signal–to–noise ratio, but are straw–men, because they usually involve some infinite task, such as taking derivatives to infinity, sampling a signal over an infinite duration, assuming perfect measurement capabilities, or ignoring the effects of the uncertainty principle (Martel & Mathews, 1961). The original concept of the ideal observer, however, intended the observation to occur in a finite time.

The concept of the ideal observer in psychophysics is too important to simply abandon. The original engineering goal of designing systems, rather than describing systems, however, should not dictate how psychophysicists use the ideal observer. In this research, the concept of the ideal observer is taken to mean the liberal view. This is because the aim in psychophysics is to understand real observers. By manipulating an ideal observer, to make it less and less ideal, it may be possible to better understand human hearing. The ideal observer then simply becomes a functional model of a system.

1.1.2 Psychophysical tasks and ROC analysis

In modern psychophysics, the experimental methods most widely used are the single–interval forced–choice (SIFC) task and the two–interval forced–choice (2IFC) task.[1] Both have been used as a basis for theories about the detectability of many types of acoustic stimuli, and there are a number of theoretical relationships linking the two tasks. For instance, Lapsley Miller, Scurfield, Drga, Galvin, and Whitmore (1999) and Scurfield (1995, 1996) have examined and extended the non-parametric relationships between the SIFC and 2IFC tasks, both theoretically and experimentally. One of their conclusions was that the SIFC task is often preferable to the 2IFC task. It is still important to consider both tasks, as a number of theoretical findings have only been derived for one or the other task. The theoretical links between the SIFC and 2IFC tasks may then be used to transform predictions for one task to the other.

The SIFC and 2IFC tasks

The fundamental detection problem can be applied, in theoretical and experimental work, as an SIFC task (Egan, 1975). The SIFC task consists of a set of independent trials, each of which contains an observation interval, a decision interval, and a pay–off interval. During the observation interval either the $\mathcal{SN}$ event occurs or the $\mathcal{N}$ event occurs. The observer never has direct access to these events, only evidence associated with the events. During the decision interval, the observer's task is to decide whether the $\mathcal{SN}$ event occurred, based on this evidence. The evidence is modelled as a univariate random variable X, conditional on each event: $X_{\mathcal{SN}}$ and $X_{\mathcal{N}}$. X is known as the decision axis. To make a decision, the observer partitions X by using a criterion; if the evidence is greater than or equal to the criterion then the observer responds "Yes" it was a signal, otherwise if the evidence is less than the criterion the observer responds "No" it was not a signal. The observer may vary the criterion depending on the prior probabilities of the events and the pay–offs. During the pay–off interval the observer may receive knowledge of results, monetary rewards, or other pay–offs, based on the four possible outcomes (hit, miss, false–alarm, and correct–rejection).

Each trial in a 2IFC task contains *two* observation intervals, a decision interval, and a pay–off interval. In one observation interval the $\mathcal{SN}$ event occurs, and in the other the $\mathcal{N}$ event occurs. During the decision interval, the observer's task is to decide in which *interval* the $\mathcal{SN}$ event occurred, or equivalently, the task may be to decide which *order* the two events occurred. Although these two formulations of the task are conceptually different, the task can be modelled the same way. The evidence the observer uses is the result of some *comparison* of the evidence from the first interval and the evidence from the second interval. This comparison may be modelled as the arithmetic difference between the evidence from each interval. This difference is a univariate random variable conditional on each 2IFC event–ordering, and forms the 2IFC decision axis. To make a decision, the observer partitions the decision axis by using a criterion; if the evidence is greater than the criterion then the observer responds "One", otherwise if the evidence is less than

[1] To reduce ambiguity, these forced–choice tasks should be named by the number of *intervals* presented in any one trial. Others refer to them by the number and type of decision alternatives, which has led to confusion. The SIFC task has been referred to as the "Yes"–"No" task (Green & Swets, 1966), the rating task, (Green & Swets, 1966), the single–interval task (Egan, 1975), the fixed interval observation experiment (Egan, Schulman, & Greenberg, 1959), and the two–alternative forced–choice ($2\mathrm{AFC}$) task—because the observer has two possible decision alternatives (Creelman, 1965; McFadden, 1970). The 2IFC task (Egan, 1975; Markowitz & Swets, 1967) is also known as the two–category forced–choice task (Marill, 1956), the forced–choice task (Green & Swets, 1966; Swets & Green, 1961), and the two–alternative forced–choice ($2\mathrm{AFC}$) task (Green & Swets, 1966; MacMillan & Creelman, 1991; Swets, 1959).

the criterion the observer responds "Two". The 2IFC task can, therefore, be analysed in the same way as an SIFC task, because the *event–orderings* of the 2IFC task can, themselves, be regarded as events. Thus, both the SIFC and 2IFC task may be regarded as two–event tasks, even though the way in which *stimuli* are presented is different for each task. The observer may vary the criterion, depending on the prior probabilities of the event–orderings, and the pay–offs. During the pay–off interval, like the SIFC task, the observer may receive knowledge of results or pay–offs based on the four possible outcomes (hit, miss, false–alarm, and correct–rejection) where the event–ordering whose associated evidence distribution has the larger mean is designated as the equivalent of the $\mathcal{SN}$ event.

If the same measure of detectability is used for both the SIFC and 2IFC tasks, the resulting performance in the 2IFC task is greater than for the SIFC task. This is because there is more information in the 2IFC task on each trial. For instance, if the underlying distributions are normal, $d'_{\mathrm{2IFC}} = \sqrt{2}d'_{\mathrm{SIFC}}$ (Egan, 1975).

The ROC curve

One of the more important concepts to come out of TSD is the receiver operating characteristic (ROC) curve, which summarises the observer's performance as a function of the observer's criterion. The ROC curve is a graph of the observer's hit rate against false–alarm rate, for all possible criteria (Egan, 1975; Green & Swets, 1966; Peterson et al., 1954). It shows the observer's ability to discriminate between the two events. Both SIFC and 2IFC ROC curves can be generated by either repeating a "Yes"–"No"(or "One"–"Two") experiment, each time with the observer using a different criterion, or using an N–point or continuous rating–scale where the observer employs multiple criteria within an experiment (Rockette, Gur, & Metz, 1992; Watson, Rilling, & Bourbon, 1964; Watson, Kellogg, Kawanishi, & Lucas, 1973; Wilcox, 1967). Some make a distinction between the "Yes"–"No" and rating tasks (e.g., Bamber, 1975; Green, 1964b), but this is unnecessary if the "Yes"–"No" task is considered as a *two–point* rating task. Green and Swets (1966) show that the theoretical SIFC ROC curve is the same for either method, and there is evidence that this is also true for empirical SIFC ROC curves (Emmerich, 1968; Hanley & McNeil, 1982; Nachmias, 1968; Watson et al., 1964, 1973). Intuitively, there should also be no difference between 2IFC ROC curves obtained by the "One"–"Two" or rating methods. This has not been pursued, however, for the preferred measure of detectability in the 2IFC task is $P(\mathrm{C})_{\mathrm{2IFC}}$, which does not require obtaining the ROC curve. There are only a few examples of theoretical or empirical 2IFC ROC curves in the literature (Drga, 1988; Egan & Clarke, 1966; Friedman & Carterette, 1964; Falmagne, 1985; Leshowitz, 1969; Perniske, 1987; Schulman & Mitchell, 1966). Luce (1997) explained that "... no one seems ever to collect 2–alternative, forced–choice ROC curves, there being a myth to the effect that this procedure, unlike the yes–no one, is unbiased."

Measures of detectability

A measure of detectability is a number that summarises an observer's performance in a psychophysical task. If the task involves a sensory system then a measure of detectability is also a measure of *sensitivity*. A measure of detectability should ideally be free from confounding by processes not associated with stimulus properties (non-sensory processes), such as the prior probabilities of the events, the pay–offs, and the decision criterion. A measure of detectability is also more useful if it is non-parametric, for comparisons may then be made among different observers

and different detection tasks, regardless of the nature of the underlying evidence distributions, or decision axes.

Scurfield (1995, 1996) suggested that a measure of detectability should also be a *metric*, that is, the measure should have a true zero, be nonnegative, be invariant to labelling of the events and decisions, and obey the triangle inequality. A metric is a distance measure (Borowski & Borwein, 1989).[2] Therefore, if a measure of detectability is also a metric, then it is a measure of the distance between the evidence–points from one event–set to evidence–points in another event–set in evidence–space. The larger the distance among the evidence points, the greater their discriminability, and vice–versa. It is important to understand that this is not the Euclidean distance, nor is it a simply a measure of the difference between the means of the evidence distributions. A distance summarises the relationship among all possible evidence values.

With particular assumptions, the SIFC and 2IFC tasks are related theoretically through their ROC curves and measures of detectability by:

1. *Parametric* relationships, such as $d'_{\mathrm{2IFC}} = \sqrt{2}d'_{\mathrm{SIFC}}$, which have limited generality, because they usually assume normal evidence distributions (Egan, 1975; Elliot, 1964; Marill, 1956; Tanner & Swets, 1954; Tanner & Birdsall, 1958; Schulman & Mitchell, 1963, 1966; Simpson & Fitter, 1973).

2. The non-parametric relationship that equates the area, $\mathcal{A}_{\mathrm{SIFC}}$, under the receiver operating characteristic (ROC) curve in the SIFC task to the proportion of correct decisions, $P(\mathrm{C})_{\mathrm{2IFC}}$, in the 2IFC task (Green, 1964b; Green & Moses, 1966; Green & Swets, 1966). Lapsley Miller et al. (1999) extended this relationship, by removing unnecessary assumptions about how the evidence is distributed, by generalising to discrete random variables, and considering evidence distributed over the x decision axis and not just the likelihood ratio decision axis $\ell(x)$.

3. Scurfield's (Lapsley Miller et al., 1999; Scurfield, 1995, 1996) new measures of detectability: $\mathcal{D}_{\mathrm{SIFC}}$ and $\mathcal{C}_{\mathrm{2IFC}}$.

4. A method of deriving the 2IFC ROC curve from the slope of the SIFC ROC curve (Green & Swets, 1966) and a method of deriving the 2IFC ROC curve from the SIFC hit rates and false–alarm rates (Lapsley Miller et al., 1999).

These relationships are important, because they predict psychophysical performance independently of any one psychophysical task (Green, 1964b; Green & Swets, 1966; Swets, 1959). This was previously impossible with classical psychophysical methods (Swets & Green, 1961).

To date, the most common measures of detectability in use are: d', which is parametric; $P(\mathrm{C})_{\mathrm{SIFC}}$ and $P(\mathrm{C})_{\mathrm{2IFC}}$, which are *both* dependent on the prior probabilities of the events and decision criterion; and $\mathcal{A}_{\mathrm{SIFC}}$ and $\mathcal{A}_{\mathrm{2IFC}}$, which although non-parametric, and independent of priors, pay–offs and criterion, are not metrics. Only Scurfield's (1995, 1996, 1998) new information–theoretic measures (see Chapter 6), $\mathcal{D}_{\mathrm{SIFC}}$ and $\mathcal{C}_{\mathrm{2IFC}}$, are metrics (or pseudo–metrics if the decision axis is not strictly monotonic with the likelihood ratio decision axis), non-parametric, and independent of prior probabilities, pay–offs, and decision criterion.

Scurfield's new measure of detectability is based on information theory and is applicable to multiple events (Lapsley Miller et al., 1999; Scurfield, 1995, 1996, 1998). It is a transform of $\mathcal{A}_{\mathrm{SIFC}}$

[2] More specifically, a distance between sets in a metric space is the infimum (or greatest lower bound) of the distance between points in one set and points in the other (Borowski & Borwein, 1989).

for the two–event SIFC case:

$$\mathcal{D}_{\mathrm{SIFC}} = \log 2 - \mathcal{H}_{\mathrm{SIFC}}, \tag{1.1}$$

where

$$\mathcal{H}_{\mathrm{SIFC}} = -\mathcal{A}_{\mathrm{SIFC}} \log(\mathcal{A}_{\mathrm{SIFC}}) - (1 - \mathcal{A}_{\mathrm{SIFC}}) \log(1 - \mathcal{A}_{\mathrm{SIFC}}). \tag{1.2}$$

$\mathcal{H}_{\mathrm{SIFC}}$ represents the *Shannon entropy* (Shannon, 1949b) of the areas above and below the ROC curve. $\mathcal{A}_{\mathrm{SIFC}}$ itself cannot be interpreted as a distance measure, but with this transformation to $\mathcal{D}_{\mathrm{SIFC}}$ it becomes a distance measure. Scurfield (1996) showed that $\mathcal{D}_{\mathrm{SIFC}}$ is (a) a pseudo–metric in general, (b) a true metric for likelihood ratio decision axes (or for axes strictly monotonic with likelihood ratio), and (c) a measure of the distance between the events $\mathcal{SN}$ and $\mathcal{N}$.

The maximum value of $\mathcal{D}_{\mathrm{SIFC}}$ is $\log(2)$. This maximum occurs when $\mathcal{A}_{\mathrm{SIFC}} = 1$, or $\mathcal{A}_{\mathrm{SIFC}} = 0$. As $\mathcal{A}_{\mathrm{SIFC}}$ and $1 - \mathcal{A}_{\mathrm{SIFC}}$ become closer in value, $\mathcal{D}_{\mathrm{SIFC}}$, decreases with zero being the minimum value (occurring when $\mathcal{A}_{\mathrm{SIFC}} = 1 - \mathcal{A}_{\mathrm{SIFC}} = 0.5$). The log base used is arbitrary, although $\log_2$ has some advantages, because the unit of $\mathcal{D}_{\mathrm{SIFC}}$ is bits, and the range is [0,1]. Similar to the relationship $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{\mathrm{2IFC}}$, $\mathcal{D}_{\mathrm{SIFC}}$ is equal to the channel capacity, $\mathcal{C}_{\mathrm{2IFC}}$, in the 2IFC task (Lapsley Miller et al., 1999).

Scurfield extended $\mathcal{D}_{\mathrm{SIFC}}$ to multi–event multi–interval tasks (see Chapter 6; Scurfield, 1995, 1996). This can be applied in simple SIFC experiments, where there are multiple $\mathcal{SN}$ events (e.g., different signal–to–noise ratios) and one $\mathcal{N}$ event. Instead of calculating $\mathcal{A}_{\mathrm{SIFC}}$ or $\mathcal{D}_{\mathrm{SIFC}}$ for each level relative to the $\mathcal{N}$ event, the global measure of detectability $\mathcal{D}_n$ summarises the observer's ability to discriminate among all n events. This notation is used from now on: $\mathcal{D}_2$ for two events (instead of $\mathcal{D}_{\mathrm{SIFC}}$) and $\mathcal{D}_6$ for six events.

1.1.3 Waveform representation

TSD takes a *stimulus–oriented* approach by using physical properties of the stimuli to determine the nature of an ideal observer. Therefore, to model an ideal observer, an appropriate mathematical representation of acoustic waveforms is necessary.

The usual way of representing an acoustic waveform is to describe the fluctuation of some physical property of the waveform (e.g., power, voltage) as a function of time in the *time domain*. By using the *Fourier transform* or *integral*, the same waveform can also be specified as a function of frequency in the *frequency domain*. Psychophysicists are most interested in the set of waveforms to which the human hearing system is sensitive. These waveforms have three defining properties: they appear to have a finite bandwidth, a finite duration, and a finite energy content (e.g., a spoken word). Properties of this subset of waveforms can be derived from the Fourier integral (Papoulis, 1962).[3] This section, and the more detailed explanation in Appendix A, describe (a) the difficulties of defining the properties of bandwidth, duration, and their product for finite energy waveforms, and (b) how finite energy waveforms may be represented by a set of $2\mathcal{WT}$ numbers. These descriptions indicate how and why $\mathcal{WT}$ is a fundamental parameter for acoustic waveforms. Finally, the nature of the energy and envelope functions, and their distributions, is introduced.

[3] For a mathematically rigorous and relatively complete coverage of waveform representation, refer to Franks (1969). For a more accessible review, see the relevant sections in Papoulis (1962) and Vakman (1968).

Acoustical uncertainty principles

In the context of Fourier representation of acoustic waveforms, it is impossible for a waveform to be limited in both time and frequency. If a waveform is limited in time, then it must have an infinitely wide bandwidth, and conversely, if a waveform has a finite bandwidth, then it must have an infinite duration. This result is known as the *acoustical uncertainty principle*. There are many forms of the acoustical uncertainty principle, because $\mathcal{W}$ and $\mathcal{T}$ can be defined in many ways. All acoustical uncertainty principles, however, specify a lower bound to the bandwidth–duration product,[4] $\mathcal{WT}$, indicating that a waveform cannot have finite spread in both domains.

The acoustical uncertainty principle contradicts what seems to be a defining property of the waveforms to which the human hearing system is sensitive: that waveforms are both band–limited and time–limited. How can this contradiction be resolved? If a small amount of energy is allowed to exist outside the bounds of bandwidth or duration, then it is meaningful to refer to band–limited and time–limited waveforms, because these waveforms have *essentially* all of their energy between two well-defined bounds. The energy outside these bounds is so small as to be virtually nonexistent. The problem is that the energy is there, and this makes defining the bandwidth and duration arbitrary. As a result there are many definitions of $\mathcal{W}$ or $\mathcal{T}$, depending on the nature of the problem or the context, and there are also many different forms of the uncertainty principle (see Bourret, 1958; Brillouin, 1962; de Bruijn, 1967; Cohen, 1989; Franks, 1969; Gabor, 1946; Hilberg & Rothe, 1971; Kay & Silverman, 1957, 1959; Lampard, 1956; Landau & Pollak, 1961, 1962; Leipnik, 1959; Slepian & Pollak, 1961; Weyl, 1931; Zakai, 1960).

The acoustical uncertainty principle was traditionally based on the analogous uncertainty principle of quantum physics. This has lead to problems of interpretation when applied to acoustics. Landau, Pollak, and Slepian have provided the best alternative approach to the classical acoustical uncertainty principle. In a series of papers entitled "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty" they have proposed new definitions of the concepts of bandwidth and duration, and derived the corresponding uncertainty principle (Landau & Pollak, 1961, 1962; Slepian & Pollak, 1961). An excellent overview of this research is in Slepian (1983).

Landau and Pollak (1961) suggested that a good definition of $\mathcal{W}$ and $\mathcal{T}$ would describe the behaviour of the waveform $f(t)$ in a given finite time interval and likewise the behaviour of its spectrum $F(\omega)$ in a given finite frequency band. They argued that although waveforms cannot be bounded in both domains, bounds that essentially constrain the waveforms can still be specified. One way to specify these bounds is to look at the *energy* content and *energy* spread of a waveform in both the time and frequency domains. Bandwidth and duration can be specified by calculating the proportion of energy constrained between two bounds, relative to the total energy of the waveform. The proportion of energy β^2, constrained by the *essential bandwidth*, Ω, is defined as

$$\beta^2 \quad = \quad \frac{\int_{-\Omega}^{\Omega} |A(\omega)|^2 \, d\omega}{\int_{-\infty}^{\infty} |A(\omega)|^2 \, d\omega} \tag{1.3}$$

where $0 \leqslant \beta^2 \leqslant 1$, and $|A(\omega)|^2$ is the energy spectrum (see Appendix A). Similarly, the proportion

[4]The bandwidth–duration product is abbreviated to $\mathcal{WT}$ if the measurement of bandwidth ($\mathcal{W}$) is in Hertz (Hz), or $\Omega\mathcal{T}$ if bandwidth (Ω) is in radians per second. The conversion is $\Omega = 2\pi\mathcal{W}$. Some authors use B instead of $\mathcal{W}$ or refer to the duration–bandwidth product, $\mathcal{TW}$. $\mathcal{T}$ is the duration in seconds.

of energy, α^2, constrained by the essential duration, is defined as

$$\alpha^2 = \frac{\int_{-\mathcal{T}/2}^{\mathcal{T}/2} |f(t)|^2 \, dt}{\int_{-\infty}^{\infty} |f(t)|^2 \, dt} \tag{1.4}$$

where $0 \leqslant \alpha^2 \leqslant 1$, and $f(t)$ is an acoustic waveform. These definitions have been used in psychophysical experiments by Ronken (1970a, 1970b). An algorithm for calculating the essential bandwidths and durations of digital or digitised signals is given in Appendix B.

The dimensionality of waveforms

Shannon's sampling theorem shows how a waveform of bandwidth $\mathcal{W}$ in an interval $\mathcal{T}$ can be approximately represented by the linear combination of $2\mathcal{W}\mathcal{T}$ weighted sines and cosines, which are orthogonal basis functions (Goldman, 1953; Shannon, 1949a). No waveform can be exactly represented by a finite number of orthogonal basis functions—in fact an infinite number of functions is required. Instead, it can be shown that a finite energy waveform can be *approximately* represented by $2\mathcal{W}\mathcal{T}$ orthogonal basis functions (Landau & Pollak, 1962), thus giving a finite number of degrees of freedom. Slepian (1983) describes these types of signals as being asymptotically $2\mathcal{W}\mathcal{T}$ dimensional. The basis functions are traditionally sinusoids, but generalised Fourier analysis shows that any other orthogonal function may also be used (Davenport & Root, 1958). Some of these functions have better properties than the sinusoids (Landau & Pollak, 1962).

Waveform representation using generalised Fourier analysis simplifies many derivations, as well as giving useful geometrical analogies of waveforms as vectors in m–dimensional space (Davenport & Root, 1958; Franks, 1969; Harmuth, 1968; Helstrom, 1960; Landau & Pollak, 1962; Vakman, 1968). An important application is in the representation of acoustic signals. In particular, Gaussian noise may be represented as $2\mathcal{W}\mathcal{T}$ samples, where each sample is actually a Gaussian random variable (see Appendix A).

Energy and envelope representation

A detector is a set of mathematical functions, a computational algorithm, or an electronic system, that transforms a waveform, and outputs a single value or waveform, that summarises some aspect of the input waveform. Therefore, the formulation of a detector inherently requires a description of the waveforms, both before and after the transformation.

Engineers, and some psychophysicists, have traditionally reserved the term 'detector' to describe the output of a rectifier. In psychophysics, however, it is more useful to use the term *detector* to refer to the entire system of filters, rectifiers, integrators, and samplers. This is how the term 'detector' is used here. Two common detectors are the energy and the envelope detector. The emphasis in this section is not on deriving these detectors, but on introducing the representation of the energy function and the envelope function, and their distributions.

An energy detector outputs the energy of the waveform, $f(t)$,

$$\mathrm{Ene}(t) \quad = \quad \int f^2(t) \, dt. \tag{1.5}$$

It can be implemented, computationally and electronically, with a square–law rectifier, and a true integrator.

An envelope detector outputs a function that is tangential to the peaks of a waveform. There are many forms of envelope functions, but, according to Rice (1982), differences are generally so small that they are only of theoretical interest. One definition is what Rice called the *natural* envelope

$$\mathrm{Env}(t) \quad = \quad \sqrt{a^2(t) + b^2(t)} \tag{1.6}$$

where $a(t)$ and $b(t)$ are the in–phase and quadrature components, respectively, of $f(t)$ (see Goldman, 1953). One way to implement an envelope detector, computationally or electronically, is to rectify the waveform with a linear rectifier (either half–wave or full–wave), then integrate it with a leaky integrator. The shape of the envelope function is varied by manipulating the time constant of the integrator. A short time constant means the envelope hugs the rectified waveform closely, a long time constant means the envelope reacts slowly to changes in the waveform (see Figure 4.3 in Chapter 4 for examples).

de Lano (1949) showed that the signal–to–noise ratio at the output of a linear rectifier is greater than that of a square–law rectifier. This finding is in contrast to many other proofs that claim to show that the energy detector is optimal. de Lano (1949) also showed that at small (less than unity) and very large (unspecified) signal–to–noise ratios the two detectors are equivalent. According to de Lano, the difference is due to the energy detector having larger signal–to–noise ratio intermodulation distortion[5] than the envelope detector; something that is usually ignored in the development of the energy detector.

Marcum (1960) suggested that people preferred working with the square–law rectifier, because the linear rectifiers were more difficult to work with, mathematically. In fact, many people used the square–law rectifier to *model* systems with a linear rectifier. According to Marcum, "The difference in results for the linear and square law [rectifiers] turns out to be so small that extreme accuracy must be used in the calculations to show the relation in its true form." (Marcum, 1960, p.189). In Marcum's evaluation of the linear and square–law rectifiers, testing appeared to be done only for particular criteria (e.g., a false alarm rate of 10^{-6}, and a hit rate of 0.5). Marcum did not give a criterion–free measure of the differences between these detectors.

Marcum (1960) showed that the two detectors were identical when $\mathcal{WT}$=1 and 70, that between these values the linear detector was better (by 0.11 dB at $\mathcal{WT}$=10), and that for $\mathcal{WT} > 70$, the square–law detector was better, and asymptotically exceeded the linear detector by 0.19 dB, as $\mathcal{WT}$ tended to infinity. The test, however, was also at only one criterion, as a function of signal–to–noise ratio. In Chapter 4, Marcum's predictions are put to the test by evaluating the performance of detectors using linear and square–law rectifiers, for Gaussian noise waveforms with $\mathcal{WT}$=1, 2 and 4, where the difference is measured using the psychometric functions of $\mathcal{A}$ and $\mathcal{D}_2$.

Although the energy function has been defined as a continuous function of time, it may also be approximately represented as a set of samples from a finite sampling space, using generalised Fourier analysis (see Appendix A). Otherwise, according to Grenander, Pollak, and Slepian (1959), to derive the exact distribution of energy in a finite time sample of band–limited white Gaussian noise, a particularly difficult integral equation needs to be solved. This has been achieved by using the methods of Slepian (1954), however, the math is inaccessible. For psychophysical purposes, it is generally considered to be adequate to work with approximations. For small $\mathcal{WT}$ waveforms, however, these approximations are not necessarily good enough.

[5] "The unwanted modulation of one frequency by another caused by a non-linear stage" (Sinclair, 1988, p.157).

Grenander et al. (1959) showed that an approximation of the energy in a finite sample of band–limited white Gaussian noise, for large $\mathcal{T}$, results in $2\mathcal{WT}$ coefficients that are statistically independent. The sum of the coefficients is distributed as a χ^2 probability distribution function, with $2\mathcal{WT}$ degrees of freedom. They state that the approximation is "quite good" even for small $\mathcal{WT}$. Their plots of the χ^2 approximation compared with the exact solution, show the discrepancy for small $\mathcal{WT}$. It is not really known how this difference affects the detection statistics, because a model of an energy detector using the exact solution has not been formally derived.

The distribution of the envelope function for narrow–band noise is a Rayleigh distribution (Jeffress, 1968, 1970), and is a special case of the χ distribution (for two degrees of freedom). The Rayleigh distribution is derived from a continuous representation of narrow–band noise. Jeffress (1968) showed that for transient noise, the optimal envelope detector integrates over a duration equal to the reciprocal of the bandwidth (i.e., $\mathcal{WT}$=1), indicating that the Rayleigh distribution is also appropriate for transient waveforms. McGill (1967) showed how the energy statistics for $\mathcal{WT}$=1 waveforms may also be derived by transforming from Rayleigh envelope statistics to energy statistics, for bursts of narrow–band noise. Both Jeffress (1968) and Green and McGill (1970) used this transformation to extend their tone–in–noise envelope and energy models, respectively, to larger $\mathcal{WT}$. Through a change of variable, the χ envelope statistics and the χ^2 energy statistics are related, and result in the same ROC curves for tone–in–noise detection. It is unclear, however, how well these statistics represent the underlying distributions, given that Marcum's (1960) results suggest there should be a difference between the two detectors for $\mathcal{WT} > 1$.

Envelope detectors, as psychophysical theories, have only been developed for tone–in–noise and for noise–in–noise with $\mathcal{WT}$=1. Chapter 4 uses computer simulations of noise–in–noise detection, for energy and envelope detectors, for signals of $\mathcal{WT}$=1, 2, and 4.

1.1.4 Summary

The results from this section indicate how ideal observers could be derived by using generalised Fourier analysis, where the waveform can be viewed as essentially time–limited and band–limited, with finite degrees of freedom, and finite energy. In representing Gaussian noise, generalised Fourier analysis, rather than Shannon's sampling theorem, is more robust and results in better approximations of the waveform. For noise–in–noise detection, ideal observers have been derived that are energy detectors and envelope detectors (for $\mathcal{WT}$=1 only). Before evaluating these detectors in more detail, it is useful to first describe the functional properties of the human hearing system, then show why the energy and envelope detectors are useful models of human hearing.

1.2 Overview of human hearing models

Mathematical representations of acoustic waveforms form the basis of theories about human hearing. For instance, modern theories of hearing using TSD are usually based on a finite Fourier waveform representation. The use of Fourier analysis to describe hearing can be traced back to Ohm and Helmholtz (Boring, 1957; Whitfield, 1967). Ohm's Acoustic Law (1843; in Boring, 1957) says that for a complex waveform, the ear hears the Fourier sine and cosine components. Helmholtz proposed that the ear has a series of tuned resonators that split a compound waveform into elementary sinusoids, like Fourier analysis (Green & Swets, 1966).

By the twentieth century, the model of the ear as a Fourier analyser was well established and accepted. As Fourier analysis itself was developed (particularly due to its use in quantum physics), more interest was paid to the details of the models of human hearing based on this analysis. In particular, the resolution of human hearing was limited by the acoustical uncertainty principle.

Stewart (1931) was the first to suggest that the limits of detectability of an acoustic signal were defined by an acoustical uncertainty principle, analogous to the Heisenberg uncertainty principle of quantum physics. He proposed the relationship $\Delta v.\Delta t = 1$, but did not define either the bandwidth Δv or the duration Δt. Kock (1935) followed on from Stewart (1931) and looked at the effect of Stewart's uncertainty principle in music. There was also interest in time–frequency analysis, most notably by Gabor (1946). Gabor gave a credible basis for studying hearing in terms of *instantaneous* time–frequency analysis, where the acoustical uncertainty principle defined the resolution. Corliss (1963, 1967) applied Gabor's findings in a simple mechanistic model of hearing, based on a threshold system, where she specified an explicit relationship between the uncertainty principle and the energy resolution of the ear.

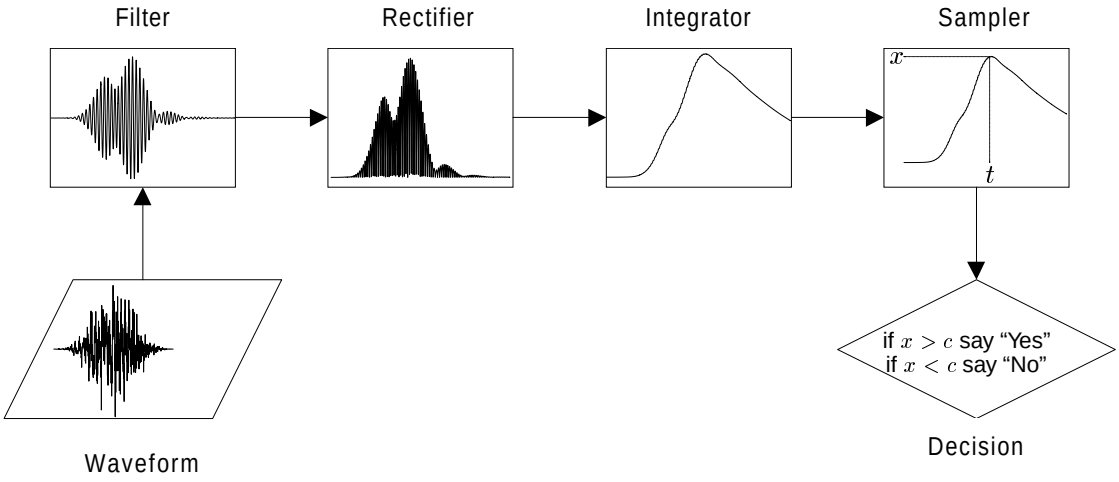

Figure 1.1: The general structure for a simple model of hearing, with illustrations for an envelope detector (in the time domain). The example shows an $\mathcal{SN}$ trial where a Gaussian noise input waveform of $\mathcal{WT}$=2 {25ms, 80Hz: 16dB} is presented to the detector. The filter is matched to the bandwidth of the signal, the rectifier is full–wave, the integrator is leaky with a time constant of 50 ms, and the sampler takes sample x at time t. A decision is made as to whether a signal was present by comparing the evidence x to a criterion c.

Further application and development of Gabor's (1946) work was overshadowed by the popularity of TSD in the 1950's and 1960's. Other research, however, focused on particular processes of the human hearing system rather than considering its overall abilities. These processes were frequency selectivity (filtering), rectification, temporal integration, and sampling. The basic structure of such a system is shown in Figure 1.1, with an example of an envelope detector processing an $\mathcal{SN}$ waveform (based on the models of Jeffress, 1964, 1967, 1968). This section considers the nature of these individual processes, to give a context for the ideal observer models, discussed in the subsequent section.

1.2.1 Frequency selectivity

The human hearing system is able to enhance the detectability of band–limited waveforms by using a process analogous to electronic filtering. This process is known as the auditory filter. It increases the signal–to–noise ratio by reducing noise that would otherwise *mask* the signal. Masking is the obscuring of one sound by another (Jeffress, 1970). The existence of a filtering process is not controversial, but the characteristics of the process are controversial. These characteristics include (a) the bandwidth of the auditory filter, especially the minimum bandwidth or *critical band*,[6] (b) the shape of the filter, (c) how the bandwidth changes with centre frequency, (d) whether the filter bandwidth is fixed or adjustable, and whether information can be combined across critical bands, and (e) whether the critical band changes over the duration of the signal.

The critical band

The concept of the critical band was devised by Fletcher to account for the filtering process of the human hearing system (Fletcher, 1940, 1953). Fletcher described the critical band as a filter with a fixed bandwidth for a specific frequency. From this, he defined the masked threshold of a tone as the level where its intensity just equalled the intensity of the noise within the critical band. This definition is now known as the critical ratio (Scharf, 1970).[7] The model Fletcher fits to his data appears to have required much imagination; luckily for Fletcher, his critical band model has been thoroughly tested and examined by many other researchers and its general form is a vital construct in explaining human hearing (Schafer, Gales, Shewmaker, & Thompson, 1950; Scharf, 1970; Greenwood, 1961).

Reed and Bilger (1973) suggested that if the critical band concept is appropriate, then different methods of estimating it should come up with similar numbers. This is not the case, indicating the auditory filtering process is not as straightforward as expected, or there are methodological problems with many of the studies. On the other hand, Kaplan (1975) warned that the nature of different experimental paradigms may result in different estimates of the critical band even if the critical band was invariant. In either case, using critical band estimates for modelling performance in a different type of task should be done with caution.

One methodological problem is that some techniques for measuring the critical band do not make direct measurements of the critical band, such as Fletcher's (1940, 1953) critical ratio method. Other techniques, however, do estimate the critical band more directly, such as Patterson's notched noise method (Patterson, 1974, 1976; Patterson & Henning, 1977; Patterson, Nimmo-Smith, Weber, & Milroy, 1982) and Houtgast's (1974) modulation method. Another methodological problem comes about from observer inconsistency. This particularly affects Fletcher's method, causing overestimation of the critical band. Scharf (1970) suggests that many critical band estimates are only reliable to within $\pm$15%, due to within and between observer variability. A third problem is that the stimuli used are usually tonal transients, whose duration cannot be manipulated independently of bandwidth. Thus, there is a confounding of the signal bandwidth and the critical bandwidth if different durations are used, either within or across experiments. de Boer (1985) summarised other problems inherent in studying the concept of the critical band. These included

[6] Not to be confused with the bandwidth of an individual hair cell.

[7] The critical ratio is a measure of the threshold in decibels. It may be converted to a critical band by dividing by 10 and taking the anti–logarithm. This estimate is usually 2.5 times smaller than estimates of the critical band using direct methods. Scharf (1970), however, points out that Fletcher's original assumption, that the threshold was at 0 dB, is probably wrong, and if the threshold is reset to -4 dB, then the discrepancy between the two estimates disappears.

nonlinearities, stimulus variability, off–frequency listening and frequency splatter, temporal factors such as the "overshoot effect", two–tone suppression, and whether the critical band was always available or needed time to work.

de Boer (1962) attempted to reconcile the critical band literature by suggesting that very narrow critical band estimates are the result of artifacts from tone–in–noise masking where the noise is very narrow–band and the tone is in the centre of the noise band. He believed this was because the fluctuations of narrow–band noise interacted with the tone in a different way to wide–band noise, so it was not just the effective power of the noise contributing to the masking. He suggested that using narrow bands of noise as a signal may be a way of analysing this problem.

As a result, it is still not clear if the critical band is fixed, and the widely ranging estimates are due to these methodological problems, or whether it is flexible, and the widely ranging estimates are due to changes of the filtering process. Assuming the auditory filter is flexible, some have suggested that the concept of the critical band should be reserved for estimates of the minimum *detection* bandwidth (Bernstein & Raab, 1990).

Auditory filter shape

Patterson's research has focused on estimating the shape and bandwidth of the auditory filter (Patterson, 1974, 1976; Patterson & Henning, 1977; Patterson et al., 1982). Determining the shape of the filter is important, because it could have a large effect on estimates of the critical band. This research showed that:

- The auditory filter is symmetric on a linear scale, when measured using a notched noise masker test.
- The auditory filter is not rectangular and is not single–tuned, but somewhere in between. The rounded–exponential (roex) filter was suggested as an appropriate shape.
- Energy detectors with more realistic filter shapes resulted in changes to the slope of the function of threshold versus bandwidth, compared with that of the ideal rectangular filter.
- The typical method of estimating the critical band, by varying the bandwidth of the masker, is confounded, because the threshold is not just a function of signal power, but also $\mathcal{WT}$. In other words, the detectability of a noise is determined by its signal–to–noise ratio *and* $\mathcal{WT}$.
- The derivations for non-rectangular filters are still only valid for large $\mathcal{WT}$. For small $\mathcal{WT}$ the Gaussian approximation is not appropriate so the mathematics is more difficult.

Despite using more realistic models, the difference in the predicted thresholds was not greatly different to those predicted from the simpler models.

Kollmeier and Holube (1992), on the other hand, considered the effect of the definition of bandwidth on estimates of critical bands, for a variety of filter shapes. Depending on which auditory filter shape they chose, the critical band varied from about 43 Hz to 159 Hz for the equivalent rectangular bandwidth (ERB).[8] The critical band also depended on the definition of bandwidth chosen, and this interacted with the assumed auditory filter shape.

It appears, therefore, that part of the problem of disparate estimates of the critical band comes from the assumptions made about the shape of the filter, which is further confounded by how

[8] See Section 3.3.1 for various definitions of bandwidth.

bandwidth is defined. There are no simple answers to these problems, except to be very clear about which filter shapes and definitions of bandwidth are being used.

Critical bands as a function of frequency

Some researchers have developed analytical expressions for the critical band, as a function of centre frequency, by fitting functions to the collective body of experimental results. If CB is the bandwidth of the critical band, and f is the centre frequency, then some of these expressions are:

- Zwicker and Terhardt (1980):

$$\text{CB} = 25 + 75[1 + 1.4(f/1000)^2]^{0.69}, \tag{1.7}$$

- Moore and Glasberg (1983):

$$\text{CB} = 6.23f^2 + 93.39f + 28.52, \tag{1.8}$$

- Moore and Glasberg (1987), based on Greenwood (1961):

$$\text{CB} = 19.5(6.046f + 1), \quad \text{and} \tag{1.9}$$

- Moore, Peters, and Glasberg (1990):

$$\text{CB} = 24.7(4.37f + 1). \tag{1.10}$$

The experiment described in Chapter 3 uses noise signals centred at 500 Hz. Equations (1.7), (1.8), (1.9), and (1.10) predict critical bands, at 500 Hz, of 117, 76.77, 78.5, and 78.7 Hz, respectively. These estimates are all similar, because the functions are similar for low frequencies. The differences are primarily due to assumptions about the shape of the auditory filter. Moore and Glasberg (1987) concluded that there was still controversy over the estimates of the filter bandwidth, especially at the lower frequencies, and that the functions do not take into account individual variability.

Is the critical band fixed or adjustable?

It is generally accepted that the critical bandwidth changes with centre frequency. More controversial is whether the critical bandwidth is fixed or adjustable at individual frequencies. So far the answer has been difficult to determine, because (a) there are large individual differences, (b) data are noise degraded with observer inconsistency, (c) estimates tend to depend on the task used and the assumptions made about the shape of the filter, and (d) that bandwidth and duration are often confounded in critical band experiments, especially when using tonal transients. The evidence for adjustable critical bands comes from research into short duration signals. Long duration signals generally provide reliable 'steady–state' estimates of the critical band. Scharf (1970) provides a good summary of critical band estimates for short durations.

Hamilton (1957) estimated critical bandwidths by measuring masked thresholds of tone–in–noise stimuli in terms of $\mathcal{WT}$. Here, $\mathcal{W}$ referred to the bandwidth of the masker transients (which varied from 19–1100 Hz, and were all 2000 ms long), and $\mathcal{T}$ referred to the duration of the 800 Hz

tonal transient (which ranged from 25 to 400 ms in duration, implying their bandwidths varied from 40 Hz down to 2.5 Hz). Unlike many other studies, he considered the effects of frequency domain spreading of the tones, as a function of tonal duration, when the tones were windowed with a rectangular pulse. That is, as duration decreased, the bandwidth of the tonal transient spread by the reciprocal of the duration. Hamilton made the assumption that although the tones had varying bandwidths, most of the spread was within one critical band. If this was the case then the hearing system would not filter out part of the tonal transient.

Hamilton's (1957) results, albeit averaged across observers, indicated that the critical bandwidth, effective at masking the tonal transient, was somewhat dependent on the duration of the transient. Between 100–400 ms the critical band estimate was the same, at about 145 Hz, but as duration decreased to below 50 ms, the critical band estimate increased considerably. Hamilton suggested that this was not due to frequency spreading of the signal, above and beyond what would be expected from the reciprocal of the duration, but instead may be described by signal "salience". He admitted that this was merely descriptive, and did not explain why the critical band would appear to widen. His estimate of 145 Hz has often been quoted since, but looking more closely, his graph indicates that depending on the interpolation used, there could be quite a range in the estimate. Assuming that the data are as noisy as psychophysical data normally are, then this estimate could easily be less than 100 Hz.

Hamilton (1957) showed that equal $\mathcal{WT}$ from different combinations of $\mathcal{W}$ and $\mathcal{T}$ resulted in the same masked thresholds. He concluded that "there was no detection improvement until the filter width reached a critical value dependent on tonal duration" (p. 511). Hamilton suggested that an increase in the energy threshold as a function of $\mathcal{WT}$ was due to the ear not integrating fully beyond 50 ms, but it is not clear that Hamilton could make this inference from the rather noisy data, based on a threshold model.

Creelman (1961) extended research into critical bands by using damped sinusoids, as they were meant to be more complex than tones, and similar to speech sounds. He found some evidence that observers could adjust their bandwidths, but found that it was difficult to make stronger conclusions, because of confounding with the spectral characteristics of the signals.

Swets, Green, and Tanner (1962) also considered that the critical band may be adjustable at each frequency. They were prompted to consider this possibility because of the many disparate estimates of the critical band in the literature. They considered that variability in the estimates may have come about because (a) the hearing system may behave differently depending on the experimental task, (b) that the critical band may be dependent on overall signal level, (c) that the assumed shape of the critical band would influence measurement, (d) that there could be interactions between external filtering of the signal and internal filtering, (e) that small differences in measurement can result in large differences in estimates, and (f) that the critical band parameters may not be fixed:

> "We would suggest a consideration—in theory construction and in research strategy—of the possibility that the parameters of the mechanism of frequency selectivity vary from one sensory task to another under intelligent control. If they do, then, of course, we cannot speak of, or measure *the* critical band." (Swets et al., 1962, p.113).

van den Brink (1964) estimated thresholds by varying the duration of the tone and the bandwidth of the combined tone–in–noise signal. The masker was continuous, and the mixed signals were passed through a very steep filter to minimise the spread of the short duration tone in the

frequency domain. He found evidence that supported a critical band that increased as the tone duration decreased, thereby supporting the findings of Creelman (1961), Green, Birdsall, and Tanner (1957) and Hamilton (1957). Northern (1967) also obtained unusually wide estimates of the critical band for short duration noise.

Researchers such as Green (1960a), Jeffress (1964), Mulligan and Elrod (1970), and Swets (1963) all thought the critical band could be flexible, although apparently Jeffress changed his mind in later research (personal communication, Whitmore, 1998). Other researchers, such as Spiegel (1979), have suggested that the critical band is fixed, but observers can attend to more than one critical band at a time depending on the task. It is difficult, if not impossible, to separate out these two possibilities experimentally, because they have the same effects.[9] Spiegel (1979) was interested in finding out whether the critical band was fixed or adjustable. He considered spectral integration, and the critical band, by varying the signal bandwidth to estimate the maximum limit of integration, and by varying the masker bandwidth to determine the critical band. At that time, no other study had manipulated both with the same observer, so differences reported in the literature may have been due to individual observer differences. He also pointed out that the variability in experiments was so great that no one experiment could be used to standardise critical band estimates. His results were consistent with an energy detector, except for an attenuation constant, and the critical band was estimated to be about 80 Hz (at a 1 kHz centre frequency). He also found evidence that frequency information could be combined over 3 kHz, indicating that either the critical band was flexible or critical band channels could be integrated. Scharf (1970), on the other hand, summarised the research into short duration critical bands as indicating that as duration decreased, detectability became independent of bandwidth and all energy contributed to masking.

There is evidence that for short durations, critical band estimates are wider than the steady–state estimates. It is difficult to say, however, whether this difference comes from a variable critical band, or a combination of information across critical bands, or whether this distinction is even meaningful.

Temporal effects on the critical band

It is possible that some of the differences in critical band estimates were a result of how the filter changed with time. For instance, Green (1969) considered whether the critical band narrowed after the onset of a stimulus—a form of sharpening that may be similar to that in touch and vision. Zwicker and Fastl (1972) also considered a related idea: is the critical band permanently on, or does it "develop" during excitation? Research had indicated there was a difference between auditory filter masking functions using continuous stimuli versus those using transient stimuli, but it was not clear whether this was due to the development of the critical band or was simply a transient artifact. Zwicker and Fastl (1972) concluded that if the transient signals were designed carefully, so that their frequency and time domain properties did not include unwanted transients, then there was little difference between the auditory filter masking functions for the two types of stimuli. Thus, the critical band did not develop. On the other hand, Dai and Green (1993) suggested, after considering the role of duration in profile analysis experiments, that the auditory filter bandwidth did decrease after stimulus onset.

It is very difficult to avoid confounding time and frequency when measuring the temporal na-

[9] Section 1.4.2 considers this research in more detail.

ture of the auditory filter. For instance, comparing long and short transients, with the same bandwidth, is not enough, because potential interactions with the temporal integrator are ignored. In general, the evidence to date is not reliable enough to distinguish between a critical band that is constant over the duration of a particular stimulus, but may vary for different stimuli, and a critical band that varies over the duration of a stimulus.

1.2.2 Rectification

If an ideal observer integrates, or sums, temporal waveform information, then the waveform must first be rectified, otherwise the integral, or summation, will tend to the mean of the waveform. In hearing models, the rectifier is usually placed between the filter and the integrator. The two forms of rectification commonly used in modelling human hearing are square–law and linear (full or half–wave) rectification. It is the form of the rectification that primarily determines whether an observer is classed as an energy (square–law) or envelope (linear) detector. Section 1.1.3 discussed the differences between the two types of rectification.

The physiological evidence strongly suggests that the auditory nerve performs half–wave rectification, because the neural firing pattern is phase–locked (Laming, 1986; Whitfield, 1967), at least for frequencies below 4 kHz. The neurons fire during rarefaction of the waveform. Many researchers tend not to use linear detectors, however, because (a) they are considered more difficult mathematically, and (b) the two types of rectification are incorrectly considered to be equivalent. The two forms of rectification only result in the same statistics for $\mathcal{WT}=1$ and 70, so in general, they should not be considered functionally identical (Marcum, 1960).

There has been little work to see if psychophysically, rather than physiologically, rectification is better modelled by a linear rectifier for human hearing. An exception is Gilkey (1981) and Gilkey and Robinson (1986) who found some evidence that a computer simulation using a half–wave rectifier was better correlated with human data than one using a square–law rectifier (see Section 1.4.1). The problem has been that most tests have compared energy versus envelope detection, where the *integrator* is also different, rather than comparing forms of rectification without confounding from other processes.

1.2.3 Temporal integration and sampling

The use of a temporal integrator in models of human hearing, according to Moore, Glasberg, Plack, and Biswas (1988), was first suggested by Munson (1947; in Moore et al., 1988) and Zwislocki (1960, 1969; in Moore et al., 1988) and was based on similar models for electronic detectors (North, 1943/1963). The integrator is typically modelled as a true integrator, which discharges completely at the end of integration, or a leaky integrator (either a rectangular running averager or a negative–exponential). Physiologically, integration is considered a higher–order process, occurring at a neural level. Gambardella and Trautteur (1966), for instance, showed that the cochlea could not be the place where temporal analysis of waveforms is conducted.

Integration time, time constants, and sampling time

True integration, or true summation, is usually modelled as occurring over the duration of the signal, resulting in a single number at the end. This number is the *sample statistic* or *evidence* used to make decisions about the events. True integration may also be modelled as a process that

sweeps over the signal giving a continuous output, thus requiring a *sampling strategy* to give the sample statistic. For true integration, an optimal sampling strategy is to sample at the end of the signal, therefore, the integration time is the same as the sampling time, and is usually the same as the signal duration.

Leaky integration is usually modelled as a process, therefore requiring a sampling strategy. For leaky integration, however, there is often confusion as to what is meant by *time constant*, *response time*, *decay time*, *integration time*, *effective duration*, and *sampling time*. The terms response time, decay time, and time constant are synonymous, and are equal to the ERD for the leaky integrator.

The time constant of a leaky integrator is defined as the length of time it takes for the integrator to reach the proportion $1 - e^{-1}$ of the final saturated value.[10] In electronic terms it is also equal to RC and is the equivalent rectangular duration (ERD) of the integrator's impulse response.[11] It is also simple to derive the time constant with respect to the essential duration, $\mathrm{Ess}\mathcal{T}_{\alpha^2\%}$, of the leaky integrator. It is a transform of the ERD:

$$\mathrm{Ess}\mathcal{T}_{\alpha^2\%} \quad = \quad -\log_e(1-\alpha^2)\mathrm{ERD}/2 \tag{1.11}$$

where α^2 is the proportion of energy constrained. This equation can be rewritten to calculate the proportion of energy constrained for any integration time, D, (assuming the integration time begins at time zero):

$$\alpha^2 \quad = \quad 1 - e^{-2D/\mathrm{ERD}} \tag{1.12}$$

For instance, after a time equal to the ERD, the proportion of energy constrained is $\alpha^2 = 1 - e^{-2} = 86.5\%$.

The *effective duration*, or integration time, of a leaky integrator is related to the time constant. Mathematically, the integration time is infinite, but practically, after a certain time, the value of the input waveform no longer contributes significantly to the current output value, because the negative exponential weighting is so small. Determining the effective duration, therefore, requires a criterion. The effective integration time is usually defined with respect to the proportion of the final saturated value. For instance, Jeffress (1968) attempted to define the effective integration time by using electronic simulations. This can also be done mathematically by considering the integral of the impulse response. For a standardised integrator, the integral goes from 0 at time zero to 1 at infinity (the saturated value). The integrator reaches $1 - e^{-1}$ after a duration of one time constant, $1 - e^{-2}$ after two time constants, and so on. Jeffress' electronic integrator took 2.8 time constants to reach about 95% of the saturated value. Mathematically, after three time constants, the integrator is 95% saturated, and after five time constants, the integrator is 99% saturated. Alternatively, the effective integration time could be specified in terms of the energy of the impulse response, and defined in terms of the essential duration. This is about 1.3 time constants for $\mathrm{Ess}\mathcal{T}_{92.4\%}$.

Sampling time refers to the time the output of the integrator is sampled, to obtain a sample statistic. There are a number of ways of sampling the output of a leaky integrator, including sampling at a particular time and peak sampling. It may be difficult to determine if the sampling strategy is optimal for a transient waveform, because it is dependent on properties of the

[10] Skilling (1965, p. 42) also discusses time constants for negative exponential decay, but in a slightly different context.
[11] See Section 3.3.2 for definitions of duration.

waveform, and its window, rather than the integrator. Specific sampling strategies for a transient waveform and a leaky integrator include (a) sampling at the time equal to the time constant, (b) sampling at a time equal to the signal duration (absolute duration, or otherwise), (c) sampling at the average time of the maximum output of the integrator, and (d) peak sampling, where the actual sampling time varies for each waveform. Jeffress (1968) suggested, however, that the statistics of the envelope maxima probably did not differ from the statistics of random samples of the envelope in a way that could be measured in a psychophysical experiment. Other sampling strategies that could be used include averaging the output, taking "multiple–looks" at the output then taking the largest or averaging them, and so on.

Experimental estimations of the temporal integrator

There have been a variety of studies on how the human hearing system deals with temporal information (Scharf, 1978). Such studies essentially estimate the maximum and minimum time the human hearing system can integrate temporal information, without degradation in performance. Estimating either, especially with tones, is difficult, because short duration stimuli spread in the frequency domain, and may interact with auditory filters. Conversely, long durations result in extreme boredom for experimental observers, and may result in unreliable data.

Green (1973) explained that for many tasks it is optimal to integrate for a long time, because this increases the effective signal–to–noise ratio. On the other hand, for tasks such as the detection of clicks, it is optimal to integrate for a very short duration. After considering the evidence, Green (1973, 1985) estimated the maximum integration time to be about 200 ms ($\pm$100 ms) and the minimum integration time to be on the order of 1–2 ms or 10–20 ms depending on which experimental paradigm was used. de Boer (1985) came to similar conclusions.

Studies that have found evidence that the time constant of the temporal integrator was short include Robinson and Pollack (1973), Robinson (1974), and Ronken (1973). Studies that have found long duration (greater than 200 ms) time constants include Makita and Miyatani (1950) and Henning and Psotka (1969). Most experimental findings have indicated, however, that the human hearing system has active control over the integration time (Garner, 1947; Garner & Miller, 1947; Green et al., 1957).

In studying maximum integration time, a time versus intensity experiment with long duration signals is usually employed (Green, 1985). The rationale of these experiments is that the intensity of a signal may be traded off for duration for constant detectability. If the human hearing system has a maximum integration time there will be a point where the intensity of a signal must be held constant (instead of decreased), as the duration increases. In other words, as fast as one end of the integrator fills up the other end empties. The resulting function will have a breakpoint at this time. In reality, the function does not look like this because of the difficulty of collecting data using very long durations. It appears that there is always some improvement with duration. This could possibly be accounted for if the observer combined information across multiple looks of the long process.

Green et al. (1957) ran three experiments to establish the effect of duration and intensity on detectability of tones in noise. Instead of estimating the durations and intensities that gave equal detectability (e.g., Garner, 1947) they came up with a three parameter equation of the 'surface of detectability' where the parameters were duration, intensity, and detectability. This equation came from three experiments: (1) duration was constant and amplitude was varied, (2) energy

was constant and duration was varied, and (3) amplitude was constant and duration was varied.

Results from the first experiment indicated that, for each duration, as the energy of the signal increased, detectability also increased as a function of intensity. In log coordinates, for both d' and E/N_0, this was a linear function. In the second experiment, detectability was constant over a region of duration (ranging from about 20 ms to 200 ms) and was attenuated for shorter and longer durations. The third experiment showed that detectability continued to increase as duration increased while power was kept constant. These results could indicate that the ear optimally integrates acoustic power over this range, but the process breaks down for longer or shorter durations. It could also indicate a bandwidth phenomenon. The frequency content of a sinusoid is spread proportional to the reciprocal of its duration. For short durations, if the critical band is narrower than this bandwidth, then not all the signal energy will be passed and performance will deteriorate. Likewise, if the critical band has a lower bound, then long duration sinusoids will include more masker thereby attenuating performance.

Finally, Green et al. (1957) compared their results to the signal–known–exactly, or cross–-correlation, detector assuming either matched bandwidths, or critical bandwidths narrower or wider than ideal. They found that this model, as expected, did not adequately describe human performance, except for the results of the second experiment if an attenuation factor was included.

Moore et al. (1988) tackled the problem of the shape of the temporal window by applying the same analogies and experimental procedures that Patterson (1974, 1976) used to study the critical band, except that they were applied in the time rather than the frequency domain. They described the temporal window as a weighted running average of the waveform energy. Moore et al. (1988) believed that the observer was able to choose the best window for the task. For longer duration signals, they suggested the observer was able to combine information across the outputs of multiple windows in the same way that, as has been conjectured, observers can combine information across critical bands. The alternative, they suggested, was that the observer was able to use windows with adjustable durations. Moore et al. (1988) found evidence that the temporal window for their experiments can be well described by an asymmetrical roex function with an ERD of 8 ms.

The shape of the temporal window suggested by Moore et al. (1988) is quite different to that of the true integrator or the negative exponential leaky integrator. The roex function has a peak in the middle of the window such that the current value of the waveform is weighted less than near values in the past. In comparison, the energy detector assumes a rectangular window to give true integration, which implicitly discharges at the end of the signal. The envelope detector assumes an exponential window that weights the current value of the waveform more than previous values.

de Boer (1985) criticised the concept of the temporal integrator as being an *ad hoc* model. This is perhaps why there are so many disparate estimates of the time constant. Gerken, Bhat, and Hutchison-Clutter (1990) also considered the evidence for the existence and nature of the temporal integrator. They concluded that many things said about the form of the integrator were due to experimental artifacts—especially when using very long duration signals. In particular, the concept of a constant integration time was not supported, indicating observers could gather temporal information for much longer durations than previously thought.

Viemeister and Wakefield (1991) presented a good summary of the temporal resolution–integration paradox. The paradox is that a number of models (with supporting empirical evidence) have indicated the temporal integration time to be short, and a number that it is long. Models

that assume short integration times are bad at predicting the results of models that assume long integration times, and vice-versa. The usual explanation is that the auditory system is flexible, and uses multiple time constants, or that the time constant mechanism is flexible, and may be matched to the signal duration. They suggested a different type of temporal integration where the observer combines information from multiple looks at the output of an integrator with a short time constant. Such a model consists of the usual critical band filter and half–wave rectifier, but the leaky integrator is short (3 ms) and the running output is stored in a memory with its own decay characteristics. Their experiments indicated that this could be happening, and were, therefore, able to account for the paradox.

Preliminary investigations, using the simulation described in Chapter 4, were run to test Viemeister and Wakefield's (1991) hypothesis, using the small $\mathcal{WT}$ noise signals from the experiments reported in Chapter 3. On a stimulus level, there was a large correlation between between full integration, and a short leaky integrator (3 ms), sampled with a long rectangular running averager (500 ms). This running averager is a way of implementing a multi–look model, where a look is taken at every sample point. These results indicated that it would be difficult to test the two theories, because they predicted similar outcomes.

Experimental estimations of the sampling strategy

Experimental estimates of the sampling strategy of the human hearing system have focused mainly on whether the hearing system uses peak sampling or samples at a particular time. The method has generally been to correlate human performance with simulations using each type of sampling strategy (see Section 1.4.1 for further studies using this method). There have been, however, only a few studies that considered sampling strategy. Part of the problem has been a lack of mathematical models for peak detection, because of difficulties with the statistics of peak detection (Jeffress, 1968).

Nichols and Jeffress (1966) considered the detection of sinusoids in wide–band noise. They calculated correlations among four observers and two electronic observers. They found that the electronic observer that used average voltage was a better predictor of both the events ($\mathcal{N}$ and $\mathcal{SN}$) and the human observer's responses, than the electronic observer that used the peak voltage.

Gaston and Jeffress (1974) considered different sampling strategies for an envelope detector: peak voltage detection and end–of–signal voltage detection. They found that while performance for end–of–signal detection plateaued at the end of 200 ms, peak detection continued to increase out to 2 seconds. Gaston and Jeffress (1974) suggested that the reason peak sampling is more effective at long durations is that the samples are less variable and, therefore, the magnitude of d' increases. Indlin (1979a, 1979b), however, produced some evidence that human detection of noise bursts may be better described by a peak amplitude type detector rather than an energy detector.

Gilkey (1981) and Gilkey and Robinson (1986) also tested sampling strategies, using simulation, and found that the best strategy partly depended on the form of integration assumed (see Section 1.4.1). The problem of testing sampling strategies was that differences in performance may be subtle under any given sampling scheme.

1.2.4 Summary

The details of the four main components of human hearing: filtering, rectifying, integrating, and sampling; are still being developed. The current study attempts to clarify some of these con-

troversial problems by using error–reducing techniques in the human experiments, and through simulation.

Research into the nature of the critical band has produced conflicting and contradictory results, partly due to methodological problems. The current study attempts to deal with some of these problems by (a) estimating the detection bandwidth using noise signals with systematically varied bandwidths and durations, (b) carefully designing the signals to be well–defined in both the time and frequency domain, and (c) using error–reducing techniques to remove the effects of observer inconsistency. Further discussion about critical band research is continued in Section 1.4 after energy and envelope detectors are introduced, and then the critical bands of human observers are evaluated in Chapter 5.

The form of rectification used by the human hearing system is evaluated in the Chapter 5 by investigating the *full–linear* detector. The only difference from the energy detector is that the full–linear detector uses a full–wave, rather than a square–law, rectifier, therefore, the confounding caused by the different integrator is removed. Although the full–linear detector is a linear detector, it should not be considered an envelope detector, in the usual sense of the term, because of the true integrator. An envelope detector will asymptotically approach the full–linear detector as the leaky integrator time constant gets large.

There is no consensus on the form of the temporal integrator. The nature of true integrators and leaky integrators, with short and long time constants, are considered in Chapter 4, by using simulations, then correlated with human performance (in Chapter 5). Because there is no clear criteria for optimal sampling of a leaky integrator, the simulations using a leaky integrator in this project will consider multiple sampling strategies to help establish what strategies result in optimal detectability.

Evaluating each component individually has caused ambiguous results, because of confounding due to interactions with other components. What is needed is a model of how the components work together as a system. The next section introduces two types of system for noise–in–noise and tone–in–noise detection: the energy detector and the envelope detector. Then, more recent research into the components of human hearing are discussed with respect to these detectors.

1.3 Energy and envelope detectors

The previous section discussed the nature of the various processes of human hearing: filtering, rectifying, integrating, and sampling. These processes have also been evaluated as a whole, and not just as individual components, by using TSD to derive ideal observers, such as the energy and the envelope detector. The energy detector consists of a band–pass filter, square–law rectifier, and a true integrator. An envelope detector consists of a band–pass filter, linear rectifier, and a leaky integrator. This section describes various ideal energy and envelope detectors. The next section looks at experimental evaluations and comparisons of these detectors, as well as modifications to these detectors to make them work more like humans. Both tone–in–noise and noise–in–noise models are considered because their development did not occur in isolation.

Two common approaches in deriving an ideal observer are by (a) using the dimensionality of sampling spaces and (b) making mathematical analogies to electronic components. The two approaches are essentially the same, but can result in confusion by taking analogies too far. For instance, the limitations of real electronic components may be incorporated in an electronic model.

Others have confused the discrete signal representation for real sampling (Pridham, 1968). Essentially, the same result can be derived from either approach.

1.3.1 Engineering models

The engineering models of tone–in–noise and noise–in–noise detection began the development of energy and envelope detectors as psychophysical models of human hearing.

Peterson et al. (1954) showed that the detectability of a deterministic signal, masked by white Gaussian noise, could be directly related to signal energy. Their measure of detectability was $d^2 = 2E/\mathrm{N}_0$, which was modified by Tanner and Birdsall (1958) to $d' = \sqrt{2E/\mathrm{N}_0}$. This SKE observer is a cross–correlation detector, derived using the likelihood ratio (Elliot, 1964). The underlying distributions are normal with equal variance, and d' is the standardised difference between the means. The biggest problem with this derivation is that it is only approximately related to signal energy for small $\mathcal{WT}$. Although the cross–correlation detector may not be a good model of human hearing, it is important to consider its properties, because it can be used as a comparison to the signal–known–statistically (SKS) ideal observers. Peterson et al. (1954) also derived a number of other detectors including an envelope detector for deterministic signals.

Urkowitz (1967) was also interested in the engineering problem of the detection of an effectively finite, deterministic, signal in band–limited Gaussian noise. Information about the exact nature of the signal was unknown, except for the bandwidth. Because the signal is unknown, a cross–correlation detector is inappropriate, so Urkowitz derived an energy detector instead. Unlike many other derivations of the energy detector, Urkowitz went to some length to explain the waveform representation used and its limitations. He also explicitly derived the band–pass case. Although he was aware of the work by Grenander et al. (1959), where the exact energy distribution of essentially band and time limited noise was derived, Urkowitz used the sampling plan approach, because he believed the difference for engineering uses was negligible for $\mathcal{WT} > 1$. He used Shannon's sampling scheme, but in an appendix used the Karhunen–Loève expansion, which relied on using prolate spheroidal wave function as the orthogonal basis functions; a more rigorous way that gives the same result for $\mathcal{WT} > 1$. To derive $\mathcal{WT}$=1, he transformed the energy statistics to envelope statistics, which resulted in the Rayleigh–Rice detection statistics (based on Davenport & Root, 1958).

1.3.2 Early SKS observer models of human hearing

Early SKS ideal observer models were either energy or envelope detectors of tone–in–noise or noise–in–noise. The energy detectors of Green, McGill, and Pfafflin and Mathews, and the envelope detectors of Marill, Jeffress, and Whitmore and Drga are considered. The subsequent section then examines how well these models describe experimental results from human observers.

Marill's envelope detector

Influenced by the early psychophysical work by Tanner and Birdsall (1958), Marill (1956) extended TSD to include ideal observers for the 2IFC task, and incorporated a variable criterion into the model to account for psychological noise. He argued that his 'response model' was superior, because it removed the unobservable psychological experience, and concentrated solely on predicting responses from stimuli without needing to specify the psychological stimulus magnitude.

He also showed how the introduction of 'catch–trials' allowed the observer's *task* to be defined also without reference to psychological phenomena. Specifically, Marill derived the SKE case and the SKS (phase) for tone–in–noise detection in the 2IFC task.

Green's energy detector

Green (1960a) derived an energy detector for 2IFC noise–in–noise detection, based on Peterson et al. (1954). He assumed that $\mathcal{WT}$ was large so he was able to use a Gaussian approximation for the $\mathcal{N}$ and $\mathcal{SN}$ distributions. From these distributions, he derived $P(\mathrm{C})_{\mathrm{2IFC}}$, assuming a difference decision rule. Green (1960a) was interested in using noise as a signal, as well as a masker, because the bandwidths of the signal and noise form a continuum. When the bandwidths are identical, the task is to detect increments in intensity. When the bandwidths are different, the detectability of the masked stimulus or the effectiveness of the masking stimulus can be studied.

Green (1960a) aimed to see how the detection of a noise signal depended upon the power, bandwidth, centre frequency, and duration of the signals. When the signal and noise bandwidths were the same, no additional filtering could improve detectability. Thus, Green's model could draw all its parameters from the physical signal, rather than the observer. Green's model is also applicable to the case where the signal bandwidth is narrower than the masker, if a filter is incorporated.

Green's (1960a) energy detector consisted of an ideal rectangular filter, a square–law rectifier, and a true integrator. Sampling occurred at the end of the input signal, thus integration was over the duration of the signal. Green used a square–law rectifier to generate a value whose average was monotonic with signal intensity. It was a mathematical convenience that the square–law was chosen over a linear rectifier, which implies Green did not set out to derive an energy detector in particular. The integrator smoothed the output of the filter and rectifier, and the sampler returned a number to be used as the evidence for a decision. The parameters that resulted were a bandwidth, which was either the bandwidth of the signal or the bandwidth of the filter (whichever was smaller), and a duration, which Green said was the signal duration, but, more accurately, it was the duration out of the filter (which for a narrow filter could be longer than the original signal duration due to ringing). Green realised this but used signal duration, because for large $\mathcal{WT}$, the two are nearly identical. The resulting detector is ideal, because it is monotonic with likelihood ratio. Green considered his model had the minimum number of elements necessary to begin to describe human hearing, and that it gave a basis to build a more realistic or detailed model.

The detection statistics of Green's energy detector are χ^2 with $2\mathcal{WT}$ degrees of freedom (Green, 1960a; Green & Swets, 1966). Because of the Gaussian approximation, this detector is only valid for $\mathcal{WT} > 30$, although some researchers believe the approximation is good for $\mathcal{WT} > 10$. The measure of detectability he derives is d'–like, and is only valid for signal–to–noise ratios less than 0 dB, where the variances of the distributions remain essentially equal. Green and Swets (1966) extended this derivation to a number of related energy detectors: both for tone–in–noise and noise–in–noise detection, and for simple and pedestal style experiments. It is unfortunate that the popularity of Green's detector has meant that many researchers ignore the limits on $\mathcal{WT}$ and signal–to–noise ratio, and apply the model regardless. This is despite the fact that McGill and his colleagues subsequently developed a similar energy detector for small and large $\mathcal{WT}$ and arbitrary signal–to–noise ratio.

Pfafflin and Mathews' energy detector

Pfafflin and Mathews (1962) derived a 2IFC energy detector that consisted of a single–tuned band–pass filter, a square–law rectifier, and a true integrator. A number of assumptions were made to simplify the mathematics. For example, the duration of the signal must be larger than the response time ($1/W$) of the filter. This implied their model was not applicable to $\mathcal{WT}$=1. They also stated that there is essentially no difference between the square–law and half–wave rectifiers if the signal is symmetric about zero. Assuming large $\mathcal{WT}$ (implicitly), they derived $P(\mathrm{C})_{2\mathrm{IFC}}$, and, therefore, came up with a normal approximation. In comparing performance of the simplified mathematical model with a Monte Carlo simulation (which had fewer assumptions) they found good agreement.

McGill's energy detector

McGill (1967, 1971) tried to reconcile ideal energy detection with a possible neural counting model in the auditory system. The aim was to mimic the continuous χ^2 energy detection statistics with counting (Poisson and negative binomial) statistics to find out if something akin to energy detection could occur within a discrete neural counting model. He achieved this aim, and showed that the resulting characteristics of the counting model were essentially the same as the energy detector, for many circumstances. McGill (1971) postulated that this type of Poisson counting process underlies the flux of information flowing through any sensory monitoring system. For noise intensity discrimination, his results for the neural counting mechanism gave essentially the same results as the energy detector of Green (1960a). This was because the neural counting model obeyed Weber's law and produced psychometric functions dependent on signal–to–noise ratio, but not overall level.

McGill (1967) and Green and McGill (1970) based their derivations on a novel formulation of Gaussian noise, which used the convolution of Rayleigh noise ($\mathcal{WT}$=1) samples, spaced $1/\mathcal{T}$ Hz apart in the frequency domain. The result was independent samples of noise. This appears to be an alternative, but equivalent, way of achieving waveform representation using the sampling space, as discussed in Appendix A. From this representation of noise, it can be shown that $\mathcal{WT}$ is related to degrees of freedom, ν, where $\mathcal{W}$ is the bandwidth of the counting process, and is unknown, and $\mathcal{T}$ is the post–stimulatory observation period, which is also unknown and not necessarily the same as the signal duration. McGill (1967) showed that if N is the total energy in the noise stimulus, then $N/\mathcal{WT} = N/\nu = \mathrm{N}_0$, where N_0 is the mean noise energy per Rayleigh component, and is comparable to the noise power per unit cycle in models like those of Peterson et al. (1954).

McGill (1968b) showed how to derive a tone–in–noise energy detector from a Rayleigh–Rice envelope detector. He called it the "energy" form of the Rayleigh–Rice distributions. He showed that this detector was equivalent to Marill's (1956) ideal observer, but that it could also be derived without reference to an ideal observer simply by considering the energy distributions of the signals. Because Jeffress (1964, 1967) showed that a detector based on Rayleigh–Rice statistics only applied to noise with $\mathcal{WT}$=1, McGill's transformed energy detector also applied only to noise of $\mathcal{WT}$=1.

McGill (1968b) realised, however, that an imperfect human observer may have a much wider bandwidth than an ideal observer, and therefore detectability would be based on statistics with a much larger $\mathcal{WT}$ (assuming the observer integrated over the duration of the signal). Therefore,

he generalised his energy detector to accommodate noise of any integer $\mathcal{WT}$, by using the model of wide band noise developed in McGill (1967). The output statistics of this detector were χ^2 with 2 degrees of freedom for $\mathcal{N}$, and non–central χ^2 with 2 degrees of freedom and a non–centrality parameter for $\mathcal{SN}$. From these statistics, he derived the psychometric function for $P(\mathrm{C})_{\mathrm{2IFC}}$ for arbitrary, integer, $\mathcal{WT}$.

Green and McGill (1970) expanded on earlier work by deriving energy detector psychometric functions for 2IFC tone–in–noise and noise–in–noise detection tasks. For the noise–in–noise case, $P(\mathrm{C})_{\mathrm{2IFC}}$, defined using a ratio decision rule, was shown to be the probability associated with the tail of an $\mathcal{F}$ distribution with ($2\mathcal{WT}$, $2\mathcal{WT}$) degrees of freedom beyond the cutoff of $\mathrm{N}_0/(\mathrm{S}_0+\mathrm{N}_0)$. Similarly, $P(\mathrm{C})_{\mathrm{2IFC}}$ for the tone–in–noise case was shown to be the probability associated with the tail of an $\mathcal{F}$ distribution with ($2\mathcal{WT}$, $2\mathcal{WT}$) degrees of freedom, and a non-centrality parameter of $2E/\mathrm{N}_0$. The psychometric function for tone–in–noise predicted that as $\mathcal{WT}$ increased, detectability decreased. This is because most of the signal is concentrated within a bandwidth of $1/T$ Hz, thus as the amount of noise masker is increased, the signal becomes harder to detect. The psychometric function for noise–in–noise predicted that as $\mathcal{WT}$ increased, detectability increased. This indicated that a noise–in–noise observer should strive to match the bandwidth and duration of the signal to optimise detectability. McGill (1968a) and Green and McGill (1970) also produced polynomial versions of the $P(\mathrm{C})_{\mathrm{2IFC}}$, noise–in–noise and tone–in–noise, psychometric functions for arbitrary $\mathcal{WT}$.

Jeffress' envelope detector

Jeffress was an advocate of the simple–filter model of the ear, based on the concept of the ideal observer. He believed that monaural phenomena could be explained by a narrow filter followed by a simple detector, because he could find no experimental evidence that the ear was a sophisticated detector (Jeffress, Blodgett, Sandel, & Wood, 1956). He considered the problem of tone–in–noise detection by studying the statistics of the stimulus using a *stimulus–oriented* approach (Jeffress, 1964, 1967, 1968). He took three themes—mathematical models, electronic models, and psychophysical experiments—and tried to build a consistent model of tone–in–noise detection.

Assuming continuous tones, masked by narrow–band noise, Jeffress (1967, 1968) presented Rice's (1954) derivation of the envelope detector, where the evidence was derived from independent, instantaneous samples of the envelope of the waveform. This process produced the Rayleigh distribution for narrow–band noise, and the Rice distribution for tone–in–noise. The detector was then implemented as an electronic model where the output was shown to be a good fit to the mathematical model, for particular values of bandwidth, integration time constant, and sampling strategy.

Jeffress (1967, 1968) then compared Rice's (1954) envelope detector with the envelope detector of Peterson et al. (1954) and claimed they were the same, because the form of the probability distributions were similar. He made the argument that Rice's parameter was the signal–to–noise ratio A/σ (where σ^2 is the rms noise voltage, $2\mathrm{N}_0\mathcal{W}$, of bandwidth $\mathcal{W}$, and A is the amplitude of the tone) and that Peterson et al. (1954) had the parameter E/N_0 (where the energy of the tone is $E = A^2\mathcal{T}/2$, and N_0 is the power per unit cycle of the noise). Therefore, $\sqrt{2E/\mathrm{N}_0}$ was equivalent to A/σ. Furthermore, $A^2/2\mathrm{N}_0\mathcal{W} = A^2\mathcal{T}/2\mathrm{N}_0$, therefore, $\mathcal{WT}$=1. This is despite the fact that Rice was dealing with continuous waveforms and Peterson et al. (1954) were dealing with gated waveforms. Jeffress claimed that this showed the envelope detector of Peterson et al. (1954) used

a filter with a bandwidth that was the reciprocal of the duration of the waveform. He did not, however, use an electronic model to test these relationships.

Jeffress (1968) stated that, ideally, there was an optimal bandwidth for a given signal duration, and an optimal time constant, and that the ideal detector would match both of these parameters using the signal duration. His ideal SKS detector for gated signal in continuous noise therefore had a filter with a bandwidth that was the reciprocal of the signal duration, followed by a matched envelope filter.

Jeffress (1967, 1968) obtained statistics of the performance of his electronic simulations by sampling the output of the integrator at the end of the gated signal for the $\mathcal{SN}$ event, and randomly for the continuous noise from the $\mathcal{N}$ event. Frequency histograms were obtained from the samples that approximated the underlying probability density functions, and the estimated *p.d.f.* moments were used to calculate measures of detectability, and to assess the form of the $\mathcal{SN}$ and $\mathcal{N}$ distributions. Jeffress found that the $\mathcal{SN}$ distributions approached normality as signal level increased, and that although the $\mathcal{N}$ distributions were skewed, they were not well fitted by the Rayleigh distribution. These results, along with experimental results of other researchers, lead Jeffress to develop the non-central χ model of tone–in–noise detection. This model was developed because it seemed that performance was somewhere between Rayleigh–Rice (which is a special case of the χ distribution) and normal–normal.

Jeffress (1967, 1968) derived the χ model using the same approach as McGill (1968a, 1968b). Because χ statistics can be derived from χ^2 statistics, McGill's generalisation to arbitrary $\mathcal{WT}$ can be used, with a change of parameter, to extend the degrees of freedom of the χ model. Jeffress noted that the χ and χ^2 models give the same ROC curves for tone–in–noise detection.

Although Jeffress' focus was on tone–in–noise detection, the envelope detector is appropriate for noise–in–noise detection too. Whitmore et al. (1968) and Drga (1988) are the only ones to have developed a psychophysical model for this case, but only for $\mathcal{WT}$=1.

Whitmore and Drga's envelope detector

Much of Whitmore's research has focused on the detection of narrow–band noise signals in both monaural and binaural tasks (e.g., Wilbanks & Whitmore, 1967; Whitmore, Williams, & Ermey, 1968; Whitmore, 1969; Whitmore, Drga, & Taylor, 1993). He thought that Jeffress' (1968) envelope model was a better model of human hearing than the energy models of Green and McGill (1970), but at the time there was no theory of noise–in–noise detection for the envelope detector.

Rayleigh–Rice statistics were a good model of an envelope detector for tone–in–noise detection, so Whitmore thought Rayleigh–Rayleigh statistics would be a good model for an envelope detector of narrow–band noise. The task was modelled as increment detection of narrow–band noise (assuming that either both the signal and masker noise were narrow–band, or a narrow–band filter [critical band] operated on a wider–band noise process).

Whitmore et al. (1968) derived the psychometric function using Egan's formulae for Rayleigh–Rayleigh ROC curves and showed that

$$\mathcal{A}_{\mathrm{SIFC}} \quad = \quad (\sigma_{\mathcal{N}}^2 + \sigma_{\mathcal{S}}^2)/(2\sigma_{\mathcal{N}}^2 + \sigma_{\mathcal{S}}^2) \tag{1.13}$$

where $\sigma_{\mathcal{N}}^2$ and $\sigma_{\mathcal{S}}^2$ were the rms voltages of the masker and signal noises. This turned out to be a special case of Green and McGill's (1970) *energy* detector for $\mathcal{WT}$=1, which implied that for

$\mathcal{WT}=1$, the noise–in–noise energy and envelope detectors were identical.

Drga (1988) derived the 2IFC ROC curves and psychometric functions for the Rayleigh mathematical model and the negative exponential mathematical model. These derivations were based on the SIFC derivations of Whitmore et al. (1968) and Egan (1975). The negative exponential distribution is a strictly monotonically increasing transform of the Rayleigh distribution, and results in the same ROC curves for the SIFC task (Egan, 1975). Drga was interested in establishing whether the 2IFC ROC curves and psychometric functions were equal for the two distributions, because the negative exponential is easier to work with mathematically.

1.3.3 Ideal observers of narrow–band Gaussian noise

Egan and Clarke (1966) and Green and Swets (1966) both concluded that the experimental results ruled out the cross–correlation (SKE) detector as a model of human performance, and that either the energy or the envelope detector were better. McGill and Teich (1991) asked why a real detector would not optimally detect tone–in–noise akin to a cross–correlation detector. One answer is that this same detector also has to detect noise as well, thus only one mechanism is used for detecting deterministic and non-deterministic signals. Birdsall (1960) also suggested a reason for a human observer to use energy detection, rather than cross–correlation, for deterministic signals, would be when the memory template is very noisy. In this case, an SKS observer would be superior.

Green and Swets (1966) made the important point that the way noise is represented is fundamental to any development of an ideal observer. Subtle changes in representation may have major effects on a detector based on that representation. Most criticisms of an ideal observer are directed towards the representation of noise. They also believe that models of human hearing should be parsimonious so that psychological mechanisms are not be invented for a phenomenon that can be explained in the stimulus domain.

All of the models discussed in the previous section were remiss in at least one of the following areas:

1. The derivations of the models were incomplete—either missing important details of the waveform representation or of the detection statistics.

2. They did not define both bandwidth and duration. These definitions are vital, because the degrees of freedom are specified entirely by their product: $\mathcal{WT}$.

3. They were restricted to only $\mathcal{WT}=1$ or very large $\mathcal{WT}$ and small signal–to–noise ratio so that a normal approximation could be used.

4. Many underlying assumptions and approximations were left unspecified.

Focusing on the positive: Urkowitz (1967) provided probably the best description of waveform representation; Green and McGill (1970) and Egan (1975) did the best job of deriving the detection statistics for arbitrary degrees of freedom; McGill (1967) and Pfafflin and Mathews (1962) came closest with how they used and defined bandwidth and duration; and Green and Swets (1966) did a reasonable job of outlining the effects of their assumptions. These positives can perhaps be joined up to form a cohesive theory of energy and envelope detection.

The previous sections have described the theory of signal detectability, ideal observers, waveform representation, and some of the vast body of research into human hearing. This research

project is a follow–on from Whitmore et al.'s (1968) noise–in–noise envelope detector, and Green and McGill's (1970) noise–in–noise energy detector. This section presents an overview of the specific mathematical detector assumed for this project. Because no theory exists for an envelope detector of narrow–band noise with $\mathcal{WT} > 1$, the energy detector will be used as the main mathematical model, with comparisons made to the envelope detector when appropriate.

The current study focuses on the detectability of essentially band and time limited Gaussian noise, which is masked by wider band (but same duration) Gaussian noise, thus a filter needs to be explicitly included. For simplicity, the filter is assumed to be an ideal rectangular filter, because incorporating more complicated filters is easier to do by simulation rather than mathematics. Later, the computational models will be compared to the mathematical models. The concept of internal noise is not included, because the use of GOC analysis (introduced in Chapter 2) reduces the effect of internal noise, and FORCE analysis allows the asymptotic performance to be estimated.

$\mathcal{WT}$ and energy detection

Section 1.1.3 and Appendix A explained how the bandwidth–duration product characterised the acoustical uncertainty principle and defined the dimensionality of the signal space. Because $\mathcal{WT}$ is a fundamental parameter in waveform representation, it is also a parameter in all *stimulus–oriented* psychoacoustic theories.

Using an electronic analogy, the energy detector of narrow–band Gaussian noise, masked by wider–band Gaussian noise, consists of a filter, square–law rectifier, and a true–integrator. It is assumed that the filter has the same bandwidth as the $\mathcal{S}$ signals, and that the signal duration is unchanged. The output of the integrator, at time $\mathcal{T}$, is proportional to an approximation of the energy of the filtered signal. This model may be approximated electronically or computationally, and, therefore, provides another way of modelling human hearing.

Evidence distributions

After filtering, the waveform can be represented approximately as a sampling of $2\mathcal{WT}$ points, where each point is independently, identically, distributed (*i.i.d.*) as standard normals $\mathcal{N}(0, \sigma^2)$. This assumption is probably the weakest, because it does not take into account the windowing of the signal in the time domain. From Section 1.1.3 and Appendix A, an essentially time–limited and band–limited signal, with finite energy can be represented approximately by $2\mathcal{WT}$ independent sample values.

As was discussed in Section 1.1.3, it is generally considered that the χ^2 approximation for the distribution of energy is adequate for psychophysical purposes. If this model is used, then an energy detector can be derived that results in a non-standard χ^2 distribution with $2\mathcal{WT}$ degrees of freedom for the $\mathcal{N}$ event and a non-standard χ^2 distribution with $2\mathcal{WT}$ degrees of freedom for the $\mathcal{SN}$ event. If $\mathcal{WT}$=1 then Urkowitz (1967) suggested transforming to envelope statistics, resulting in distributions that are Rayleigh. This distribution is meant to be the exact solution for $\mathcal{WT}$=1. It is in fact, however, a transform of a χ^2 distribution with two degrees of freedom, that Grenander et al. (1959) showed to be not equal to the exact solution for the distribution of energy.

Egan (1975) showed how a vector of $2\mathcal{WT}$ normally distributed variates may be transformed into a non-standard χ^2 distribution with $2\mathcal{WT}$ degrees of freedom. It is non-standard, because the variances of the $\mathcal{SN}$ and $\mathcal{N}$ events are different.

ROC curves and psychometric functions

Egan derived the χ^2 ROC curves from the appropriate χ^2 distributions, using likelihood ratio, along with related measures of detectability such as $\mathcal{A}_{\mathrm{SIFC}}$ (Egan, 1975, section 5.6). He also showed that the hit rate and the false–alarm rate have the same form, but differ in the values of the parameters (Egan, 1975, Appendix C.4, p. 239) . Specifically, the rate is given by one minus the distribution function $H(c)$, as a function of the criterion c:

$$1 - H(c|\sigma, \mathcal{WT}) \quad = \quad \exp\left(\frac{-c}{\sigma^2}\right) \sum_{i=0}^{\mathcal{WT}-1} \frac{(c/\sigma^2)^i}{i!}. \tag{1.14}$$

Egan only derived $\mathcal{A}_{\mathrm{SIFC}}$ for the case of $\mathcal{WT}$=1, where he showed that the relationship between hit rate and false–alarm rate was power–law. For larger $\mathcal{WT}$, he derived another measure of detectability D_c that is specific to the χ^2 detection problem.

Green and McGill (1970) derived the more general measure of detectability, $P(\mathrm{C})_{\mathrm{2IFC}}$, for the χ^2 detector. This may be equated to $\mathcal{A}_{\mathrm{SIFC}}$ through the relationship $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{\mathrm{2IFC}}$, which holds for a ratio decision rule in the 2IFC task for continuous random variables. There are indications that this relationship also holds empirically for $\mathcal{WT}$=1 noise–in–noise signals (Lapsley Miller et al., 1999; Whitmore et al., 1968). Green and McGill (1970) gave the formula for the psychometric function, $P(\mathrm{C})_{\mathrm{2IFC}}$ as a function of S_0 and N_0, as

$$\begin{aligned} \mathcal{A}_{\mathrm{SIFC}} &= P(\mathrm{C})_{\mathrm{2IFC}} \\ &= \sum_{j=0}^{j=\mathcal{WT}-1} {}^{2\mathcal{WT}-1}\mathrm{C}_j \left(\frac{\mathrm{N}_0}{\mathrm{S}_0 + 2\mathrm{N}_0}\right)^j \left(\frac{\mathrm{S}_0 + \mathrm{N}_0}{\mathrm{S}_0 + 2\mathrm{N}_0}\right)^{2\mathcal{WT}-j-1} \end{aligned} \tag{1.15}$$

where C is the combinatoric operator. If N_0 is assumed to be unity then the signal–to–noise ratio is equal to $10\log(\mathrm{S}_0)$. A family of psychometric functions, for different $\mathcal{WT}$, is shown in Figure 1.2(a) for $\mathcal{A}_{\mathrm{SIFC}}$.

The psychometric function for $\mathcal{D}_2$, as a function of signal–to–noise ratio, may also be derived by simply transforming $\mathcal{A}_{\mathrm{SIFC}}$ using equation (1.1). The corresponding family, for different $\mathcal{WT}$, is shown in Figure 1.2(b), which has never before appeared in the literature.

There is contradictory evidence that the envelope and energy detectors result in the same performance for $\mathcal{WT} > 1$. For the tone–in–noise task, Jeffress (1964, 1967, 1968) indicated that the two detectors are equivalent for arbitrary $\mathcal{WT}$. It is not clear whether this equality holds for noise–in–noise detection, because there is no equivalent theory for arbitrary $\mathcal{WT}$. The results from Marcum (1960), however, indicate that there should be a difference between the two detectors, for both tone–in–noise and noise–in–noise detection, because there is a difference in the output of square–law and linear rectifiers. This result, however, was based on the assumption that χ and χ^2 were the underlying distributions. It is true that one may be transformed into the other. The question is whether these distributions are accurate models for the noise process, especially for small $\mathcal{WT}$. At least for energy detection, the χ^2 distribution is only an approximation.

A mathematical model of envelope detection for noise–in–noise is, therefore, more of a problem, because the theory has not been explicitly developed for this task for $\mathcal{WT} > 1$. By analogy, it is likely that the energy and envelope ROC curves will be the same for noise–in–noise. This is because the $\mathcal{N}$ evidence distribution is the same for tone–in–noise and noise–in–noise, and the $\mathcal{SN}$ noise–in–noise evidence distribution is of the same form as the $\mathcal{N}$ evidence distribution. Thus, it

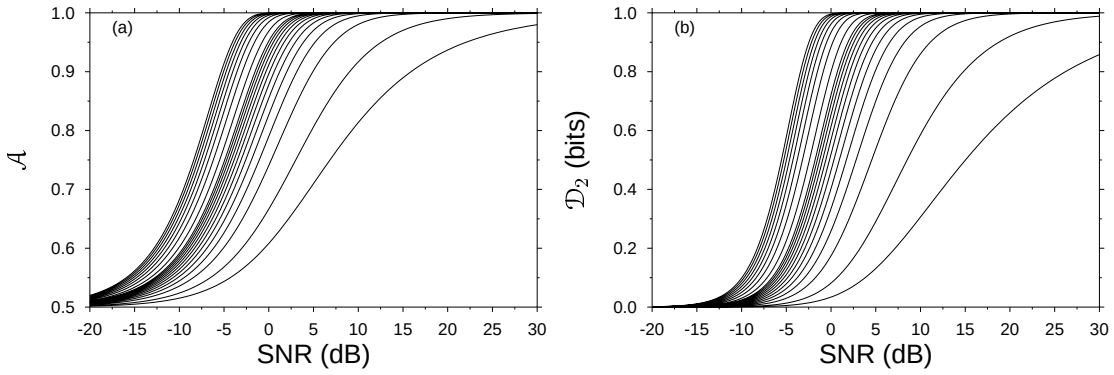

Figure 1.2: Psychometric functions of (a) $\mathcal{A}$, (after Green & McGill, 1970; McGill & Teich, 1991) and (b) $\mathcal{D}_2$, from equation (1.1), for the χ^2 energy detector, as a function of $\mathcal{WT}$ and signal–to–noise ratio. The lowest function is for $\mathcal{WT}=1/2$, the next twelve are for $\mathcal{WT}=1$–12, and the rest are in steps of 5 for $\mathcal{WT}=15$ to 50.

will be assumed, for the moment, that the same ROC curves underly both energy and envelope detection. Deriving this mathematically is outside the scope of the current study. Instead, the envelope detector for noise–in–noise detection is evaluated with computer simulations (Chapter 4).

1.4 Ideal observers as models of human hearing

All of the energy and envelope detectors introduced in the previous section have $\mathcal{WT}$ as a fundamental parameter. In other words, these detectors have bandwidth–duration reciprocity, because detectability is the same regardless of the individual values of $\mathcal{W}$ and $\mathcal{T}$. In studying real observers, estimates of $\mathcal{W}$ and $\mathcal{T}$ are also of interest, but only $\mathcal{WT}$ may be estimated without making large assumptions about how the observer filters and integrates waveforms. As a result, researchers have had to be very crafty in their experimental designs and methods.

The energy and envelope detectors, as models for human hearing, have been modified as more evidence comes in about how the human hearing system works. One way the energy detector has been modified, for instance, was by incorporating an internal noise process. This process attenuates performance to match that of a noisy human. Detectors have also been modified to have fixed or unmatched filter bandwidths, various non-ideal filter shapes, fixed temporal integrators, and various sampling strategies. This section firstly describes the initial testing of the ideal observer models, then describes how some of these models have been modified, and compared.

Marill's envelope detector

Marill's (1956) envelope detector predicted the scale and slope of the human psychometric function better than that predicted by the SKE model, although both models were unable to account for the location of the function. In other words, the human psychometric function was attenuated relative to the theoretical function. Marill attempted to explain the location problem by appealing to internal and external noise, and suggested the internal filter of the ear was wider than the implicit filter of the ideal observer. His 2IFC experiments were confounded, however, because he used no interval between the two observation intervals. This potentially caused interactions between the signals in each interval.

Green's energy detector

Green (1960a) conducted two experiments to test his energy detection model: one that manipulated duration for a constant bandwidth of 3800 Hz, the other that manipulated bandwidth as a function of one duration and a variety of centre frequencies. Results from the first experiment indicated that except for the longest duration, performance was similar for all durations. In general, performance was attenuated 5–6 dB from the theoretical, but the psychometric functions were steeper. The data, however, are extremely noisy and it is difficult to pick any shape out. Analysis at threshold indicated that performance was constant with duration. Results from the second experiment indicated that performance was the same at all bandwidths, and all centre frequencies, except for the highest, implying the ear is capable of manipulating its critical band if it proves useful to do so. Green suggested that the small critical bands, estimated with tone–in–noise studies, could possibly be added together to make filter banks that would effectively widen the bandwidth.

Green suggested that his model was useful, because, except for one variable (attenuation), performance was consistent with the model across most bandwidths, durations, and centre frequencies. He inferred that the human hearing system could therefore match bandwidth, and duration, and that the attenuation was due to some internal noise process. He could not, however, account for the steeper psychometric functions. Green, however, was only able to test large $\mathcal{WT}$ signals. Considering Green and McGill's (1970) subsequent psychometric functions for the energy detector, the shape for large $\mathcal{WT}$ values is very similar as $\mathcal{WT}$ changes, thus, Green (1960a) would not have been able to identify more subtle variations in detectability.

Although the energy detector has been subsequently modified and extended, this study is still important. It was one of the first to systematically consider detectability of noise signals, as a function of bandwidth and duration, and the data have been used as a baseline for comparison ever since.

Pfafflin and Mathews' energy detector

Pfafflin and Mathews's (1962) research was influenced by Jeffress et al. (1956) and Sherwin, Kodman, Kovaly, Prothe, and Melrose (1956), and they anticipated much of the research to come in the 1960's and later. For the tone–in–noise pedestal task, they compared the performance of their ideal energy detector, implemented mathematically and computationally, with human experimental data. They found that the energy detector could account for many of the phenomena associated with tone–in–noise pedestal experiments. They compared their energy detector model with Green (1960a), and could account for the attenuated results two ways: (a) an energy detector with a very wide bandwidth and (b) an energy detector with a bandwidth nearer the critical band estimates in the literature of around 200 Hz that had internal noise added to the test statistic. Pfafflin and Mathews (1962) decided that the latter was more likely and used this model to test a number of pedestal phenomena.

Jeffress' envelope detector

Jeffress primarily used electronic simulations to test and develop his tone–in–noise envelope detector, but he also compared his findings to the experimental results of other researchers. His initial findings (Jeffress, 1964) indicated (a) that the SKS observer, which cannot use phase information, was a better model of tone–in–noise detection for human observers than the SKE ob-

server, (b) that there was empirical evidence against threshold models, and (c) there was weak evidence that the ear adjusted the filter bandwidth with longer signal durations (commensurate with Hamilton, 1957).

In considering the task of detecting a gated signal masked by continuous noise, Jeffress used an electronically simulated envelope detector to account for the results of Green et al. (1957). The simulation consisted of a narrow–band filter, half–wave rectifier, and a leaky integrator with a short time constant. The best performance of the simulation occurred at a duration equal to the reciprocal of the bandwidth of the filter, but the results were not consistent with Green et al. (1957). When he increased the time constant of the integrator, performance became similar to Green et al. (1957). According to Gilkey (1981), Jeffress (1964) considered an adjustable critical bandwidth, but by his later papers felt he could explain the data of Green et al. (1957) without appealing to the 'questionable' hypothesis of an adjustable bandwidth.

Jeffress (1968) suggested that the differences between ideal and human performance were that (a) the human observer used different auditory bandwidths and time constants, (b) the human data were noise degraded from observer inconsistency, and (c) humans used different sampling strategies. He suggested that the statistics of sampling at the end of the integrator, compared with the maximum of the integrator, would not differ very much, and would probably not show up in a psychophysical experiment. The observer would, however, respond differently to the same stimuli under each strategy.

Jeffress (1968) considered the role of duration in tone–in–noise detection by fitting the χ model to the output of the electronic simulation, and to human performance. It was relatively straightforward to establish the appropriate $\mathcal{WT}$, but next to impossible to establish $\mathcal{W}$ or $\mathcal{T}$ without making gross assumptions about one or the other. Jeffress assumed that duration was determined by the integration time, and not by the duration of the original waveform for long durations (where the integrator has ample time to saturate). For short duration signals, he admitted the effective duration was more dependent on the original waveform, but that it was difficult to specify. Jeffress had trouble working out an appropriate value for $\mathcal{T}$, because the leaky integrator output was not necessarily proportional to duration. If $\mathcal{T}$ was equal to the original signal duration, performance should have improved as signal duration decreased, for constant signal–to–noise ratio. This did appear to occur, but it was not known whether it was due to a reduction of $\mathcal{WT}$, the narrow filter, or both.

Jeffress (1968) concluded that (a) monaural detection of tone–in–noise in continuous noise was best modelled with a modified non-central χ distribution (of which the Rayleigh–Rice distributions are a special case), (b) the time constant for monaural detection was about 50 ms, which implied an effective integration time of about 140 ms, (c) the *effective* signal duration for signals shorter than the effective time constant, was shorter than the absolute signal duration, which meant the effective signal–to–noise ratio was less than that computed using absolute signal duration, and (d) the assumption of a 50 ms time constant gave 50–70 Hz estimates of the critical band.

Whitmore and Drga's envelope detector

Whitmore et al. (1968) evaluated the noise–in–noise envelope and energy detectors two ways: by comparing the tone–in–noise and noise–in–noise detector in SIFC and 2IFC experiments, and by comparing the noise–in–noise detector for three different values of $\mathcal{WT}$ in a 2IFC experiment.

The ROC curves for the tone–in–noise versus noise–in–noise experiment were well fitted by the traditional normal–normal equal variance model (they did not show if the Rayleigh model fitted), but the empirical psychometric functions were definitely not well fitted by the normal model. Instead they found that the human data were well described by the $\mathcal{WT}$=1 psychometric functions, for tone–in–noise and noise–in–noise, except for an attenuation factor. John Whitmore has kindly let me reanalyse his data. Figure 1.3 shows these data transformed from $\mathcal{A}_{\mathrm{SIFC}}$ and $P(\mathrm{C})_{\mathrm{2IFC}}$ into d' compared with the theoretical functions. The SIFC empirical psychometric functions are less noisy, because the data were collected using a rating scale experiment. This decreases the error when calculating $\mathcal{A}_{\mathrm{SIFC}}$ using the trapezoidal rule.

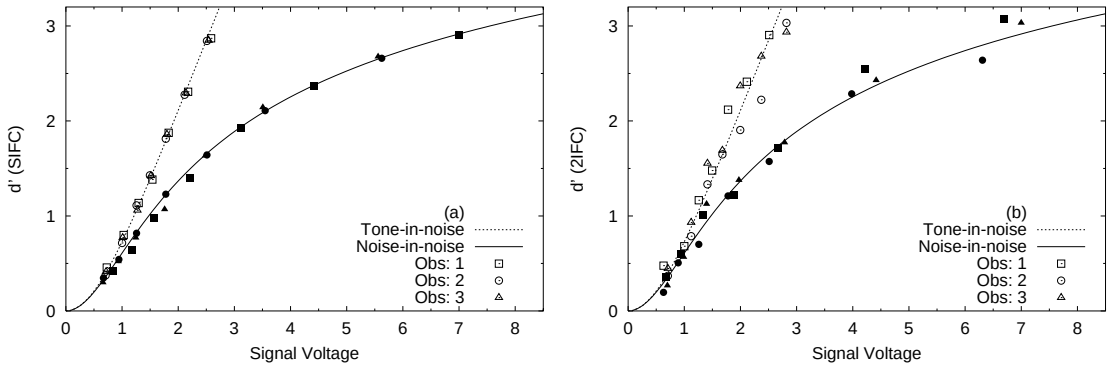

Figure 1.3: Psychometric functions of d' as a function of signal voltage (assuming σ_{N}=1) for each observer in (a) the SIFC rating experiment and (b) the 2IFC binary–choice experiment in Whitmore et al. (1968). Psychometric functions for each observer are translated with a multiplicative constant until they fall on the theoretical functions (dashed line and open symbols are tone–in–noise, solid line and closed symbols are noise–in–noise).

Whitmore (1969) ran a 2IFC noise–in–noise experiment, where the duration was varied for the same bandwidth, to produce narrow–band noise signals with nominal values of $\mathcal{WT}$=1/2, 1, and 2. A surprising result was that all the empirical psychometric functions were best fitted by $\mathcal{WT}$=1, regardless of whether the signals were $\mathcal{WT}$=1/2, 1, or 2. This suggested that the human hearing system could not accommodate the changes in duration—perhaps implying that it used a fixed bandwidth and a fixed integration time—regardless of the signal. This finding was not consistent with the literature, but Whitmore (1969) did not offer a reason why.

John Whitmore kindly allowed me to reanalyse his data from this experiment as well. Plotted in Figure 1.4 are the theoretical psychometric functions for each $\mathcal{WT}$, attenuated to best fit the data for each observer. It does indeed appear to be the case that $\mathcal{WT}$=1 is the best fitting model for all data sets. The data set for the longest duration (representing a nominal $\mathcal{WT}$=2) fall between the theoretical for $\mathcal{WT}$=1/2 and 1. This is strange because it theoretically should have resulted in the steepest psychometric function.

Some computer simulations were run, using SIM IO (Chapter 4; Lapsley Miller, 1998c) to see if the signal generation method Whitmore (1969) used could result in signals with the nominal parameters of $\mathcal{WT}$=1/2, 1, and 2. No evidence was found that the technique used by Whitmore (1969) could have resulted in signals with $\mathcal{WT} < 1$. This is backed up by Laming (1986) who said that $\mathcal{WT}$=1/2 existed only as a mathematical formula and did not have any physical interpretation. The most likely explanation was that when the signal duration was reduced, using a rectangular gate, the frequency spread, resulting in a waveform with $\mathcal{WT} > 1$. Whitmore then

filtered this short transient to shape it in the time domain. Because the signal was so short, the filter possibly warped the signal even more. This occurred because filters and gates are only capable of resolving a waveform to a minimum $\mathcal{WT}$, which is generally greater than unity. For the longer duration signals, the filter–shaping procedure may have reduced its effective duration. Indications from the simulations were that all the signals had a very similar $\mathcal{WT}$, between 1 and 2, but with different bandwidths and durations.

If the bandwidths of the observers' auditory filters were dependent on duration, then the longer duration condition may have been filtered using a narrower filter, and, therefore, reduced the effective $\mathcal{WT}$. This would also decrease the signal–to–noise ratio. In support of this possibility, the attenuation for this condition was about 5 dB greater than for the 17.8 ms ($\mathcal{WT}$=1) condition. Further computer simulations, this time modelling non-ideal energy and envelope detectors, using parameters estimated from a human observer in the current project, resulted in somewhat similar results. In particular, one combination of parameters did result in a shallower psychometric function for the long duration signals. The purpose of the simulations was to show there was at least one other possible set of circumstances that could give Whitmore's (1969) results, aside from the ear operating with a fixed $\mathcal{WT}$.

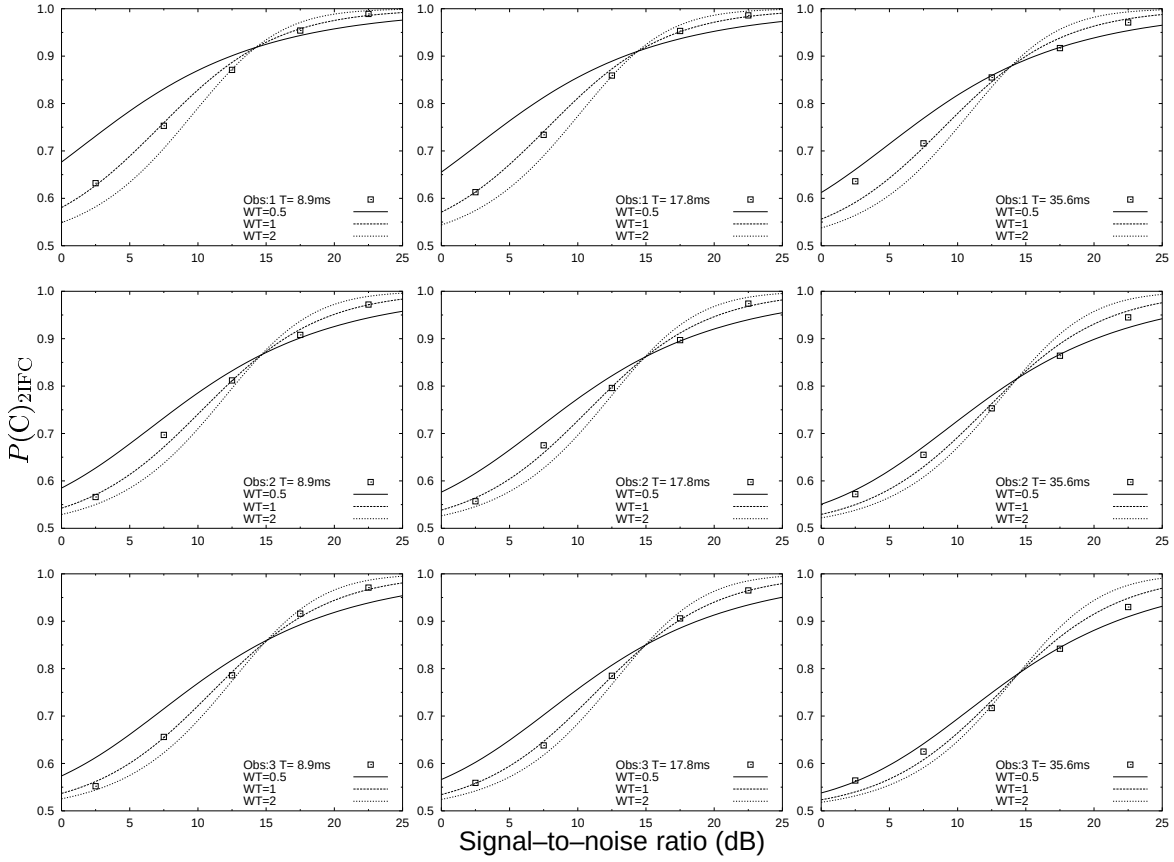

Figure 1.4: Attenuated theoretical psychometric functions, for each observer, and for three different signal durations, from Whitmore (1969).

1.4.1 Stimulus–level analysis

Green (1964a) made a distinction between what he called *molar* psychophysics and *molecular* psychophysics. Molar psychophysics uses averages of sets of responses to large numbers of stim-

uli, whereas molecular psychophysics studies performance on a stimulus–by–stimulus basis. He described the advantages of the molecular approach as including the ability to test a hypothesis (a) that may not predict a difference in detectability at the molar level, or (b) when not enough data can be collected to test at a molar level. Gilkey (1981) and Gilkey and Robinson (1986) added that the molar approach to psychophysics assumes that different models can be distinguished solely on the shape of their ROC curves and psychometric functions. But, Gilkey argued, this assumption does not hold up because (a) differences between the evidence distributions may be small, (b) variability in human data swamp small differences between models, (c) ROC curves and psychometric functions are not necessarily sensitive to changes in distribution shape, (d) two different distributions may produce the same ROC curve, and (e) that many models perform at a much higher level than humans. He suggested that a molecular approach may be better able to address these issues.

Green (1964a) thought that the problem facing molecular psychophysics was that observers were not consistent in their responses, when presented with the same stimulus. He tested some of his predictions experimentally using taped waveforms so observer consistency could be measured, and estimated that the ratio of internal to external noise was unity. Molecular analysis required, according to Green, that the effects of non-stimulus variables on the response must be predictable. He believed that predictability was limited by internal noise, because its influence could not be directly measured. An alternative approach is to remove these effects. To do this, stimuli are presented multiple times and responses are averaged across the same stimulus to remove error (Ahumada, 1967; Gilkey, 1981; Gilkey & Robinson, 1986). The responses can then be compared with responses from an ideal observer—usually a computer simulation—that has been presented the same stimuli. By using simulations, the model can be tweaked until both the simulation and the human are responding in a similar way.

Sherwin et al. (1956) were the first to construct an electronic–analog–model (EAM) simulation for human hearing. Their true–power detector consisted of a band–pass filter and a square–law rectifier, followed by a leaky integrator with various time constants. Sherwin et al. (1956) attempted to measure hit rates and false–alarm rates, but, because they did not have well defined observation intervals, they had to guess when $\mathcal{N}$ events occurred. Although they talked about correlations, in fact they built up histograms of responses for both the EAM, and the humans, and compared the hit and false–alarm rates. They found that although the EAM and humans had a similar hit rate (because the criterion for both was set to achieve this aim), the false–alarm rate for the humans was half that of the EAM. They concluded that their model was an appropriate one for human hearing, but their filter bandwidth was probably too wide.

Pfafflin and Mathews (1966) used digital computers to generate sets of reproducible noise—some of which had tones added. Prior to this time, noise waveforms were irreproducible, or taped, which had unacceptable noise and distortion characteristics on playback. Spectral analysis was performed on the waveforms so that the energy could be calculated within a 100 Hz width around the tone. The relative energy difference was calculated between pairs of stimuli, and this statistic was used as the independent variable. Human experiments using the same stimuli were run repeatedly and the proportion of correct detections for each stimulus pair was plotted against the energy statistic. Results from the different conditions all indicated that $P(C)$ increased with increasing energy difference, thereby discounting the cross–correlation detector, but supporting the energy detector.

Ahumada's filter bank theory

Ahumada (1967) varied the filter bandwidth of an EAM energy detector to establish which bandwidth had the highest correlation with human performance for the same tone–in–noise stimuli. He repeated the experiment four times using the same stimuli and calculated the sum–of–ratings for each stimuli over each, and all, observers. The sum–of–ratings for each stimuli was then correlated with the energy–statistic from the EAM. The stimuli were 500 Hz tonal transients.

He estimated the energy passed by the EAM for each stimuli, and the sum–of–ratings to a "Yes"–"No" rating scale for each stimuli. He then calculated hit and false–alarm rates and d'_e. [12] Two observers performed worse than the EAM with no filter, and the other two performed as well as the EAM with a bandwidth of around 100 Hz. Ahumada used only 33 $\mathcal{SN}$ stimuli and 27 $\mathcal{N}$ stimuli, therefore, sampling variability was high.

Ahumada found the curious result that human data, for $\mathcal{SN}$ stimuli, were best correlated with an EAM bandwidth of about 40 Hz, but the $\mathcal{N}$ stimuli were best correlated with a much wider bandwidth. Similar results were obtained in a replication of this experiment with new stimuli (Ahumada & Lovell, 1969). Ahumada suggested this may be because the ear is better modelled by a bank of filters. For the $\mathcal{SN}$ stimuli, the bulk of the energy would always be in one filter, and thus would be more highly correlated with a narrow bandwidth. But for $\mathcal{N}$ stimuli, the best correlation would be for the bandwidth that covered the range of filters being monitored in the filter bank. He also suggested that instead of a filter–bank, there was frequency sharpening at higher signal–to–noise ratios, that is, the bandwidth of the filter decreased with increasing signal levels.

Ahumada continued the same line of investigation, but with more stimuli and with a deeper level of analysis (Ahumada & Lovell, 1971; Ahumada, Marken, & Sandusky, 1975). They found higher correlations, using an energy model, for $\mathcal{SN}$ stimuli compared with $\mathcal{N}$ stimuli, again suggesting the observers' bandwidths may be wider on $\mathcal{N}$ trials. They also found evidence that some energy bands had negative correlations for $\mathcal{SN}$ stimuli. They did a similar analysis to Ahumada (1967) by correlating the output of non-optimal energy detectors (single–tuned filter, square–law rectifier, and full integrator) with the observers' sum–of–ratings. The detectors were non-optimal, because the filter bandwidth was not matched to the signal, and the integration time was not necessarily over the entire duration of the signal. They found that (a) $\mathcal{N}$ stimuli resulted in higher correlations if the energy detector had a wide bandwidth and long integration time compared with the $\mathcal{SN}$ stimuli, (b) $\mathcal{SN}$ stimuli resulted in much higher correlations than $\mathcal{N}$ stimuli, but (c) if stimuli from both events were analysed together, then the correlations were slightly higher again. The best energy detector was the one most correlated with the $\mathcal{N}$ stimuli—a surprising result given the $\mathcal{SN}$ stimuli had higher correlations, but some of the differences in the correlations were tiny.

They then calculated FFTs of subsections of the stimuli (in a time and frequency matrix) and correlated the energy in each region with the observer sum–of–ratings, over eight replications. They found negative correlations in the regions outside of the signal in time and frequency for $\mathcal{SN}$ stimuli, but not for $\mathcal{N}$ stimuli. This implied a lateral suppression mechanism. They also found the extremely wide bandwidth estimates for $\mathcal{N}$ trials only occurred if the noise was gated, rather than continuous. They surmised that the observer searches for a pattern of temporal and spectral changes, rather than just an overall increase in levels, by negatively weighting the information

[12] d'_e is a measure of detectability estimated from the negative diagonal of an ROC curve (Green & Swets, 1966). It is equal to d' if the underlying distributions are normal with equal variance.

preceding the tone in time, and outside of the frequency band where a tone could be presented. Thus, the filter bandwidths and integration times for $\mathcal{N}$ stimuli were wider and longer than for $\mathcal{SN}$ stimuli, because the observer was searching for these changes through time and frequency.

Wilcox (1968) followed up Ahumada's (1967) suggestion of modelling tone–in–noise detection with a filter bank and energy detector, and compared it to the linear–uncertain model, which Wilcox had derived. The idea was to model the uncertainty of the observer as to the exact form of the signal to be detected. The model consisted of a weighted sum of a linear and a quadratic function of the input waveform. Special cases were the energy and envelope detectors.

The linear–uncertain model assumed large $\mathcal{WT}$, and incorporated observer uncertainty using an internal noise model. As uncertainty increased, if the internal specification of the signal was poor, the observer behaved more like an energy detector; as uncertainty decreased, the observer behaved as a pure linear (SKE) detector. An important difference between the models was that the correlation between a noisy linear–uncertain observer and an electronic energy detector that received only noise, was never less on $\mathcal{N}$ trials than on $\mathcal{SN}$ trials. This was the opposite of what was predicted by the filter–bank model.

Wilcox (1968) compared the filter–bank and the linear–uncertain models by correlating the human responses with the output of computer simulations. He found that the linear–uncertain model, and its special cases, could not account for the human data for monaural experiments. Ahumada's (1967) filter–bank model, however, was better able to account for the human results, though not in all cases.

Wilcox (1968) also compared the performance of an electronic cross–correlation detector, an electronic energy detector, and three human observers in a tone–in–noise task. There was no correlation between the human observers and the cross–correlation model. Correlations were higher for $\mathcal{SN}$ trials, compared with $\mathcal{N}$ trials, for human observers and the energy detector, thereby giving support to the filter–bank model. The filter–bank model could not predict human performance when the signal presentation interval was shortened, but Wilcox did not take frequency spread into account.

There appears to be some empirical support for the filter bank model, however, it is possible the evidence may be a result of interactions between a filter with a fixed bandwidth (for $\mathcal{N}$ and $\mathcal{SN}$ signals), and the $\mathcal{SN}$ signal bandwidth and spectrum shape. Also, only a few filter bandwidths were usually tested in these simulations, leaving the possibility that the maximum correlated bandwidth is somewhere between the tested values. Chapters 4 and 5 follow up this alternative scenario in more detail.

Gilkey's comparison of tone–in–noise detectors

Gilkey modelled human tone–in–noise detection with computer simulations (Gilkey, 1981; Gilkey, Robinson, & Hanna, 1985; Gilkey & Robinson, 1986). He hoped that using molecular psychophysical methods to study human hearing would enable him to make distinctions between competing models that predicted the same or similar performance at a molar level. He used a quasi–molecular approach where the same stimulus was presented on several trials and the average response calculated and compared to various computational detectors. By averaging responses for the same stimulus he was able to reduce internal noise such that the parameters of the stimulus could be used to predict performance. His approach was influenced by molecular and quasi–molecular research of Ahumada (1967), Ahumada and Lovell (1971), Ahumada et al.

(1975), Green (1964b), Pfafflin and Mathews (1966), Sherwin et al. (1956), Watson (1962 / 1963) and Williams and Jeffress (1967).

Gilkey observed that all these previous studies assumed only one type of model—usually the energy detector—so he compared empirical results from human observers with envelope, energy, and other detectors. To do this he used a flexible computer simulation for tone–in–noise detection, where the signal duration was shorter than the noise duration, and the signal bandwidth was narrower than the noise bandwidth. He used a narrow–band single–tuned digital filter, a half–wave or square–law rectifier, a leaky integrator with variable time constant, and various sampling strategies (average over the signal or noise duration, end of signal or noise duration, or maximum over noise duration). By choosing different combinations, different detectors were implemented.

The human data were collected from a binary–choice, SIFC, tone–in–noise experiment using reproducible stimuli presented multiple times at two different signal levels (Gilkey, 1981; Gilkey et al., 1985). The number of stimuli per event was 25, but each stimulus was presented between 78 and 198 times. Thus, at the molar level, sampling variability, due to the small number of stimuli per event, would be so huge as to negate any conclusive analysis, but at the molecular level, the reduction of noise greatly enhanced the ability to predict stimulus–by–stimulus performance.

The measure of detectability used for the human experiment was the proportion of "Yes" decisions for each, individual, noise stimulus, with and without the addition of a tone. The hit rate (with the tone) or false–alarm rate (without the tone) for each noise signal could then be calculated and plotted as a single point in the ROC space. The decision variable of the simulation was converted to the predicted proportion of "Yes" decisions by using a logistic approximation to the cumulative normal. The least–squares fit of the obtained proportion of "Yes" decisions and the predicted proportion of "Yes" decisions, was used to get a goodness–of–fit statistic.

As expected, the results of the molar analysis on the human data were hugely variable. Gilkey presented the mean ROC results, but did not use the sum–of–ratings to generate GOC curves (see Chapter 2). He then compared the molar ROC points with the molecular ROC points (for each stimulus) and noted the huge variability of the latter. He suggested that the average ROC point did not summarise the data. This is, presumably, because he compared error degraded points to points where error had been removed (see Chapter 2). If the observer had no internal noise then the ROC points for each stimulus would fall only on corners of the ROC space. Instead the points were spread throughout the ROC space.

Gilkey compared virtually every combination of filter bandwidth, rectifier, integrator time constant, and sampling strategy with the human data, and the $\mathcal{N}$ and $\mathcal{SN}$ events were analysed together. The results showed (a) an interaction between type of integration and type of sampling strategy, (b) that the best correlation was when the bandwidth was around 50 Hz, and (c) that half–wave rectification was more effective at predicting the results than square–law rectification. The two best used integrators and sampling strategies that either (a) had a short integrator time constant and then averaged over the integrator output, or (b) had a long integrator time constant and took the maximum of the integrator output.

Gilkey then focused on analysing specific models from the literature. The envelope detector consisted of a narrow filter, half–wave rectifier, leaky integrator with a short time constant (1 ms), and sampling at the end of the signal interval. The leaky integrator detector differed from the envelope detector by having a longer time constant for the integrator of 50–100 ms. The energy detector consisted of a filter, square–law rectifier, leaky integration with a very short time constant, and average sampling. This approximated true integration.

Results from the envelope detector with a short time constant did not agree at all with the experimental results. Better fits were obtained by changing the sampling strategy of the integrator to coincide with the end of the signal duration, or to average the integrator output over the signal duration. Results from the leaky integrator detector were improved by changing the sampling strategy from the end of the signal to the maximum value over the entire duration of the noise masker. The energy detector's performance was also comparable to the other simulated detectors.

Gilkey also found that the performance of two other detectors was more correlated with the performance of the human observers than any of the other standard detectors. Model 1 was better than the energy detector (which differed only in the form of the rectification) and the envelope detector (which differed only in the time constant). It used a 50 Hz band–pass filter, a half–wave rectifier, a leaky integrator with a 0.1 ms time constant, and sampling strategy of the average value of the integrator output over the signal duration (thereby approximating true integration). Model 2 consisted of a 50 Hz filter, half–wave rectifier, leaky integrator with a 200 ms time constant, and a sampling strategy of taking the maximum output over the masker duration.

The performance of the two best detectors was very similar. Gilkey did not think this was surprising, because his research showed high correlations between the predictions from these different models. None of the results of the goodness–of–fit tests with the human data, however, were very high. Gilkey acknowledged this, so performed split–half correlation tests, which indicated that much of the human variability was potentially explainable. Therefore, he looked at multiple detector models where spectral and temporal weighting functions were calculated. If only the part of the signal around the centre frequency and within the signal duration was used, then the other spectral or temporal weights would be zero. They were not, in fact some were negative, suggesting a process analogous to lateral inhibition or suppression.

A further finding was that changes in the filter bandwidth did not markedly change the correlation between observers. Gilkey suggested that the human critical band may vary from trial–to–trial. Using simulations, however, Gilkey showed that even a fixed bandwidth model could produce similar results. Further analysis, using the best fitting simulation to test patterns of correlations against other models, indicated that unless the model fitted the data perfectly, there would likely be a difference found between $\mathcal{SN}$ and $\mathcal{N}$ events. Particularly that $\mathcal{SN}$ filter bandwidths would be underestimated and $\mathcal{N}$ bandwidths overestimated.

1.4.2 Experimental evaluations of the simple detection models

A number of studies have shown that the ideal energy or envelope detector models are not sufficient to explain human hearing. On one hand they appear too simple because they are deaf to fine details in acoustic waveforms and, on the other hand, too complex because they are able to adjust their bandwidths and time constants to optimally fit the signal. This is not surprising, because the original aim was to develop parsimonious explanations for the simplest detection problems. This section covers a variety of studies that focused on the detectability of complex or Gaussian noise signals by humans, and considers whether the simple detection models are sufficient to explain the experimental findings.

Studies that manipulated both $\mathcal{W}$ and $\mathcal{T}$

Creelman (1961) was interested in studying the combined effect of $\mathcal{W}$ and $\mathcal{T}$ on detection. The signals he chose were repeated damped sinusoids with a damping period of 10 ms, durations that

were multiples of this period (10–200 ms), and five damping rates. The pre-damped frequency was 500 Hz, but he does not specify the frequency characteristics of the actual waveforms presented to the observers. By examining Creelman's graphs, the $\mathcal{W}$–$\mathcal{T}$ matrix of conditions and results can be created (studying $\mathcal{WT}$ was not the focus of his study). There were only a few conditions where the same $\mathcal{WT}$ was used. The first experiment resulted in two observers producing the same performance for the two $\mathcal{WT}$=1 and $\mathcal{WT}$=2 conditions, but $\mathcal{WT}$ was not traded in the larger $\mathcal{WT}$ conditions. The third observer was more variable. In the second experiment, constant performance was found for $\mathcal{WT}$=2 and 16 for one observer, and $\mathcal{WT}$=2 and 4 for another observer; otherwise performance decreased as duration increased and bandwidth decreased.

Michaels (1961) suggested that an envelope detector would be a good model for his intensity discrimination experiment using narrow–band, short–duration, noise. He had sixteen conditions of four bandwidths (0.5, 2, 8, and 32 TdB Hz) and four durations (32, 128, 512, and 2048 ms). It is unlikely the resulting signals had these properties because it is extremely difficult to generate signals with $\mathcal{WT} < 1$. He implicitly claims to have $\mathcal{WT}$=0.016, 0.064, and 0.0256, but this is prior to creating the transients with a relay system. This would have increased the actual bandwidth presented to the observers considerably. Michaels did not analyse his data with respect to $\mathcal{WT}$, however, it is possible to pick out $\mathcal{WT}$ combinations for comparison. The conditions with $\mathcal{WT} < 1$ and $\mathcal{WT} > 16$ were ignored, because they did not include enough combinations for comparison. The conditions for $\mathcal{WT}$=1 and 4, indicated however, that the difference limen tended to decrease with increasing duration, and to decrease with increasing $\mathcal{WT}$. These patterns are only apparent when averaged across observers.

Majerník, Konečný, and Nehnevaj (1978) studied detection of noise signals by varying both bandwidth and duration. They used a Same–Different task, but although the two stimuli in a trial were contiguous, there was no inter–stimulus interval. Therefore, like Marill (1956), there was a confounding of forward and backward masking. The transients were gated with a rectangular gate, then filtered with bandwidths equal to integer critical bands, but the way the critical bandwidths were measured was not mentioned. Majerník et al. (1978) found a strong duration effect, where short durations had much larger difference limens than long durations. This was interpreted as due to the increase in $\mathcal{WT}$ as duration increased. The difference between the one–critical band and three–critical band stimuli was a small constant difference in the difference limen over all durations, with no interaction. They also found a small effect at short durations indicating that the observers' critical bandwidth may have widened.

Raab and Goldberg (1975) studied intensity discrimination of reproducible noise as a function of intensity, bandwidth, and duration. Their aim was to compare three models of intensity discrimination (Green's (1960a) energy detector, McGill's (1967, 1971) counting model, and Penner's (1972) counting model) with the performance of real observers. The three models all predicted bandwidth–duration reciprocity, but they found that there was no reciprocity for the conditions they studied, and concluded disparagingly that the results were "...an embarrassment for all three models..." (Raab & Goldberg, 1975, p. 442). Bandwidth and duration were not treated symmetrically by the auditory system, with results being largely independent of bandwidth, but dependent on duration.

The values of $\mathcal{WT}$, however, were all very large (5, 50, 158, and 500) and may have exceeded the critical band or temporal integration time of the ear. Their conclusions about bandwidth–duration reciprocity were also made only for $\mathcal{WT}$=50, and only for two combinations: $\mathcal{W}$=5000 Hz and $\mathcal{T}$=10 ms, and $\mathcal{W}$=500 Hz and $\mathcal{T}$=100 ms. Despite these issues, it does seem that there is some

evidence that bandwidth and duration were treated differently by the hearing system, because although the noise signals were all equal in energy, they did not result in equal detectability.

Hanna (1984) also looked at reproducible noise as a function of bandwidth and duration. He hypothesised that the discrimination of two noise bursts should increase as either bandwidth or duration increased (while keeping signal–to–noise ratio equal; energy increases). Hanna found that performance increased as duration increased up to 25 ms, but that an increase in bandwidth increased performance only slightly. For the very short duration signals (0.1 ms–6.4 ms), the nominal $\mathcal{WT}$ was 0.005, 0.02, 0.08, and 0.32. The problem was that duration was defined before filtering. After filtering, the effective $\mathcal{WT}$ would be considerably bigger; at least unity. Hanna also used the normal model, which is inappropriate for small $\mathcal{WT}$. Therefore, Hanna's conclusions are unfounded.

van den Brink and Houtgast (1990) considered spectro–temporal integration of windowed tones, and other more complex signals, by estimating temporal integration as a function of bandwidth, and spectral integration as a function of duration. Their first experiment indicated that the long–duration, wide–band signal was the most difficult to detect and that the short–duration, narrow–band signal was the easiest. The second experiment considered six different combinations of $\mathcal{WT}=1$ band–pass, impulse response, signals. They found that as bandwidth increased, and duration decreased, that the masked threshold increased. It appeared that for the three narrow–band and long–duration signals that the masked threshold was about the same. The final experiment tested the two extremes from the second experiment by widening the bandwidth of the narrow–bandwidth long–duration condition and lengthening the duration of the short–duration wide–band condition. They achieved these signals by grouping windowed tone components. The overall effect was the required spread in duration and frequency (either by randomly choosing or sweeping the position of the tones). They hoped to measure the critical bandwidth for temporal integration and the critical time window for spectral integration. For the fixed duration condition, they found that once the bandwidth increased beyond 1/3 octave that thresholds increased. For the fixed bandwidth condition they found a critical masking interval of about 30 ms. They found that efficient integration occurred if the signal was restricted to 1/3 octave bandwidth and was within 30 ms. Integration still occurred for signals outside of this region, but it was less efficient. Their results are difficult to interpret, because they did not use a theoretical observer to compare performance for these complex signals, although they assume that energy is the appropriate quantity for detection.

Formby, Heinz, Luna, and Shaheen (1994) used noise bursts to study detection of speech–like sounds. They were particularly interested in how noise signals were integrated over time. They used Green's (1960a) model of noise detection, but modified it to include the case when the signal bandwidth was narrower than a critical band. They compared fixed and random level gated maskers and tested detectability of noise signals with bandwidths from 62 to 4000 Hz (in octave steps), centred at 2500 Hz, and durations from 10 to 480 ms. They did not say how and when bandwidth was measured. It is likely bandwidth was measured prior to gating (with a raised cosine window), thus their estimates of $\mathcal{WT}$ would be wrong—especially for the short duration signals. The condition of 10 ms and 62 Hz would probably have a bandwidth over 100 Hz, for instance. Unfortunately, without this information it is difficult to interpret their findings. In general, they found evidence of a critical band of around 500 Hz. They also found that the human hearing system does not act like an ideal constant–$\mathcal{WT}$ detector, because performance is attenuated as duration increases and as bandwidth increases, but with no major interaction between

them. A further problem is in using Green's model, which is only applicable to large $\mathcal{WT}$, about 30% of their conditions had a $\mathcal{WT}$ nominally smaller than this requirement. Thus, their multi–cue model is suspect because it is based on faulty premises.

Adjustable critical bands revisited

In extending the work by Raab and Goldberg (1975), Schacknow and Raab (1976) found evidence that the critical band was flexible, but it did not match the signal bandwidth. They developed Green's (1960a) large–$\mathcal{WT}$ noise detection model to account for the case where the signal bandwidth was narrower than the masker bandwidth. This extension, however, relied on a normal approximation, and was, therefore, valid only for large $\mathcal{WT}$. Schacknow and Raab (1976) realised this, but still felt obliged to offer many other reasons as to why human observers did better than the theory in the $\mathcal{WT}$=10 condition.

Bernstein and Raab (1990) further extended the work of Schacknow and Raab (1976) by considering the effect on critical band estimates as signal bandwidth (but not duration) was varied. They used a modified energy detector to compare results where the energy detector's performance was degraded due to internal noise. From experiments using tones and noises in gated and continuous maskers, they found that the critical bandwidth was rather wide (on the order of hundreds of hertz) and varied as the signal bandwidth varied. The critical band did not, however, vary to *match* the bandwidth of the signal. It tended to always be wider than necessary. The critical band estimates were also narrower in the continuous masker condition. Bernstein and Raab (1990) suggested that this may be due to off–frequency listening, which could only be beneficial with continuous maskers.

Bernstein and Raab (1990) interpret their findings as indicating that the filtering inherent in the human hearing system is adjustable. They suggest that the term "detection bandwidth" should be used to refer to the internal filter bandwidth estimates, and that the term "critical bandwidth" should, therefore, be reserved for the smallest estimate of the detection bandwidth. They also suggest that the wide detection bandwidths are made up by summing across critical bandwidths. The reason why they are wider than the signal bandwidth, they postulate, is to enable some form of lateral suppression or inhibition perhaps similar to that postulated by Houtgast (1974).

Bernstein and Raab (1990) consider only Green's (1960a) energy detector model. This model is only suitable for large $\mathcal{WT}$ and small signal–to–noise ratio. They claim that the Gaussian approximation is not badly violated by small $\mathcal{WT}$, but they give no reason why they do not use the χ^2 model instead.

Models incorporating internal noise

de Boer (1966) extended Green's (1960a) energy detector model for noise signals, by including the concept of internal noise. He stated that although frequency resolution and temporal integration are modelled as different processes, the slowly modulating temporal aspects of narrow–band noise maskers detract from locking onto the appropriate frequency range. This makes separating out the effects of time and frequency processes difficult. Instead, Boer suggested that if narrow–band noise signals and wide band noise maskers were used, frequency resolution was easily shown. At the time it was not well known how the auditory system dealt with wide–band noise signals, so this study compared performance with both types of signals.

de Boer (1966) derived an average intensity detector, which was essentially the same as the

energy detector of Green (1960a), except that it was not applicable for very large $\mathcal{WT}$ (but did not specify how large), as it implied too fine a discrimination compared with experimental evidence. de Boer also included internal noise as a way of accounting for the human experimental data. The statistics of his model were based on normal distributions, with either equal or unequal variance. His data analysis was somewhat dubious, because it relied on a number of arbitrary assumptions relating to assumed integration times, and a threshold difference limen based on d'. His analysis also assumed that the human hearing system matched the signal bandwidth.

de Boer (1966) tried to explain both his own data and those from the literature that gave rise to psychometric functions with different slopes, but again the analysis assumed a normal model. He also suggested that for wide–band signals, the critical band is 'turned–off', and that for short duration signals, the temporal integrator is switched off too. Overall, his data are consistent with his theory for long duration noise, but not for short noise bursts.

Henning (1967) also considered a different energy detector; one that did not predict perfect performance for very large signal–to–noise ratios and arbitrarily small differences in frequency (in a frequency discrimination task). The resulting model slowly varied the centre of the filter such that it could be modelled as a random variable. This variation is a form of internal noise. Henning was able to account for some psychophysical data for both frequency and amplitude discrimination, but could not account for amplitude detection. His data were so noisy that it was unclear if the model was useful or not.

Ronken (1969) considered the detectability of Rayleigh noise, which he considered a compromise between tones and wide–band noise. The noise was generated as tonal transients, with each sample having a random amplitude, because Rayleigh noise was considered to have a constant amplitude over short durations. He compared his results to Green and McGill's (1970) energy detector and found only a 0.7 dB attenuation over a wide range of signal levels. His signal–to–noise ratio measure, however, was not $10\log_{10}(\sigma_{\mathrm{S}}^2/\sigma_{\mathrm{N}}^2)$; instead Ronken defined signal–to–noise ratio as $10\log_{10}(\sigma_{\mathrm{SN}}^2/\sigma_{\mathrm{N}}^2)$. The difference in attenuation is not constant with changes in signal power, but at higher signal levels the difference is minimal.

Hautus and Irwin (1992) followed up the research by Ronken (1969) and Whitmore et al. (1968), which had given contradictory results: Ronken found attenuations of human data from the theoretical of only about 1 dB, whereas Whitmore et al. found a 5–6 dB attenuation. The difference between the two studies was in the type of signal used. Whitmore et al. used narrow–band noise generated using electronic filters. Ronken used sinusoids distributed with a Rayleigh distribution. This required the assumption that the amplitude of transient Rayleigh noise is constant, but is variable *across* a set of transients. A further difference was that Whitmore et al.'s signals were 10 ms in duration and 100 Hz wide whereas Ronken's were 100 ms in duration and therefore approximately 10 Hz wide (assuming reciprocal spreading in the frequency domain). These differences aside, theory would predict approximately the same level of performance. Hautus and Irwin (1992) replicated these experiments and found that the best fitting psychometric function, averaged over three observers, was when $\mathcal{WT}=1/2$. The ROC curves, especially for the noise signals, were best fitted with a much larger $\mathcal{WT}$. Furthermore, attenuation in each experiment was on the order of 4 dB.

Distractor and probe experiments

Kaplan (1975) compared Green's (1960a) energy detector model with Ahumada's (1967) energy–contrast model by using tone distractors in either the $\mathcal{SN}$ or $\mathcal{N}$ trials of a 2IFC 500 Hz tone–in–noise experiment. His aims were to estimate the critical band by varying the distance of the distractor from the tone, and differentially test the predictions of the two energy models. One distractor was presented on each trial at an intensity 3 dB down from the observer's threshold. Distractors, which changed the threshold for the 500 Hz–200 ms tone were interpreted as indicating the width of the critical band. The estimate of the critical band was 80 Hz, but that depended on interpolation. Simulations using the two models showed that Green's model was able to account for the findings, but the energy–contrast model could not.

Moore, Hafter, and Glasberg (1996) used a probe signal method to test the auditory filter width. The results indicated that probes were more detectable if only one auditory filter was attended. (Dai, 1991; in Moore et al., 1996) made a distinction between the filter characteristics measured using the probe signal method, and that using notched noise. The former he called the "attention" bandwidth, because it represented the effective bandwidth through which the observer listened. This may be the result of the combination of multiple auditory filters.

Wright and Dai (1994b) used the probe signal method to estimate the auditory filter bandwidth, as signal duration was manipulated. The resulting probe–signal–contour (PSC) function (in decibels as a function of frequency) resembled the shape of the auditory filter for long durations. Short durations, however, resulted in much wider PSC functions than long durations, but a notched–noise experiment indicated that the auditory filter was in fact narrower than indicated by the PSC function for the short duration. Wright and Dai (1994b) interpreted this as indicating the observer is able to listen through multiple filters, but does not necessarily do so. A second probe signal experiment used probes of unexpected durations. They found that performance deteriorated for probe durations unmatched with the target duration, for both short and long target durations (5 and 295 ms). They discounted the possibility of energy splatter in the frequency domain for short durations, because the estimated bandwidths were much wider than the spread. They concluded that the observer matched the duration of the integration interval to the duration of the *expected* target signal, and that one way of viewing this is that observers listen through a flexible time–frequency window.

Wright and Dai (1994a) called the PSC function the probe–signal–contour filter, but explained that it was not a model of the peripheral auditory filter. Instead, they suggested it was perhaps a result of combining multiple auditory filters, and that this process occurred much higher up in the auditory system. Under certain circumstances the PSC filter may resemble the auditory filter shape if the observer is only listening through one channel. They also considered changes in the PSC filter as a result of continuous or gated maskers. They found that gated maskers broadened the PSC filter for long signals for some observers, and that the gated data were more variable. Their data suggested that different observers used different listening strategies in these experiments, which indicated that combining data across observers should be done cautiously.

Dai and Wright (1995) compared a multi–look model to the energy detector by getting observers to detect (a) signals where the duration was unexpected, because they were attending to signals of a fixed duration, and (b) signals where the duration was uncertain on each trial. In both cases the onset of the signal was marked. The experimental results indicated a decrement in performance in (a) as the unexpected duration varied away from the expected duration, but no

decrement as duration was varied in (b) where the duration was uncertain. This indicated that observers could selectively listen for a particular interval, or could use a flexible temporal integration interval if the situation warranted it. Predictions for the energy and multi–look models were essentially identical for these tasks. The data were fitted in case (a), but real observers did better than the models in case (b). Dai and Wright (1995) concluded that the energy detector was more parsimonious for their experiment, but there was no way of testing the two models in this context.

Dai and Wright (1996) compared predictions based on Fletcher's (1940) critical band model and Green and Swets's (1966) energy detector for tone–in–noise tasks, where duration was varied from 2 ms up to 1000 ms, and found results fell somewhere in between the two models. Because the normal assumption is not appropriate for small $\mathcal{WT}$, they modelled the data using χ^2 random number simulations. Their findings indicated that the human observers behaved like an energy detector, but not ideally.

1.4.3 Summary

Although many studies have shown that the human hearing system is capable of complex tasks, many of them have also shown that the evaluation of the simple energy and envelope detectors is still needed, because results are still contradictory.[13]

There is evidence that the hearing system can use energy or envelope cues in detection, but there is no evidence that the filtering and integrating processes of the ear are optimally matched to an arbitrary waveform. Thus, although an observer may be using energy information, the resulting performance would not be optimal. To test whether this is the case requires more subtle investigation than just comparing ROC curves or psychometric functions.

The general approach of Ahumada and Gilkey is an excellent way of assessing different models and their relationship to human performance, but what stands out is how few stimuli were used. Even if observer inconsistency was reduced by repeated presentations, the number of stimuli would result in large sampling variability. In the current study, there are enough stimuli and enough replications so both molar and quasi–molecular psychophysics can be used. Human hearing may then be tested from both perspectives.

1.5 Overview

According to Boer, models of human hearing tend to only explain time domain phenomena, or frequency domain phenomena, but not both: "This is a deplorable situation, the more so since 'meaningful' sounds necessarily include variations in frequency *and* time" (de Boer, 1985, p. 142).

The aim of this study is to estimate some properties and parameters of the human hearing system, by conducting a parametric study that compares human performance to simulated and theoretical performance of energy, full–linear, and envelope detectors. The task is amplitude discrimination of small–$\mathcal{WT}$ Gaussian noise signals, masked by wide–band Gaussian noise. Noise signals are used, because, unlike tones, it is possible to manipulate their bandwidth and duration relatively independently. The parameters $\mathcal{W}$ and $\mathcal{T}$ are varied for specific values of $\mathcal{WT}$. From

[13] Other models of human hearing have also been developed by Durlach and Braida (1969), Laming (1986), Moore (1975), Richards, Heller, and Green (1991) and Richards (1992). It is beyond the scope of the current study to examine these models in detail.

the results, estimates of the parameters of the human hearing system's filter, rectifier, integrator, and sampler are obtained.

This chapter examined the literature on the properties of the human hearing system, and how these properties have been modelled. It seems there is some evidence that the filtering and integrating processes of the hearing system may be fixed, and other evidence that they may be flexible. Assumptions, flaws, and simplifications in the mathematical models of energy and envelope detection were pointed out. These models incorporate the bandwidth–duration product $\mathcal{WT}$, which implies that detectability of a waveform is dependent on the product, but not on the individual values of $\mathcal{W}$ and $\mathcal{T}$. This is counterintuitive when it comes to human experience and the human data in the literature.

Most of the problems with the mathematical models come about from the way the acoustic waveforms are represented mathematically. Section 1.1.3 and Appendix A looked more closely at the waveform representation literature and found that the traditional Fourier integral approach is somewhat limited, and that there are generalisations to the Fourier integral that could result in better models. The acoustical uncertainty principle is directly related to $\mathcal{WT}$, such that $\mathcal{WT}$ has a minimum value, and is dependent on how $\mathcal{W}$ and $\mathcal{T}$ are defined. This problem of definition underpins much of the controversy about the mathematical models and waveform representation.

There is still a number of holes in the literature. In particular, it is not clear how, or even if, the human hearing system is able to trade off bandwidth for duration to optimise signal detectability. This project puts the theories to the test with computer simulation and human experimentation. It can be seen as a three–way comparison among the mathematical models, the computational models, and the data from human experiments. It is expected that human performance will not be well modelled by the ideal energy and envelope detectors over all conditions and levels. The following chapters present the details and results of these comparisons.

Firstly, the problem of observer inconsistency is explained, and the new analyses for dealing with this problem are introduced in Chapter 2. Chapter 3 describes the experimental methodology, and the generation of essentially band and time limited Gaussian noise signals with the desired $\mathcal{W}$, $\mathcal{T}$, $\mathcal{WT}$, and signal–to–noise ratio. The experiment is replicated six times to allow estimation of asymptotic error–free performance. Results from these analyses are then compared with the linearly attenuated mathematical energy detector.

Chapter 4 focuses on computational simulations. The aim is to establish whether the mathematical models described in this chapter are good models for the type of essentially band and time limited Gaussian waveforms used in the experiment. The results of energy, full–linear, and envelope detector simulations are presented. The pattern of the attenuation of human performance from theory is then examined by running hundreds of sub–optimal energy and envelope detector simulations and correlating the simulation evidence with the sum–of–ratings from the human experiment. These results are presented in Chapter 5.

Finally, in Chapter 6, Scurfield's (1995, 1996) multi–event information–theory measure of detectability, $\mathcal{D}_n$, is used to compare asymptotic human performance to the simulated optimal full–linear detector. $\mathcal{D}_n$ allows relative comparisons so performance is described in terms of the proportion relative to the ideal observer. Contour plots of $\mathcal{WT}$–detectability space are presented that show how the human hearing system trades off time and frequency information.

Chapter 2

Unique and common noise

"Like warmed-up cabbage served at each repast,
The repetition kills the wretch at last."

(Decimus Junius Juvenalis) Juvenal.

.....

Variability in the responses of observers in psychophysical experiments, which have been repeated multiple times with the same stimuli, may be categorised into *unique noise* and *common noise*. Unique noise is due to random perturbations in the observer, and its environment, on each presentation of the same stimulus. This noise causes the observer to make inconsistent decisions about the same stimulus. Common noise is the result of random perturbations of the events (and is why the evidence for each event is represented as a random variable), that are the same across repeated presentations of the stimuli. If an experiment is not repeated then the effects of unique and common noise are indistinguishable.

According to Taylor, Boven, and Whitmore (1991), common noise is the complement of unique noise. It is the statistical component that is consistent from replication to replication, and ideally comes from using identical stimuli in each replication. If only a subset of possible stimuli are used in an experiment (which is the usual case), then the common noise component is affected by sampling variability, and the observer's performance will not necessarily reflect the underlying population statistics, even if unique noise is removed. Both unique noise and sampling variability make it difficult to interpret psychophysical results, therefore, this chapter focuses on methods for reducing unique noise and common noise sampling variability in psychophysical data.

2.1 Unique noise variability

Observer inconsistency is the biggest problem still facing experimental psychophysics. As Green and Luce (1974, pp. 373) stated: "Perhaps the single most pervasive characteristic of psychophysical data is the inconsistency of subjects when answering most questions we ask them about simple stimuli". If observers were consistent, their ratings (decisions) would be identical for each presentation of the same stimulus. Observers, however, are not consistent, and their ratings vary considerably across repeated presentations.

To deal with the effects of observer inconsistency, either the error must become part of the model, by explicitly describing all forms of noise, or the error must be eliminated. Observer inconsistency has generally been characterised by using the concepts of *internal noise* and *external noise* where the emphasis was on the *sources* or *causes* of noise, or error, in experimental tasks (e.g., de Boer, 1966; Green, 1964a; Pfafflin & Mathews, 1962; Raab & Goldberg, 1975). Drga (1988) says, however, that these concepts are neither mathematically precise, nor mutually exclusive.

Boven (1976), Taylor (1984) and Taylor et al. (1991) replaced the concepts of internal and external noise with those of *unique* and *common* noise, which emphasised the statistical components or *effects* of error rather than sources of error. For any single replication of an experiment, the effects of common and unique noise are inseparable, but when more than one replication is run, ratings can be averaged across replications for the *same* stimulus. Averaging removes some of the unique noise while retaining common noise. This process, when applied to ROC analysis, is GOC analysis.

GOC analysis was introduced by Watson (1962 / 1963). The technique has been used by Ahumada (Ahumada, 1967; Ahumada & Lovell, 1969, 1971; Ahumada et al., 1975) in tone–in–noise detection experiments; Boven (1976) for studying observer inconsistency in multiple–observer experiments; McAulay (1978) in frequency discrimination experiments; Taylor (1984) in studying the auditory psychophysics of birds; Metz and Shen (1992) in medical diagnostics; Galvin, Podd, Drga, and Whitmore (1999) and Whitmore and Galvin (1993) for comparing the Type I and Type II tasks; Lapsley, Scurfield, Drga, Galvin, and Whitmore (1993) and Lapsley Miller et al. (1999) for studying whether $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{\mathrm{2IFC}}$ for both discrete and continuous random variables; and Drga (1999) and Whitmore et al. (1993) for studying amplitude discrimination. Taylor et al. (1991) summarised the earlier research and provided a comprehensive coverage of applying GOC analysis. Recently, Drga (1993a, 1993b, 1995, 1999) has reported on his development of the *theory* of GOC analysis, and the development of FORCE analysis, which can predict asymptotic performance.

The results of GOC analysis—from experiments on humans and other animals, and computer simulations—show that data from any single replication of an experiment could not support any theory with confidence. This is because unique noise can change the shape and attenuate an ROC curve or psychometric function. Observer inconsistency does not just affect psychophysical data; all psychological data suffers from error due to observer inconsistency.

2.1.1 Group operating characteristic (GOC) analysis

GOC analysis can be viewed as removing error in the *ordering* of the stimuli, where the ordering is implied by the observer's rating. Because of error, the ratings for each stimulus can be modelled as a random variable. The statistical principle on which GOC is based is *stochastic ordering* (Drga, 1999). If X and Y are the random variables of the ratings associated with two stimuli, then X and Y are stochastically ordered if the cumulative distribution function (*c.d.f.*) of X is greater than or equal to the *c.d.f.* of Y. That is, if $F_X(t) \geqslant F_Y(t)$, for all *t*, and $F_X(t) > F_Y(t)$ for some non-zero interval, then X is stochastically less than Y (Drga, 1999). By using stochastic ordering, the GOC curve can be shown to be based on the expected stimulus ordering of the unique–noise degraded stimuli, as replications are added and variability due to unique noise removed. For example, for the j^{th} stimulus, the common noise component is x_j. If this component is degraded by unique noise, represented by the random variable U_{ji}, the unique–noise degraded

stimulus is $Y_{ji} = x_j + U_{ji}$. The asymptotic GOC curve is the curve for stimuli ordered by $\mathrm{E}(R_j)$, where R_j is the rating of the j^{th} stimulus. Drga (1999) showed that this is the same as the GOC curve ordered by $\mathrm{E}(Y_{ji})$ if and only if Y_{ji} and R_j are stochastically ordered. Stochastic ordering is necessary to project back from the rating scale to the underlying evidence distributions. The focus here, however, is on the practical calculations of a GOC analysis.

One way to do GOC analysis, which principally involves calculating the GOC curve, is described by Taylor et al. (1991) and Watson (1962 / 1963). Drga (1999) suggested a different method, which is computationally more memory efficient, and can be used for real valued ratings and not just integers, but is functionally the same as Taylor et al. (1991) and Watson (1962 / 1963). An example of Drga's method is shown in Tables 2.1 and 2.2 for four replications of a 9–point rating scale experiment, with one $\mathcal{N}$ event and two $\mathcal{SN}$ events, and five $\mathcal{SN}$ stimuli and ten $\mathcal{N}$ stimuli.

Table 2.1: Example sum–of–ratings table for four replications of an experiment.

Stimulus	Event	Rating				Sum–of–rating
		Rep 1	Rep 2	Rep 3	Rep 4	
1	$\mathcal{N}$	3	5	8	1	17
2	$\mathcal{N}$	1	2	2	4	9
3	$\mathcal{N}$	4	3	5	2	14
4	$\mathcal{N}$	2	4	5	2	13
5	$\mathcal{N}$	6	1	2	6	15
6	$\mathcal{N}$	1	2	1	1	5
7	$\mathcal{N}$	5	4	6	3	18
8	$\mathcal{N}$	1	1	4	1	7
9	$\mathcal{N}$	9	3	4	1	17
10	$\mathcal{N}$	1	1	4	2	8
1	$\mathcal{SN}1$	4	1	1	3	9
2	$\mathcal{SN}1$	4	5	5	6	20
3	$\mathcal{SN}1$	8	6	5	1	20
4	$\mathcal{SN}1$	9	4	2	1	16
5	$\mathcal{SN}1$	4	5	4	2	15
1	$\mathcal{SN}2$	6	1	3	4	14
2	$\mathcal{SN}2$	9	8	5	8	30
3	$\mathcal{SN}2$	7	7	9	9	32
4	$\mathcal{SN}2$	9	7	7	7	30
5	$\mathcal{SN}2$	8	6	4	8	26

1. Calculate the sum–of–ratings across replications for each stimulus (see Table 2.1).

2. The sum–of–ratings may then be converted to an average,[1] and analysis proceeds in exactly the same way, with the average rating being sorted instead of the sum–of–ratings. If the number of replications is not the same for each stimuli then the ratings must be averaged so each total is weighted appropriately.

[1]The sum is a type of average, and is computationally faster. The arithmetic average and the sum–of–rating GOC curve is the same, if the number of replications for each stimulus is the same. Any form of averaging may be used, e.g., arithmetic, geometric, harmonic, sine, or z–transform averaging (Drga, 1993a).

Table 2.2: Example GOC hit rates and false–alarm rates, calculated from the sum–of–ratings.

Sum–of–rating	Tally			Cum. Tally			Rate		
	$\mathcal{N}$	$\mathcal{SN}1$	$\mathcal{SN}2$	$\mathcal{N}$	$\mathcal{SN}1$	$\mathcal{SN}2$	FAR	HR1	HR2
5	1	0	0	10	5	5	1.0	1.0	1.0
7	1	0	0	9	5	5	0.9	1.0	1.0
8	1	0	0	8	5	5	0.8	1.0	1.0
9	1	1	0	7	5	5	0.7	1.0	1.0
13	1	0	0	6	4	5	0.6	0.8	1.0
14	1	0	1	5	4	5	0.5	0.8	1.0
15	1	1	0	4	4	4	0.4	0.8	0.8
16	0	1	0	3	3	4	0.3	0.6	0.8
17	2	0	0	3	2	4	0.3	0.4	0.8
18	1	0	0	1	2	4	0.1	0.4	0.8
20	0	2	0	0	2	4	0.0	0.4	0.8
26	0	0	1	0	0	4	0.0	0.0	0.8
30	0	0	2	0	0	3	0.0	0.0	0.6
32	0	0	1	0	0	1	0.0	0.0	0.2
Total	10	5	5						

3. All the resulting sum–of–ratings, over all events and stimuli, are sorted into ascending order, and separate tallies are recorded for each event (see the first four columns of Table 2.2). The same sum–of–rating may occur more than once for the same event (e.g., sum–of–rating 17 in Table 2.2), and across events (e.g., sum–of–rating 15 in Table 2.2). The minimum possible sum–of–rating is equal to the minimum rating times the number of replications. The maximum sum–of–rating is equal to the maximum possible rating times the number of replications. In the example, the theoretical minimum is 4 and the maximum is 36, but the actual minimum is 5 and the maximum is 32.

4. The sum–of–ratings are now treated as cut–offs, like in a rating scale experiment. The tallies for each event are cumulated from the largest sum–of–rating value to the smallest across all events (see the columns 5–7 in Table 2.2). If a sum–of–rating value does not have a tally for a particular event then zero is added to the cumulative tally.

5. The cumulative tallies are then divided by the number of stimuli for that event and the result is the hit rates and false alarm rates for the GOC curve (see the last three columns in Table 2.2), which may then be plotted in the ROC space.

6. From the GOC curve, analysis proceeds like ROC analysis for rating–scale data to obtain the desired measures of detectability.

As more replications are included in the GOC analysis, more unique noise is removed and the GOC curve tends towards the theoretical curve.[2] If there is no common noise then the GOC curve does not converge to a stable position as more replications are added (Drga and Whitmore, 1999, personal communication).

The striking difference between the results of ROC analysis and GOC analysis can be seen by comparing the GOC curve to the mean ROC (mROC) curve. An example is given in Figure 2.1.

[2] Or sample–theoretical curve if, as in most experimental cases, only a sample of stimuli were used.

The mROC curve summarises the multiple ROC curves from a multiple replication experiment, and gives the average, unique–noise degraded, ROC curve. It is calculated by averaging the hit rates and false–alarm rates for each criterion over each replication (Dorfman & Berbaum, 1986; MacMillan & Kaplan, 1985). This does not eliminate observer inconsistency—it just gives the expected performance for a single replication.

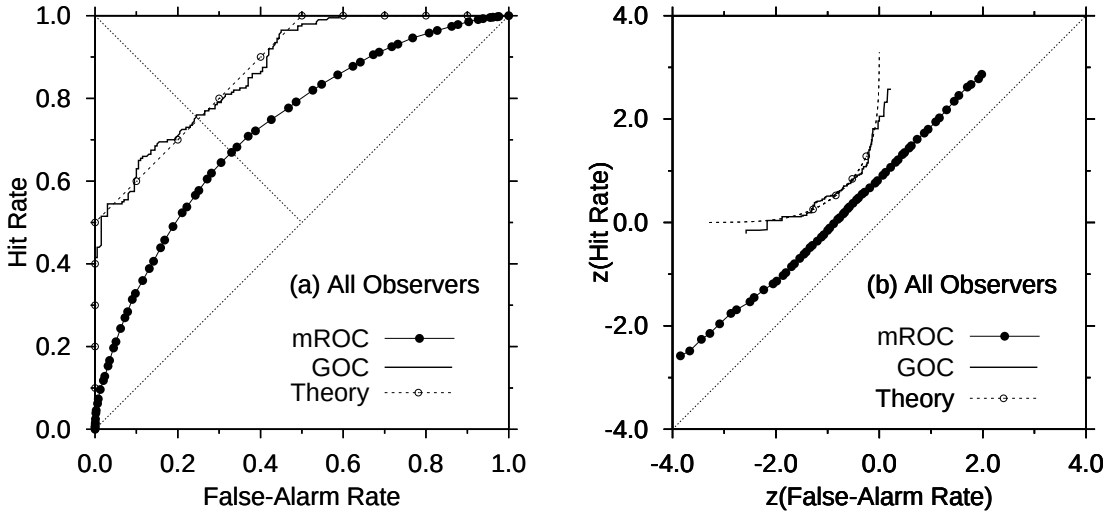

Figure 2.1: The GOC curve and the mROC curve for all observers, and the theoretical ROC curve in the SIFC condition in Experiment I from Lapsley Miller et al. (1999) (reprinted with permission). (a) is in linear probability coordinates and (b) is in $z-$coordinates.

The results of GOC analysis may help explain some psychophysicists' preoccupation with normal–normal models. Hanley and McNeil (1982) comment that the majority of psychophysical data can be fitted very well on the assumption that the underlying random variables are normal—regardless of the appropriateness of this assumption. There is, unfortunately, a belief that TSD is only applicable to detection tasks where the evidence is normally distributed (Hodos, 1970; Simpson & Fitter, 1973). Hanley and McNeil (1982) suggest that many researchers have been turned off TSD because of this mistaken belief. *TSD does not need the assumption that the evidence is distributed as a normal density.* There are, in fact, many psychophysical models that use other distributions (e.g., Egan, 1975; Green & Swets, 1966; Jeffress, 1968). Non-parametric measures of detectability, such as $\mathcal{A}$ and $\mathcal{D}_n$, then allow comparisons to be made among detectors without needing to know, or assume, how the evidence is distributed.

Taylor et al. (1991) demonstrated, using known, finitely discrete, evidence distributions, that once GOC analysis has been performed the apparent normality may disappear. This phenomenon was also found by Lapsley Miller et al. (1999) where it was known that the theoretical distributions were overlapping discrete uniform distributions. The theoretical ROC curve was, therefore, a straight line in linear ROC space, but very curvilinear in the z–transformed ROC space. The experimental data showed that (a) the mROC curve was curvilinear in linear ROC space and was well fitted by a straight line in z–coordinates—inconsistent with the theoretical model—implying the normal model may be appropriate, but that (b) after GOC analysis, the GOC curve tended towards the theoretical ROC curve. Figure 2.1 shows the SIFC mROC and GOC results combined across all observers compared with the *known* SIFC ROC curve. This is an extreme example; many theoretical models, based on distributions other than the normal, predict ROC curves that

are not too dissimilar from the normal–normal model. The effect of GOC analysis in these cases may be to bring out underlying asymmetries (e.g., for Rayleigh–Rice models) and to decrease the attenuation of the empirical psychometric function from the theoretical. In summary, the results from GOC analyses show that any interpretation of noise–degraded data must be made cautiously, because it is too easy to draw erroneous conclusions about the underlying processes.

2.1.2 Function of replications combined estimation (FORCE) analysis

FORCE analysis is a data–modelling procedure developed by Vit Drga, Alan Taylor, and John Whitmore (Drga, 1999). The analysis estimates asymptotic measures of detectability from GOC analysis, by fitting a data model of the measure of detectability as a *function of replications added* (FORA) to the empirical FORA, and extrapolating to infinity. This process works because as replications are added, performance generally improves asymptotically.

FORCE analysis is calculated on the measures of detectability, for instance, $\mathcal{A}$, $\mathcal{D}_2$, and d'. This procedure gives excellent estimates from only a few replications. A rule–of–thumb recommended by Drga (1999) is a minimum of 8–10 replications for good asymptotic estimates, however, 6–7 replications may also give reasonably reliable results. FORCE analysis is possible with 3–5 replications, but the resulting estimates may be biased and have large sampling variability. Drga (1999) gives some estimates of the variability in the asymptotic measure, as a function of replications–added, but it is calculated from only one data set (by calculating the variability of the estimates of multiple sub–samples of n replications from 75 replications of an experiment). Prior to the development of FORCE analysis, over 16 replications was necessary before performance neared asymptotic levels. One GOC experiment (Drga, 1999) was repeated 100 times to assess how performance improved with the addition of more replications.

The empirical function of replications–added (FORA)

The first step in FORCE analysis is to calculate the (arithmetic) average measure of detectability as replications are added. This process is easiest explained by example. Assume that an experiment was replicated six times. First, the GOC curves for all combinations of one replication are calculated, then the measures of detectability for each GOC curve are calculated and averaged. This first point is merely the average of the measures from the six ROC curves, because the GOC curve of one replication is just the ROC curve. Next all the combinations of two replications are calculated. Fifteen GOCs of two replications can be calculated from a total of six replications. Then twenty combinations of three replications, fifteen combinations of four replications, six combinations of five replications, and finally one combination of six replications. The number of combinations as replications are added is of course the binomial coefficient: ${}^{n}\mathrm{C}_{r}$, where $n = 6$ and r is replications added. When plotted, Drga calls these points the empirical FORA.

An example of the effect of this process, for real data, is shown in Figure 2.2(a). These data comes from six replications of one level from the experiments in this thesis (see the FORA for the experimental condition {25ms, 40Hz: 8dB} in Figure E.1(e) and Figure E.2(e) in Appendix E). Note how both the error and the rate of increase in $\mathcal{A}$ decrease as replications are added.

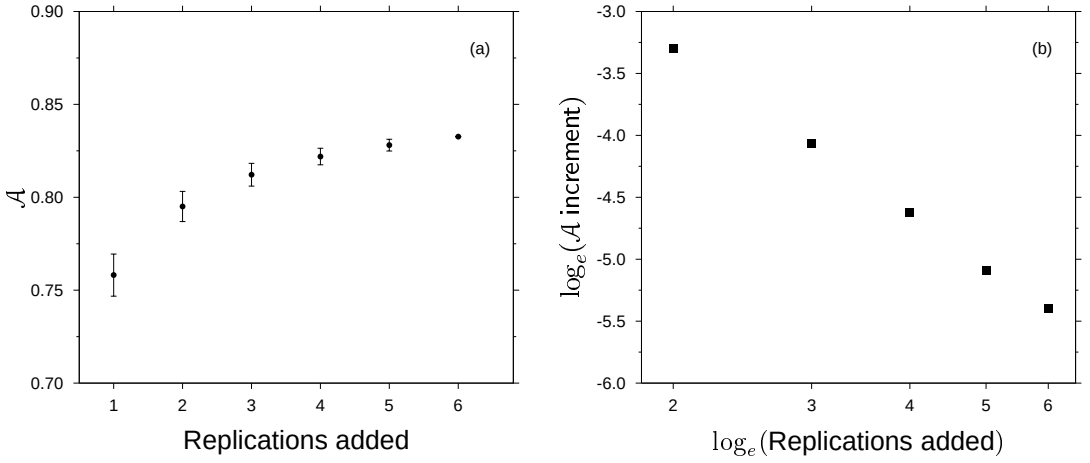

Figure 2.2: An example of (a) a six replication empirical FORA, and (b) the same FORA in log–increment coordinates. The error bars represent ± 1 standard deviation. Note the ordinate does not cover the entire range of $\mathcal{A}$.

The empirical function of detectability–increments and replications–added in log–increment coordinates

Plotting the average measure of detectability against replications–added generally (but not necessarily) gives a monotonically increasing set of points. Drga discovered that by plotting the log of the *increment* in the measure of detectability against the log of replications added, that all the data sets he examined[3] could be fitted extremely well by a straight line with a negative slope. In fact, most of the data sets Drga examined, from a variety of experiments, have Pearson's correlation coefficients of less than -0.99. The results obtained in this research project (see Appendix E) also showed correlations of less than -0.99, in most cases.

Figure 2.2(b) shows an example of the increment of detectability as a function of replications added in log–increment coordinates. The Pearson's correlation coefficient is -0.9998 indicating the data are indeed linear in these coordinates. These straight lines indicate how a data model can be fitted to the empirical FORA in linear coordinates. The point is not to fit a model to the log–increment data, because they only represent *increments* in detectability, and, therefore, can not be used directly as a model to estimate asymptotic detectability. The log–increment plot is still useful, however, because it shows biases. If the plot turns up at the end, relative to the fitted FORA, then the asymptotic estimate will be under–estimated; if it turns down then the estimate will be over–estimated (Drga, 1999).

Modelling the detectability–increments as a function of replications–added

Drga (1999) derived a regression function to fit to the detectability–increments as a function of replications–added. To model the increments as a function of replications–added, consider n replications of an experiment. If $\delta_j = \mathcal{A}_j - \mathcal{A}_{j-1}$, where $2 \leqslant j \leqslant n$, then δ_j is the $(j-1)^{\text{th}}$ increment in mean–$\mathcal{A}$ values and may be approximated by a straight line, as suggested by the log–increment

[3] In developing FORCE analysis, Drga has examined over 70 data sets, from six different observers, and six experiments (the six observers did not participate in every experiment).

plot, of the form

$$\begin{aligned} \log_e(\delta_j) &\approx m\log_e(j) + c \\ &= \log_e(j^m) + \log_e(k) \\ &= \log_e(kj^m). \end{aligned} \tag{2.1}$$

where $k = \exp(c)$. This implies that $\delta_j \approx kj^m$. If the empirical log–increment plot is exactly linear then this inequality is exact. Note that any log base may be used, but the development of FORCE has used the natural logarithms. If $\mathcal{A}_1$ is the first value then

$$\begin{aligned} \mathcal{A}_i &= \mathcal{A}_1 + \sum_{j=2}^{i} kj^m \\ &= \mathcal{A}_1 + k\sum_{j=2}^{i} j^m, \qquad \text{where } i \geqslant 2. \end{aligned} \tag{2.2}$$

This is the fitted FORA function, which may be extrapolated to infinity:

$$\begin{aligned} \mathcal{A}_\infty &= \mathcal{A}_1 + k\sum_{j=2}^{\infty} j^m \\ &= \mathcal{A}_1 + k(\zeta(-m) - 1) \end{aligned} \tag{2.3}$$

where $\zeta(x) = \sum_{j=1}^{\infty} j^{-x}$ is Riemann's ζ function. This function does not converge quickly because it sums from below. Instead, an alternative formulation, which sums alternately from above and below, is used

$$\zeta(x) = \frac{1}{1 - 2^{(1-x)}} \sum_{j=1}^{\infty} (-1)^{(j+1)} j^{-x}. \tag{2.4}$$

It should be noted that if the first form of the ζ function is used, then Equation (2.3) only converges if $m < -1$. If the second form of the ζ function is used then Equation (2.3) converges if $m < 0$. Equation (2.3) gives an estimate of *the asymptotic measure of detectability if an infinite number of replications were run*.

To estimate the asymptotic measure of detectability, a parameter triplet ($\mathcal{A}_1$, k, m) is required, which minimises the total least–squares error between the fitted FORA, of the form in equation (2.2), and the empirical FORA. The *initial* values for these parameters are $\mathcal{A}_1$, which is the average $\mathcal{A}$ from the ROC curves, and k and m calculated from equation 2.1 by simple regression. A gradient–descent method is then used to calculate the final values (Drga, 1999). Note that the initial parameter values may be any number, but convergence is faster and more reliable if the initial values are close to the final values.

An example of a FORA *fitted* to the example data set from Figure 2.2 is shown in Figure 2.3(a) where the dashed line on the plot is the asymptotic measure of detectability. The function of line segments is not connecting the points of the empirical data set; they are connecting the points of the fitted FORA derived as above. Line segments are used as a visual aid, because the fitted FORA is generally on top of the empirical FORA. In virtually all cases examined so far, the fit is excellent for the measures of $\mathcal{A}$, $\mathcal{D}_2$, $\mathcal{D}_6$, and d'. The results from $P(\mathrm{C})$, however, tend to be unreliable, because $P(\mathrm{C})$ values are more variable due to measurement error. The fact that $P(\mathrm{C})$

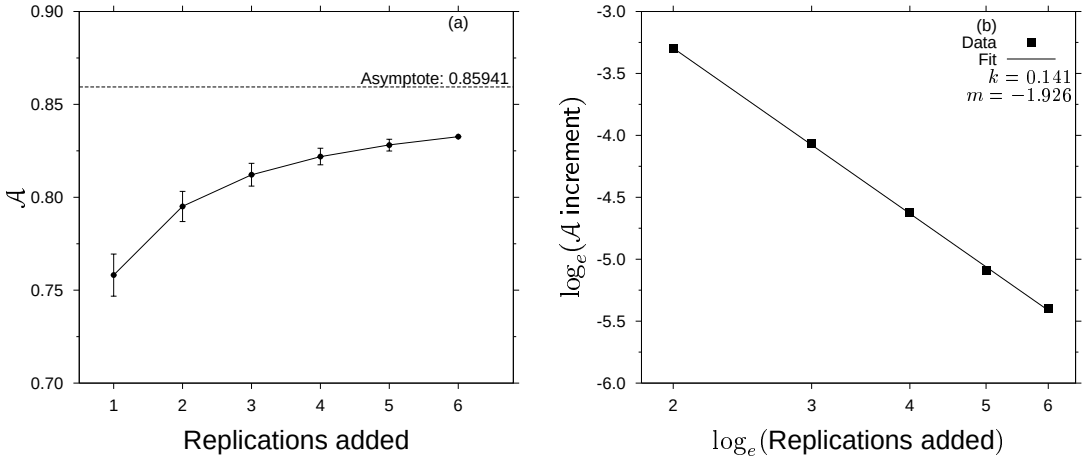

Figure 2.3: An example of a FORA fitted to the empirical FORA (a) using the method described by Drga (1999). The dashed line shows the asymptotic estimate of $\mathcal{A}$. (b) shows the empirical log–increment FORA (solid points), and the linear function associated with the parameters from the fitted FORA.

is confounded with the criterion may also be a factor, because this also increases variability.

The linear equation based on the parameters estimated from Drga's fitting method may also be plotted in log–increment coordinates. This is illustrated in Figure 2.3(b). The examples shown are for a relatively ordinary data set. Appendix E shows both types of plots for all the data collected in this research project, including some data sets that do not behave quite as well.

According to Drga (1999), the estimate of asymptotic performance has sampling error itself, and may be biased for small numbers of replications. The bias shows itself by either over–estimating or under–estimating the asymptotic value. This bias decreases as the number of replications increases. Bias may be assessed by considering the fitted and empirical FORA. If the last few empirical points fall below the fitted FORA then the asymptotic measure will be overestimated, likewise, if they fall above then the asymptotic measure will be underestimated. In the log–increment plot, the empirical points will tend to curve down, or up, respectively. This may be difficult to assess for small numbers of replications, which is unfortunate because these cases are more likely to be biased. After considering the FORCE analysis in the current project, it appeared that another way of assessing biases could be to consider the linearity of the empirical log–increment FORA. The less linear the empirical log–increment FORA, the more likely the asymptotic estimate was biased; in some cases the estimate was an impossible number (e.g., greater than unity for $\mathcal{A}$).

2.2 Common noise sampling variability

Even if unique noise is removed from psychophysical data, interpretation of the results may still be difficult, because of common noise sampling variability. Generally, the more stimuli that are used, the lower the sampling variability. It is important, however, to have some estimate of the degree of variability for a given sample size.

Derivations of ideal observers consider the population statistics of the stimuli. This is analogous to saying that the ideal observer is presented with every possible stimulus in the stimulus

set of interest. But in an experiment or simulation, generally only a sample of stimuli can be taken from the (perhaps infinite) stimulus set. What is the effect of this sampling variability and how should it be measured?[4]

Although sampling variability comes about from using only a subset of possible stimuli, it is easier to model the variability of the evidence, rather than the variability of the stimulus. The evidence is a result of a stimulus, and may be modelled as a sample from a probability distribution. This approach comes from the theory of ideal observers, which shows how the *distribution* of a physical property of class of waveforms may be used to calculate the detectability of those waveforms. For example, the distribution of the energy of narrow–band Gaussian noise may be modelled as χ^2. If the detection task is intensity discrimination, then the detectability statistics are based on two χ^2 distributions (for $\mathcal{N}$ and $\mathcal{SN}$ events).

In an experiment, the subset of waveforms used as stimuli can be modelled as a sampling from the evidence distributions, even when these distributions are unknown. The samples from the $\mathcal{N}$ and $\mathcal{SN}$ evidence distributions form the *sample theoretic* ROC curves, which are also unknown. It is important to realise that the GOC curve tends towards the sample theoretic ROC curve, not the population theoretic ROC curve. If a different sampling of waveforms were chosen as stimuli, then the GOC curve would tend towards a different sample theoretic ROC curve. It is simplest to model the variability of the sample–theoretic ROC curve by measuring the variability of a measure of detectability derived from the ROC curve, for instance $\mathcal{A}$.

The problem of determining the amount of sampling variability of $\mathcal{A}$, for a given number of stimuli, lends itself to simulation. The general procedure (Lapsley Miller, 1995a, 1995b; Lapsley Miller et al., 1999) is to

1. Choose $\mathcal{N}$ and $\mathcal{SN}$ evidence distributions, with known parameters, and calculate the theoretical measures of detectability and ROC curves.

2. Generate M random deviates from these distributions to form the sample set of evidence.

3. Either generate the sample–theoretic ROC curve directly from the samples (by using the procedure described in Section 2.1.1), or convert each sample to a rating on an N–point rating scale (by binning the data) then generate the sample–theoretic ROC curve.

4. Calculate the measures of detectability.

5. Repeat the sampling procedure R times.

6. Calculate the sample statistics (mean, variance, etc.) of each measure of detectability, over the R replications.

By systematically varying the sample size (M), variability as a function of samples size may be assessed (Lapsley Miller, 1995a). In comparison to the simulation method, the maximum variability of $\mathcal{A}$ may be estimated assuming binomial sampling (Bamber, 1975; Pollack & Hsieh, 1969).

Pollack and Hsieh (1969) ran Monte Carlo simulations, using a similar method, to assess sampling variability of $\mathcal{A}$ for very small sample sizes. They used three different evidence distributions: normal with unequal variance, uniforms with unequal variance, and negative exponentials with unequal k, and sample sizes of 10 to 40. They found that:

[4] In an experiment common noise may also occur from systematic error across replications.

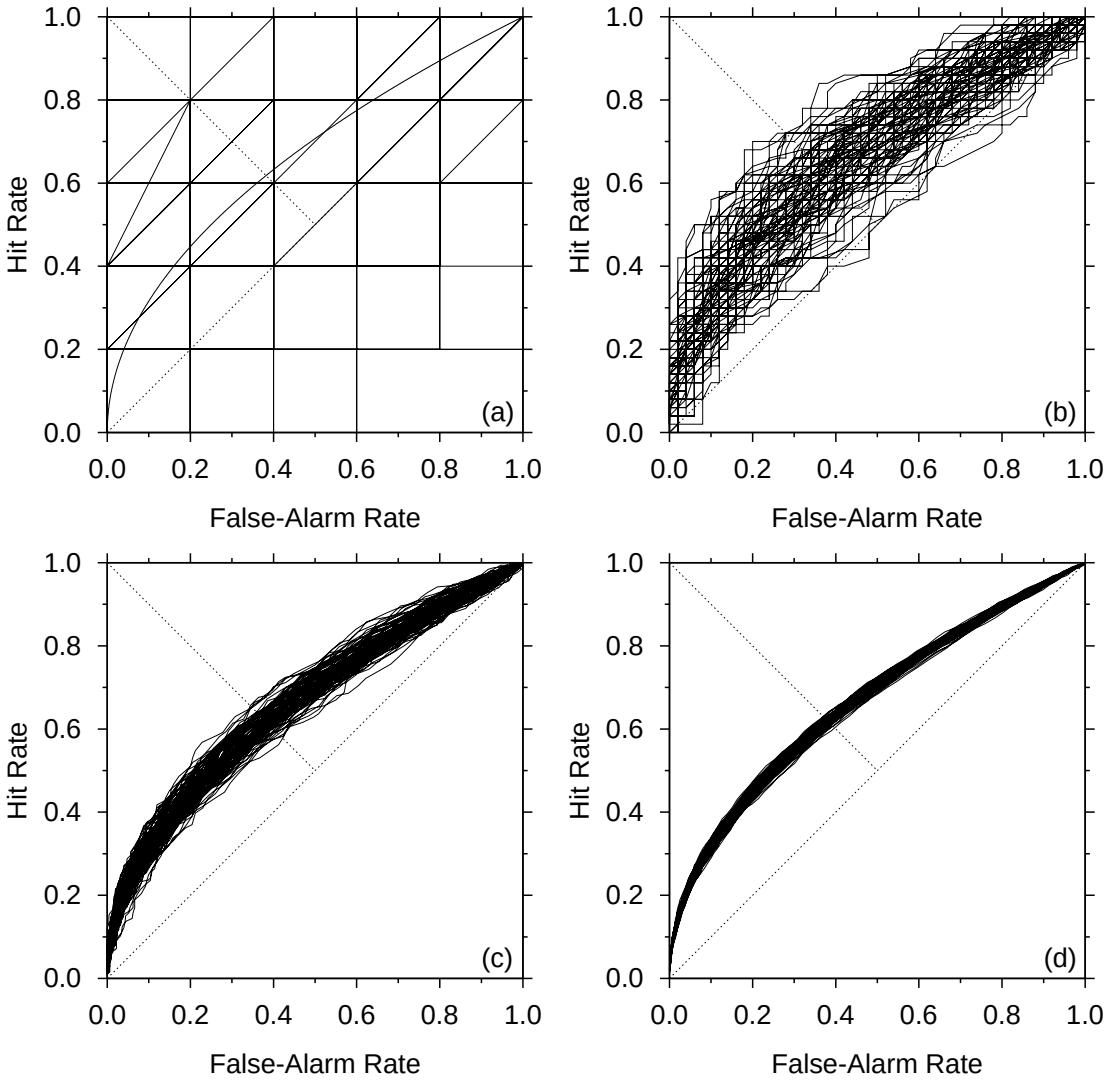

Figure 2.4: Simulated sampling variability of a negative exponential model for four sample sizes (a) 5, (b) 50, (c) 500, and (d) 5000. The population theoretic ROC curve (smooth solid curve) is overlayed on all plots, but is only visible in (a). It should be noted that the lines in (a) are not a grid, but are in fact the ROC curves. The grid pattern occurs because a sample size of 5 severely restricts the resolution in the ROC space.

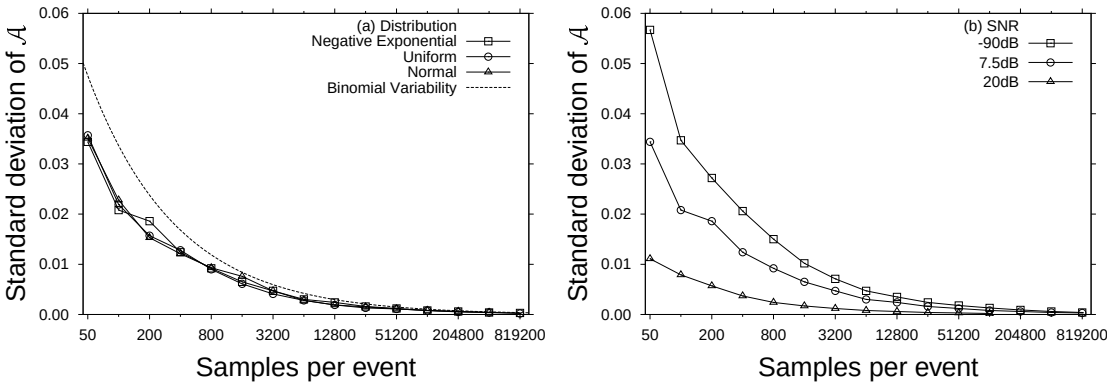

Figure 2.5: Simulated sampling variability of $\mathcal{A}$ for (a) three models with $\mathcal{A}$=0.87 compared with the maximum binomial sampling variability, and (b) for the negative exponential ROC curve, at three signal–to–noise ratios.

- The sampling variability of $\mathcal{A}$, defined as the standard error of the arithmetic mean, after 100 repeats, increased as the overlap between the distributions decreased (either because the difference between the means of the $\mathcal{N}$ and $\mathcal{SN}$ distributions decreased or the variance increased).

- For different classes of probability distribution, similar standard deviations were obtained when $\mathcal{A}$ was similar and the number of samples were the same. This implies that sampling variability of $\mathcal{A}$ is relatively independent from the conditions leading to that value. It is an important result, because estimates of sample size or variability may be made without needing to know the form of the underlying distributions.

- Variability was slightly less than that predicted by binomial sampling, especially for larger $\mathcal{A}$. The maximum variance, $\sigma^2_{\rm bin}$, assuming binomial sampling is

$$\sigma^2_{\rm bin} = \mathcal{A}(1-\mathcal{A})/n. \tag{2.5}$$

These findings were verified and extrapolated to larger stimuli per event (for it is unlikely that any substantive experiment would use fewer than 40 stimuli per event) by running similar simulations using the program RAYSIML (Lapsley Miller, 1995a).

To indicate the problem of sampling variability, Figure 2.4 shows the effect of sampling variability in the ROC space of samples from negative exponential evidence distributions (equivalent to a signal–to–noise ratio of 0 dB and a theoretical $\mathcal{A}$=0.667) repeated 100 times. Figure 2.4(a) (b), (c), and (d) are based on 5, 50, 500, and 5000 samples per event respectively. It should be noted that the lines in (a) are not a grid, but are in fact the ROC curves. The grid pattern occurs because a sample size of 5 severely restricts the resolution in the ROC space. Sample–theoretic $\mathcal{A}$, from this number of stimuli, could range from zero to one. The variability illustrated in Figure 2.4(b), for a sample size of 50, is also of concern, because there are many papers in the literature that use fewer than 50 stimuli per event.[5]

[5] A quick perusal of some recent journals produced the following examples. The Journal of the Acoustical Society of America, Vol. 97(6): 10, 25, and 50 stimuli per event, and 48 stimuli in total; Perception and Psychophysics Vol. 55(2): 24, 72, 100; and Memory and Cognition Vol. 19: 20–40 stimuli per event.

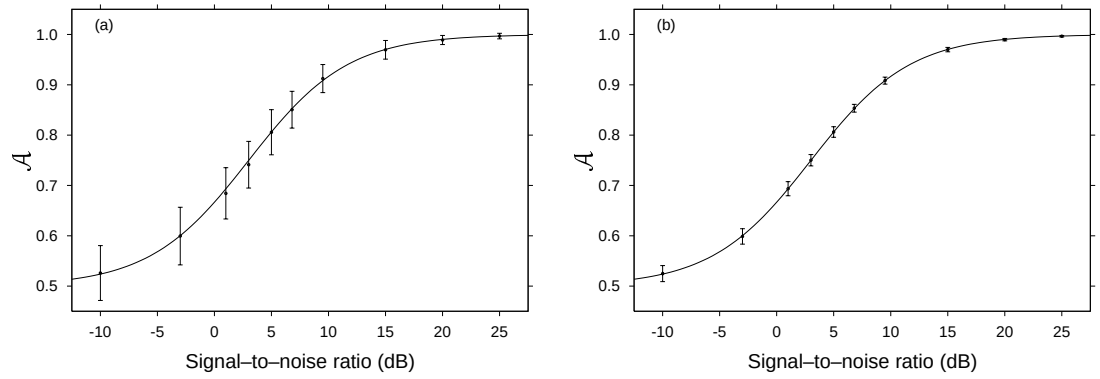

Figure 2.6: Sampling variability of $\mathcal{A}$ for the negative exponential psychometric function for (a) 50 samples per event, and (b) 800 samples per event. Error bars are ± 1 standard deviation.

Figure 2.5(a), shows the standard deviation of the arithmetic average of $\mathcal{A}$ as a function of samples per event for three models: normal–normal equal variance, uniform–uniform equal variance, and negative exponential. The parameters of each model have been carefully chosen to give approximately $\mathcal{A}$=0.87 despite their ROC curves being quite different. It seems as though the nature of the distribution has little effect on the resulting sampling variability. The binomial variability, which was suggested by Pollack and Hsieh (1969) as a good approximation to the sampling variability in the ROC space, is also included for comparison, but it consistently overestimates the error. Figure 2.5(b) shows that, for the negative exponential model, as signal–to–noise ratio increases, sampling variability of $\mathcal{A}$ decreases. The relationship between $\mathcal{A}$ and signal–to–noise ratio is given in equation (1.15). An argument could be made therefore, to use fewer stimuli at larger signal–to–noise ratios, and thereby have the same variability for all signal levels. This is not necessarily a good idea, because the phenomenon is possibly an artifact, because $\mathcal{A}$ is bounded at one. If the sampling variability of d' is measured instead, the variability *increases* with signal–to–noise ratio. Figure 2.6 shows an example of the effect of sampling variability in the negative exponential psychometric function for 50 and 800 samples per event.

Bamber supports the third result of Pollack and Hsieh (1969) by citing two proofs (Dantzig, 1951, in Bamber, 1975; Birnbaum and Klose, 1957, in Bamber, 1975) that give a smaller maximum sampling variability ($\sigma^2_{\max}$) on the assumption that the evidence distributions are continuous random variables, and the ROC curve is totally above or totally below the chance line (Bamber uses the term "stochastically comparable"). If $N_G > N_L$, where N is the sample size for each event, and A_G is the greater of $\mathcal{A}$ and $1-\mathcal{A}$, then the maximum sampling variability is

$$\begin{aligned}\sigma^2_{\text{stoc}} &= \frac{1}{N_G N_L}((2N_L-1)A_G(1-A_G)-(N_G-N_L)(1-A_G)^2 \\ &\quad +\frac{1}{3}(N_G-2N_L+1)[1-(2A_G-1)^{\frac{3}{2}}]).\end{aligned} \tag{2.6}$$

Bamber then proves that if, additionally, the decision axis is strictly monotonic with likelihood ratio (Bamber uses the term "monotonic posterior") then the maximum sampling variability is smaller:

$$\sigma^2_{\text{mono}} = \frac{1}{3N_G N_L}[(2N_G+1)A_G(1-A_G)-(N_G-N_L)(1-A_G)^2]. \tag{2.7}$$

To evaluate these formulae, the program RAYSIML (Lapsley Miller, 1995b) was used to run the following simulations:

1. Overlapping uniform distributions of equal mean and unequal variance, giving $\mathcal{A}$=0.5 (see Figure 2.7(a)). This approximates the case of Fig. 7a. in Bamber (1975, p. 404). The ROC curve is a vertical line with a false–alarm rate=0.5 (see Figure 2.7(e)).

2. Overlapping uniform distributions of unequal mean and unequal variance, giving $\mathcal{A}$=0.87 (see Figure 2.7(b)). The ROC curve is a vertical line with a false–alarm rate=0.13 (see Figure 2.7(f)).

3. Negative exponential distributions of equal variance, giving $\mathcal{A}$=0.5 (see Figure 2.7(c)). This should fulfil the requirements of being monotonic with likelihood ratio. Note that these two distributions are the same so the ROC curve falls along the chance line (see Figure 2.7(f)).

4. Negative exponential distributions of unequal variance, giving $\mathcal{A}$=0.87 (see Figure 2.7(d)). The ROC curve is strictly monotonically increasing (see Figure 2.7(h)).

Each simulation generated 100 ROC curves, each with 200 points.

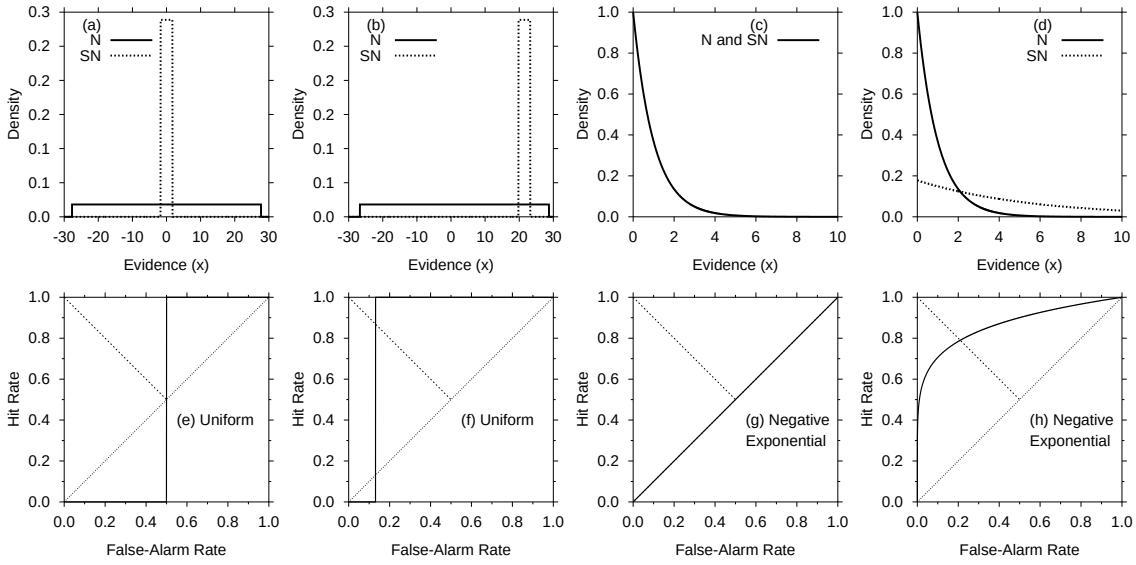

Figure 2.7: The theoretical evidence distributions (a)–(d), and associated ROC curves (e)–(h), used for the four simulations run to test Bamber's (1975) sampling variability formulae. (a) and (e) uniform distributions with equal mean and unequal variance ($\mathcal{A}$=0.5), (b) and (f) uniform distributions with unequal mean and unequal variance ($\mathcal{A}$=0.87), (c) and (g) negative exponential distributions with signal–to–noise ratio=$-\infty$ dB ($\mathcal{A}$=0.5), and (d) and (h) negative exponential distributions with signal–to–noise ratio=7.5 dB ($\mathcal{A}$=0.87).

Figure 2.8 shows variability as a function of the number of samples for the three definitions of $\sigma_{\max}$, from Equations (2.5), (2.6), and (2.7), for equal samples per event compared with the simulations described above. The resulting ROC curves for simulation (1) should, according to Bamber, maximise Equation 2.5. This appears to be the case as shown in Figure 2.8. The variability in simulation (3) is overestimated by the binomial variance, but both the other definitions fit the simulation well. In fact when $\mathcal{A}$=0.5, Equations (2.6) and (2.7) are equal, because the functions

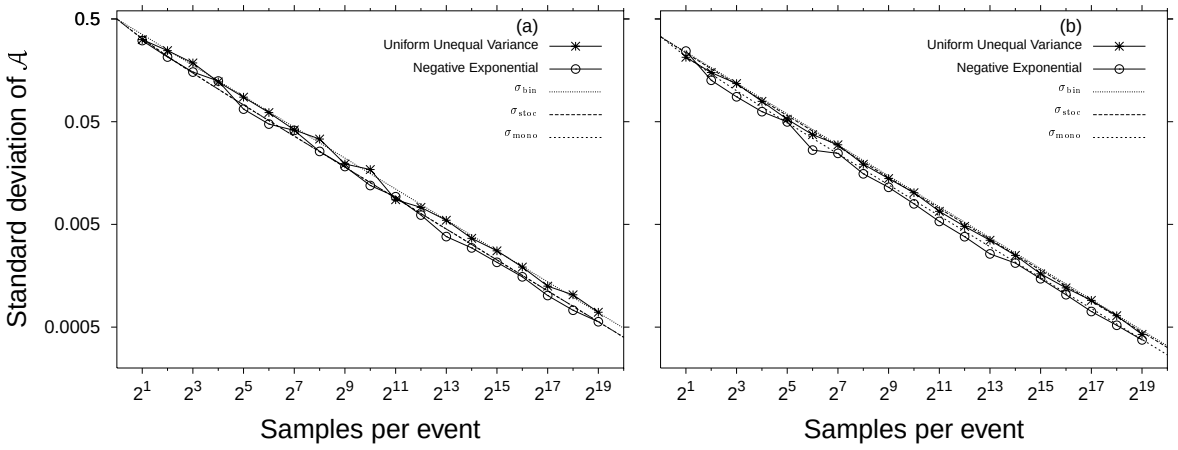

Figure 2.8: Sampling variability using the three definitions of $\sigma_{\max}$, from Bamber (1975), compared with simulations using uniforms and negative exponentials. (a) shows the standard deviation of $\mathcal{A}$ for simulations designed to give $\mathcal{A}$=0.5, and (b) shows the standard deviation of $\mathcal{A}$ for simulations designed to give $\mathcal{A}$=0.87. Note that both axes are scaled using $\log_2$.

fall on top of one another. This can also be shown algebraically by substituting $\mathcal{A}$=0.5 into the equations.

Figure 2.8(b) shows $\sigma_{\max}$ for the three definitions, for equal samples per event, compared with simulations (2) and (4) described above. The sampling variability of the uniform distributions are well described by Equation 2.5. For negative exponentials, the binomial variability overestimates the variability as expected, and both the other definitions fit the simulation well. When $\mathcal{A}$=0.87 Equations (2.6) and (2.7) are no longer equal and instead $\sigma^2_{\text{bin}} > \sigma^2_{\text{stoc}} > \sigma^2_{\text{mono}}$, as predicted by Bamber.

Note, too, that the relationship between sampling variability and sample size is linear in log coordinates—something that neither Pollack and Hsieh (1969) nor Bamber (1975) appear to have noticed. They also have the same slope, indicating that the rate of improvement in sampling variability is constant, but the initial amount of variability is dependent on $\mathcal{A}$.

To estimate the maximum sampling variability given an empirical $\mathcal{A}$, Bamber suggests replacing $\mathcal{A}$ and A_G with the sample $\mathcal{A}$ in all three equations; changing N_L for $N_L - 1$ in Equation (2.5), and changing the denominator from $3N_GN_L$ to $(3N_GN_L - N_G - N_L + 1)$ in Equation (2.7). From a knowledge of $\mathcal{A}$ and the number of stimuli per event, the variability may therefore be estimated without running simulations.

2.3 Implications for experiments

Unique noise and common noise sampling variability cause uncertainty in the interpretation of experimental results. It is therefore important in an experiment to minimise *both* types of error. To decrease common noise sampling variability, more stimuli per event, and, therefore, more trials per replication, are required. To decrease unique noise variability, more replications, and, therefore, more repeats of the same stimuli are required. Both these requirements increase the total number of trials needed in an experiment.

Obviously, it is important to reduce both the common and unique noise variability, but re-

ducing both cannot be achieved without increasing the number of trials, and, therefore, the time spent doing the experiment. The less time that is spent per condition, the better, because more conditions can be attempted. At the very least an educated decision about the number of stimuli, trials, replications, and conditions may be made, even if the results are still not ideal.

Chapter 3

Experiments

"All our science, measured against reality,
is primitive and childlike—and yet
it is the most precious thing we have."

Albert Einstein.

.....

Mathematically, the energy and envelope detectors show bandwidth–duration reciprocity, because they predict the same performance for a given $\mathcal{WT}$, regardless of the values of $\mathcal{W}$ or $\mathcal{T}$. Practically, even if the human hearing system is capable of ideal performance, there would still be physiological limitations that restrict the range over which time and frequency information may be integrated. Previous experiments reported in the literature have been inconclusive and contradictory, with (a) some indicating that the critical band or detection bandwidth may or may not be fixed, (b) that rectification may be square–law or linear, (c) that temporal integration may or may not occur over fixed durations or with fixed time constants, and (d) that sampling may be peak detection, end of signal, or some other strategy.

Many of these experiments were not able to fully test the nature of the auditory filter, rectifier, temporal integration, and sampler, because they were not considered together as a system. Because experimental manipulations in the frequency domain affect the time domain, and vice–versa, it is necessary to consider both temporal and frequency domain phenomena together. These experiments were also limited due to the abiding problem of observer inconsistency.

To try and reconcile these problems, the experiments in this project were designed to parametrically, and independently, vary the bandwidth and duration of the signals, for three values of $\mathcal{WT}$. The experiments were also repeated multiple times so GOC and FORCE analysis could be applied.

3.1 Overall Design

The experimentation was designed to show whether the human hearing system is capable of bandwidth–duration reciprocity.

The stimuli chosen for the experiments were narrow–band short–duration Gaussian noise waveforms, because it is possible to manipulate both the bandwidth and the duration of such stimuli relatively independently of each other. Chapter 1 reported that both the energy and envelope detector models have been proposed as optimal for the detection of this class of waveform. Given the same type of waveform, can the human hearing system also perform optimally (under any or all circumstances), and be modelled by an energy or envelope detector?

It is unlikely that any ideal mathematical model would be appropriate for the human hearing process, except perhaps over a small range. Based on previous research, it is more likely that these ideal models may only be appropriate if an attenuation constant is included. It is unclear from the previous research, however, if this attenuation is due to unique noise degradation, or whether there is some limitation in the hearing system's ability to detect such waveforms. This will be assessed by using GOC and FORCE analysis.

The hearing system will only optimally detect a signal if it can match the bandwidth, and integrate over the entire duration, of the signal, for only then can the signal–to–noise ratio be optimised, and all the degrees of freedom used. For instance, if the hearing system matches the bandwidth, but only integrates over half the duration of the waveform then $\mathcal{WT}$ will halve and performance will drop. Likewise, if the hearing system matches the duration, but uses a bandwidth wider than that of the signal, then the $\mathcal{WT}$ of the noise will increase more than for the signal–plus–noise, and so the signal–to–noise ratio will decrease making the signal harder to detect. It may also be the case that the hearing system is inflexible and uses the same bandwidth (at a particular frequency) regardless of the duration of the signal; or that the ear always integrates for the same length of time, but can tune the bandwidth.

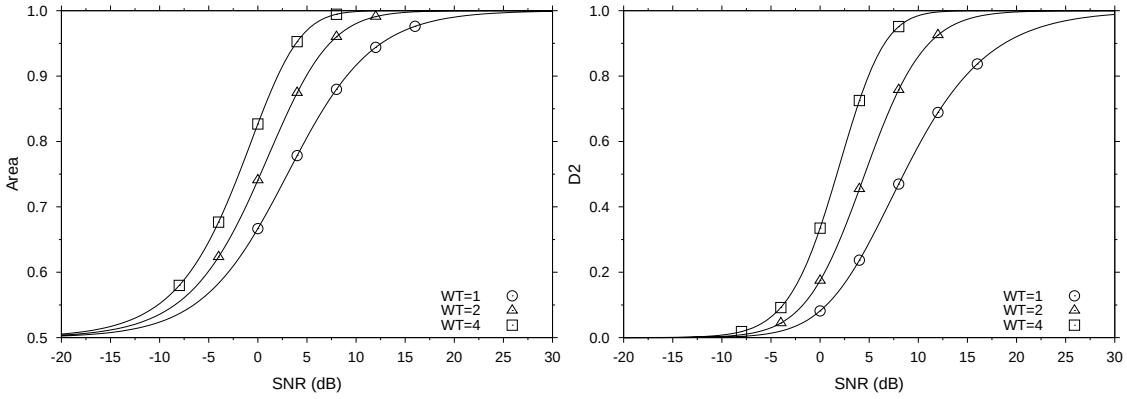

Figure 3.1: Theoretical energy detector psychometric functions for (a) $\mathcal{A}$, and (b) $\mathcal{D}_2$. The points indicate the experimental signal–to–noise ratios.

Bandwidth–duration reciprocity of the human hearing system can be explored parametrically by varying $\mathcal{W}$ and $\mathcal{T}$ for different bandwidth–duration products. The mathematical model of energy detection predicts (see Figure 3.1 and 3.2) that detectability (a) should be the same for the same value of $\mathcal{WT}$, (b) should increase as $\mathcal{WT}$ increases, and (c) should increase as the signal–to–noise ratio increases. If the hearing system cannot perform in this manner, then detectability will be attenuated, thereby indicating its limitations. The mathematical model for envelope detection is only developed for $\mathcal{WT}$=1, where it predicts the same performance as the energy detector. Simulations of full–linear and envelope detectors for larger $\mathcal{WT}$ are considered in Chapter 4.

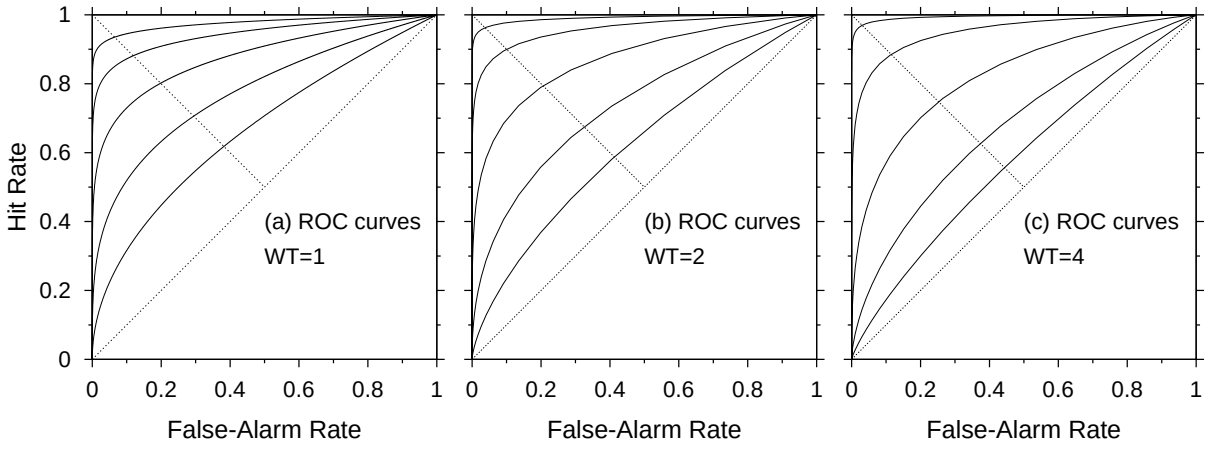

Figure 3.2: Theoretical ROC curves based on the energy detector for the experimental values tested.

Performance using this parametric design can be visualised as a surface in three–dimensional $\mathcal{WT}$–space. Two axes are bandwidth and duration. Although bandwidth and duration take on real values their product is integer under the χ^2 model of detectability, therefore the two–dimensional bandwidth–duration plane results in diagonal strips of possible $\mathcal{WT}$ values when plotted in log coordinates (see Figure 3.3). The third axis is a measure of detectability (for example: $\mathcal{A}_{\mathrm{SIFC}}$, $\mathcal{D}_2$, $\mathcal{D}_6$, equivalent signal–to–noise ratio, or attenuation). Mathematical theory predicts that this surface should be constant for the same $\mathcal{WT}$. By examining how human performance differs from this model we should gain some insight as to how the human hearing system works.

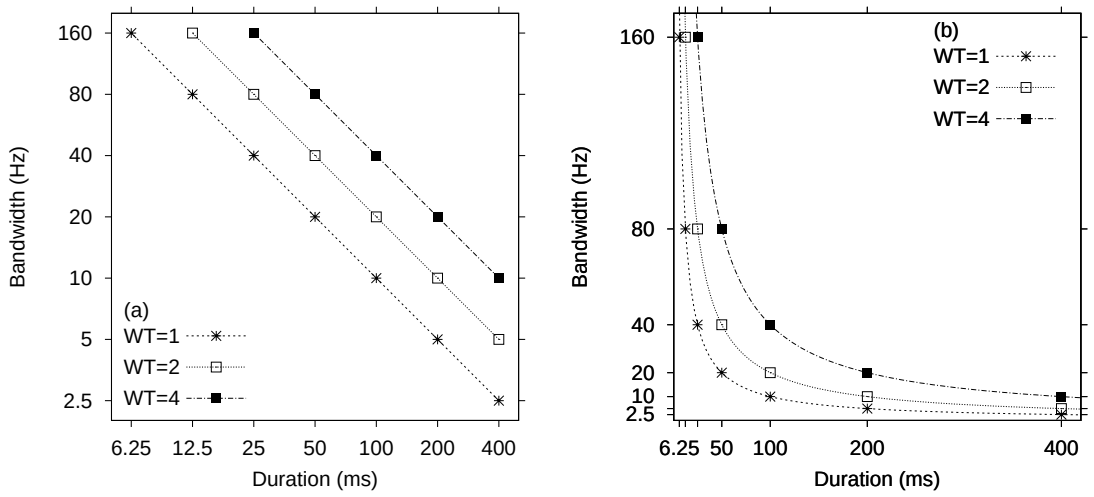

Figure 3.3: Experimental parameter space for $\mathcal{W}$, $\mathcal{T}$, and $\mathcal{WT}$ in (a) $\log_2$ and (b) linear coordinates.

Lapsley Miller et al. (1999) showed that the SIFC task is generally preferable to the 2IFC task because (a) it requires fewer assumptions about how the evidence is used to make a decision, (b) the supposed benefit of the 2IFC task, that it is unbiased, is shown to be false, (c) the effect of the inter–stimulus–interval is unclear (i.e., decay from the first interval affecting the second

interval, time–order effects, etc.), and (d) SIFC experiments are generally faster to run, because there is only one observation interval, and no inter–stimulus–interval. It is also the case that both $P(\mathrm{C})_{\mathrm{2IFC}}$ and the 2IFC ROC curve may be derived or estimated from the SIFC hit rates and false–alarm rates, but it is impossible to derive or estimate the SIFC ROC curve from the 2IFC ROC curve (Lapsley Miller et al., 1999). The SIFC task is also preferable over the Same–Different task, because none of the theory for noise–in–noise detection has been derived for this task, and much of existing theory requires the assumption of normality (e.g., Irwin, 1989). Therefore, the SIFC task was used for the experiments in this project.

Table 3.1: Defining parameters for each signal set. Bandwidth, duration, and $\mathcal{WT}$ are defined by $\mathrm{Ess}\mathcal{W}_{92.4\%}$, $\mathrm{Ess}\mathcal{T}_{92.4\%}$, and $\mathrm{Ess}\mathcal{WT}_{92.4\%\%}$ (see Section 3.3). The code is used for labelling graphs in figures.

$\mathcal{WT}$	$\mathcal{W}$ (Hz)	$\mathcal{T}$ (ms)	Signal–to–noise ratios (dB)	Code
1	2.5	400.00	$-\infty, 0, 4, 8, 12, 16$	1A
1	5.0	200.00	$-\infty, 0, 4, 8, 12, 16$	1B
1	10.0	100.00	$-\infty, 0, 4, 8, 12, 16$	1C
1	20.0	50.00	$-\infty, 0, 4, 8, 12, 16$	1D
1	40.0	25.00	$-\infty, 0, 4, 8, 12, 16$	1E
1	80.0	12.50	$-\infty, 0, 4, 8, 12, 16$	1F
1	160.0	6.25	$-\infty, 0, 4, 8, 12, 16$	1G
2	5.0	400.00	$-\infty, -4, 0, 4, 8, 12$	2H
2	10.0	200.00	$-\infty, -4, 0, 4, 8, 12$	2I
2	20.0	100.00	$-\infty, -4, 0, 4, 8, 12$	2J
2	40.0	50.00	$-\infty, -4, 0, 4, 8, 12$	2K
2	80.0	25.00	$-\infty, -4, 0, 4, 8, 12$	2L
2	160.0	12.50	$-\infty, -4, 0, 4, 8, 12$	2M
4	10.0	400.00	$-\infty, -8, -4, 0, 4, 8$	4N
4	20.0	200.00	$-\infty, -8, -4, 0, 4, 8$	4O
4	40.0	100.00	$-\infty, -8, -4, 0, 4, 8$	4P
4	80.0	50.00	$-\infty, -8, -4, 0, 4, 8$	4Q
4	160.0	25.00	$-\infty, -8, -4, 0, 4, 8$	4R

3.1.1 Terminology

The terminology used in the experimental design is summarised as follows.

Condition: There were three experimental conditions: $\mathcal{WT}$=1, $\mathcal{WT}$=2, and $\mathcal{WT}$=4. Each experimental condition consisted of a number of experimental levels with different bandwidths and durations but the same bandwidth–duration products.

(Experimental) Level: An experimental level consisted of six Signal Levels, each at a different signal–to–noise ratio (including $-\infty$ dB where only noise was presented) If "Level" is used on its own, it means "Experimental Level".

Signal Level: The six signal levels within an experimental level, were set at specific signal–to–noise ratios, and each consisted of 500 waveforms (stimuli).

Replication: A replication refers to a single presentation of the 3000 waveforms (500 waveforms $\times$ 6 Signal Levels) belonging to a specific Condition and Level. Each replication was completed in 8–15 sessions.

Session: In a session, a random sample (200, 300, or 375) of waveforms were presented from one replication of one specific Condition and Level in a series of trials.

Trial: Each trial consisted of a warning interval, an observation interval where a waveform was presented, a decision interval where the observer made a rating (decision) about the evidence presented, and a reset interval.

{**Xms, YHz: ZdB**}**:** is used to concisely summarise a condition, experimental level, and signal level. If there is no signal level specified then the reference is to all signal levels.

3.1.2 Details of the experimental design

Each experimental condition used a different $\mathcal{WT}$. Values for $\mathcal{W}$ and $\mathcal{T}$ were chosen by determining the approximate maximum and minimum durations and frequencies that the experimental system could accommodate (these restrictions were determined by length of time to generate and analyse the stimuli, the cost of storage for long duration signals, the earphone response to short duration signals, and earphone characteristics over the signal frequency band). After determining the minimum bandwidth and maximum duration, bandwidths and durations for the other levels were calculated by doubling bandwidth and halving duration until the upper bounds were reached. Within each condition, there were six signal levels, at six different signal–to–noise ratios. Figure 3.3 shows these parameters in $\log_2$ and linear coordinates and Table 3.1 shows the parameters for every signal set. Figure 3.1 illustrates where these values fall on the theoretical energy detector psychometric functions, and similarly, in Figure 3.2 for the theoretical energy detector ROC curves.

A number of analyses were used. Firstly, ROC analysis (including mROC, GOC and FORCE analysis) was used to estimate detectability (Section 3.4). Secondly, analysis with the Bester correlation method examined the relationship between the sum–of–ratings from the human observers, and the performance of various ideal, and not–so–ideal, simulated observers on a stimulus–by–stimulus basis (Chapter 5). Finally, the six different signal levels were treated like a six–event experiment so that multiple–event ROC analysis could be done (Chapter 6).

3.2 Method

The experiments were run in a psychophysics laboratory consisting of a sound–attenuated chamber, with five observer booths, and a control room. The experimental system was based on both commercial and purpose–built equipment and software.

3.2.1 Observers

The three observers were all adults with normal hearing over the range of frequencies being tested. Observer 1[1] had participated in one other amplitude discrimination experiment, Observer 2 was a novice observer, and Observer 3 had over thirty years experience as an observer in psychoacoustic experiments. Observer 1 and Observer 2 trained for the task for approximately two months, completing over 30 000 practice trials, using a variety of $\mathcal{WT}=1$ signals (not the final

[1] Observer 1 was taking the potentially ototoxic drug *hydroxy–chloroquine* throughout the experiment. An independent audiogram, conducted by an ENT surgeon at the end of the experiment, indicated she had clinically normal hearing over the range of frequencies in the experiment, but 40–50 dB hearing loss at very high frequencies. Observer 2 also had clinically normal hearing as measured by an ENT surgeon six months after the experiment finished. Observer 3 had age–related hearing loss at higher frequencies.

signal set) with and without feedback. This practice time was also used to finalise experimental parameters. Observer 3 completed at least one practice session of each experimental level before beginning the experiment.

3.2.2 Experimental equipment

Figure 3.4 shows a block diagram of how the experimental system was configured. The components of the system were:[2]

Brüel & Kjær Frequency Analyzer, Type 2121: The frequency analyser was used for calibration and system analysis. The internal noise floor of the frequency analyser was estimated by reading the output on the smallest scale with no input. This self measurement read approximately 8.2 μV. The analyser was calibrated before each set of measurements.

Sound Chamber: The sound chamber is a sound–attenuated and electrically–shielded room, on the middle floor of a three–story wooden house. There are six booths with headsets and observer interface boxes, although only booth three was used for the experiments. The sound chamber characteristics were measured by Taylor (1984) using the Brüel & Kjær frequency analyser and 1 inch Brüel & Kjær condenser microphone (cartridge type 4144). Attenuation followed a negative exponential function of frequency with -15 dB at 28 Hz, -45 dB at 400 Hz, and -65 dB at 2000 Hz for airborne sound.

486DX–33 MHz Computer: The 486DX–33 MHz Computer runs the MS–DOS 6.22 operating system and is the host for the HP 68030 coprocessor (communicating on a PC–ISA bus) and Turtle Beach soundcard. In the experiment it was responsible for signal output on interrupt from the timers.

HP 68030 Co–processor: The HP 68030 Co–processor (16 MHz) is a computer on a board that is hosted in the 486DX–33 MHz Computer. It is the experimental control computer, and runs the HP BASIC 5.1 operating system. It is connected to the Programmable Timers by an HPIB and to the HP 6940B Multiprogrammer by a GPIO bus.

Programmable Timers: The six programmable timers were purpose built by the School of Psychology. They were used to control the timing of the intervals and for triggering events in the experiments, communicating over an HPIB. The timers were based on the Intel 8254 programmable timer chip. Drivers for the Timers were written in HP BASIC 5.1 by Linton Miller.

Observer Interface: The observer interface boxes were built by the School of Psychology and are mounted in each sound chamber booth. Mounted on the front of the box are two buttons and a 12 cm Systron Donner slider for observer responses, and six LED lights to indicate the timing of the intervals and observer feedback. Each light is controlled by a different timer and are software programmable to be either on or off. The entire box may be rotated and raised or lowered to suit the observer.

The slider controls a linear potentiometer. At the end of each response interval the voltage across the potentiometer is read by the Respondometer (with 2048 possible values) and converted into a rating on a 2000 point rating scale.

[2] Calibration and measurement of the equipment was not done under ideal circumstances. The Brüel & Kjær frequency analyser used for some of the analyses required a BNC input so the testing of any equipment with balanced–line output required going through a Line I/O before measurement, which added noise. The Turtle Beach ADC–DAC, used for some measurements, was also not independent of the system because the DAC was used to generate test signals and the ADC to take samples.

Noise floors were estimated using the Brüel & Kjær frequency analyser, with 8 kHz low–pass filtering, so the noise–floor was assumed to be distributed over only 8 kHz. Harmonic distortion was estimated by generating a tone with the Turtle Beach DAC, repeatedly sampling the system with the Turtle Beach 16 bit ADC, then calculating the spectral–averaged FFT. The harmonics were then estimated graphically. This was not an ideal test, because the test signal was not completely independent from the system. The 16 bit ADC, however, allowed much finer measurement than the other 12 bit ADCs available.

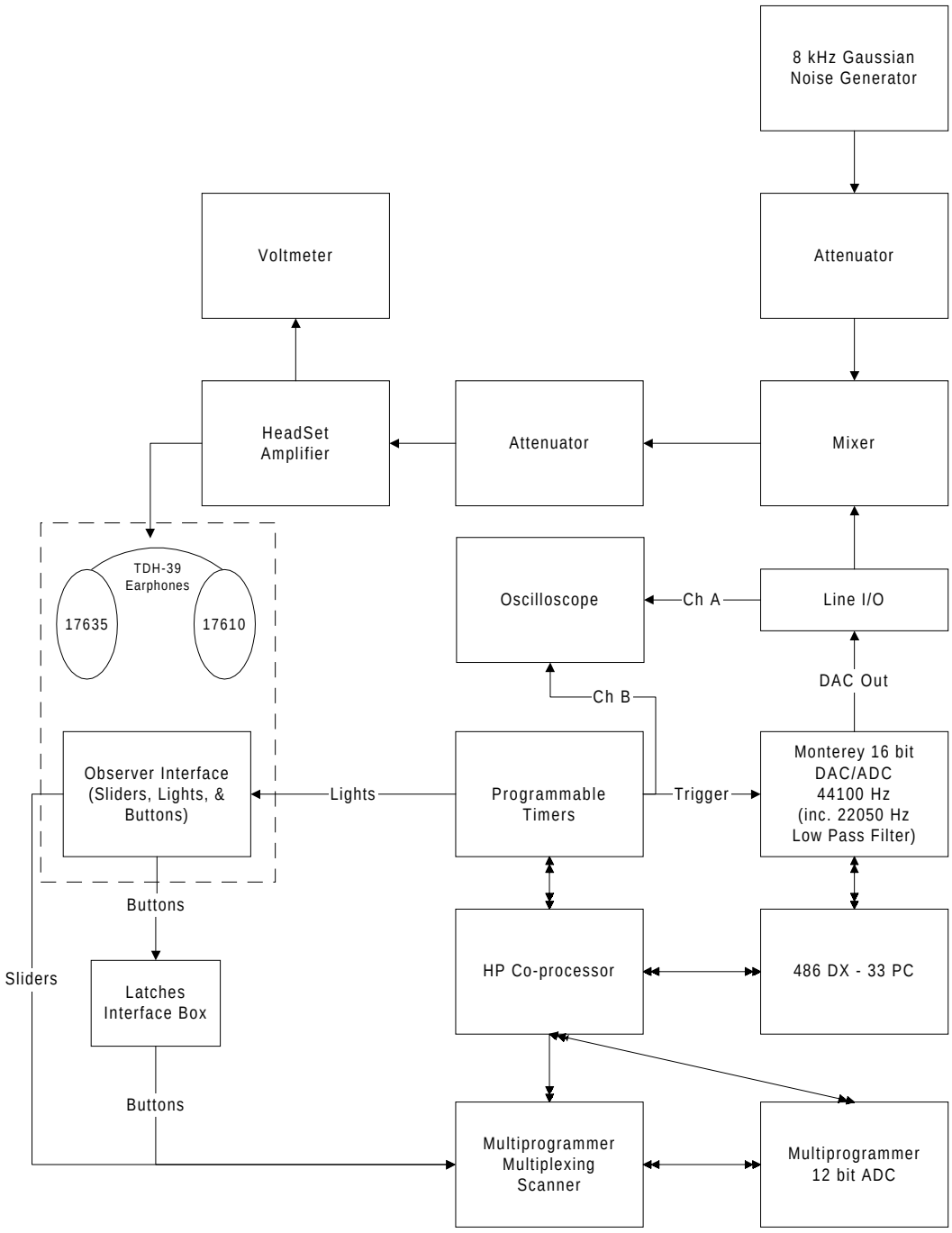

Figure 3.4: Block diagram of the experimental equipment. Note that the equipment within the dashed box is in the sound–attenuated chamber. The other equipment is in the control room.

Respondometer: The Respondometer reads voltages from the buttons and sliders in the sound chamber. It consists of an HP 69442A High Speed A/D Voltage Converter Card (12 bit ADC, clocked at 20 kHz), an HP 69336A High Speed Scanner Card (a sixteen channel multiplexer reading a maximum of 20 000 channels per second), both housed in an HP 6940B Multiprogrammer, and interface circuitry (latches) in a purpose built box. Drivers for the Respondometer were written in HP BASIC 5.1 by Linton Miller.

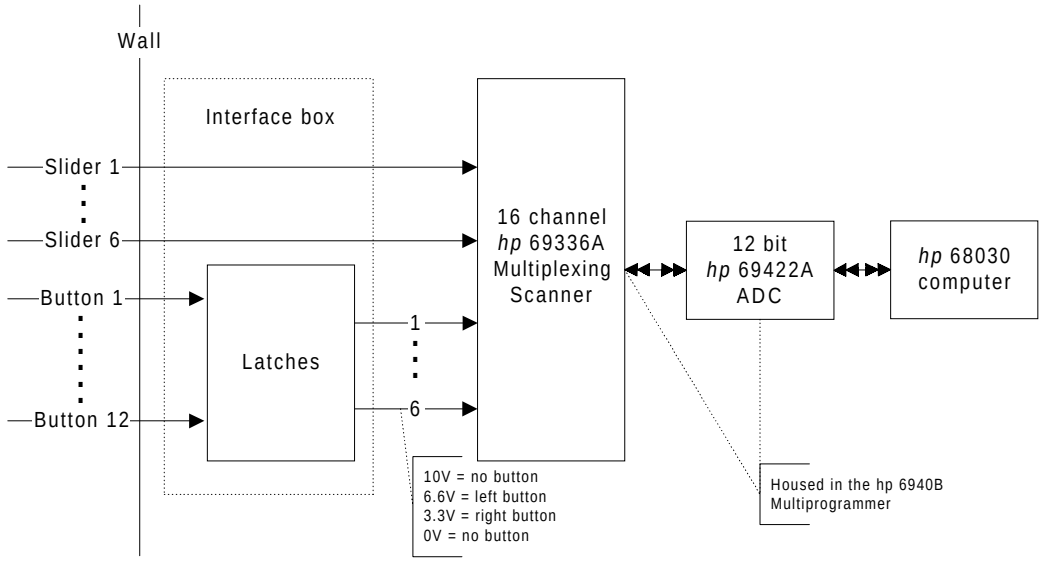

Figure 3.5: Functional relationships among the components of the Respondometer.

Figure 3.5 shows the functional relationship among the components. The cable from the observer interface to the scanner carries voltages from the sliders (6 lines) and buttons (12 lines). Voltages from the buttons are latched until read and then they are cleared. The scanner multiplexes across each line so the ADC can read each voltage.

Gaussian Noise Generator: The 8.1 kHz (AERB) low–pass Gaussian Noise Generator was purpose built by the School of Psychology. As shown in Figure 3.6, the passband is flat out to about 8 kHz and the roll–off after 8 kHz is about 24 dB per octave. The temporal–averaged phase spectrum tended to zero, indicating random phase components.

Turtle Beach *Multisound Monterey* sound card: The *Monterey* sound card system consisted of the Turtle Beach *Tahiti* sound card with a *Rio* midi synthesiser daughter board. The *Rio* was removed because it was not required and introduced unnecessary noise into the system. The *Tahiti* sound card consisted of 16 bit ADC and DAC chips. It was housed in the 486DX–33 computer and was connected on the PC–ISA bus. The DAC system consisted of a 128× interpolator, a fourth order Delta–Sigma modulator, a 16–bit DAC, and an analog low–pass filter. The board was controlled by Motorola 65001 DSP. Software drivers for both the ADC and DAC were written in Motorola 65001 code and BORLAND PASCAL 7.0 by Linton Miller.

In the experiments, the DAC was used to convert the signal buffers to an analog waveform at a clocking rate of 44 100 Hz (note that the sound card automatically low–pass filtered the output to 20 kHz using a second–order Butterworth filter). The ADC was used for system analysis.

Line I/O: The Line I/O, used to convert BNC lines to three–pin balanced lines and vice–versa, was purpose built in the School of Psychology. The Line I/O balanced line to BNC unit had a noise floor of approximately 13–14.5 μV. The BNC to balanced line unit, however,

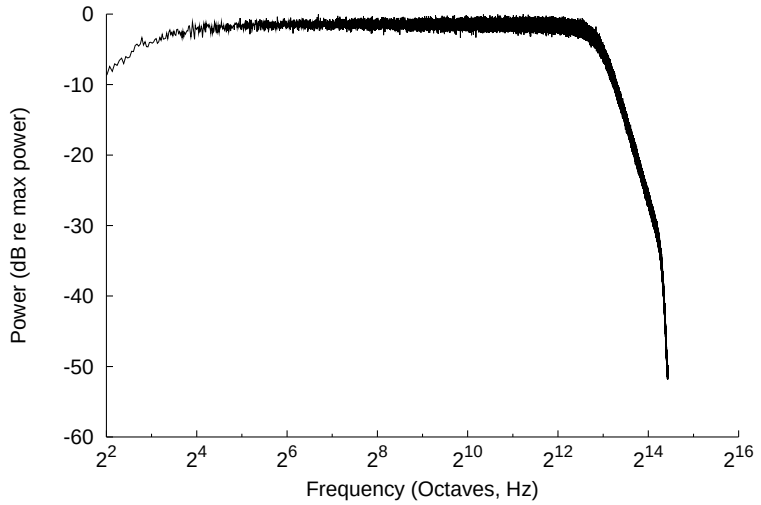

Figure 3.6: Amplitude spectrum of the Gaussian noise generator set at 8 kHz. 107 six–second samples were taken at a clocking rate of 44 100 Hz and analysed with a 2^{18} point FFT.

could not be measured independently, because the frequency analyser required a BNC input. Thus, the combined BNC to balanced line to balanced line to BNC gave a noise floor of approximately 17.5 μV. This part of the system is usually driven at 2 V-rms, thus the noise floor is at least 140 dB down (or 134 dB down re 1 V-rms).

Hatfield Step Attenuators, Model 2050: Two attenuators were used: one to attenuate the background masker prior to mixing, and the other to attenuate the mixed signal and continuous masker. The step attenuators added no measurable noise to the system.

Mixer: The mixer was purpose built by the School of Psychology. It was used to mix the signal and the continuous background masker. The free–floating noise floor of the Mixer was measured at 40 μV-rms (-133 dB re 2 V-rms, -127 dB re 1V-rms) using the Brüel & Kjær frequency analyser via the Line I/O.

Headset Amplifier: The diotic/dichotic headset amplifier was purpose built by the School of Psychology. In the experiments, the amplifier was run in diotic mode meaning the same signal was output to both headphones with no phase shift. The free–floating noise floor of the headset amplifier was measured as approximately 14–15 μV at the test point, with the output loaded with the TDH–39 100 Ω earphones, and 16 μV at the measure output (used for calibration). This equates to a noise floor of approximately -113 dB relative to 73 mV (the voltage of the calibration signal) and -135 dB re 1 V-rms.

TDH-39 100 Ω Earphones: The earphones were fitted with Rudmose Tracor RA125 Otocups and MX41/AR cushions. Rudmose Otocups were rated the best for using with TDH-39 earphones and MX-41/AR cushions, out of those tested by Copeland and Mowry (1971). The earphones were driven by the Headset Amplifier. The two earphones were matched for frequency response and calibrated such that gain for each earphone was equal at the frequency of interest (500 Hz).

The earphones were calibrated using a Brüel & Kjær Type 4152 artificial ear, the Brüel & Kjær Type 2121 frequency analyzer, and a 1 inch Brüel & Kjær Cartridge Type 4144 condenser microphone. Test signals were generated using the 16 bit DAC. To calibrate, the earphone was removed from the otocup, then plugged into the headset amplifier, and clamped to the artificial ear using 500 g of pressure. A continuous 500 Hz tone was generated, attenuated to 316 mV-rms, and passed to the headset amplifier. The appropriate channel (i.e., left ear or right ear) of the headset amplifier was trimmed until the desired SPL of 106 dB was obtained. The procedure was then repeated for the other earphone and channel.

Figure 3.7 shows the sound pressure level from each earphone over a number of test frequencies after trimming.

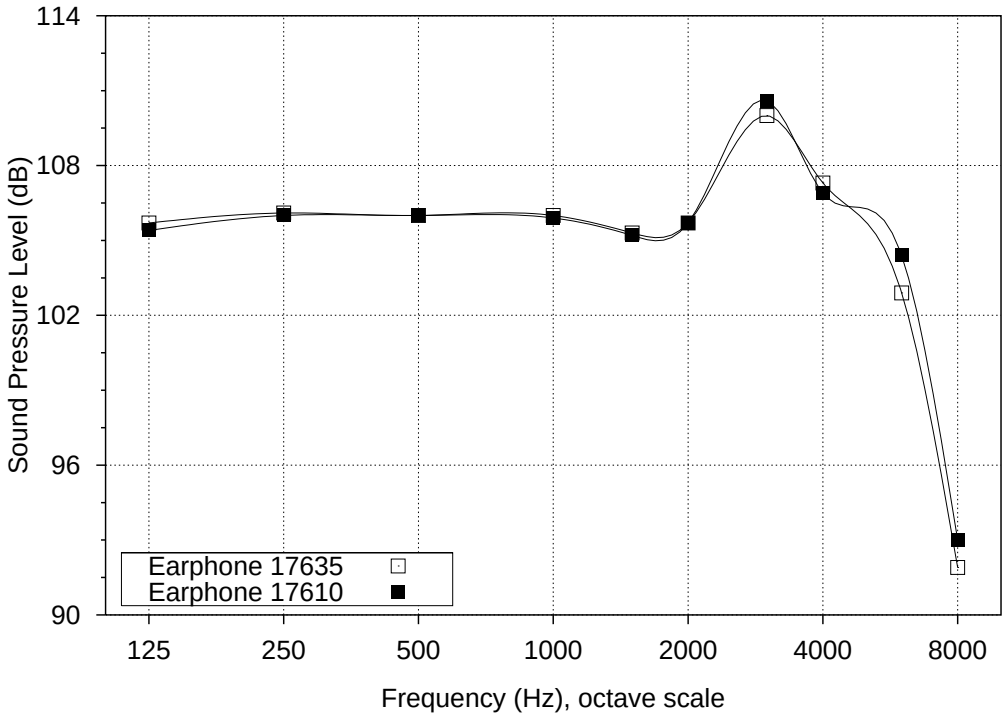

Figure 3.7: Calibration curves for the two TDH-39 earphones used in all experiments. Points are interpolated with a cubic spline.

Systron Donner Digital Multimeter, Model 7344A: The multimeter was used to calibrate the test signal and masker voltage. The signal was calibrated first, using one Hatfield step attenuator, then it was unplugged and the continuous masker was calibrated using another attenuator (without changing settings on the first attenuator).

Entire signal output system: When driving the complete system with a 2.8 V-peak, 2.5 kHz sinusoid, stopped down to 0.4868 V-rms at the headset amplifier, the first harmonic was measured at 79 dB down from the fundamental. Total harmonic distortion was estimated to be -76 dB (0.0158%) and the noise floor was approximately -108 dB relative to the fundamental.

3.2.3 Experimental control

Four separate, but interrelated, computer programs controlled the running of the experiments: RANDOMBOT (Lapsley Miller, 1997), RUNITBOT, (Lapsley Miller & Miller, 1997), SIGNALS (Miller, 1997c), and ANALBOT (Lapsley Miller, 1998a).

RANDOMBOT was written to (a) create all the data files that the programs RUNITBOT, SIGNALS, and ANALBOT used to run the experiment, (b) create the signal and event combinations for each trial, then randomise the presentation order (with event run–limiting), and (c) create a randomised or counterbalanced order for replications in an experiment. A linear congruent pseudo–random–number–generator was used for randomisation (Scurfield, 1994).

RUNITBOT was the main experimental control program. Its functions were to (a) read in the files that described the current experimental session, (b) establish experimental parameters and

calculate the trial sequence, (c) program the timers to sequence the intervals, (d) spawn the program SIGNALS so the appropriate waveforms were output during the observation interval, (e) convert the slider voltage read from the respondometer into a rating between 1 and 2000, adjusted for maximum and minimum voltage on the line, (f) determine when the slider was correctly reset, by repeatedly polling the respondometer, before beginning the next trial, (g) calculate the measures of detectability and display ROC curves of the current *session*, for observer feedback, and (h) save the observer's ratings and other data to file.

The functions of SIGNALS were to (a) output to the DAC a continuous, unwindowed, sine wave of a specific frequency, amplitude and phase for calibrating the system before starting an experimental session, (b) load the appropriate random signal sequence using the information passed by RUNITBOT, (c) read in the signal buffer for the next trial from file (during the previous trial's decision interval, or before the session began if it was the first trial) then output it to the DAC during the observation interval of a trial, when triggered by the programmable timers. For practice trials this may also entail randomly choosing a waveform on–the–fly.

ANALBOT performed the majority of the data analyses, including (a) ROC, mROC, and GOC analyses, and (b) the calculation of attenuations from the theoretical (χ^2) energy detector, and the fitted polynomial full–linear psychometric functions. Results were output to file so they could be easily used for LaTeX tables and GNUPLOT graphics (Williams & Kelley, 1997).

Timing within a trial

An experimental session consisted of a set number of trials. Each trial consisted of a series of intervals: warning, observation, decision, and reset (possibly with feedback). The timing of the intervals on each trial was controlled by the programmable timers.

The first timer unit, which controlled the warning interval light, was triggered by RUNITBOT. Subsequent software triggering of each experimental interval occurred on interrupt from the timers at the end of the previous interval. Hardware triggers from each Timer unit were used to trigger other events, such as the output of the signal (under control of the SIGNALS program) in the observation interval, the activation of the appropriate light on the Observer Interface box, and oscilloscope displays. RUNITBOT and SIGNALS ran independently, but were synchronised before each trial began.

Figure 3.8 illustrates the functional timing relationships among components in the course of an experimental trial.

Replication and session design

Observer 1 and Observer 2 completed conditions $\mathcal{WT}$=1, $\mathcal{WT}$=2, and $\mathcal{WT}$=4 consecutively, but, within a condition, the order of sessions was randomised over experimental level with the constraints that (a) all sessions of a replication for all levels were completed before starting on the next replication and (b) within a replication, each session was a different level from the previous session. For Observer 3, sessions were also randomised over condition. The differences between the observers occurred because there had been delays with signal generation so only the $\mathcal{WT}$=1 signals were available when Observer 1 and Observer 2 were due to start data collection. Observer 3 started data collection after Observer 1 and Observer 2 had finished so all signals were available. From the observers' perspective, they listened to a different experimental level every time they ran a session.

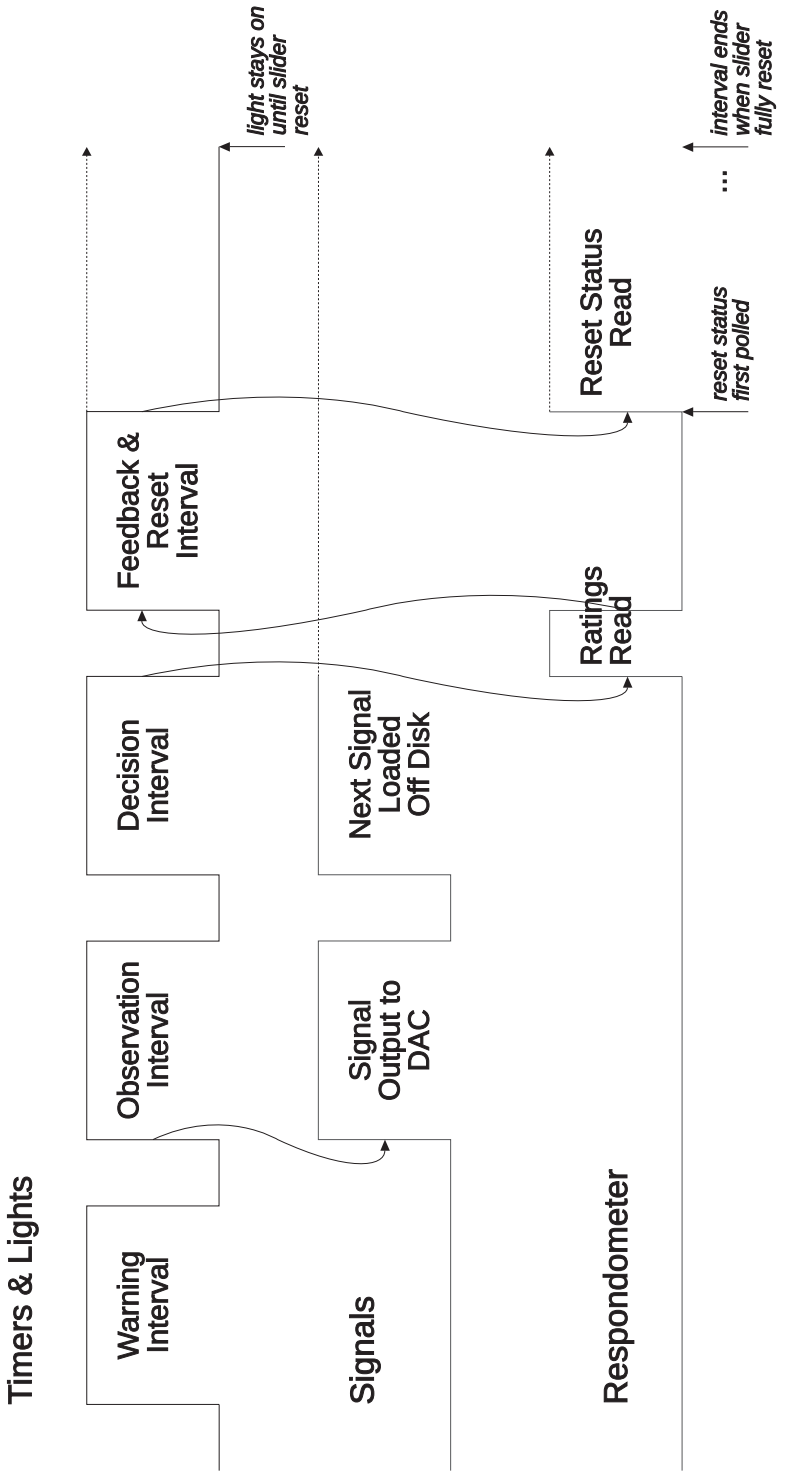

Figure 3.8: Functional timing relationships within one trial among the intervals (controlled by the timers), signal loading and output (controlled by SIGNALS), and the reading of responses (controlled by RUNITBOT).

Trials for the six signal levels (one $\mathcal{N}$ event and five $\mathcal{SN}$ events) were intermixed within a session using the multi–signal–level rating–scale procedure described in Tucker, Evans, and Jeffress (1967). Thus, fewer $\mathcal{N}$ trials were needed if all signal levels were run together, instead of separately. The only potential problem was that the resulting false–alarm rate was then common across the hit rates for each signal level.

The number of trials in a session was tailored to the observer to minimise fatigue and boredom while still maximising the number of trials completed daily (the more sessions run, the longer each replication took, because there was some overhead with starting and stopping each session). Observer 1 and Observer 2 completed a level of each $\mathcal{WT}$=1 replication in eight sessions of 375 trials, and each $\mathcal{WT}$=2 and $\mathcal{WT}$=4 replication in ten sessions of 300 trials. Observer 3 completed a level of each replication in fifteen sessions of 200 trials. A session took 10–20 minutes to complete depending on the duration of the current set of waveforms, the number of trials, the length of the trial intervals, and whether the observer took small pauses (micro–breaks).

Each session began with a set of practice trials. Observer 1 and Observer 2 listened to 24 practice trials in $\mathcal{WT}$=1 (two signals at each signal–to–noise ratio starting with the highest, stepping down to the lowest and then up again), and 12 practice trials in the subsequent experiments (only one signal at each signal–to–noise ratio). Observer 3 listened to seven practice trials in all sessions: alternating the highest signal–to–noise ratio signals with noise–alone signals (starting and ending with a signal–plus–noise).

The observers typically completed 800–1500 trials a day. Data collection continued for about eight months for each observer, requiring at least a couple of hours commitment per day, every day.

Observer 1 and Observer 2 both completed six replications of each level and condition, resulting in 324 000 trials in total per observer. Observer 3 completed three replications of eight levels: $\mathcal{WT}$=1: {25ms, 40Hz} and {6.25ms, 160Hz}, $\mathcal{WT}$=2: {400ms, 5Hz}, {50ms, 40Hz}, and {12.5ms, 160Hz}, and $\mathcal{WT}$=4: {400ms, 10Hz}, {100ms, 40Hz}, and {25ms, 160Hz}, resulting in 72 000 trials. In total, 720 000 decisions were collected. These decisions (ratings) formed the basic data set.

Trial Design

The key to running multiple replications of an experiment is to use different randomised trial sequences for each replication. Order effects then become part of the unique noise and not the common noise.

The program RANDOMBOT was used to pre-calculate these trial sequences. All event and signal combinations were calculated then a pseudo–random number generator was used to randomly sort the sequence. The last number at the end of a replication was used to seed the generator for the next replication. Additionally, two tests were made. Firstly, a check was made at the end of each replication to see if the seed chosen for the next replication had been used before. Secondly, every trial was event run–limited.

Event run–limiting is needed because humans tend to be poor random number generators. That is, very long sequences of one event are not seen as random, and problems such as the gambler's fallacy may result. The run–limit should not, however, be a fixed number—especially if trial–by–trial feedback is used (because the observer may realize this and count). To stymie the observer, the maximum run–limit should be randomised *on every trial*. For instance, on one

trial the event run–limit may be three, but the next trial may be five. Thus, the observer cannot predict the next event. In these experiments the minimum run–limit was three and the maximum run–limit was five.

3.2.4 Procedure

Table 3.2: Interval durations for each observer.

	Interval Durations (ms)		
Observer	Warning	Decision	Reset
1	50	1250	300
2	100	1250	300
3	300	1350	250

Table 3.2 presents the durations for each interval in a trial for each observer. At the beginning of each trial, an amber LED indicated the warning interval. This interval was immediately followed by an observation interval (indicated by a red LED). The length of the observation interval was equal to the absolute duration of waveform, which varied with experimental level. On each trial a randomly chosen stimulus was presented to the observer diotically[3] through the earphones, simultaneously with the observation interval light. The observer then made a rating by setting the slider in the decision interval (indicated with a green LED) and then a flashing red LED indicated that the rating had been recorded and that the slider should be reset. The reset interval had a minimum duration to help pace the observer so that the trials were not rushed. The next trial began once the slider was correctly reset to the left. The observer could take micro–breaks at any stage by not immediately resetting the slider. Breaks were encouraged to reduce fatigue and to minimise repetition strain injuries.

The Observers' Task

The observers' task, in each trial, was to detect if a signal was presented in the observation interval. The observer was asked to move the slider to the right–hand–side of the scale if they were certain that a signal was presented and to the left–hand–side of the scale if they were certain that only noise was presented. They were encouraged to use as much of the scale as possible and to be consistent in how they used the rating scale.

3.2.5 Stimuli

The signal generation and analysis is detailed in Section 3.3 and Appendix C. Briefly, signals for each $\mathcal{WT}$ combination were generated and mixed with gated wide–band Gaussian noise (4300 Hz ERB, 4000 Hz $\mathrm{Ess}\mathcal{W}_{92.4\%}$) of the same duration as the signal. All signals were presented to the observers mixed with a continuous wideband Gaussian noise masker (8102 Hz AERB, 8310 Hz $\mathrm{Ess}\mathcal{W}_{92.4\%}$).[4]

[3] Langhans and Kohlrausch (1992) considered the difference between monaural and diotic tasks and found that in some circumstances the diotic masked threshold is lower than the monaural threshold by a few decibels. They surmised that this was due to less internal noise in the diotic case, because a comparison could be made across ears. This would be a good experiment to replicate using GOC analysis to see whether asymptotic performance is the same.

[4] See Section 3.3.1 for definitions of these bandwidth measures.

Signal sets for the six signal levels of each experimental level and condition were stored as separate files on CD–ROM and transferred onto the hard–disk of the experimental computer at the beginning of each experiment. The points making up each waveform were stored as 16–bit codes such that they could be output directly to the sound card. During a trial, a pre–chosen random waveform was loaded off disk, and output to the 16–bit DAC at a clocking rate of 44 100 kHz, then mixed with a continuous background Gaussian–noise masker (8 kHz, low–pass).

Calibration

The noise–alone masker level was set to 60 dB SPL per unit cycle, and the continuous background noise masker level to 20 dB SPL per unit cycle . Thus, the masker to continuous background masker signal–to–noise ratio was 40 dB. The maximum signal signal–to–noise ratio was 16 dB so the maximum SPL produced was 96 dB per unit cycle. Although this SPL may seem high, the amount of energy involved was small, because the durations were relatively short.

At these levels the continuous background masker was able to effectively mask much of the unwanted environmental noise passed by the sound–attenuated chamber while contributing less than 0.01% to the total masking power.

Prior to each session, the signal generation system was manually calibrated by (a) generating a continuous 500 Hz sinusoid and setting the attenuators so the sinusoid was 73 mV-rms at the headset amplifier (this resulted in noise of 10 mV-rms). The continuous masker was calibrated to 1.4 mV-rms.

3.3 Signal generation and analysis

Chapter 1 considered previous experimental and theoretical research into the detectability of Gaussian noise waveforms, discussed ways of representing this type of waveform, and considered the resulting limiting case or uncertainty principle. To investigate the detectability of this class of waveforms experimentally, a procedure was needed to generate Gaussian noise waveforms of a given $\mathcal{WT}$, but for a variety of bandwidths, durations, and signal–to–noise ratios.

The section starts by describing how to define and measure $\mathcal{W}$, $\mathcal{T}$, $\mathcal{WT}$, and signal–to–noise ratio for digital waveforms. Then a novel signal generation method (Drga, 1997b, 1998) that was used to generate the signals for the experiments and simulations, is described. Finally these signal sets are analysed to determine whether they fulfil the requirements of approximating band and time limited Gaussian waveforms of the desired bandwidth, duration, $\mathcal{WT}$, and signal–to–noise ratio.

3.3.1 The frequency domain

Characterising the spectrum of a waveform, or set of waveforms, generally means estimating the frequencies over which the waveform is most concentrated. This estimate is not necessarily straightforward. If the spectrum is rectangular then the bandwidth is the width of the spectrum. Finite duration spectra, however, can not be rectangular and many definitions of bandwidth may justifiably be used.

Measures of bandwidth

There are a number of measures of bandwidth. Some are more useful for theoretical problems (e.g., the Heisenberg bandwidth), and some are more useful for practical problems (e.g., the TdB bandwidth). Choosing which one to use for a particular application can be difficult, especially when the application has both theoretical and practical components.

Three dB bandwidth: The three dB (TdB) bandwidth is calculated by finding the frequency with maximum power, then to the right (and left for band–pass spectra) finding at which frequency the spectrum is 3 dB down from the maximum. Obviously this only makes sense if the spectrum has a single peak and does not have much ripple in the passband. Likewise, bandwidths for six, or any other, dB points may be measured but three is the most common standard, because the TdB points are at half the power of the maximum.

Equivalent rectangular bandwidth (ERB): The ERB is defined as the sum of the absolute values of the power spectrum divided by the maximum power value and multiplied by the fundamental frequency. In other words, a rectangle is calculated that has the same maximum value and the same area as the spectrum, and the bandwidth is then derived from the definition of that rectangle.

Average equivalent rectangular bandwidth (AERB): For real signals the ERB can be distorted if the passband is noisy. This is because the *maximum* value is used to determine the height. If we can assume that the spectrum is a result of sampling from a distribution with a smooth spectrum, then the AERB may be used instead. This method was developed in conjunction with Vit Drga. It can be seen as a way of estimating the population spectrum from sample spectra by averaging out the error.

Instead of taking the peak value, the average value of the *flat part* of the pass band is used instead. The flat section could be determined heuristically, but at the moment it is estimated through visual inspection of the spectrum. Calculation is then done like the ERB, but using the average value instead.

The spectrum of a Gaussian noise waveform, windowed with the Kaiser window, is flat through part of the passband for waveforms of $\mathcal{WT} > 1$ so the AERB may be more appropriate as a measure of bandwidth. For $\mathcal{WT}$=1, the shape of the spectrum is dominated by the spectrum of the window and is smoother with a single peak so the ERB may be more appropriate. See Appendix C for examples.

Essential bandwidth (Ess$\mathcal{W}_{\beta^2\%}$): There seems to be no definition of the essential bandwidth for discrete–time in the literature. Based on the definition of essential bandwidth for continuous frequency (Equation 1.3) by Landau and Pollak (1961), for a given proportion of energy constrained (β^2), a heuristic algorithm was developed. It assumes that the spectrum has a single peak (although there may be small local maxima) and that the spectrum is concentrated in frequency. According to Miller (1998, personal communication) this algorithm will result in the minimum bandwidth. The algorithm will work for spectra of other shapes but the interpretation of the resulting bandwidth may be more difficult. The steps are:

- Decide on the proportion of energy to be constrained, e.g., 95%.

- Calculate the total energy in the power spectrum by summing the discrete values in the spectrum.

- Find the bin with the largest value and calculate its proportion of energy relative to the total energy. If it already contains more than requested, then the bandwidth may be determined by linear interpolation.

- If more energy is required, check the point to the left and to the right of the maximum point and choose the largest. Repeat the previous step to determine whether all the energy is contained in those points. If not, set a left/right marker on that index and a right/left marker on the original bin. Look to the bin right of the right marker and left of the left marker and choose the largest. Repeat until the energy equals that required.

- The bandwidth is the width between the current left point and the current right point (or proportion of that point that results in the energy required).

The algorithm, in pseudo–code, is presented in Appendix B. Depending on the application, the DC component may, or may not be, important. The algorithm, as presented, includes the DC component by looping from zero, but this may be changed to one if necessary.

3.3.2 The time domain

Presenting a timelimited waveform to an observer is not as simple as turning a waveform on and off, because the process of switching causes undesirable spread of the waveform's spectrum (Harris, 1978; Nuttall, 1981). This is a serious problem for psychophysicists, for instance, Zwicker and Fastl (1972) showed that signals that were poorly designed in the time domain, led to erroneous conclusions about the nature of the critical band.

The extent of the spread in the frequency domain, however, can be minimised by using non-rectangular switching, or *windowing* in the time domain. There are many different windows and the best one depends on the application. According to Harris (1978) and Geckinli and Yavuz (1978) the Kaiser window (Kaiser, 1966) has a number of desirable properties in both the time and frequency domain.

Measures of duration

There have been a number of measures of duration suggested in the literature including the Heisenberg duration, essential duration ($\mathrm{Ess}\mathcal{T}_{\%}$), equivalent statistical duration (Bendat & Piersol, 1986), equivalent rectangular duration (ERD), squared–equivalent rectangular duration (ERD2), and according to Gerken et al. (1990): the time interval from stimulus onset to offset, the half–power points, the temporal interval containing 90% of full amplitude, and the on–off interval used to control an electronic switch. Gerken et al. (1990) concluded that the controversy is how to include the transitional rise/fall portion of the waveform. Algorithms for the calculation of the various duration measures, when using digital signals, are not always as forthcoming. The algorithms used in this project are described, then the duration measures are compared. Note that all these duration measures may be calculated on a data window or a waveform, but the waveform must be rectified first.

Equivalent rectangular duration (ERD**):** The continuous time ERD is defined as the integral of the absolute values of the time series divided by maximum power. It may also be defined in terms of the cross–correlation function. In discrete time the sum of the absolute values of the time series is divided by the maximum power value and multiplied by the time–base. In other words, a rectangle is calculated that has the same maximum value and the same area as the time series and the duration is then defined as the base of the rectangle.

Drga's ERD**2:** Drga (personal communication) has suggested a new measure of duration that is symmetrical with the ERB (see Section 3.3.1). Like the ERB, this measure is based on the total power in the time domain divided by the maximum power and multiplied by the time–base. The difference between the two measures is the rectification assumed. The result is a narrower measure of duration compared with the ERD.

Essential duration ($\mathrm{Ess}\mathcal{T}_{\alpha^2\%}$**):** Like the essential bandwidth, there appears to be no definition of the essential duration for discrete–time in the literature. Based on the definition of essential duration for continuous time (Equation 1.4) by Landau and Pollak (1961), for a given proportion of energy constrained (α^2), a heuristic algorithm was developed for discrete time. This algorithm is virtually identical to that for the essential bandwidth, except that the discrete values in the time–series are squared first. The algorithm, in pseudo–code, is presented in Appendix B.

Absolute duration: The absolute duration of a discrete time series is simply the number of points in the time series times the time–base. It is not necessarily a good measure of duration, because much of the time series could be essentially zero. It may be used for defining the total number of points in a digital waveform. For the Kaiser window, the ERD and $\mathrm{Ess}\mathcal{T}_{\%}$ may be calculated from the absolute duration, or vice–versa, by multiplying with a scalar.

The Kaiser window

In Chapter 1, it was suggested that a useful definition of duration for psychophysical applications may be the essential duration, because it is defined with respect to the spread of energy in the time domain. Landau and Pollak (1961) showed that the essential bandwidth and duration could be used to define an acoustical uncertainty principle, and, as a result, the prolate spheroidal wave function family constrained the most energy in both domains. Kaiser (1966) suggested that a prolate spheroidal wave function may, therefore, be useful as a data window, because it would maximise the energy contained in a waveform, compared with other data windows.

Kaiser has provided one solution by deriving an approximation to the prolate spheroidal wave function window. This approximation is accurate and easy to use, because it has only one parameter (see Kaiser, 1966; Geckinli & Yavuz, 1978; Harris, 1978; Nuttall, 1981; Rabiner & Gold, 1975). Other solutions include a window based on the discrete prolate spheroidal wave function (Barbosa, 1986). The characteristics of Barbosa's window are apparently very similar to the Kaiser data window, but it is more difficult to calculate.

The continuous time Kaiser window (Harris, 1978; Kaiser, 1966; Nuttall, 1981) is a family of weighting functions, $k(t, \kappa)$,

$$k(t,\kappa) \quad = \quad \frac{1}{L} I_0(\kappa\sqrt{1-(2t/L)^2}), \qquad |t| < L/2. \tag{3.1}$$

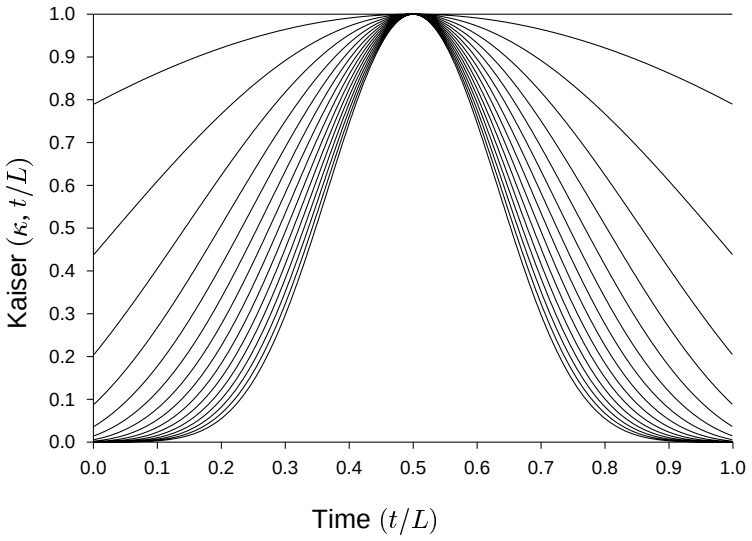

Figure 3.9: Examples of the Kaiser window family for $\kappa = 0$ to $\kappa = 15$.

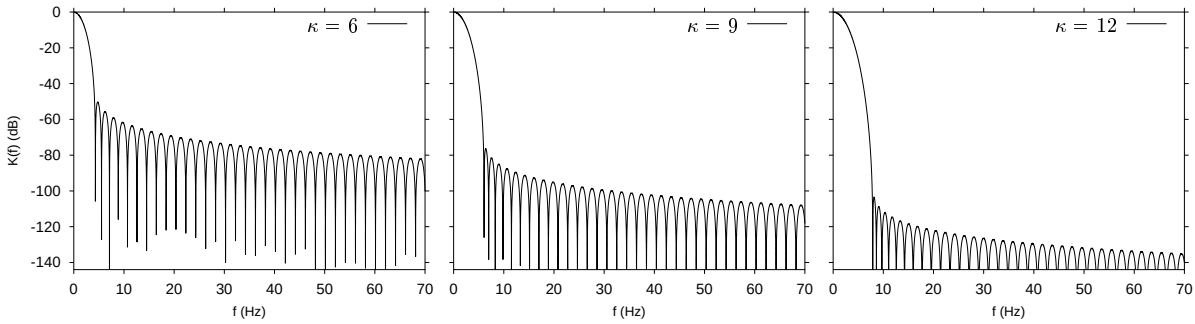

Figure 3.10: Examples of the Kaiser window spectra for $\kappa = 6$, $\kappa = 9$, and $\kappa = 12$, for a 500 ms window. Note how the main lobe widens and the side lobes drop as κ increases.

where I_0 is the modified Bessel function of the first kind and order zero, L is the absolute duration of the window and κ is a parameter taking on useful values from 0 to about 15.[5] In the frequency domain, the spectrum $K(f)$ is

$$K(f,\kappa) \quad = \quad \frac{\sin(\sqrt{\pi^2 L^2 f^2 - \kappa^2})}{\sqrt{\pi^2 L^2 f^2 - \kappa^2}}, \qquad \text{for all } f. \tag{3.2}$$

Figure 3.9 shows the Kaiser window family for $\kappa = 0$ (horizontal line) to $\kappa = 15$ (narrowest window) for a standardised duration. Figure 3.10 shows the spectrum of the Kaiser window for three values of κ. By varying κ, time domain and frequency domain characteristics may be traded-off. As κ gets larger the window gets narrower, and the main lobe of the spectrum gets wider as the side lobes get lower. This is a direct effect of the acoustical uncertainty principle. Choosing κ depends on the application, but in general, values between 8–12 give good results in both domains.

[5]The parameter κ is more commonly known as β in the literature but this would cause confusion with β^2 in the definition of essential bandwidth.

In discrete time, the Kaiser window $k(n)$ is defined as (Galvin & Whitmore, 1986)

$$k(n,\kappa) = \frac{I_0\left[\kappa(1-\{2n/(N-1)\}^2)^{\frac{1}{2}}\right]}{I_0(\kappa)} \qquad |n| \leqslant \frac{N-1}{2} \tag{3.3}$$

Note that continuous time and discrete time versions are equivalent if $\tau = \frac{N-1}{2}$, $n = t$, and $\kappa = \omega_a \tau$. Similar definitions of the discrete–time window are given by Harris (1978) and Nuttall (1981).

ERD and $\mathrm{Ess}\mathcal{T}_{\%}$ of the Kaiser window

Galvin and Whitmore (1986) derived the ERD of the Kaiser window because they needed a way of defining the duration of a set of signals that had been windowed. They stated that "For stationary processes, windowing reduces the energy of a time series in proportion to the area under the window". They also suggest that the ERD is the most appropriate measure of duration for calculating the signal–to–noise ratio, or the bandwidth–duration product. This may be true for practical applications, but, at least historically, there is little evidence in the literature of the use of ERD for deriving or measuring $\mathcal{WT}$. In fact, in the psychophysical literature, duration is often never defined, or is defined as the absolute duration of the signal.[6]

Their derivation showed that the ERD is a constant proportion of the duration of the window for a given value of κ. They did not, however, explicitly relate this to the energy content, which they claimed was proportional to the area under the window. The derivation itself is clear. They start, however, from the discrete Kaiser window, but then derive the ERD of the continuous Kaiser window without saying how they are related. A table of proportions of duration, for a number of κ, is presented. They then make the assumption that the ERD of a windowed stationary time–series is equivalent to the ERD of the window alone.

Despite these concerns, the derivation is extremely useful, because it shows that to calculate the ERD, the proportion for a given κ is simply multiplied with the absolute duration. In practical terms it works because the ERD of the window is, on average, the same as the rectified temporal–averaged time–series of a set of signals.

It is difficult to derive the essential duration for a given α^2 using similar techniques to Galvin and Whitmore (1986). Therefore, numerical approximations were used to compare the absolute duration, the ERD, and the essential duration (for α^2= 0.05 to 1.0 in steps of 0.05) for the discrete–time Kaiser window.

The standardised ERD, as shown in Table 3.3, for each value of κ (0–15) for the Kaiser window was the same, to 5 decimal places, of those tabulated in Galvin and Whitmore (1986). The standardised $\mathrm{Ess}\mathcal{T}_{\%}$ varied in a similar way, except for when $\kappa = 0$ (where $\mathrm{Ess}\mathcal{T}_{\%}$ equalled α^2, but the ERD equalled 100%). To calculate the $\mathrm{Ess}\mathcal{T}_{\%}$, the proportion for a given α^2 and κ are simply multiplied with the absolute duration. These relationships are shown graphically in Figure 3.11, and are tabulated in Tables C.1 and C.2 in Appendix C.

There is an interesting comparison also, which can be seen in Figure 3.11, between essential duration and ERD: for $4 < \kappa < 12$, the ERD constrains a similar amount of energy (about 0.924) giving a very similar performance to an essential duration with $\alpha^2 \approx 0.92$. For most practical pur-

[6] According to Whitmore (1999, personal communication), the Bartlett window was often used, where the absolute duration was equal to the ERD. The absolute duration was not that of the signal, however, but the time between the onset and offset of the gate.

Table 3.3: The ERD and $\mathrm{Ess}\mathcal{T}_{92.4\%}$ of the Kaiser window, as a proportion of absolute duration, and as a function of κ. The ERD values are the same as stated in Galvin and Whitmore (1986) to 5 d.p.

κ	ERD	$\mathrm{Ess}\mathcal{T}_{92.4\%}$
0	1.00000	0.92400
1	0.92823	0.89955
2	0.79551	0.81477
3	0.68417	0.70051
4	0.60365	0.61110
5	0.54481	0.54825
6	0.50002	0.50188
7	0.46461	0.46578
8	0.43574	0.43657
9	0.41164	0.41228
10	0.39113	0.39165
11	0.37340	0.37384
12	0.35788	0.35826
13	0.34414	0.34449
14	0.33187	0.33219
15	0.32082	0.32112

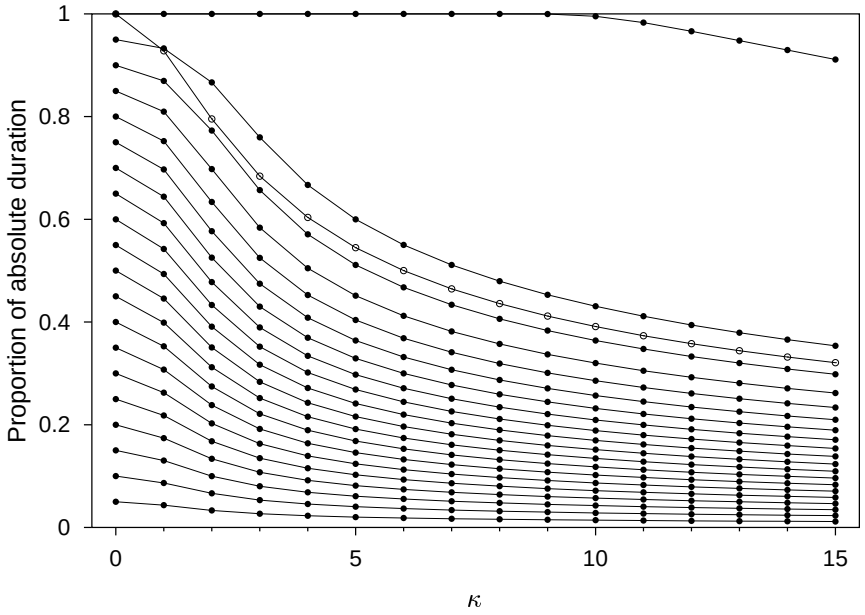

Figure 3.11: The essential duration of the Kaiser window, as a proportion of absolute duration, and as a function of κ. Each filled circle line represents a different proportion of energy constrained for the essential duration with the bottom line representing $\mathrm{Ess}\mathcal{T}_{5\%}$, increasing in steps of 5% to $\mathrm{Ess}\mathcal{T}_{100\%}$ at the top. The open circle line represents the proportion of absolute duration for the equivalent rectangular duration.

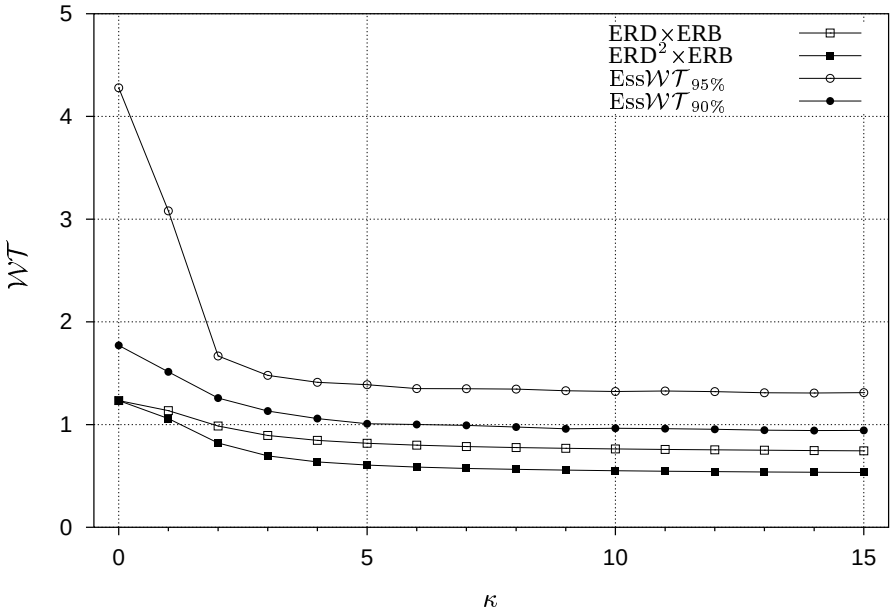

Figure 3.12: $\mathcal{WT}$ of the Kaiser window, as a function of κ. Each line represents a different definition of $\mathcal{WT}$. This suggests that to obtain a $\mathcal{WT}$ of one, the proportion of energy to be constrained in both domains should be slightly larger than 90%.

poses, then, ERD is equivalent to an $\mathrm{Ess}\mathcal{T}_{92.4\%}$. Essential duration is perhaps better, theoretically, when parameters such as $\mathcal{WT}$ are important, because it can be directly manipulated and related to the results of Landau and Pollak (1961, 1962) and Slepian and Pollak (1961) (see Chapter 1 and Appendix A).

3.3.3 Defining $\mathcal{WT}$

The definitions of bandwidth, duration, and $\mathcal{WT}$ are arbitrary. It seems that the definitions used in both experimental and theoretical work are chosen as much for tradition, rather than theoretical considerations. For psychophysical purposes, the definitions that appear to best bridge the gap between theoretical and practical purposes are the essential bandwidth and duration measures of Landau and Pollak (1961, 1962), Slepian and Pollak (1961) and Slepian (1976, 1983).

The ERD for $\kappa = 9$ constrains approximately 92.4% of the energy. It seems reasonable then to use 92.4% as the proportion of energy constrained by the essential duration, because then these measures may be compared directly. It is also reasonable, if the aim is to treat the time and frequency domains equivalently, to expect the bandwidth measure to constrain the same amount of energy as the duration measure. Therefore, the definition of $\mathcal{WT}$ used to define the signals in this project was $\mathrm{Ess}\mathcal{W}_{92.4\%} \times \mathrm{Ess}\mathcal{T}_{92.4\%}$. Many other measures of bandwidth and duration were also measured but the standard was $\mathrm{Ess}\mathcal{WT}_{92.4\%}$. This definition results in a minimum $\mathcal{WT}$ of approximately one for $\kappa = 9$. The corresponding ERD×ERB was about 0.71.

Figure 3.12 compares $\mathcal{WT}$ as a function of κ. This figure shows the minimum $\mathcal{WT}$ obtainable when using the Kaiser window for four definitions of $\mathcal{WT}$. Table C.5 shows the effect of these various definitions on the resulting $\mathcal{WT}$ for the sets of $\mathcal{S}$ waveforms used in the experiments.

3.3.4 Signal generation with the IFFT

Drga (1997b, 1998) has developed an approach of generating time and band limited Gaussian noise by a method originally suggested by John Whitmore. A similar approach has been described in Hsueh and Hamernik (1990), however, the current method was developed independently. The steps in Drga's approach are to (a) create a spectrum with a particular bandwidth, and a large number of spectral components; each component having a random amplitude and phase, (b) calculate the inverse fast Fourier transform (IFFT) to get a very long time series, (c) chop sections out of the time series to get the individual *transients*, (d) scale the transients so they have a mean of zero and a variance of one, (e) scale and mix the transients to make *buffers* of the desired signal–to–noise ratio, then (f) window the buffers to get the desired duration. This approach generates waveforms that are particularly well defined in the frequency domain—far better than those created with a digital filter approach.

This new method was used to generate the signals for the experiments and simulations in this project. Thus, the rest of this section is devoted to describing this method and analysing the resulting experimental signals.

The process begins by randomly sampling complex frequency components, from the input bandwidth, to be transformed into a time series by the IFFT. The components make up random amplitudes and phases, and are selected from a bivariate distribution. Drga (1998) prefers the circular uniform distribution (uniform amplitude and uniform phase). According to Drga, the normality of the final waveforms is not dependent on the distribution of amplitude and phase of each component in the input spectrum. The central limit theorem indicates that if many components are included, the distribution will eventually tend to a Gaussian distribution.

The final quality of the waveforms depends on the fundamental frequency (which is the reciprocal of the number of points in the IFFT times the discrete time–base (clocking rate) in Hz). The smaller the fundamental frequency, the larger the IFFT for the same duration waveform. It requires lots of computer memory to store these very large arrays.

Practically, to generate a signal set with a particular $\mathcal{WT}$ requires a bit of trial–and–error and some rules–of–thumb. In this project, a Kaiser window with parameter $\kappa = 9$ was used. By successive approximation, the input bandwidth to the IFFT was estimated for each combination of $\mathcal{W}$ and $\mathcal{T}$ so that the resulting time series would have the desired bandwidth. It was serendipitously found that if the bandwidth was narrower than the spectrum of the window, the resulting $\mathcal{WT}$ would be near unity regardless of the duration. This is because only the window was contributing to the bandwidth spread. By varying the duration of the chopped transients, all required bandwidths were obtained. For larger $\mathcal{WT}$, the required $\mathcal{W}$ and $\mathcal{T}$ were obtained iteratively by generating small signal sets of various input bandwidths and calculating the resulting bandwidth and duration.

The IFFT used was based on the algorithm in Numerical Recipes (Press, 1988) which was implemented in assembly code by Linton Miller. This implementation was an extension of BIGARRAYS, a set of methods written by Miller (1997a), to enable BORLAND PASCAL 7.0 to access all available computer memory. BORLAND PASCAL 7.0 for MS–DOS normally has the limitation that variables cannot be greater than 64 Kb, because memory addresses are only 16 bit. BIGARRAYS beats this 64 Kb limit by providing 32 bit arrays.

All IFFTs were radix 2 with 2^{22} points. This size was chosen because it was the largest that could be achieved on the computer available (with 80 Megabytes RAM) without paging (which

slowed the calculations down too much). The larger the IFFT the better, because it increases the number of spectral components that can be included within the input bandwidth.

The duration of a time series from a 2^{22} point IFFT, with clocking rate 44 100 Hz, is 95 seconds. From this very long time series, unique transients of the desired duration were selected. It is important to (a) leave enough time between each successive transient so the correlation between successive transients is small and (b) include random jitter to remove any periodicity effects.

Firstly, some autocorrelation functions were calculated for the very long time series. The autocorrelation functions were symmetrical about the midpoint with a duration of 95 seconds. The width of the main lobe was approximately $2/W$ where W was the bandwidth of the input spectrum. From the main lobe, the functions decayed rapidly as a function of the input bandwidth. For wide bandwidth signals the autocorrelation was essentially zero outside of the main lobe. For very narrowband signals, the autocorrelation remained significant throughout the duration of the function. Thus, no matter what lag is included, consecutive pairs of transients will be correlated. The rule–of–thumb was to make the minimum distance between consecutive pairs of transients at least $4/W$ seconds. For example, if the input spectrum had a bandwidth of 0.5 Hz at 500 Hz, the time between pairs would be a minimum of 8 seconds. Further to this, the random jitter factor had a minimum of zero and a maximum of the number of points in the transient. From these numbers, the maximum number of transients that could be chopped from one long time series was calculated. Any left–over points were spread between the transients if there were more points than there were transients being chopped, otherwise the extra points were left at the end. Finally, the total number of IFFTs to be run was calculated by dividing the number of transients required by the maximum number of transients that could be selected per IFFT. These numbers for each signal set generated are specified in Tables C.3 and C.4.

It is important to note that, at this point, the transients are gated using a rectangular window. This causes unacceptable spreading in the frequency domain; in fact these transients have a much wider bandwidth than the input spectrum. By windowing the transients over the entire duration of the waveform, with the Kaiser window, the spreading in the frequency domain is minimised. It is important, due to the nature of the present study, that signals are windowed over the entire duration to minimise frequency domain spreading. As described in Section 3.3.2, this window gives very near optimal constraint of energy in both the time and frequency domains. The bandwidth will still be wider than the input spectrum, but the window will ensure that it is as constrained as possible (with respect to the energy content). As shown in Harris (1978), windowing only over the beginning and end of a waveform increases the height of the side lobes.

Before mixing the $\mathcal{S}$ and $\mathcal{N}$ transients to make signal buffers that can be output to a digital–to–analog converter (DAC), the transients are scaled so both the $\mathcal{S}$ and $\mathcal{N}$ transients have variance, or power, of one. Given this constraint it is possible to calculate the scalar needed to give the desired signal–to–noise ratio. The signals and noises were randomly and uniquely mixed, scaled, then converted into 16–bit buffer codes suitable to be output to the DAC.

When the transients were converted to buffers, an appropriate scaling was determined by mapping the absolute maximum value of all the transients to 16 bits. Not all of the signal sets were generated before starting the experiments, therefore, there had to be some allowance for subsequent extreme signal sets, otherwise the signals could not all be scaled to the same level. A maximum value of 7 was chosen, because the largest value obtained in the practice signal sets was 6.8 standard deviations. The maximum standard deviation of the signal sets, however, tended to be no more than 5.5, so in most cases the full scale of the DAC was not used.

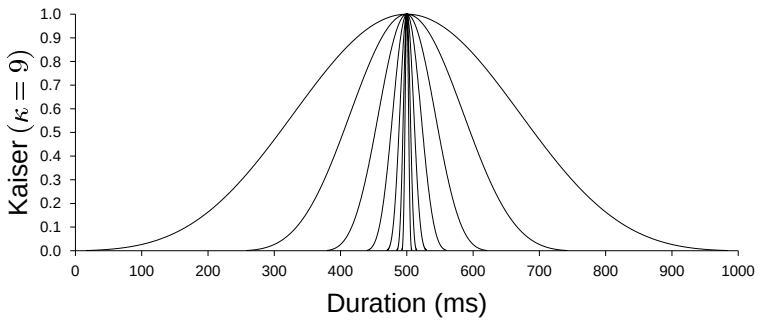

Figure 3.13: The Kaiser windows ($\kappa = 9$) used to gate the signals for the seven different durations required in the experiments.

3.3.5 Signal analysis of experimental transients

The sets of transients were analysed in both the time and frequency domains to ensure each set were close to the desired parameters.

The transients in the time domain

Figure 3.13 shows the scale of the Kaiser data window for the seven durations used in the experiments. The shape is the same, because $\kappa = 9$ was used for all signals.

The transients in the frequency domain

The waveform sets were fast–Fourier–transformed (FFT), then spectral-averaged, to obtain frequency domain information. The size of the FFT depended on the number of time domain points: 2^{16} point FFTs were used for signals less than 20 000 points (giving frequency domain resolution of 0.67 Hz) and 2^{18} point FFTs for signals greater than 20 000 points (giving 0.17 Hz resolution). Thus, all waveforms were zero-padded to some degree. The size of the FFT was chosen to maximise the effect of interpolation in the bandwidth estimation procedure while still minimising calculation time.

The power spectra for all waveform sets are presented in in Figures C.1 and C.2, with power in decibels. The spectra were scaled by the maximum power within each transient set. In all cases, the $\mathcal{S}$ spectra are plotted between 0 and 1000 Hz and the $\mathcal{N}$ spectra are plotted between 0 and 8000 Hz. All spectra extend to 22 050 Hz, but there is little or no detail beyond the points plotted.

The spectral–averaged phase spectra, for each $\mathcal{S}$ and $\mathcal{N}$ transient set, over the frequencies where the power was greater than -72 dB, tended to zero, because the random phases cancelled out for all transient sets.

Measuring signal–to–noise ratio

In signal *generation*, the signals and maskers are mixed to give a desired signal–to–noise ratio. In signal *analysis*, the signal–to–noise ratio is estimated from the power spectrum.

Subsets of 500 $\mathcal{S}$ and $\mathcal{N}$ transients were selected, without replacement, to form buffers for each signal–to–noise ratio. The estimates of bandwidth were, however, taken over the entire $\mathcal{S}$ and $\mathcal{N}$ transient sets, thus the actual signal–to–noise ratio was slightly different from that desired due to sampling variability.

In Drga's (1997b) signal generation program, the ERB of the $\mathcal{S}$ and the AERB of the $\mathcal{N}$ transients were used to specify the scalar, when mixing the signal with the noise maskers. Once the scalar is determined, it is possible to estimate what the signal–to–noise ratios would be for any of the other bandwidth measures. Subsequently, Drga (personal communication, 1997) derived a method of estimating the actual signal–to–noise ratio from the spectrum of the buffers, assuming the ERB and an ideal rectangular spectrum. From these estimates, the signal–to–noise ratios for the other bandwidths may be estimated by considering the near constant difference in signal–to–noise ratio. This is shown in Tables C.11, C.12, and C.13 for the signal sets used in the experiments.

Drga's method is to estimate the difference between the *peak* of the $\mathcal{SN}$ signal power, r_{peak}, and the average $\mathcal{SN}$ signal power outside the $\mathcal{S}$ signal passband, so only masker is included, r_{floor}, (1 kHz–4 kHz in this case). The signal–to–noise ratio is related to this difference by the relationship

$$\text{SNR}_{\text{peak}} \quad = \quad 10\log_{10}\left[10^{(r_{\text{peak}}/r_{\text{floor}})/10} - 1\right]. \tag{3.4}$$

This method is prone to variable estimates due to noisy passbands. To remedy this, the method was amended to (a) estimate the masker floor by including the *average* power of the entire $\mathcal{N}$ passband as well as r_{floor}, notated $r_{\mathcal{SN}+\mathcal{N}}$, and (b) by taking the *average* power of some of the $\mathcal{S}$ passband, notated $r_{\mathcal{SN}\text{ave}}$, if the top was flat (this requires visual inspection of the spectrum), or the peak if the top was rounded. This gives a signal–to–noise ratio estimate of

$$\text{SNR}_{\text{ave}} \quad = \quad 10\log_{10}\left[10^{(r_{\mathcal{SN}\text{ave}}/r_{\mathcal{SN}+\mathcal{N}})/10} - 1\right]. \tag{3.5}$$

For $\mathcal{WT} > 1$, the passband does not necessarily have a single peak (for $\mathcal{WT}$=1, the passband is forced into a smooth curve by the spectrum of the window) and tends to be flat, with random ripple, over a particular bandwidth. If the power in the band is not averaged, the signal–to–noise ratio may be biased by a number of decibels because of sampling variability (up to 2–3 dB difference was noted in some cases) especially for small signal–to–noise ratios. Tables C.14, C.15 and C.16 show this difference for the signal sets used in the experiments. Note that for $\mathcal{WT}$=1, except for the very wideband signals, there was no flat part to the spectrum so the peak power was used for both.

Descriptive statistics of the transients

Tables C.7 and C.8 present the descriptive statistics (mean, standard deviation, skewness, and kurtosis) of each pre-mixed waveform set, before and after scaling. The scaling is part of the signal generation procedure where each waveform set is rescaled by the sample mean and standard deviation so that they are distributed with a mean of zero and a standard deviation of one. This rescaling aids in the mixing of the waveforms to obtain a specific signal–to–noise ratio and does not affect the skewness or kurtosis.

A standard normal distribution has a mean of zero, a variance of one, a skewness of zero (symmetrical), and a kurtosis of three. In all cases, the rescaled waveform sets can be considered normally distributed on the basis of their skewness and kurtosis measures.

Histograms

Normality may also be tested by plotting the standardised z-histogram for each waveform set against the normal distribution that has the same mean and standard deviation. For the set of signals generated for this project the $\mathcal{S}$ histograms were normal out to 2.5–3.5 standard deviations and the $\mathcal{N}$ histograms were normal out to 3.5–4 standard deviations. The divergence from normality outside of these deviations is only slight out to 4 standard deviations. The best fits tend to be for the longer duration signal sets because there are many more points contributing to the distributions, and for larger $\mathcal{WT}$ because there are many more spectral components contributing to the waveforms.

Correlation

Ideally there should be no correlation between consecutive pairs of transients (chopped from the same IFFT) but this is impossible to achieve for very narrow–band processes. This is because their auto–correlation function does not tend to zero very quickly for any time lag.

Pearson's product–moment correlation coefficient was used for all correlation tests. The following tests were performed on the $\mathcal{S}$ and $\mathcal{N}$ transient sets for each experimental level:

global–r: is the correlation over all pairs of points, selected as described below, in the transient set.

mean–r: is the average correlation for each transient pair over all transient pairs where the transients are selected as described below. This should tend to zero, because no provision is made for positive and negative correlations.

z–test: tests whether the measured correlation deviates from the expected correlation of zero.

Overlapping, ordered pairs: This test measures sequential dependency by correlating consecutive pairs of transients: transient–one with transient–two, transient–two with transient–three, etc. The global–r coefficient is akin to measuring the auto–correlation for the specific offset of one transient duration and ideally should result in a value of zero—that is there is no correlation between consecutive pairs.

The top half of Tables C.9 and C.10 report the results of this test on each transient set. The largest absolute correlation is a mere 0.018665. This indicates that the number of points–between–signals (see Tables C.3 and C.4) were suitably large for all signal sets.

For virtually all the $\mathcal{S}$ sets, and some $\mathcal{N}$ sets, the z–test indicated that the correlations, although minute, were large enough to be considered significantly different from the ideal of zero. This should be taken with a grain of salt, because the number of points contributing to each correlation are on the order of millions—it is easy to find significant differences if the set size is large enough. The practical significance of a global correlation of less than 0.0187 is negligible.

Independent, random pairs: This test measures the correlation of pairs of transients chosen randomly, unlike in the overlapping, ordered pair test where the pairs were taken in order. Thus, global–r is no longer a point on the autocorrelation function. In general, this results in similar global–r correlations as the overlapping, ordered pair test, but there are fewer pairs contributing to the correlation.

The bottom half of Tables C.9 and C.10 reports the results of this test on each transient set. The largest absolute correlation is a mere 0.019031 and does not come from the same signal set that gave the largest overlapping, ordered pair correlation.

3.3.6 Signal analysis for buffers

The digital codes of the buffers were analysed, then some selected buffers were analysed as analog waveforms at the headset amplifier, and out of an artificial ear, using an analog–to–digital converter.

In Figure 3.14, the spectrum of the digital buffer codes for signal set {6.25ms, 160Hz: 16dB} was used to compare the same buffers at the headphone amplifier, through the earphone, and through an artificial ear. This signal set was used, because the very short duration would highlight any problem with the response time of the system. Note that a 100 Hz high–pass filter was used when sampling through the artificial ear (to reduce the effect of low frequency vibrations) which attenuated the low frequencies in the spectrum. Note too that a different earphone from the experimental earphones was used for this test but it had similar frequency characteristics. The blip in the spectra at about 3 kHz is an artifact of this earphone.

The shape of the spectrum at the headset amplifier closely reflects the original digital buffer spectrum. The noise at the low frequencies are harmonics from the mains frequency. The biggest difference is the noise floor, which is much greater due to the small voltage of the signal at the headset amplifier.

The shape of the spectrum through the earphones and artificial ear is still recognisable as the original buffer spectrum, but it is also shaped by the earphone frequency response, which is only flat out to 1 kHz (see Section 3.2.2 and Figure 3.2.2). Thus, the final quality of the signals is determined mainly by the spectral characteristics of the earphones. The frequency response of the earphones, however, has the same shape for all the signal sets so can be considered a constant source of error.

3.4 Results

In this section, the experimental data are analysed using ROC analysis (Section 1.1.2), mROC analysis (Section 2.1.1), GOC analysis (Section 2.1.1), and FORCE analysis (Section 2.1.2). More advanced analyses are detailed in Chapters 4, 5, and 6. The ROC, mROC, and GOC analyses were calculated using ANALBOT (Lapsley Miller, 1998a). The FORCE analysis was calculated using ACA (Drga, 1997a), which was modified slightly for the current experiment. All graphs were produced by the program GNUPLOT (Williams & Kelley, 1997).

Observer 3's data were used only as additional corroboration that the findings for Observer 1 and Observer 2 were similar for another observer. Because there were not enough replications completed for reliable FORCE analysis, only ROC, mROC, and GOC results are presented for Observer 3.

3.4.1 Model fitting

For most of the analyses in this section, comparisons are made to a mathematical ideal observer. The ideal observer chosen was the χ^2 energy detector, specified in Chapter 1. The energy

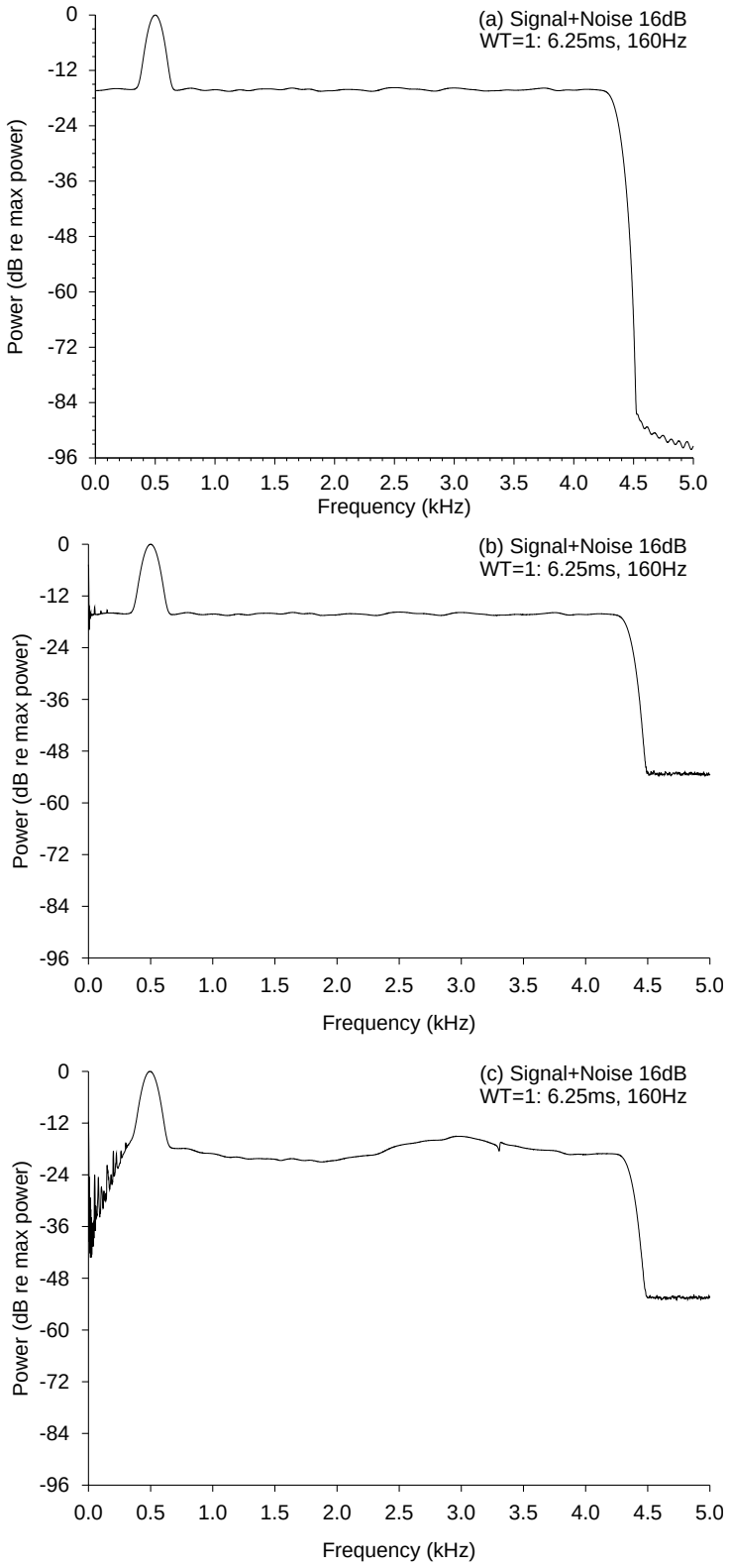

Figure 3.14: The spectrum of the buffer set {6.25ms, 160Hz: 16dB} (a) as the original digital codes, (b) at the headset amplifier, (c) through the earphones and an artificial ear. The fall–off at low frequencies in (c) is due to a 100 Hz high–pass filter, which was used to remove low frequency environmental vibrations.

detector was chosen over the envelope detector, because there are theoretical ROC curves and psychometric functions for the three $\mathcal{WT}$ products considered. The energy detector is also a simpler model, because there are only two parameters: $\mathcal{WT}$, and signal–to–noise ratio. The value of the $\mathcal{WT}$ parameter was $\mathcal{WT}$=1 for condition $\mathcal{WT}$=1, $\mathcal{WT}$=3 for condition $\mathcal{WT}$=2, and $\mathcal{WT}$=5 for condition $\mathcal{WT}$=4. The values for $\mathcal{WT}$ were chosen from the signal analysis of the simulations in Chapter 4, which used the same waveforms. Results from these simulations indicated that the actual $\mathcal{WT}$ was larger than that indicated from the signal analysis on the original waveforms (see Section 3.3). Because the χ^2 model has integer degrees of freedom, $\mathcal{WT}$ was rounded to the smallest integer larger than the average estimated $\mathcal{WT}$. The other parameter, effective signal–to–noise ratio, was estimated by considering the linear attenuation from the energy detector.

Some graphs also have results from a data model, to be introduced in Chapter 4, which are polynomial psychometric functions for a full–linear detector.

3.4.2 ROC and mROC analysis

Each experimental level consisted of one noise–alone event and five signal–plus–noise events. The experiment also used a 2000 point rating–scale, so each level produced one false–alarm rate array, and five hit rate arrays, each with a maximum of 2000 points.[7] The ROC curves were plotted in the usual way, as parametric functions of hit rate versus false–alarm rate, as a function of the criterion (which was each possible rating from 1 to 2000). The false–alarm rates were common to each signal–to–noise level within an experimental level.

The ROC curves for each observer's six replications for each experimental level are presented in Appendix D, and an example in Figure 3.15(a). On each ROC space are thirty ROC curves (five signal–to–noise ratios $\times$ six replications). The most immediate property of all the ROC curves is the variability of the six replications from the same observer, experiment condition, level, and signal–to–noise ratio. This degree of observer inconsistency is similar to previous multiple replication experiments (e.g., Lapsley Miller et al., 1999). There does not seem to be any major differences among observers.

As $\mathcal{WT}$ increases, the ROC curves tend to become more symmetrical about the negative diagonal, but are too noisy to fit to any particular model. In comparing experimental levels, the long duration–narrow bandwidth levels from all conditions appear to be attenuated for all observers.

Tables D.10–D.18 (in Appendix D) show the measures of detectability, $\mathcal{A}$ and $\mathcal{D}_2$, averaged over the six replications, and their standard deviations. The standard deviations indicate the amount of observer inconsistency over the six replications. The average measure is used as an initial value in the FORCE analysis.

The mROC curve shows the average performance of an observer in the ROC space. It is obtained by averaging the hit rates and false–alarm rates for the same criterion over the replications for a specific signal–to–noise ratio and specific condition and level. The averaging in mROC analysis may be of any form, for example, arithmetic, z–transform, or arcsine. Arithmetic averaging is simplest, but distorts extreme hit rates and false–alarm rates because they are bounded by zero and one. Averaging with the z–transform has the opposite effect because it is undefined when hit rates and false–alarm rates are zero or one. Arcsine averaging is therefore preferable, for the arcsine function has a sigmoidal shape but is finite for zero and one. The results from these ex-

[7] The overall minimum rating was 1, and the overall maximum was 1968, rather than 2000. This was because the conversion was calibrated off the minimum and maximum voltages on the line.

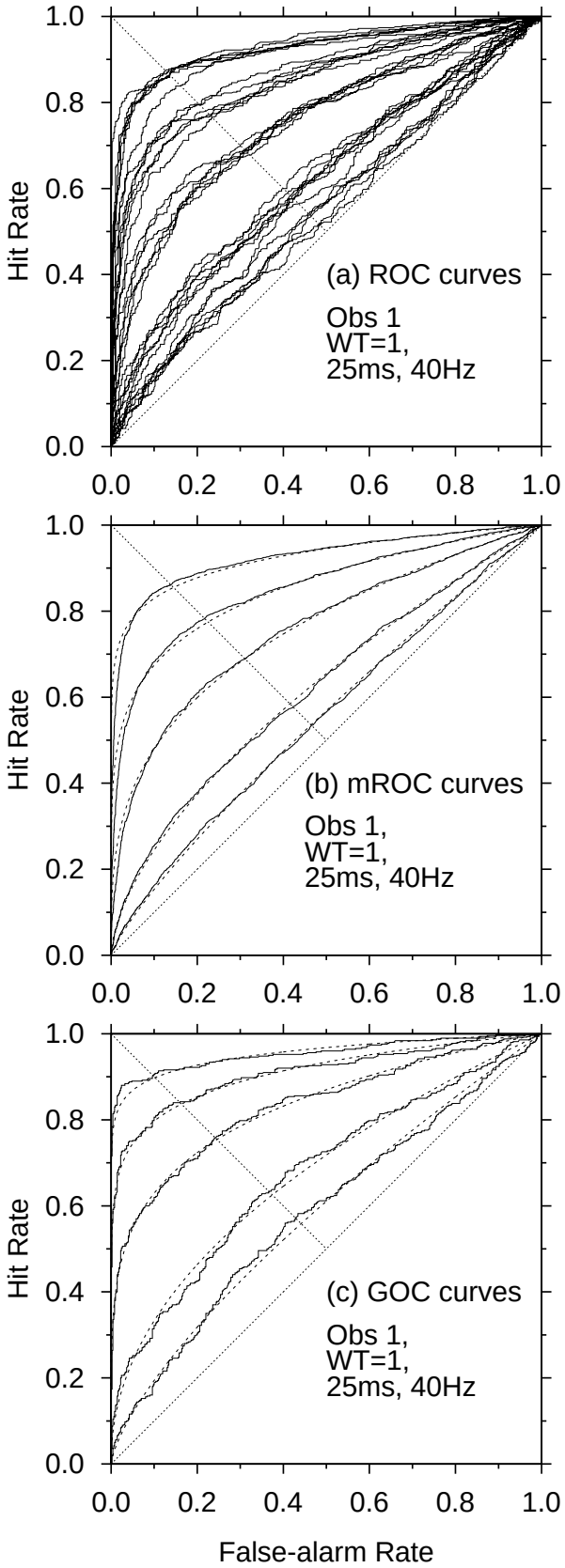

Figure 3.15: Examples of the results of (a) ROC analysis, (b) mROC analysis, and (c) GOC analysis, for Observer 1 in level {25ms, 40Hz}. The theoretical functions plotted in (b) and (c) are the *attenuated* energy detector ROC curves for $\mathcal{WT}=1$.

periments show little difference between arithmetic and arcsine mROC curves so only the results from arcsine averaging are presented.

The mROC curves for each observer's six replications of each experimental level are presented in Appendix D, and an example in Figure 3.15(b). Results averaged over all observers are not presented. Although the analysis is possible, observers are not necessarily comparable (e.g., their critical bands could be different). On each ROC space are the five mROC curves; one from each signal level. It is now clearer that as $\mathcal{WT}$ increases, the mROC curves tend to become less asymmetrical, because variability in the hit rates and false–alarm rates has been reduced. In comparing experimental levels, the long–duration, narrow–bandwidth levels from all conditions are attenuated for all observers. Within an experimental level, in all cases, detectability increases as a function of signal–to–noise ratio—something that was unclear in the ROC curves, because many of them overlapped.

Each empirical mROC curve in Appendix D is also overlayed with an attenuated theoretical χ^2 ROC curve. This curve is fitted by calculating the area under the empirical curve then finding the signal–to–noise ratio that gives the same area on the χ^2 psychometric function (for the appropriate $\mathcal{WT}$). This is done by solving Equation (1.15) for the given $\mathcal{WT}$. The resulting equation is a polynomial. A bisection root–finding algorithm was used to numerically solve specific cases. The χ^2 hit rates and false–alarm rates for that $\mathcal{WT}$ and signal–to–noise ratio were then determined using Equation (1.14). This comparison indicates that in virtually all cases, the empirical mROC curve is less asymmetrical than the mathematical theory predicts, especially for larger signal–to–noise ratios. Additionally, in all cases, the mROC curve is attenuated from the ideal mathematical theory. Both these findings can be explained by unique noise that not only diminishes performance, but changes the underlying evidence distributions such that they are less asymmetrical than expected.

Measures of detectability from the mROC curve were calculated, and compared to the average measures calculated from the ROC analysis above. The results were similar. For instance the average values of $\mathcal{A}$ (from the tables in Appendix D) may be compared with the mROC values in Tables D.1–D.9. This comparison shows agreement between the two estimates of $\mathcal{A}$, generally to 2 decimal places.

3.4.3 GOC analysis

As described in Chapter 2, GOC analysis removes much of the observer inconsistency by averaging out unique noise in the ratings for each stimulus. Figures D.15–D.21 in Appendix D, and an example in Figure 3.15(c), show the GOC curves calculated from the sum–of–ratings from the six replications, at each of the five signal–to–noise ratios, completed by each observer. Like the mROC analysis, *attenuated* theoretical ROC curves (dashed curves) are fitted to each GOC curve.

The GOC curves for $\mathcal{WT}$=1 for all three observers are strikingly close to the attenuated theoretical ROC curves, however, the attenuations are not constant over level. For the $\mathcal{WT}$=2 condition, the attenuated $\mathcal{WT}$=3 energy detector ROC curves also fit well, but there are more cases of discrepancies in shape. The discrepancies become slightly larger again in the $\mathcal{WT}$=4 condition with the $\mathcal{WT}$=5 model. The discrepancies are greater if the $\mathcal{WT}$=2 or $\mathcal{WT}$=4 theoretical curves are used instead of $\mathcal{WT}$=3 and $\mathcal{WT}$=5.

Observer 1 appeared to find the extreme long duration–narrow bandwidth levels particularly difficult, especially {400ms, 10Hz} illustrated in Figure D.17(n). The levels at the other extreme,

short–duration wide–bandwidth also proved slightly more difficult than those in the middle. Observer 2 was better at long duration–narrow bandwidth levels than Observer 1, but his performance was still more attenuated for these levels, compared with performance for the less extreme levels. The shortest–duration widest–bandwidth level also proved slightly more difficult than those in the middle. Observer 3 did not do as many levels, but his data do indicate similar patterns to the other observers except for the {6.25ms, 160Hz} where his performance was far more attenuated from the middle. Unfortunately there are no data from the {400ms, 2.5Hz} level for comparison.

Although the improvement over the mROC is considerable, there is not enough GOC evidence to say for sure that the energy detector is inappropriate, given the number of stimuli per event (e.g., compare the discrepancies to the sampling variability indicated in Figure 2.4).

3.4.4 FORCE analysis

FORCE analysis was calculated on the six replications of each experimental level for Observer 1 and Observer 2 to estimate asymptotic performance. Analysis for Observer 3 was not included because reliable FORCE analysis requires more than three replications. The tables in Appendix E show all the results from the FORCE analyses, including the parameters of the fitted FORA, the asymptotic measure of detectability, and the linearity (correlation) of the empirical FORA in log–increment coordinates.

For $\mathcal{WT}$=1, the FORCE analyses for each level and signal–to–noise ratio all converged to an asymptotic measure for both $\mathcal{A}$ and d'. All the slopes were well below -1. The improvement in $\mathcal{A}_1$ to $\mathcal{A}_\infty$ was not constant over experimental levels or signal levels, thus the slope of the FORCE psychometric function could potentially change shape compared with the mROC or GOC psychometric function. Two signal levels did not converge for $\mathcal{D}_2$—{25ms, 40Hz: 0dB} for Observer 1 and {200ms, 5Hz: 0dB} for Observer 2.

For $\mathcal{WT}$=2 all the FORCE analyses converged to an asymptotic measure for $\mathcal{A}$ and d'. The $\mathcal{D}_2$ FORCE analysis for Observer 1, however, did not converge or converged to impossible values for all the signal levels in {400ms, 5Hz} despite the FORCE analyses for $\mathcal{A}$ and d' behaving well.

For $\mathcal{WT}$=4, the FORCE analysis of $\mathcal{A}$ for Observer 1 showed problems in level {400ms, 10Hz}. {400ms, 10Hz: -8dB} did not converge and {400ms, 10Hz: 8dB} converged to an impossible value. FORCE analysis for $\mathcal{D}_2$ emphasised this level as problematic, because {400ms, 10Hz: -8, -4, 0dB} did not converge and {400ms, 10Hz: 4, 8dB} converged to impossible values. The d' analysis did not converge for {400ms, 10Hz: -8dB}. The problems at this level are not surprising given the poor performance compared to other levels. There is likely still a large amount of unique noise in the data. All but one FORCE analysis for Observer 2, on the other hand, converged. The exception was for $\mathcal{D}_2$ {25ms, 160Hz: -8dB} that did not converge.

Overall, the absolute correlation coefficients, for levels that converged validly, were bigger than 0.9900, indicating that the data are indeed linear in log–increment coordinates (remember this is *not* a measure of how well the fitted FORA fits the data in log coordinates but a measure of the *linearity* of the data in these coordinates). The levels that did not converge, or converged to impossible values, tended to be from experimental levels in which the observer did particularly badly, or signal levels for very low signal–to–noise ratios.

The d' results were also more likely to converge than $\mathcal{A}$ or $\mathcal{D}_2$, presumably because d' is positively unbounded. It is unclear why $\mathcal{D}_2$ is not as reliable as $\mathcal{A}$ given it is merely a transform of $\mathcal{A}$

(as is d'). The other non-parametric measure that is not a transform of $\mathcal{A}$ is $P(\mathrm{C})_{\max}$, however, no results are reported here for it is extremely unreliable in FORCE analysis. This is possibly because it contains more error (because it is measured at the negative diagonal and does not gain any smoothing effect of the integration of area), and because it is confounded with criterion.

Figures E.1, E.3, E.5, E.7, E.9, and E.11 show the empirical FORA (with ± 1 standard deviation error bars), the fitted FORA, and the asymptotic measure of detectability for each signal–to–noise ratio and for $\mathcal{A}$. Points greater than unity are not plotted, because they are impossible values.

Figures E.2, E.4, E.6, E.8, E.10, and E.12 shows the empirical FORA in log–increment coordinates for the increments in $\mathcal{A}$, as well as the fitted FORA. The parameters for each fitted FORA may be read from the corresponding tables in Appendix E. The ordinate of all the FORAs in log–increment coordinates have been scaled onto the same range to allow easy comparisons among them. Remember if the slope is greater than -1 then the FORA do not converge. These figures illustrate the tables for $\mathcal{A}$, enabling comparisons among the rates of improvement. They are presented in full, because to date, there are no published results using this analysis.

The FORA in linear coordinates show:

1. that the variability (indicated by the error bars) tends to decrease as replications are added—this is not entirely intuitive because the number of *combinations* of replications increases then decreases as a function of replications added,

2. that rogue FORA do not necessarily have larger error bars,

3. that some FORA are quite close to their asymptotic value but others are quite far away possibly indicating that more replications would be required to achieve the same reduction of unique noise in the final data set. This criterion could be used as a halting rule in an experiment, instead of completing the same number of replications for each level.

4. the improvement, graphically, from average performance (not mROC, but average $\mathcal{A}$, which was shown to give essentially the same value), to GOC performance, to asymptotic performance.

For the FORA in log–increment coordinates there is

1. a tendency for the $\mathcal{A}$ from the higher signal–to–noise ratios to converge more quickly than the lower signal–to–noise ratios (indicated by lines with a steeper slope that often cut across the other fitted FORA),

2. a tendency for d' FORA to be more neatly lined up (i.e., similar slopes) in order of largest signal–to–noise ratio to smallest, than for $\mathcal{A}$ (not illustrated).

3. a tendency for the last point to not fall on the fitted FORA. According to Drga (1998, personal communication) this is because the final point is not averaged, because only one combination (all six replications) contributes to it.

3.4.5 Psychometric functions

Appendix F summarises the analyses conducted so far by presenting psychometric functions for mROC, GOC, and FORCE analyses and comparing them to the energy detector mathematical model. Figures F.1, F.3,F.5, F.7, F.9, and F.11 show the psychometric functions from the FORCE,

GOC, and mROC analyses, calculated from the six replications completed by Observer 1 and Observer 2 for each experimental condition and level. Each figure shows:

1. a theoretical psychometric function (the χ^2 energy detector for $\mathcal{WT}$=1 for condition $\mathcal{WT}$=1, the χ^2 energy detector for $\mathcal{WT}$=3 for condition $\mathcal{WT}$=2, and the χ^2 energy detector for $\mathcal{WT}$=5 for experimental $\mathcal{WT}$=4),

2. the empirical psychometric points for each signal–to–noise ratio from the FORCE, GOC, and mROC analyses for each observer, and

3. the *attenuated* theoretical psychometric function for the FORCE, GOC, and mROC analyses where the attenuation is the linear average of attenuation in dB for each five signal–to–noise ratios relative to the model being tested. For the FORCE analysis, a point was only included in the attenuation calculation if the point had converged sensibly.

Results from condition $\mathcal{WT}$=1 indicate, as anticipated from the GOC and FORCE results, that the amount of attenuation is not constant as $\mathcal{W}$ and $\mathcal{T}$ are varied. There is more attenuation at the extremes. Both Observer 1 and Observer 2 have the least attenuation in level {25ms, 40Hz} after FORCE analysis. Observer 2 has less over–all attenuation than Observer 1.

The attenuation is the arithmetic average attenuation in decibels. This, of course, tells us nothing about the shape of the psychometric function. Looking at Figure F.1 there are good FORCE fits to the attenuated theory for all levels except {400ms, 2.5Hz} and to a lesser extent {200ms, 5Hz}. It appears that {400ms, 2.5Hz} results in a steeper psychometric function and would probably be better described by an energy detector with a larger $\mathcal{WT}$. This is consistent with the idea that the detection bandwidth has a minimum value. Once that minimum is reached, the auditory filter must pass more noise, so there is more attenuation, but, as a result, the $\mathcal{WT}$ must increase because $\mathcal{W}$ has increased (this of course assumes that $\mathcal{T}$ stays the same). Level {400ms, 2.5Hz} also illustrates how FORCE has changed the shape of the psychometric function compared with the mROC results. In fact, going by the mROC results alone could be very misleading. Observer 2's results also indicate that $\mathcal{WT}$=1 is an appropriate model, because all his FORCE results fall on or near the attenuated $\mathcal{WT}$=1 energy detector psychometric function. His {400ms, 2.5Hz} level is not as steep as Observer 1, indicating, perhaps, a narrower detection bandwidth.

For the condition $\mathcal{WT}$=2, Observer 1's results are generally consistent with the attenuated theoretical $\mathcal{WT}$=3 χ^2 psychometric functions, except for the long duration–narrow bandwidth conditions {400ms, 5Hz} and {200ms, 10Hz}. Both levels indicate problems with the smallest signal–to–noise ratio of -4 dB in the FORCE psychometric functions, with one converging to a very high number and the other converging, but showing negligible improvement over the mROC results. Both have the effect of skewing the average attenuation. Observer 2 produces slightly more variable results with more points falling above and below the psychometric functions compared with Observer 1 but his performance overall is again better than Observer 1. Again, performance for both observers is more attenuated at the extremes.

There was also justification for using the $\mathcal{WT}$=3 model, rather than the $\mathcal{WT}$=2 model, because Observer 2 consistently performed better than the $\mathcal{WT}$=2 model in level {50ms, 40Hz}. This may also indicate that the energy detector is not an appropriate model for human hearing.

For $\mathcal{WT}$=4, Observer 1's performance for the long duration {400ms, 10Hz} deteriorated further, compared with {400ms, 2.5Hz} and {400ms, 5Hz}, with two signal–to–noise ratios not converging after FORCE analysis. All the other levels, however, are closer to the theoretical than in

the other experimental conditions with performance in level {50ms, 80Hz} falling on-top of the theoretical. Observer 2 also shows less attenuation than in other conditions with both {100ms, 40Hz} and {50ms, 80Hz} falling onto the theoretical.

3.4.6 Attenuation analysis

The attenuation analysis summarises the information from the psychometric functions by plotting the average linear attenuation of the data from the model as a function of bandwidth and as a function of duration. A signal level only contributed to the attenuation of the FORCE function if it converged successfully. Rogue data points from the FORCE analysis for Observer 1 in levels {400ms, 5Hz: -4, 12dB}, {200ms, 10Hz: -4dB}, and {400ms, 10Hz} were dropped, because they biased the attenuations. This is not unreasonable as it was obvious that these levels converged with difficulty.

If the ear can trade $\mathcal{W}$ and $\mathcal{T}$ such that it is an optimal energy or envelope detector, then the attenuation function should be flat with 0 dB attenuation. This analysis, however, is very specific to the assumed parameters. If an energy detector with a different $\mathcal{WT}$ was chosen, for instance, then the shape of the attenuation functions may be slightly different (due to the change in slope of the psychometric function for small $\mathcal{WT}$).

Figure 3.16 shows the *average attenuation* in decibels from two types of psychometric function: the mathematical energy detector (dashed lines and open symbols) and the data model for the full–linear detector (solid lines and solid symbols), to be introduced in Chapter 4. The values of the $\mathcal{WT}$ parameter used for the energy detector were the same as for the psychometric function analysis in the previous section. Attenuation functions are shown for the FORCE, GOC, and mROC analyses, calculated from the six replications completed by each observer.

Further attenuation analysis is in Section 4.6.2, where attenuation is relative to the full–linear detector. For the two models, attenuation functions were of similar shape but the actual attenuations were different. It is not surprising that the shape is similar, because the differences between the shape of the psychometric functions for the energy and full–linear detector is subtle.

For each observer and experimental condition, the attenuation function for mROC, GOC, and FORCE analysis followed roughly the same shape with an approximately equal improvement in attenuation after each analysis. The difference in mROC to GOC tends to be on the order of 2–3 dB and from GOC to FORCE about 0.5–1.5 dB (ignoring the rogue points of Observer 1 in levels {400ms, 5Hz} and {400ms, 10Hz}).

The shape of the attenuation functions tends to be an asymmetrically bowl shaped with a steeper left side that corresponds to long duration levels. The bottom of the bowl tends to be around 40–80 Hz and 25–100 ms.

For reference, the Tables F.1–F.6 give the actual attenuations for each signal–to–noise ratio as well as the average attenuations on which the psychometric functions from Appendix F are based.

3.5 Summary

The first two sections of this chapter described the experimental methodology. The third section discussed the signal generation and analysis and considered how to define the properties of essentially band and time limited Gaussian noise in a practical context. The bandwidth, duration, and $\mathcal{WT}$, were defined as the $\mathrm{Ess}\mathcal{W}_{92.4\%}$, $\mathrm{Ess}\mathcal{T}_{92.4\%}$, and $\mathrm{Ess}\mathcal{WT}_{92.4\%}$ respectively. This does not

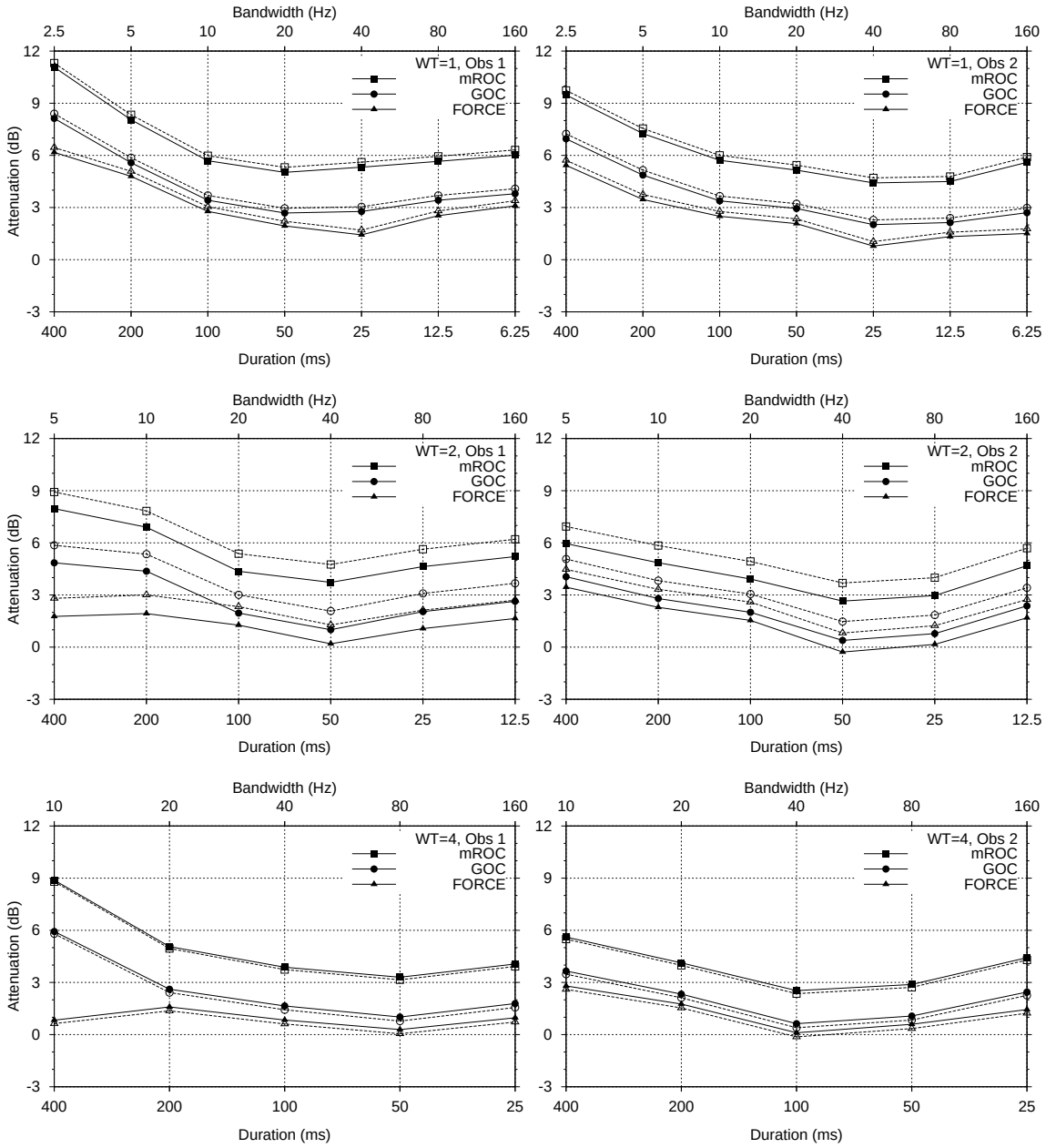

Figure 3.16: Attenuation functions after FORCE, GOC and mROC analyses, for each observer in each condition. Solid lines and solid symbols represent attenuation from the full–linear detector; dashed lines and open symbols from the energy detector.

stop any set of waveforms being defined by other methods—in fact it can prove useful to compare the relationships among the different definitions of bandwidth and duration. A new method of generating and analysing such signals—specifically for simulations and experiments—was described and thirty–six sets of transients were generated then converted into experimental stimuli. The signal analysis indicated that the sets of signals that were generated had the properties they were designed for within reasonable bounds of variability.

The fourth section reported on the initial findings of the experiments—primarily with respect to the energy detector mathematical model. It appears that human detection of small $\mathcal{WT}$ noise signals *cannot* be modelled as

1. an *ideal energy* detector, because even after FORCE analysis, performance is still attenuated from the theoretical energy detector for many experimental levels,

2. an *ideal envelope* detector, because that model predicts performance equal to the optimal energy detector for $\mathcal{WT}=1$, and similar performance across all levels,

3. an energy or envelope detector that is linearly attenuated from the theoretical, because attenuation is not constant for the same $\mathcal{WT}$.

4. an energy detector with a fixed critical band, because although the psychometric functions are steeper than the theoretical energy detector for long duration signals, they are not, in general, correspondingly shallow for short duration signals.

Although the human hearing system does not operate either with optimal, varying parameters or fixed parameters, it seems that for some cases that the hearing system is able to operate optimally. These cases tend to be when the signal bandwidth is around 40–80 Hz and the duration is around 25–100 ms. There is also no indication that it is appropriate to combine results across observers. The results so far indicate that each observer solves the time–frequency problem slightly differently for the same stimuli.

Chapter 4 presents the next step which is to compare the performance of humans to simulated detectors where the stimuli are the same. This eliminates one source of potential differences—the difference between ideal and real signals. To analyse further how the hearing system is using time and frequency information, Chapter 5 analyses simulations of energy, full–linear, and envelope detectors where the filter, rectifier, integrator, and sampling strategy parameters are systematically varied in sub–optimal ways. The evidence output from these simulations is then correlated with the human sum–of–ratings. Chapter 6 then takes advantage of some of the useful properties of Scurfield's (1995, 1996) information theoretic measures of detectability to assess performance in three–dimensional $\mathcal{WT}$–space.

Chapter 4

Simulated observers

"In theory,
there is no difference between practice and theory,
but in practice there is."

Anon.

.....

Time is a problem with human psychophysical experiments, because it takes hours of work to collect even a few, useful, data points. Simplifying assumptions are a problem with mathematical models to the extent that many models are not practically realisable. Thus, a mathematical model may not be a good model of real world detection problems. Simulation—electronic or computational—can be used to bridge the gap by modelling both human and mathematical observers.

Computer simulations have advantages over mathematical models and electronic simulations:

- They can avoid the problems of intractable mathematics by numerical approximation and estimation.
- It is generally easier to manipulate the parameters of computer simulations than of mathematical models.
- It may be easier to model physical devices, such as electronic filters, computationally rather than mathematically.
- Computer simulations are nowadays cheaper and easier to use than electronic components.

The main problem with computer simulations is the temptation to generate data for every possible combination of parameters. This quickly becomes impossible to analyse. As such, only selected simulations are dealt with in this chapter. Undoubtedly there are others that could, or even should, be run.

The computer simulation, SIM IO (Lapsley Miller, 1998c), was designed to simulate the mathematical ideal observers discussed in Chapter 1, and to model the human observer performances in the experiments from Chapter 3. Simulations that were run included:

- Simulations using the same digital waveforms that were used in the experiments on humans, so ideal observer performance could be compared directly to human performance on a stimulus–by–stimulus basis, as well as over all stimuli:
  - for simulated detectors with optimal parameters, and
  - for simulated detectors with non-optimal parameters, including systematically varying (a) the filter bandwidth, (b) the leaky integrator time constant, and (c) the sampling time.
- Simulations of different ideal observers detecting the same signals to enable direct comparisons of performance.
- Simulations using a very large number of stimuli per event, something impossible with humans, allowing estimation of the underlying distributions.

The design of SIM IO was formulated after contemplation of previous simulations discussed in the literature (see Chapter 1) including Ahumada (1967), Gaston and Jeffress (1974), Gilkey (1981), Gilkey and Robinson (1986), Green (1988), Indlin (1979a, 1979b), Jeffress and Gaston (1967), Jeffress (1968), Makita and Miyatani (1950), Pfafflin and Mathews (1962, 1966), Raab and Leshowitz (1968), Sherwin et al. (1956), Whitmore and Bowden (1985), and Williams and Jeffress (1967).

4.1 General description of SIM IO

SIM IO (Lapsley Miller, 1998c) is a flexible simulated ideal observer, which filters, rectifies, integrates, then samples digital waveforms. It is modular in form, such that each component is a separate entity, receiving an input waveform and delivering an output waveform. The output may be captured and saved to file at any stage in the system so that signal analysis can be performed after each transformation. The modules may be configured for many types and variations of ideal observer by changing the filter, rectifier, integrator, and sampling strategy characteristics.

The simulations were run as SIFC experiments (see Chapter 1), but it is relatively straightforward to *post hoc* analyse the evidence as if it were from a two–interval forced–choice (2IFC) experiment if required.[1] Simulations may be run as multiple batch–jobs or run interactively.

The main components of SIM IO are the:

Signal Generation Module: that reads in the experimental buffers or transients from disk, or generates some simple signals on–the–fly.

Digital Filter Module: including the classic infinite impulse response (IIR) filters such as the Butterworth, and the single–tuned or resonator filter.

Rectifier Module: that gives the choice of half–wave, full–wave, or square–law rectification.

Integrator Module: that implements true, leaky (negative exponential), and leaky (running average) integration.

Sampling Strategies Module: including peak sampling, sampling at a specific index or time, sweep sampling, averaged sampling, and end of waveform sampling.

[1] For example, by pairing up $\mathcal{SN}$ and $\mathcal{N}$ stimuli, and taking the difference or the ratio of the evidence associated with each pair. This assumes no interactions due to forward or backward masking, or leakage of one signal into the next.

After a simulation was finished, various analyses were performed including ROC analysis (see Section 4.4.6), signal analysis of the output of each component (see Section 4.3), parameter estimation of the output of the simulation (see Section 4.4), and correlation analysis with human performance (see Chapter 5). ROC analysis was always performed, and the other analyses depended on the context of a particular simulation.

SIM IO outputs results to a number of files including:

`*.SUM`: a summary of the characteristics of the signals, parameters of each component of the simulation, the evidence statistics, and the running time.

`*.ROC`: hit rates and false–alarm rates for each signal level.

`*.STT`: measures of detectability and attenuations from the energy detector for each signal level.

`*.EVD`: the output of the simulation (the evidence) for each signal.

`*.POW`: the spectral–averaged power spectra out of the filter for each signal level [optional].

`*.RCT`: the temporal–averaged rectifier output for each signal level [optional].

`*.INT`: the temporal–averaged integrator output for each signal level [optional].

`*.BAD`: the batch file for batch–job sessions that defines every parameter required by SIM IO.

4.1.1 Signals

SIM IO can read in stored signals or, for testing purposes, it can generate a number of test waveforms, e.g., random numbers from particular distributions with or without added deterministic signals. For the stored signals it can read in and convert the digital buffers (used in the experiments) to the digital equivalent of an analog voltage value, or it can read in the $\mathcal{S}$ and $\mathcal{N}$ transients used to create the buffers and mix them to a desired signal–to–noise ratio on–the–fly, using pre-calculated bandwidths. Multiple signal–to–noise ratios may be tested at one time and the number of stimuli per event may be unequal.

4.1.2 Digital Filters

SIM IO has a variety of filters, including the classical IIR filters, and the single–tuned filter. The implementation of the filters (FILTERIT, Miller, 1997b) was generated on–the–fly, allowing the filter design and implementation to be seamlessly integrated with SIM IO. FILTERIT was based closely on the programs GENCODE and MKFILTER by Fisher (1996). For a good description of the characteristics of the classical IIR filters see Jeruchim, Balaban, and Shanmugan (1992).

The filter used for the majority of simulations was a third order Butterworth filter (see Figure 4.1). The Butterworth filter (Stearns, 1975) is characterised by a smooth power gain function, which is maximally flat in the passband and the stop–band, and has a relatively sharp cutoff. The transition band rapidly decreases as a function of the squareness of the power–gain characteristics and the smallness of the order of the filter.

Jeruchim et al. (1992) state if the filter is normalised to the 3 dB bandwidth, it is completely specified by the *order* of the filter. As the order tends to infinity, the filter approaches the ideal

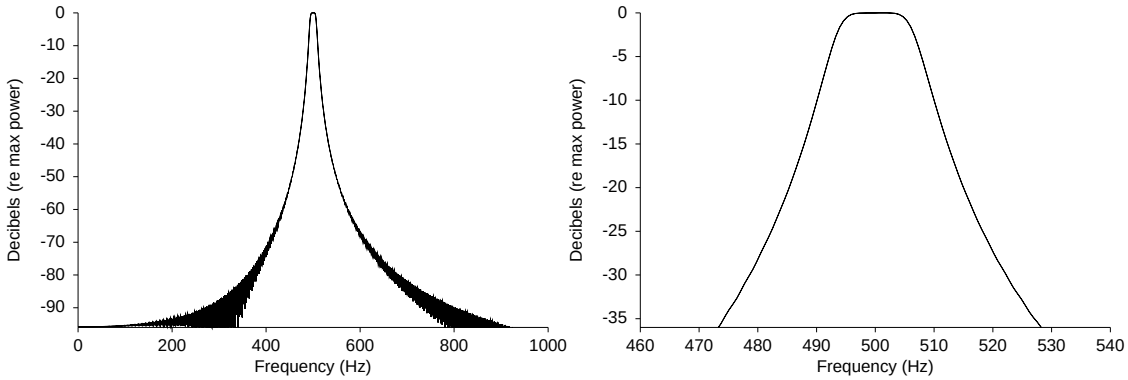

Figure 4.1: The power spectrum of the Butterworth filter impulse response. The power spectrum (in decibels relative to the maximum power) is for a 14 Hz band–pass, third–order, Butterworth filter, centred on 500 Hz. The right graph zooms in on the passband.

rectangular low–pass filter. The TdB bandwidth can be converted to $\mathrm{Ess}\mathcal{W}_{92.4\%}$ by the equation:[2]

$$\mathrm{Ess}\mathcal{W}_{92.4\%} \quad = \quad 1.16 \times \mathrm{TdB} \; + 0.4 \tag{4.1}$$

Similarly for ERB:

$$\mathrm{ERB} \quad = \quad 1.05 \times \mathrm{TdB} \; + 0.11. \tag{4.2}$$

4.1.3 Rectifiers

Three types of rectification are common in the literature: half–wave, full–wave, and square–law. Either the half–wave or the full–wave rectifier may be used in an envelope detector. Half–wave is said to be the most likely type of rectification in the human hearing system for physiological reasons (Jeffress, 1968). There appears to be little difference between the performance of the two rectifiers when used in an envelope detector. The square–law rectifier is also important, because it is a defining part of the energy detector. All three types of rectification have been implemented in SIM IO. If $f(t)$ is the input to the rectifier, as a function of time, then, then the rectifier output $r(t)$ is defined as:

- full–wave rectifier: $r(t) = |f(t)|$,
- square–law rectifier: $r(t) = f(t)^2$, and
- half–wave rectifier: $\begin{cases} r(t) = f(t), & f(t) > 0 \\ r(t) = 0, & f(t) \leqslant 0 \end{cases}$.

4.1.4 Integrators

Integration is a type of low–pass filtering. A *true* or *full* integrator sums each waveform value without weighting, thus each value contributes equally to the output. A *leaky* integrator, on the other hand, weights the current value of the waveform, and the previous $N-1$ values, by negative

[2]This equation was determined empirically for a variety of $\mathcal{W}$, $\mathcal{T}$, and $\mathcal{WT}$, and is good for TdB bandwidths greater than 6 Hz.

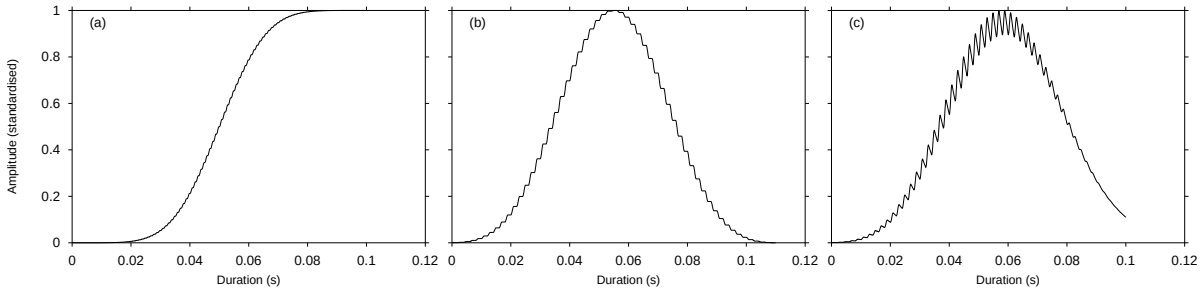

Figure 4.2: Example output for a true integrator (left), a 10 ms box–car integrator (middle), and a 10 ms leaky integrator (right) to a 100 ms windowed sinusoid input.

exponential coefficients and then sums. The points in the waveform do not contribute equally to the output at any one time. Between the true and leaky integrator is the box–car integrator. It is a leaky integrator, because it does not integrate over the entire duration of the waveform, but the weightings are equal (like a true integrator).[3] SIM IO employs true, leaky, and box–car integrators.

True integration

True integration sums *every* value of the waveform without weighting. This process is fundamental to the energy detector but physiologically it is not likely, because there is no mechanism to discharge the integrator (Jeffress, 1968). Figure 4.2(a) shows an example of the output of a true integrator where the input was a 500 Hz windowed sinusoid of 100 ms duration.

Box–car integration

The box–car integrator (also known as the moving–weighted–average integrator and the running–average integrator) averages the current value of the waveform, and the $N-1$ previous values of the waveform over N points. Box–car integration is a type of low–pass filtering, but, according to Stearns (1975, p. 105), it does not have a good low–pass transfer function. The box–car takes $N-1$ values to fill up and $N-1$ values to empty thereby increasing the length of the waveform by $N-1$ values.

Figure 4.2(b) shows an example of the output of a box–car integrator of 10 ms duration with an input of a 500 Hz windowed sinusoid of 100 ms duration.

Leaky integration

Leaky integration, using the negative–exponential function, may be used to implement the envelope detector. With a short time constant, the envelope closely follows absolute changes in the waveform. As the time constant increases, the envelope reacts only to larger changes in the waveform (Jeffress, 1968).

SIM IO uses the recursive algorithm for negative-exponential leaky integration by Brignell and Rhodes (1975). The input to the algorithm is the time constant (in ms). This time constant is also the ERD of the integrator. Equation (1.11) defines the time constant with respect to $\mathrm{Ess}\mathcal{T}_{\%}$. The

[3] The term *leaky* always refers to a negative exponential integrator here.

effective duration of the integrator, however, is dependent on the duration of the input waveform if it is shorter than the time constant.

Brignell and Rhodes (1975) derived a discrete–time leaky integrator from the classic electronic leaky integrator (smoother). It is also known as the RC integrator, for its properties depend on the value of its resistor and capacitor. This electronic integrator has the transfer function

$$H(s) = \frac{1}{1 + RCs} \tag{4.3}$$

and an impulse response of

$$H(t) = \frac{1}{RC} \exp \frac{-t}{RC}. \tag{4.4}$$

Brignell and Rhodes (1975) showed that the output of the equivalent *discrete–time* integrator is

$$y_i = kx_i + \exp(-aT) y_{i-1}. \tag{4.5}$$

and that unity gain may be achieved if k is set to $1 - \exp(-aT)$. This gives

$$y_i = (1 - b)x_i + by_{i-1} \tag{4.6}$$

where T is the clocking rate, $a = 1/RC$, $b = \exp(-aT)$, and the time constant is $-T/\log_e b = RC$.

This algorithm is easy to implement, is very efficient, and the time constant is its only parameter. Figure 4.2(c) shows the output of a leaky (negative–exponential) integrator, with a time constant of 10 ms, to a 500 Hz, 100 ms windowed sinusoid.

4.1.5 Sampling Strategies

As described in Gilkey and Robinson (1986), sampling strategies are required if the type of integration does not return a single number. They defined their sampling strategies in terms of the observation interval duration. This is not the best way of defining the sample point, because the duration of the waveform increases as it passes through the filter and the integrator, so the waveform may extend some way past the nominal observation interval.

It is not clear exactly what sampling time results in optimal detectability, so SIM IO uses a number of sampling strategies: peak sampling, sampling at a specific index or time, sweep sampling (where samples are taken from evenly spaced points over the waveform), averaged sampling, and end of waveform sampling. Any one simulation may be set up to use multiple sampling strategies so their effects can be compared.

4.1.6 Analysis of the evidence

The output of the integrator under a particular sampling strategy is the evidence. It can be seen as a sample from the underlying probability distributions of the $\mathcal{SN}$ and $\mathcal{N}$ events. As such, SIM IO records the mean, variance, skewness, and kurtosis of the evidence for each event.

4.1.7 ROC analysis

For each signal level SIM IO's performance is analysed using ROC analysis. The evidence is treated like a rating, but it is not binned into an N–point rating scale. This is because binning

results in a decrease of information. If the evidence distribution is very asymmetrical, such as a negative exponential, then binning may result in most of the evidence falling into only a few bins. Instead the evidence for all events was *sorted*, then tallied and transformed into hit rates and false–alarm rates using Event–Decision matrices. This process is essentially the same as described for calculating GOC curves from sum–of–ratings in Chapter 2.[4]

From the hit rates and false–alarm rates the ROC curves were plotted and measures of detectability ($\mathcal{A}$, $\mathcal{D}$, d', $P(\mathrm{C})_{\max}$) calculated. Additionally, the linear attenuation of the psychometric function from a stated χ^2 model were calculated if appropriate.

4.2 Energy, full–linear, and envelope detector simulations

This section describes simulating the performance of three types of detector: energy, full–linear, and envelope (see Figure 4.3). The energy and envelope detectors were chosen because they have been suggested as good candidates for noise–in–noise detection (see Chapter 1). The full–linear detector was chosen as a compromise between energy and envelope detection.[5] Firstly, mathematical theory was compared with simulations of an energy detector. Secondly, the full–linear and envelope detectors were compared with the energy detector, and each other. Because mathematical theories for full–linear and envelope detectors do not exist for arbitrary $\mathcal{WT}$, these comparisons were solely computational in nature.

The general similarities and differences among these observers are illustrated in Figure 4.3. Firstly, note that the duration of the time domain is the same in all the illustrations, but the range of the amplitude varies to accentuate changes in the shape of the waveform, rather than overall levels. The band–pass filter is identical for all detectors. The square–law rectifier, however, accentuates the main peaks, and suppresses smaller peaks, compared with full and half–wave rectification of the same waveform. Comparison of full and half–wave rectification shows that the shape is very similar, but the half–wave rectified waveform is not as dense, because negative components are transformed to zero. For integration, there is a subtle difference in the shape of the true integrator output, for square–law and full–wave rectification. The effect of the time constant on the leaky integrator is also illustrated, with the short time constant closely following changes in the rectified waveform, and with the long time constant tending towards true integration. Sampling is not illustrated.

To evaluate the detectability of small–$\mathcal{WT}$ Gaussian noise by humans, it is useful to first consider systems that result in optimal detectability. Estimating the maximum detectability of a class of waveforms provides a baseline for comparisons with other systems. Chapter 1 described how the energy and envelope detectors are considered to be ideal observers of small $\mathcal{WT}$ Gaussian noise. These ideal observers make particular assumptions about the nature of the waveforms that cannot be realised in physical systems. It is, therefore, not clear that the same ideal observer would produce the same level of performance with the current digital signals, compared with the idealised signals. It would be difficult, and possibly infeasible, to design the ideal observer to optimally detect these digital signals, because of the difficulty in specifying these signals math-

[4] For practical reasons, binning was used for the few simulations where there were hundreds of thousands of stimuli. The resolution of the binning was the maximum rating minus the minimum rating divided by 500 000.

[5] Running simulations for every other possible detector (by considering every combination of filter, rectifier, integrator, and sampling strategy) was not feasible. Preliminary investigations of leaky–energy detection (equal–bandwidth filter, square–law rectifier, and leaky integrator) in particular, did not indicate marked differences from the energy detector. This detector, however, is probably worthy of further investigation.

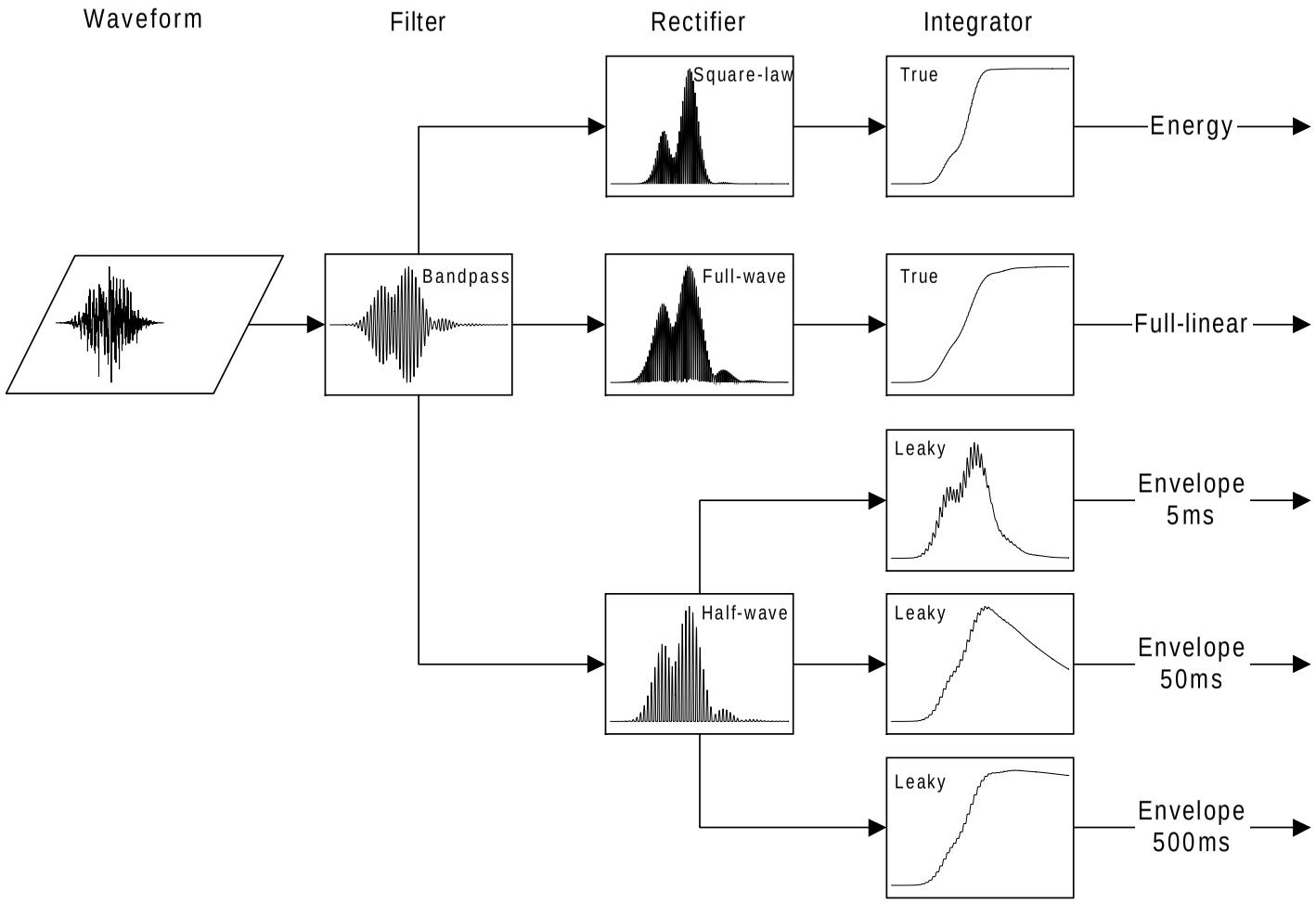

Figure 4.3: Examples of the energy, full–linear and envelope detector models (not including the sampler). The input waveform (a signal from level {25ms, 80Hz: 16dB}) and the band–pass filter are the same for all models, but the rectifiers and integrators are different. The parameter of the envelope detectors is the time constant for the leaky integrator. Note that the domain is time (and is the same for all figures), but the range is amplitude and is arbitrarily scaled so that changes in shape can be seen easily.

ematically. Such a model would need to incorporate the window, non-rectangular filtering, and waveform samples that may not be independently, identically, distributed.

The simulated ideal observer allows a compromise where the mathematical model may be tested using a variety of waveforms. Comparisons may then be directly made with the mathematical models to see if the differences between the digital waveforms and the mathematically ideal waveforms result in differences in their detectability.

The main aims are to:

- simulate an ideal energy detector and compare it to mathematical theory by
  - doing signal analysis on the output of the filter and rectifier of the energy detector, to measure the effect these processes have on the subsequent bandwidth, duration, and $\mathcal{WT}$,
  - modelling the distribution of the evidence as a gamma probability density function to estimate $\mathcal{WT}$ and signal–to–noise ratio, and to see if the gamma model can be justified for these types of signals (where the χ^2 model is a special case of the gamma model, with fixed, integer, $\mathcal{WT}$),
- compare the ideal energy, full–linear, and envelope detectors to establish any differences or similarities in performance,
- fit polynomial psychometric functions to the energy detector and full–linear detector psychometric functions. This establishes a data model to be used as a baseline comparison with the human performances, and other simulated performances, for the same signals. Although the theoretical energy detector psychometric functions are known, they are not necessarily appropriate for the experimental signals, and there are no theoretical functions for the full–linear detector.

The first series of simulations used the same buffers as those used in the experiments.[6] The second series of simulations used the transients from the experiments (see Chapter 3), but with many more signal levels and stimuli. The aims were not to model human performance directly, but to generate psychometric functions for arbitrary signal–to–noise ratios. The third series of simulations used very large sets of signals (VLS signals). The aim here was to generate enough sample points to estimate the evidence distributions produced by particular combinations of signal parameters and simulated ideal observer parameters. These simulations were not done using signal sets for all the experimental levels. Only a subset were chosen, because the signals take gigabytes of disk, and the simulations take days to run.

Further simulations are introduced in Chapter 5 where an attempt is made to model the human data from Chapter 3 with non-ideal detectors.

4.2.1 General form of the optimal detectors

The general form of the detectors is illustrated in Figure 4.3. The simulated energy detector consisted of an equal–bandwidth filter (where the bandwidth of the filter was equal to the

[6] Although it should be remembered that the digital buffers, i.e., DAC codes, are used in the simulations but the humans are presented with analog waveforms, generated from these buffers, which are mixed with a background masker, attenuated, amplified, and passed through earphones.

bandwidth of the $\mathcal{S}$ waveforms), a square–law rectifier, and a true integrator. The sampling strategy was to take the last value from the integrator. This value constituted the evidence, and was approximately proportional to the energy of the filtered waveform.

The simulated full–linear detector was identical to the energy detector except it used a full–wave rectifier. It consisted of an equal–bandwidth filter (where the bandwidth of the filter was equal to the bandwidth of the $\mathcal{S}$ waveforms), a full–wave rectifier, and a true integrator. The sampling strategy was to take the last value from the integrator. This value constituted the evidence. Although the full–linear detector is a *linear* detector (because of the linear rectifier), it is not an envelope detector, because it employs a true integrator. The true integrator means the output of the integrator does not follow the shape of the input waveform.

The simulated envelope detector consisted of an equal–bandwidth filter, a half–wave rectifier, and a leaky integrator. The time constant of the integrator was equal to the ERD of the input waveform, therefore, the time constant was different for different signal sets. Two sampling schemes were used: peak detection and sampling at the time equal to the absolute duration of the waveform (the average peak of the integrator tended to occur just before the end of the waveform). The latter is abbreviated to *ATT sampling*.

The equal–bandwidth filter was a third–order band–pass Butterworth filter, centred on 500 Hz.[7] The TdB bandwidth of the $\mathcal{S}$ transients was used as the filter bandwidth.[8] The filter was allowed to ring for an additional duration of four times the reciprocal of the TdB bandwidth.

4.2.2 Simulations using experimental buffer waveforms

Eighteen simulations were run for each detector, using the same digital buffers that were used in the experiments. These buffers were integers, coded specifically to be output to a 16 bit DAC, so were converted on–the–fly in the simulation to reals. As described in Chapter 3, the 500 buffers for each signal set were sampled without replacement from the appropriate $\mathcal{S}$ and $\mathcal{N}$ event transient sets. The parameters of the 500 waveforms forming each signal set are described in Table 3.1. The desired signal–to–noise ratios were obtained by using the ERB for the $\mathcal{S}$ transients and the average ERB (AERB) for the $\mathcal{N}$ transients. Tables C.14, C.15, and C.16 show the corresponding signal–to–noise ratio measured by other means.

An additional energy detector simulation was run on the {25ms, 160Hz} buffers, using five different filter bandwidths. The aim was to compare the effect of the equal and unequal filter bandwidths on the subsequent measures of $\mathcal{SN}$ bandwidth and duration.

4.2.3 Simulations using experimental transient waveforms

These simulations used more stimuli per event (2500 for the $\mathcal{SN}$ events and 3000 for the $\mathcal{N}$ event), and more signal–to–noise ratios, than the experimental buffer simulations in Section 4.2.2.

Eighteen simulations were run for each detector, using signals with the same bandwidth, duration, and $\mathcal{WT}$ as the experiments. The stimuli were created from the transients used in making the experimental signal buffers. $\mathcal{SN}$ transients were created for each signal–to–noise ratio by mixing each $\mathcal{S}$ transient with a randomly selected $\mathcal{N}$ transient. For each signal–to–noise ratio, sampling was done without replacement and a unique random seed was used. Thus, although the

[7] As the order of the Butterworth filter increases, the more rectangular it becomes. The larger the filter order, the more spread there is in the time domain, and the filter takes longer to calculate and run. A third–order filter was chosen as a compromise among the shape of the spectrum, spread in the time domain, and running time.

[8] These values may be read from Table C.5 on page 224.

transients were reused for each signal–to–noise ratio, they were mixed differently. The same seed was used across ideal observers, however, so that the same signals were presented, eliminating one source of variability in the comparisons.

Signal levels were chosen so that the signal–to–noise ratios increased in 3 dB steps, and that the corresponding $\mathcal{A}$ approximately covered its entire range from 0.5 to 1.0. There were fifteen signal levels for the $\mathcal{WT}$=1 signals (-18–21 dB), fourteen signal levels for the $\mathcal{WT}$=2 signals (-20–16 dB), and thirteen signal levels for the $\mathcal{WT}$=4 signals (-20–13 dB), including $-\infty$ (the $\mathcal{N}$ event).

The desired signal–to–noise ratios were obtained by using the AERB for both $\mathcal{S}$ and $\mathcal{N}$ transients. Because both transient sets were scaled to give a variance of one, it is a simple matter to rescale the $\mathcal{S}$ transients to give the desired signal–to–noise ratio when mixed with the noises, using the formula

$$\text{Scalar} \quad = \quad \sqrt{10^{\text{SNR}/10} \times \frac{\text{AERB}_{\mathcal{S}}}{\text{AERB}_{\mathcal{N}}}}. \tag{4.7}$$

4.2.4 Simulations with very large sets (VLS) of signals

To address the problem of sampling variability, due to small samples of stimuli per event, energy detector simulations, similar to those in Section 4.2.3 were done on very large sets (VLS) of signals.

Using the same methods as were used to generate the transients for the experiments, VLS signals were generated. The size of the set was determined by how many signals could fit on a 650 Mb CD-ROM. The input statistics to the IFFT were exactly the same as for the experiments except that different random seeds were used.

Only one combination of $\mathcal{W}$ and $\mathcal{T}$ were used for each $\mathcal{WT}$: {6.25ms, 160Hz: -∞, 0, 4, 8, 12, 16dB} with 200 000 $\mathcal{N}$ transients and 100 000 $\mathcal{S}$ transients, {50ms, 40Hz: -∞, -4, 0, 4, 8, 12dB} with 25 000 $\mathcal{N}$ and $\mathcal{S}$ transients, and {50ms, 80Hz: -∞, -8, 4, 0, 4, 8dB} with 25 000 $\mathcal{N}$ and $\mathcal{S}$ transients. The $\mathcal{SN}$ stimuli were made by different random combinations of the $\mathcal{N}$ and $\mathcal{S}$ transients.

4.3 Signal analysis

For the energy detector buffer simulations, the output of the filter and rectifier were captured and spectral or temporal–averaged to allow further signal analysis. The resulting spectra and waveforms were analysed to estimate the effect of these transforms on bandwidth, duration, $\mathcal{WT}$, and signal–to–noise ratio.

The size of the FFT was 2^{16} for all levels except for {400ms, 2.5Hz} and {400ms, 5Hz}, where an FFT size of 2^{17} was used instead. This was because the narrow filters used in these levels caused spread in the time domain, thereby increasing the duration of the waveforms. It was not practical to run all the FFTs at larger sizes due to time and storage constraints. At a clocking rate of 44 100 Hz and a size of 2^{16}, the resolution was 0.67 Hz. All the waveforms were zero–padded to bring their actual size to the size of the FFT (Brigham, 1974).[9]

An FFT was calculated on the filter output for each waveform in a particular buffer set, then spectral–averaged to give one spectrum for each signal level. The ERB, $\text{Ess}\mathcal{W}_{92.4\%}$, $\text{Ess}\mathcal{W}_{95\%}$, and TdB bandwidths were estimated using the algorithms from Chapter 3. The signal–to–noise ratio,

[9]For efficiencies of speed, a radix–2 FFT was used, therefore, the size of the data set had to be a power of two.

relative to the $\mathcal{N}$ spectra, was estimated using Equation (3.5), where the passband cut–offs were the same as used for the signal analysis of the $\mathcal{SN}$ buffer spectra in Chapter 3.

The output of the rectifier for each waveform from a signal level was temporal–averaged to give one time–series. The duration of the time–series was estimated using $\mathrm{Ess}\mathcal{T}_{92.4\%}$, $\mathrm{Ess}\mathcal{T}_{95\%}$, and ERD. Because the rectifier was square–law, the waveform was already proportional to its energy so the waveform points were not squared again in calculating the essential durations. Likewise, for the ERD, the square–root of each waveform point was calculated first. This resulted in duration estimates consistent with those taken before rectification, and allowed for direct comparisons.

For each signal level, $\mathcal{WT}$ was estimated four ways:

1. Essential 92.4%: $\mathrm{Ess}\mathcal{WT}_{92.4\%} = \mathrm{Ess}\mathcal{W}_{92.4\%} \times \mathrm{Ess}\mathcal{T}_{92.4\%}$.
2. Essential 92.4%: $\mathrm{Ess}\mathcal{WT}_{95\%} = \mathrm{Ess}\mathcal{W}_{95\%} \times \mathrm{Ess}\mathcal{T}_{95\%}$.
3. Equivalent rectangular: $\mathcal{WT}_{\mathrm{AERB}\times\mathrm{ERD}} = \mathrm{AERB} \times \mathrm{ERD}$.
4. ERD–TdB: $\mathcal{WT}_{\mathrm{TdB}\times\mathrm{ERD}} = \mathrm{TdB} \times \mathrm{ERD}$.

These four estimates were chosen to see how the definition of $\mathcal{W}$ or $\mathcal{T}$ interacted with the estimate of $\mathcal{WT}$—especially as a function of signal–to–noise ratio. Estimates using the ERB based on maximum power were not used because they tended to be very similar, but more variable, than those based on average power in the passband. Tables 4.1, 4.2, and 4.3 show the bandwidth, duration, $\mathcal{WT}$, and signal–to–noise ratios, estimated from the spectral–averaged filter output and the temporal–averaged rectifier output. The results indicate that regardless of the definition of bandwidth or duration, $\mathcal{WT}$ decreases as signal–to–noise ratio increases. Although there is some time–domain leakage, duration is relatively constant as a function of signal–to–noise ratio.

The bandwidth decreases because of an interaction between the filter shape and the input waveform's spectrum. The $\mathcal{N}$ spectra are flat so the TdB bandwidth matches that of the filter (compare the 3 dB values in Tables 4.1, 4.2, and 4.3, with the 3 dB values in Table C.5). The $\mathcal{SN}$ spectra, on the other hand, are not flat within the TdB points. If the $\mathcal{SN}$ signal–to–noise ratio is larger than 3 dB, then the TdB bandwidth of the $\mathcal{S}$ spectra does not match that of the filtered $\mathcal{SN}$ spectra, because the shape of the spectrum has changed; likewise the $\mathrm{Ess}\mathcal{W}_{92.4\%}$ of the $\mathcal{S}$ spectra do not quite match the filtered $\mathcal{SN}$ spectra for the same reason.

The results indicated, for $\mathcal{WT}$=1, that there was spread in the time domain. As a result of the filtering, the durations were quite a bit longer and the bandwidths narrower than the original $\mathcal{S}$ signals. The estimates of $\mathcal{WT}$ varied from around 1.34 for the $\mathcal{N}$ signals to 1.06 for the largest $\mathcal{SN}$ signals, except for {400ms, 2.5Hz}, which had larger estimates—presumably because the very narrow filter rang for much longer. For the $\mathcal{WT}$=2 signals there was slight time domain leakage from the filtering, but the result was estimates of $\mathcal{WT}$ that were smaller than the original signals for the larger signal–to–noise ratios. The estimates of $\mathcal{WT}$ varied from around 2.37 for the $\mathcal{N}$ signals to 1.88 for the largest $\mathcal{SN}$ signals. $\mathcal{WT}$=4 showed the least time domain leakage, and, again, $\mathcal{WT}$ decreased as signal–to–noise ratio increased. For $\mathcal{WT}$=1, only the 16 dB signal levels resulted in a effective $\mathcal{WT}$ similar to the original $\mathcal{S}$ transients but with narrower bandwidths and longer durations than the original signals. This indicates that the Butterworth filter has a minimum $\mathcal{WT}$ resolution of around 1.

Care is needed when comparing across the experimental conditions of $\mathcal{WT}$=1, $\mathcal{WT}$=2, and $\mathcal{WT}$=4, because the $\mathrm{Ess}\mathcal{W}_{92.4\%}$ bandwidth is constant, but the TdB bandwidth gets larger. For

instance experimental signals for levels {6.25ms, 160Hz}, {12.5ms, 160Hz}, and {25ms, 160Hz} have an $\mathrm{Ess}\mathcal{W}_{92.4\%}$ of 160 Hz but the 3 dB bandwidth changes from 110 to 141 to 166 Hz. This is because as $\mathcal{WT}$ changes from 1 to 2 to 4, the shape of the spectrum becomes more rectangular—forcing the 3 dB points further apart—resulting in a steeper spectrum (e.g., compare the $\mathcal{S}$ spectra in Figures C.1(g), (m), and (r)).

4.3.1 Effect of filter bandwidth on estimates of signal bandwidth

If a $\mathcal{SN}$ waveform is filtered, the subsequent estimate of the waveform's bandwidth can vary considerably, depending on (a) the signal–to–noise ratio of the input signal, (b) the bandwidth of the filter relative to that of the signal–alone, and (c) the definition of bandwidth used. It is important to consider this problem, because models of detectability depend on the properties of the waveform out of the filter. Therefore, the effect on the signal bandwidth after filtering, as a function of signal–to–noise ratio, was estimated by varying the bandwidth of the filter from very narrow to very wide (relative to the bandwidth of the original signals).

Repeated energy detector simulations were run on the {25ms, 160Hz} signal set. Firstly, a very narrow filter bandwidth of 28 Hz was used, which was well within the signal band. A slightly larger bandwidth of 108 Hz was then used, then an equal–bandwidth of 166.38 Hz. Finally, two very wide bandwidths of 280 and 360 Hz, were used that were much wider than the $\mathcal{S}$ bandwidth. All other parameters were held constant.

$\mathcal{W}$, $\mathcal{T}$, and $\mathcal{WT}$, were estimated as a function of signal level. As the bandwidth of the filter increased, the $\mathcal{WT}$ estimated at each signal level diverged: when the filter was very narrow the $\mathcal{WT}$'s were virtually identical, but when the filter was very wide $\mathcal{WT}$ varied from being still rather small at large signal–to–noise ratios, to being *very* large for $\mathcal{N}$ signals. As above, $\mathcal{WT}$ changed because bandwidth changed. This phenomenon is evident in the spectra of these filtered waveforms shown in Figure 4.4 where the spectra are scaled relative to their maximum power. As signal–to–noise ratio increases, the shape of the passband becomes rounder at the edges compared with the $\mathcal{N}$ spectra and the transition band becomes steeper. Compare these spectra with the original $\mathcal{S}$ spectrum in Figure C.1(r). This phenomenon helps explain the findings presented in Tables 4.1, 4.2, and 4.3 that indicate that regardless of the definition of bandwidth or duration, $\mathcal{W}$ and $\mathcal{WT}$ decreased as signal–to–noise ratio increased.

This phenomenon has important ramifications. If the internal filtering process of the ear does not exactly reflect the shape of the spectrum of the signals (and all indications are that this is the case) then the bandwidth of the signals will change after filtering as a function of signal–to–noise ratio. Thus, the mathematical model of energy detection is not totally appropriate, because it assumes that $\mathcal{WT}$ is constant with signal–to–noise ratio. Even if signals were rectangular then $\mathcal{WT}$ would still vary if the shape of the filter was not rectangular. It is only when both the filter and the signal spectra are rectangular, and the filter bandwidth is less than or equal to the signal bandwidth, that $\mathcal{WT}$ would necessarily be constant with increasing signal–to–noise ratio.

4.3.2 Equal–bandwidth filtering

Because the signal sets do not have ideal rectangular spectra, there is no specific point where the bandwidths are equal, so an arbitrary decision must be made as to the point at which the filter matches the signal. If the criterion is set so the filter bandwidth is well within the passband of the signal, the result may be equivalent $\mathcal{WT}$ products at each signal level, but the spirit of matching

the bandwidth is not met—and for $\mathcal{WT}$=1 there is no flat part to the $\mathcal{S}$ spectrum. Choosing the criterion based on the $\mathcal{S}$ signal 3 dB points at least provides an objective method, as well as the benefit of having the filter *specified* by its 3 dB points. If the criterion is chosen using a different definition of bandwidth, then the bandwidth measure must be converted into 3 dB points—something that is a function of $\mathcal{WT}$ and possibly other variables as well. For instance Equation (4.1) gives the conversion from TdB to the $\mathrm{Ess}\mathcal{W}_{92.4\%}$ of the third–order Butterworth filter, and Equation (4.2) gives the conversion from TdB to the ERB. Both were estimated from the simulated impulse response of the digital filter over the range of bandwidths used in the experiments.

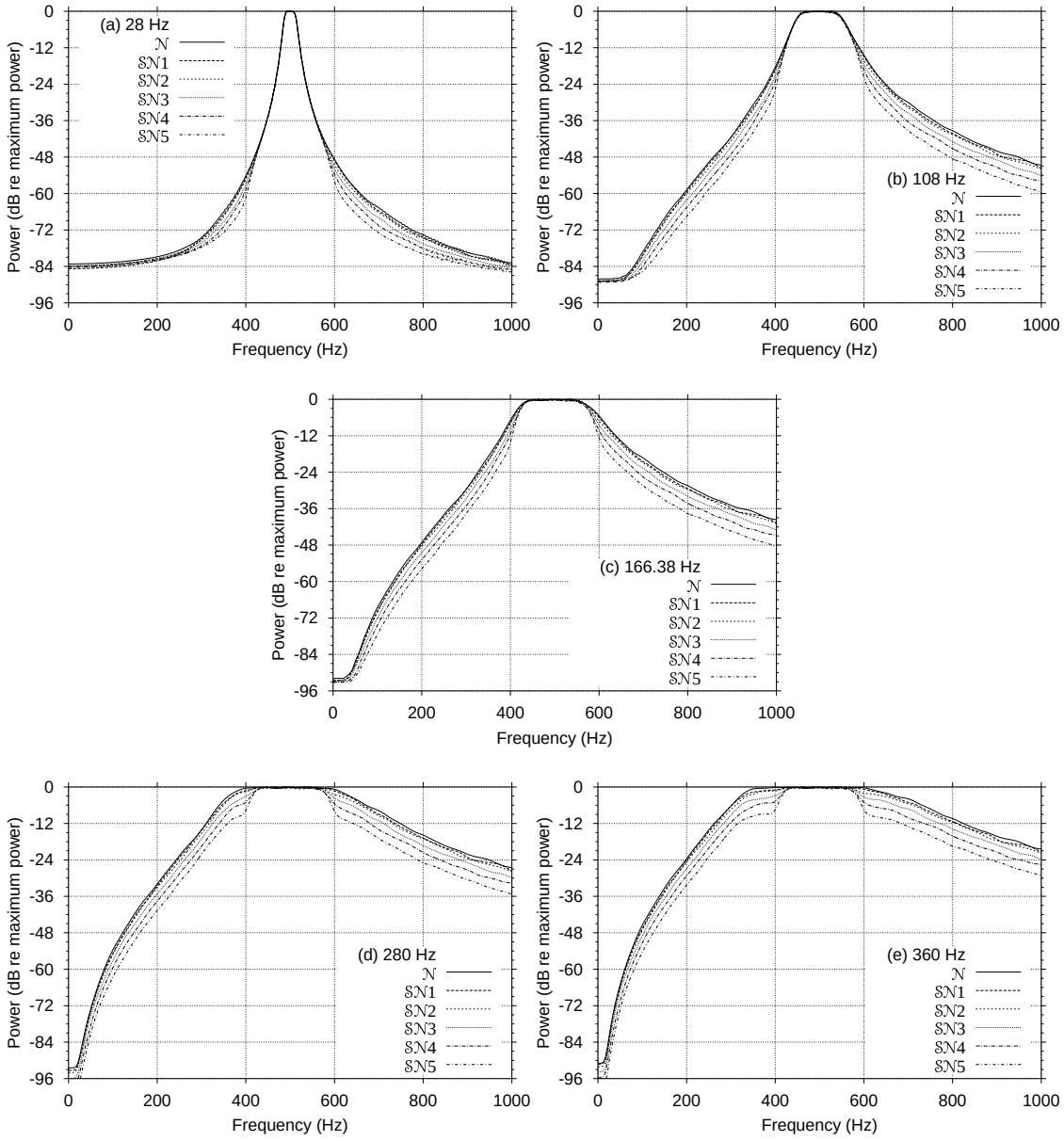

Figure 4.4: The spectral–averaged filter output of the $\mathcal{WT}$=4, {25ms, 160Hz: 16dB}, signals after Butterworth filtering with TdB bandwidths of (a) 28 Hz, (b) 108 Hz, (c) 166.38 Hz (equal–bandwidth), (d) 280 Hz, and (e) 360 Hz.

Table 4.1: Estimates of $\mathcal{W}$, $\mathcal{T}$, $\mathcal{WT}$, and signal–to–noise ratio from equal–bandwidth energy detector simulations using the $\mathcal{WT}$=1 buffers.

Signal Set	SNR		$\mathrm{Ess}\mathcal{WT}_{92.4\%}$			$\mathcal{WT}_{\mathrm{AERB}\times\mathrm{ERD}}$			$\mathrm{Ess}\mathcal{WT}_{95\%}$			$\mathcal{WT}_{\mathrm{TdB}\times\mathrm{ERD}}$		
	Orig.	Est.	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 2.5Hz	$-\infty$	$-\infty$	2.02	1049.17	2.12	2.69	891.07	2.40	2.23	1195.70	2.67	0.71	891.07	0.63
	0	0.300	1.83	1026.98	1.87	2.33	857.88	2.00	2.08	1172.13	2.43	0.74	857.88	0.63
	4	4.694	1.67	1010.93	1.69	2.14	814.36	1.75	1.96	1155.22	2.26	0.73	814.36	0.60
	8	8.784	1.62	1000.57	1.62	2.06	831.78	1.71	1.85	1144.40	2.12	0.73	831.78	0.61
	12	12.277	1.58	1002.54	1.58	2.02	825.95	1.67	1.76	1146.57	2.02	0.73	825.95	0.60
	16	16.109	1.57	998.94	1.57	1.99	827.61	1.65	1.73	1142.86	1.97	0.73	827.61	0.60
200ms, 5Hz	$-\infty$	$-\infty$	4.19	325.66	1.36	3.61	337.61	1.22	4.53	362.42	1.64	3.26	337.61	1.10
	0	−0.933	3.90	318.49	1.24	3.24	324.79	1.05	4.30	353.18	1.52	3.05	324.79	0.99
	4	3.769	3.64	314.04	1.14	3.04	318.39	0.97	4.08	347.49	1.42	2.92	318.39	0.93
	8	7.612	3.42	307.09	1.05	2.92	310.14	0.91	3.91	339.33	1.33	2.84	310.14	0.88
	12	11.807	3.33	306.13	1.02	2.85	315.60	0.90	3.77	337.87	1.27	2.80	315.60	0.88
	16	15.923	3.32	304.59	1.01	2.82	309.83	0.87	3.71	336.21	1.25	2.78	309.83	0.86
100ms, 10Hz	$-\infty$	$-\infty$	8.18	163.97	1.34	7.26	165.07	1.20	8.96	182.26	1.63	6.90	165.07	1.14
	0	−1.322	7.67	159.64	1.22	6.57	162.71	1.07	8.44	176.98	1.49	6.38	162.71	1.04
	4	3.996	7.21	155.16	1.12	6.05	162.68	0.98	7.92	171.65	1.36	5.93	162.68	0.96
	8	8.141	6.99	154.27	1.08	5.80	158.77	0.92	7.53	170.34	1.28	5.73	158.77	0.91
	12	11.813	6.86	152.81	1.05	5.69	157.70	0.90	7.36	168.57	1.24	5.65	157.70	0.89
	16	16.052	6.79	152.37	1.03	5.62	154.99	0.87	7.29	168.07	1.22	5.59	154.99	0.87
50ms, 20Hz	$-\infty$	$-\infty$	16.01	81.82	1.31	14.26	83.39	1.19	17.61	90.98	1.60	13.67	83.39	1.14
	0	−0.433	14.90	79.87	1.19	12.74	80.88	1.03	16.32	88.47	1.44	12.50	80.88	1.01
	4	2.766	14.28	78.99	1.13	12.05	80.71	0.97	15.61	87.32	1.36	11.93	80.71	0.96
	8	7.632	13.75	77.45	1.07	11.51	79.06	0.91	14.99	85.51	1.28	11.45	79.06	0.91
	12	11.425	13.46	77.01	1.04	11.24	79.65	0.90	14.62	84.94	1.24	11.23	79.65	0.89
	16	15.825	13.30	76.59	1.02	11.12	78.30	0.87	14.43	84.46	1.22	11.13	78.30	0.87
25ms, 40Hz	$-\infty$	$-\infty$	32.05	41.90	1.34	28.61	42.65	1.22	35.28	46.49	1.64	27.31	42.65	1.16
	0	0.278	29.89	39.59	1.18	25.57	41.24	1.05	32.76	43.87	1.44	25.12	41.24	1.04
	4	3.975	28.59	39.28	1.12	24.10	40.18	0.97	31.25	43.45	1.36	23.89	40.18	0.96
	8	8.112	27.42	38.79	1.06	22.98	40.00	0.92	29.87	42.84	1.28	22.95	40.00	0.92
	12	12.348	26.88	38.39	1.03	22.51	39.24	0.88	29.18	42.37	1.24	22.55	39.24	0.88
	16	16.128	26.59	38.29	1.02	22.27	38.88	0.87	28.85	42.22	1.22	22.34	38.88	0.87
12.5ms, 80Hz	$-\infty$	$-\infty$	64.53	20.81	1.34	57.69	21.37	1.23	71.05	23.10	1.64	55.18	21.37	1.18
	0	−0.105	59.57	20.05	1.19	50.98	20.77	1.06	65.28	22.21	1.45	50.08	20.77	1.04
	4	3.648	56.80	19.77	1.12	47.82	19.99	0.96	62.05	21.85	1.36	47.46	19.99	0.95
	8	7.455	54.90	19.44	1.07	46.10	19.74	0.91	59.77	21.47	1.28	46.00	19.74	0.91
	12	11.913	53.81	19.18	1.03	45.12	19.47	0.88	58.47	21.16	1.24	45.21	19.47	0.88
	16	15.920	53.12	19.19	1.02	44.51	19.95	0.89	57.66	21.17	1.22	44.67	19.95	0.89
6.25ms, 160Hz	$-\infty$	$-\infty$	127.64	10.37	1.32	113.57	10.72	1.22	140.40	11.56	1.62	109.13	10.72	1.17
	0	0.635	118.15	10.10	1.19	100.58	10.26	1.03	129.44	11.22	1.45	99.06	10.26	1.02
	4	3.705	113.70	9.93	1.13	95.88	10.45	1.00	124.21	10.99	1.36	95.16	10.45	0.99
	8	7.873	109.36	9.79	1.07	91.70	9.79	0.90	119.04	10.79	1.28	91.63	9.79	0.90
	12	11.793	106.92	9.71	1.04	89.62	10.03	0.90	116.14	10.74	1.25	89.87	10.03	0.90
	16	15.906	105.90	9.67	1.02	88.70	10.03	0.89	114.93	10.68	1.23	89.10	10.03	0.89

Table 4.2: Estimates of $\mathcal{W}$, $\mathcal{T}$, $\mathcal{WT}$, and signal–to–noise ratio from equal–bandwidth energy detector simulations using the $\mathcal{WT}$=2 buffers.

Signal Set	SNR		Ess$\mathcal{WT}_{92.4\%}$			$\mathcal{WT}_{\text{AERB}\times\text{ERD}}$			Ess$\mathcal{WT}_{95\%}$			$\mathcal{WT}_{\text{TdB}\times\text{ERD}}$		
	Orig.	Est.	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 5Hz	$-\infty$	$-\infty$	5.11	455.22	2.33	4.62	431.33	1.99	5.63	504.13	2.84	4.31	431.33	1.86
	-4	-4.523	4.91	448.27	2.20	4.48	432.47	1.94	5.41	497.08	2.69	4.10	432.47	1.77
	0	-0.294	4.66	450.27	2.10	4.14	442.94	1.83	5.08	498.46	2.53	3.99	442.94	1.77
	4	4.190	4.37	445.31	1.95	4.01	433.48	1.74	4.78	492.76	2.36	3.87	433.48	1.68
	8	7.723	4.24	448.90	1.90	3.87	442.56	1.71	4.58	495.80	2.27	3.70	442.56	1.64
	12	11.886	4.15	446.41	1.85	3.88	448.09	1.74	4.43	492.56	2.18	3.70	448.09	1.66
200ms, 10Hz	$-\infty$	$-\infty$	10.48	227.54	2.38	9.43	227.42	2.14	11.53	251.53	2.90	8.78	227.42	2.00
	-4	-4.006	9.87	228.14	2.25	8.80	229.67	2.02	10.90	250.88	2.73	8.34	229.67	1.92
	0	-0.165	9.48	224.34	2.13	8.60	220.91	1.90	10.35	248.23	2.57	8.20	220.91	1.81
	4	4.001	8.87	225.54	2.00	8.18	230.86	1.89	9.71	248.19	2.41	7.83	230.86	1.81
	8	7.690	8.58	225.47	1.93	8.08	224.31	1.81	9.31	248.91	2.32	7.74	224.31	1.74
	12	12.105	8.31	221.82	1.84	7.79	223.04	1.74	8.90	245.00	2.18	7.48	223.04	1.67
100ms, 20Hz	$-\infty$	$-\infty$	21.13	114.53	2.42	18.93	112.42	2.13	23.32	126.80	2.96	17.86	112.42	2.01
	-4	-3.520	19.85	112.09	2.22	18.24	110.53	2.02	21.76	123.75	2.69	17.28	110.53	1.91
	0	0.324	19.19	112.30	2.16	17.55	111.90	1.96	20.98	124.06	2.60	16.64	111.90	1.86
	4	3.860	18.06	112.50	2.03	16.74	111.80	1.87	19.64	124.14	2.44	15.95	111.80	1.78
	8	8.143	17.25	111.67	1.93	16.07	110.83	1.78	18.64	123.14	2.30	15.45	110.83	1.71
	12	11.921	16.87	110.98	1.87	16.17	108.59	1.76	18.12	122.63	2.22	15.62	108.59	1.70
50ms, 40Hz	$-\infty$	$-\infty$	41.58	57.22	2.38	38.43	58.11	2.23	45.80	63.17	2.89	35.66	58.11	2.07
	-4	-2.597	39.49	56.39	2.23	34.65	56.55	1.96	43.45	62.22	2.70	33.31	56.55	1.88
	0	1.040	37.30	55.45	2.07	33.54	53.82	1.81	40.77	61.40	2.50	32.35	53.82	1.74
	4	4.027	35.68	55.99	2.00	32.58	56.60	1.84	38.80	61.82	2.40	31.71	56.60	1.79
	8	8.235	34.32	55.66	1.91	31.99	56.15	1.80	37.00	61.36	2.27	31.44	56.15	1.77
	12	12.304	33.06	55.28	1.83	30.00	53.88	1.62	35.62	61.03	2.17	29.77	53.88	1.60
25ms, 80Hz	$-\infty$	$-\infty$	83.89	28.20	2.37	75.94	28.34	2.15	92.41	31.15	2.88	71.66	28.34	2.03
	-4	-4.742	80.16	28.12	2.25	72.89	28.53	2.08	88.20	31.08	2.74	68.77	28.53	1.96
	0	-0.183	76.18	28.02	2.13	69.31	27.37	1.90	83.36	31.03	2.59	66.20	27.37	1.81
	4	3.965	72.39	27.76	2.01	66.50	27.75	1.85	78.61	30.67	2.41	64.59	27.75	1.79
	8	7.476	69.88	28.27	1.98	65.68	28.23	1.85	75.38	31.16	2.35	63.71	28.23	1.80
	12	11.788	67.24	27.61	1.86	61.93	27.51	1.70	72.35	30.47	2.20	61.26	27.51	1.69
12.5ms, 160Hz	$-\infty$	$-\infty$	164.34	14.33	2.35	149.66	14.20	2.13	180.93	15.85	2.87	141.22	14.20	2.01
	-4	-3.561	155.32	14.20	2.21	135.87	14.00	1.90	170.89	15.75	2.69	131.63	14.00	1.84
	0	-0.233	149.86	14.32	2.15	134.55	14.10	1.90	164.00	15.78	2.59	128.02	14.10	1.81
	4	4.548	138.65	14.00	1.94	123.67	13.90	1.72	150.58	15.49	2.33	121.85	13.90	1.69
	8	7.851	134.76	14.07	1.90	123.31	13.75	1.70	145.57	15.59	2.27	121.72	13.75	1.67
	12	11.897	132.25	13.99	1.85	121.92	13.98	1.70	142.27	15.47	2.20	120.21	13.98	1.68

Table 4.3: Estimates of $\mathcal{W}$, $\mathcal{T}$, $\mathcal{WT}$, and signal–to–noise ratio from equal–bandwidth energy detector simulations using the $\mathcal{WT}$=4 buffers.

Signal Set	SNR		Ess$\mathcal{WT}_{92.4\%}$			$\mathcal{WT}_{\text{AERB}\times\text{ERD}}$			Ess$\mathcal{WT}_{95\%}$			$\mathcal{WT}_{\text{TdB}\times\text{ERD}}$		
	Orig.	Est.	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 10Hz	$-\infty$	$-\infty$	12.24	416.44	5.10	10.87	399.16	4.34	13.46	456.40	6.14	10.03	399.16	4.00
	-8	-8.350	11.77	411.51	4.85	10.89	410.29	4.47	12.91	450.64	5.82	10.12	410.29	4.15
	-4	-3.886	11.38	409.60	4.66	10.56	400.54	4.23	12.54	450.71	5.65	9.88	400.54	3.96
	0	-0.149	10.68	404.43	4.32	10.10	394.45	3.99	11.71	446.38	5.23	9.57	394.45	3.77
	4	3.726	10.06	408.45	4.11	9.88	386.73	3.82	10.91	448.71	4.89	9.56	386.73	3.70
	8	7.921	9.66	401.51	3.88	9.50	389.72	3.70	10.21	443.17	4.52	9.41	389.72	3.67
200ms, 20Hz	$-\infty$	$-\infty$	24.08	203.45	4.90	21.95	193.06	4.24	26.48	223.41	5.92	20.44	193.06	3.95
	-8	-8.230	23.66	204.79	4.85	21.31	197.25	4.20	26.13	225.66	5.90	19.51	197.25	3.85
	-4	-4.140	22.60	204.64	4.63	20.84	194.56	4.05	24.83	224.90	5.58	19.52	194.56	3.80
	0	-0.345	21.38	205.29	4.39	20.50	203.89	4.18	23.38	225.58	5.27	19.25	203.89	3.93
	4	4.030	19.89	203.16	4.04	19.60	194.14	3.81	21.43	222.66	4.77	18.83	194.14	3.65
	8	7.977	19.10	203.20	3.88	19.09	193.68	3.70	20.34	224.32	4.56	18.31	193.68	3.55
100ms, 40Hz	$-\infty$	$-\infty$	48.14	103.27	4.97	44.72	101.07	4.52	53.02	113.58	6.02	41.61	101.07	4.20
	-8	-7.020	47.04	102.21	4.81	43.00	98.33	4.23	51.81	112.63	5.83	39.82	98.33	3.92
	-4	-4.071	45.02	103.05	4.64	41.55	98.71	4.10	49.66	113.33	5.63	38.99	98.71	3.85
	0	0.210	42.59	102.81	4.38	39.88	100.48	4.01	46.67	113.04	5.28	38.19	100.48	3.84
	4	3.954	40.13	101.86	4.09	39.05	97.92	3.82	43.16	112.01	4.83	37.73	97.92	3.69
	8	8.004	38.40	101.93	3.91	38.31	98.49	3.77	40.86	112.29	4.59	37.52	98.49	3.70
50ms, 80Hz	$-\infty$	$-\infty$	95.38	51.32	4.90	89.24	50.54	4.51	105.06	56.68	5.95	83.01	50.54	4.20
	-8	-7.224	93.13	51.54	4.80	86.14	49.05	4.22	102.90	56.64	5.83	78.98	49.05	3.87
	-4	-4.262	91.32	51.44	4.70	85.22	49.69	4.23	100.94	56.46	5.70	78.64	49.69	3.91
	0	-0.021	85.51	51.33	4.39	82.60	48.87	4.04	93.26	56.47	5.27	77.05	48.87	3.77
	4	3.921	79.46	51.06	4.06	79.02	50.57	4.00	85.54	56.25	4.81	75.28	50.57	3.81
	8	8.076	76.25	51.54	3.93	77.34	49.52	3.83	81.19	56.46	4.58	73.84	49.52	3.66
25ms, 160Hz	$-\infty$	$-\infty$	194.03	26.23	5.09	177.40	25.26	4.48	213.02	28.79	6.13	165.57	25.26	4.18
	-8	-7.345	188.86	25.56	4.83	175.34	25.16	4.41	207.63	28.10	5.83	162.18	25.16	4.08
	-4	-3.814	181.86	25.80	4.69	169.08	24.13	4.08	200.84	28.35	5.69	158.65	24.13	3.83
	0	0.261	170.42	25.61	4.37	162.38	25.62	4.16	186.23	28.17	5.25	153.68	25.62	3.94
	4	4.207	160.15	25.88	4.14	156.73	25.06	3.93	172.69	28.40	4.90	149.86	25.06	3.76
	8	8.139	152.43	25.73	3.92	152.58	24.82	3.79	162.17	28.27	4.59	149.29	24.82	3.70

4.4 Analysis with the gamma model

The signal analysis from the energy detector simulations in Section 4.3, showed that $\mathcal{WT}$ decreased as signal–to–noise ratio increased, for all experimental levels. This implies that the χ^2 model of energy detection does not fully capture the nature of energy detection of real signals with real filters, because it assumes $\mathcal{WT}$ is constant with signal–to–noise ratio. The χ^2 distribution is, however, a special case of the gamma distribution. The parameters of the gamma distribution can be related to signal–to–noise ratio and $\mathcal{WT}$, where the latter can take on real, rather than integer, values. The aim, therefore, of the gamma model analysis is to evaluate whether it is possible, or useful, to model noise–in–noise detection using real–valued $\mathcal{WT}$, by (a) estimating the effective $\mathcal{WT}$ and signal–to–noise ratios from the output of the energy detector simulations, then (b) comparing these estimators with the original signal statistics, and (c) the estimates from the signal analysis of the simulations. Using the gamma density function, although intuitively appealing, has not been justified or derived from the properties of the stimuli, thus it can only be considered a data model.

4.4.1 The gamma probability density function $Ga(x|\lambda, n)$

The gamma probability density function, $Ga(x|\lambda, n)$ is defined (Borowski & Borwein, 1989; Larson, 1982) as

$$Ga(x|\lambda, n) = \frac{\lambda^n x^{n-1}}{\Gamma(n)} e^{-\lambda x}, \qquad x, n, \lambda > 0 \tag{4.8}$$

with mean n/λ and variance n/λ^2. The parameters n and λ are sometimes referred to as the shape and scale parameters (some authors define the scale parameter as $\alpha = 1/\lambda$). The gamma *p.d.f.* has a maximum at $(n-1)/\lambda$ when $n \geqslant 1$ but is unbounded as x tends to zero for $0 < n < 1$.

For half integer n and $\lambda = 1/2$, the gamma *p.d.f.* is known as the χ^2 *p.d.f.* with degrees of freedom, $\nu = 2n$. The non-standardised χ^2 allows for λ to range over $[0, \infty)$, and is the form traditionally used to model the energy detector, because λ can be related to the signal–to–noise ratio.

4.4.2 Justification of gamma model for energy detection

Because of the relationship between the χ^2 and gamma *p.d.f.'s*, it is a logical step to allow non-integer n and use the gamma *p.d.f.* to model the energy detector. This approach has not been justified or derived from the representation of the waveforms, like the χ^2 model, but there is no obvious reason why it would not be a useful representation.[10]

Others have also modelled the output of an energy detector with the gamma distribution. For instance, Steenson and Stirling (1965) used Monte Carlo simulations to model the transformation of white Gaussian noise through a narrow–band Butterworth filter, a square–law rectifier, and a leaky integrator. Their motivation for using the gamma *p.d.f.* came from work by Bussgang, Nesbeda, and Safran (1959; in Steenson and Stirling, 1965) who suggested that the amplitude output of the post detection filter could be modelled using the gamma distribution. They found an excellent approximation to their data with the gamma *p.d.f,* where the degrees of freedom

[10] A different form of waveform representation would be required that results in real degrees of freedom. The form of waveform representation described in Chapter 1 is only one of many possibilities.

parameter, N, was fitted to the right-hand tail of the samples. With respect to the amplitude distribution, Steenson and Stirling (1965) say they are not simple mathematical functions. The gamma distribution is an approximation for when the detector (rectifier) is square–law. There does not appear, however, to be any mathematical development of the gamma model for energy detection.

4.4.3 Fitting the gamma model

The general procedure used to fit the gamma model was to take the evidence from the energy detector simulations, and estimate n and λ for each signal level. From these two parameters, $\mathcal{WT}$ and signal–to–noise ratio were estimated.

The relationship of $\mathcal{WT}$ to gamma parameters is straightforward because $\mathcal{WT} = n$. Relating signal–to–noise ratio to the gamma parameters requires more inference. It is assumed that this relationship is analogous to that for the χ^2 model. From Egan (1975) and Parzen (1960), the relationship between λ of the non-standardised χ^2 model and signal power is $\lambda = 1/2\sigma^2$, where σ^2 is the variance of the normal distribution (which is related to the signal power of the waveforms) that gave rise to the χ^2 distribution. Because χ^2 is a special case of the gamma, it is inferred that the relationship holds for the gamma parameter λ. For the $\mathcal{N}$ evidence $\sigma^2 = \sigma^2_{\mathcal{N}}$, and for the $\mathcal{SN}$ evidence $\sigma^2 = \sigma^2_{\mathcal{SN}}$, however, $\sigma^2_{\mathcal{S}}$ is needed to estimate the signal–to–noise ratio. Using the relationship $\sigma^2_{\mathcal{SN}} = \sigma^2_{\mathcal{S}} + \sigma^2_{\mathcal{N}}$,

$$\text{SNR} \quad = \quad 10 \log \left(\frac{\sigma^2_{\mathcal{SN}} - \sigma^2_{\mathcal{N}}}{\sigma^2_{\mathcal{N}}} \right). \tag{4.9}$$

So in terms of the estimators $\hat{n}$ and $\hat{\alpha} = 1/\hat{\lambda}$,

$$\widehat{\mathcal{WT}} \quad = \quad \hat{n} \tag{4.10}$$

and

$$\widehat{\text{SNR}} \quad = \quad 10 \log \left(\frac{\widehat{\alpha_{\mathcal{SN}}} - \widehat{\alpha_{\mathcal{N}}}}{\widehat{\alpha_{\mathcal{N}}}} \right). \tag{4.11}$$

There are a number of methods of parameter estimation, such as the method of moments (MOM) (Larson, 1982), the method of maximum likelihood (MLE) (Bowman & Shenton, 1988), and Bayesian methods (Robert, 1994). Bayesian methods are preferable because they allow the analyst to use prior knowledge. There does not seem to be, however, any easily available algorithms for gamma estimation using Bayesian methods so the next best method, MLE estimation, was used instead. Although perhaps preferable to MOM estimation, MLE estimation has been criticised as being unstable for small variation on x, at least for small sample sizes (Robert, 1994).

4.4.4 Method of maximum likelihood

The method of maximum likelihood estimation (MLE) is to find the likelihood function of the sample values, then estimate the unknown parameters by finding the values that maximise the likelihood function (Larson, 1982). The gamma distribution does not give a closed form solution for MLE estimators, so numerical methods are used instead. Bowman and Shenton (1988) show how this is done and give the algorithms required.

According to Bowman and Shenton (1988), the maximum likelihood estimators for the two parameter gamma distribution are:

$$\log_e(\hat{n}) - \psi(\hat{n}) \quad = \quad y \tag{4.12}$$

where $\psi(x) = d \log_e \Gamma(x)/dx$ is the polygamma function, and

$$\hat{n}/\hat{\lambda} \quad = \quad A \tag{4.13}$$

where $y = \log_e(A/G)$, with

$$A \quad = \quad \frac{1}{r}\sum_{i=1}^{r} x_i \tag{4.14}$$

the arithmetic mean, and

$$G \quad = \quad \sqrt[r]{\prod_{i=1}^{r} x_i} \tag{4.15}$$

the geometric mean. The program GAMMAIT (Lapsley Miller & Miller, 1998) implemented their FORTRAN algorithm in BORLAND PASCAL 7.0.

4.4.5 Gamma parameter estimation of the energy detector simulations

Because parameter estimation works better on large sample sets, the first set of analyses considered the VLS signals. They were generated using the same SIGGEN input parameters as the experimental levels with parameters {6.25ms, 160Hz}, {50ms, 40Hz}, and {50ms, 80Hz}. The two gamma estimators were calculated for each signal level, and the corresponding gamma distributions were then compared with the histograms of the evidence to visually assess their fit.

The estimation procedure was then run on the equal–bandwidth energy detector simulations using the buffer waveforms, despite the much reduced sample sizes. This enabled comparisons between all the different conditions and with the signal analysis.

There are a number of ways to assess the goodness–of–fit of the estimated parameters including the χ^2 test and assessing the fit in the ROC space compared with the χ^2 mathematical model and the energy detector simulation. Unfortunately, it is not appropriate to directly do a Kolmogorov–Smirnoff test to see if the gamma *c.d.f.* estimated from the data is significantly different from the empirical *c.d.f.* This is because the estimators from the data are used to derive the theoretical distribution against which the data are tested. Thus, the function used to test the significance of the Kolmogorov–Smirnoff statistic is no longer given by its standard formula; instead Monte–Carlo methods are required to find the new distribution (Press, 1988, p. 627). It would be appropriate, however, to use Kolmogorov–Smirnoff testing to see if the simulation data differ from the *theoretical* parameters, but we can already easily tell that they are not the same by visually inspecting the data.

The assumptions for the χ^2 test[11] can be met because a degree of freedom is lost for every estimated parameter. The χ^2 test is conservative, because it tends not to reject the null hypothesis. The arbitrariness of the significance level means that it should not be used—without thought—to

[11] Not to be confused with the χ^2 energy detector model!

reject the null hypothesis. Instead, it can be used to find patterns across signal–to–noise ratio or $\mathcal{WT}$, or patterns of the χ^2 value itself can be assessed.

VLS energy detector simulations

These simulations were described in Section 4.2.4. Figures 4.5, 4.6, and 4.7 shows histograms (represented with circles, because bars occluded too much detail) of the evidence output. Shown are the results from the $\mathcal{N}$ event and five $\mathcal{SN}$ events. Because the distributions are so asymmetrical, unequal bin widths were used. Each bin in the histogram contains at least 4000 tallies for $\mathcal{WT}$=1 and 1000 tallies for $\mathcal{WT}$=2 and $\mathcal{WT}$=4. Comparisons could only be made with the theoretical parameters and the simulation performance, because no signal analysis was done on these simulations.

Overlaying the histograms are the gamma density functions fitted using MLE estimation. The MLE estimators for the $\mathcal{N}$ events do not seem to describe the histogram particularly well; the evidence distributions appear to be more peaked. The largest $\mathcal{SN}$ event, however, is visually well fitted by the gamma estimator for $\{6.25\text{ms}, 160\text{Hz}\}$.

Table 4.4: The gamma estimators for the evidence distributions from ideal energy detector simulations on the VLS signals, and results of the χ^2 tests.

Signal Set	SNR (dB)	$\widehat{\mathcal{WT}}$	$\hat{\alpha}$	$\widehat{\text{SNR}}$ (dB)	χ^2	*df*	Significance
6.25ms, 160Hz	$-\infty$	1.5144	3.5008	$-\infty$	2887.2746	47	$p < 0.05$
	0	1.3251	7.0575	0.0687	1160.9994	22	$p < 0.05$
	4	1.2211	12.6567	4.1753	858.6533	22	$p < 0.05$
	8	1.1337	27.1890	8.3036	389.4510	22	$p < 0.05$
	12	1.0699	64.9117	12.4407	129.8899	22	$p < 0.05$
	16	1.0381	160.8218	16.5261	70.8782	22	$p < 0.05$
50ms, 40Hz	$-\infty$	2.7701	4.9813	$-\infty$	325.2306	22	$p < 0.05$
	-4	2.6070	7.0039	-3.9141	311.5102	22	$p < 0.05$
	0	2.4840	10.0096	0.0409	251.5399	22	$p < 0.05$
	4	2.3675	17.5445	4.0176	179.6901	22	$p < 0.05$
	8	2.2559	36.9563	8.0747	242.4668	22	$p < 0.05$
	12	2.2244	84.7393	12.0444	195.7136	22	$p < 0.05$
50ms, 80Hz	$-\infty$	5.7534	5.5463	$-\infty$	4773.4083	22	$p < 0.05$
	-8	5.6484	6.4073	-8.0899	3445.8468	22	$p < 0.05$
	-4	5.3646	7.9705	-3.5943	2255.6730	22	$p < 0.05$
	0	5.1125	11.5426	0.3388	1132.1888	22	$p < 0.05$
	4	4.8230	20.8801	4.4165	400.5370	22	$p < 0.05$
	8	4.6800	43.8103	8.3879	184.2035	22	$p < 0.05$

Table 4.4 shows the results of the χ^2 goodness of fit tests on the MLE gamma estimators. All are significant at the 5% level indicating that the underlying distributions are not well fitted by a gamma density function.

Although the signal analysis is based on the buffer signal sets, the estimates from the VLS signal sets may still be compared, because the same process was used to generate both sets. Comparison of the $\mathcal{WT}$ estimators, in Table 4.4, with the signal analysis in Tables 4.1, 4.2, and 4.3 shows similar trends, as $\mathcal{WT}$ decreases as signal–to–noise ratio increases. The overall size of $\mathcal{WT}$

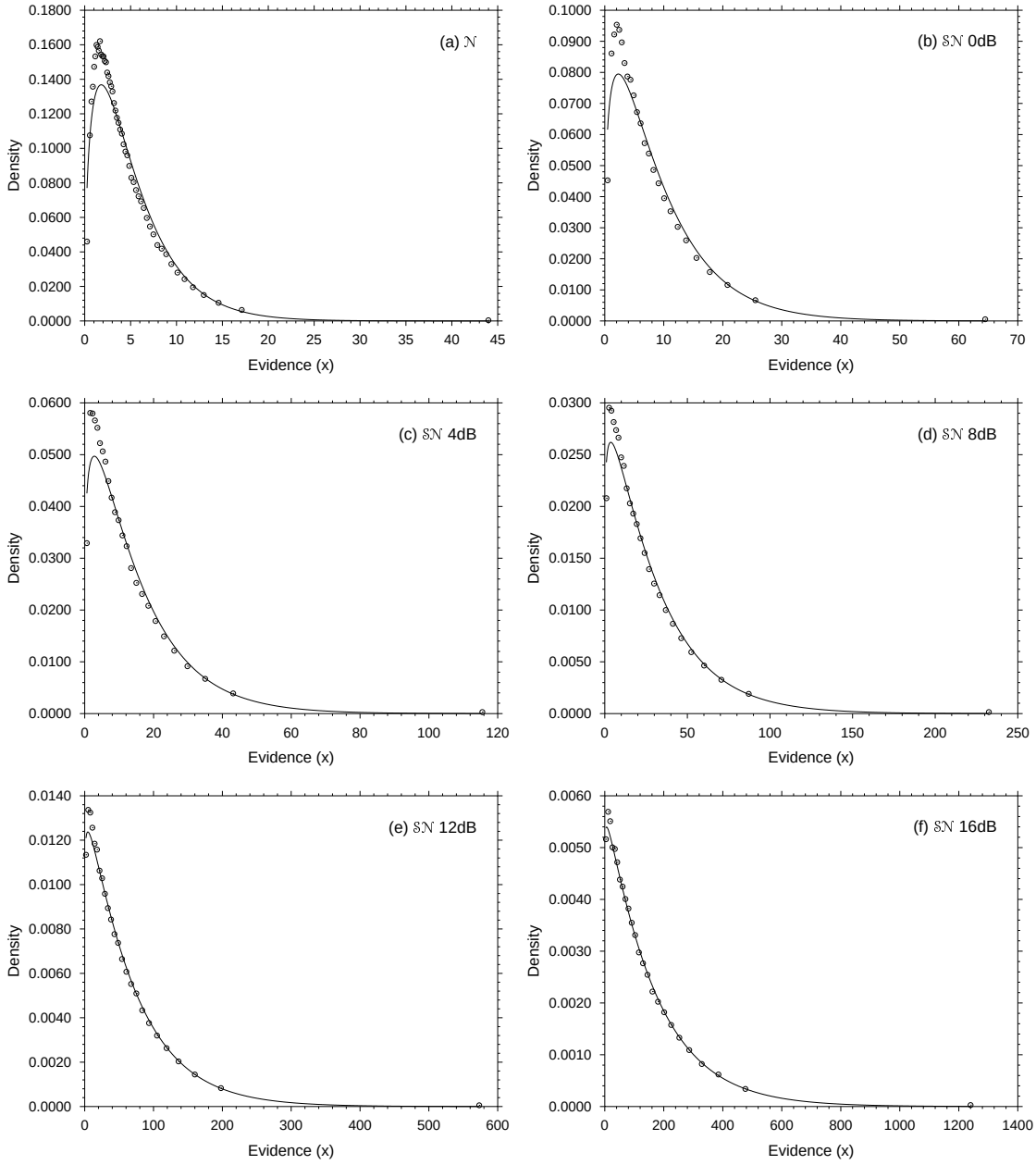

Figure 4.5: Histograms with unequal bin–widths (represented with circles instead of bars for clarity) of the output of energy detector simulations for the {6.25ms, 160Hz} VLS waveforms compared with gamma probability density functions, fitted using MLE estimation. (a) is the $\mathcal{N}$ event based on 200 000 stimuli and (b–f) are the $\mathcal{SN}$ events, for 0, 4, 8, 12, and 16 dB respectively, based on 100 000 stimuli.

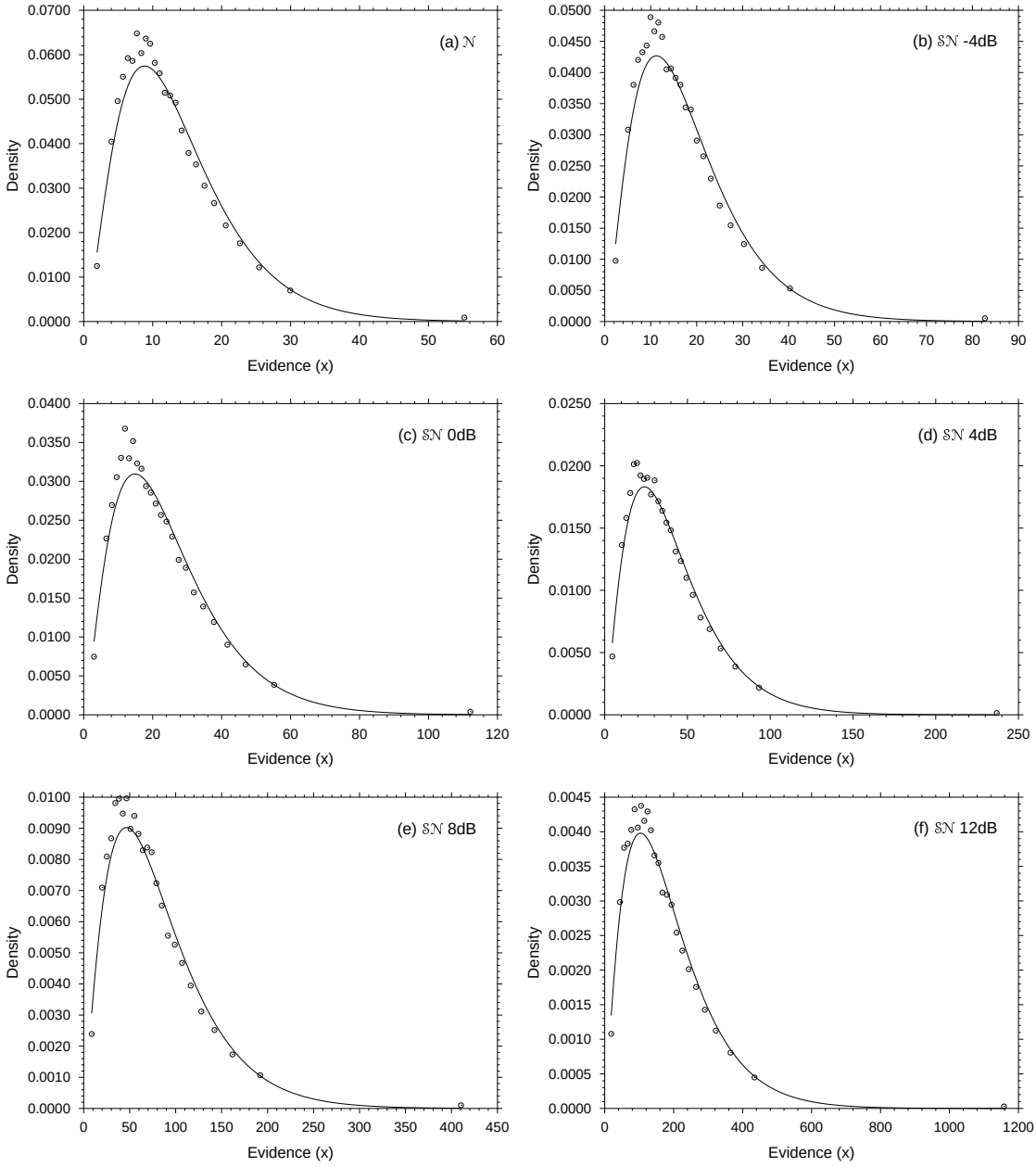

Figure 4.6: Histograms with unequal bin–widths (represented with circles instead of bars for clarity) of the output of energy detector simulations for the {50ms, 40Hz} VLS waveforms compared with gamma probability density functions, fitted using MLE estimation. (a) is the $\mathcal{N}$ event based on 25 000 stimuli and (b–f) are the $\mathcal{SN}$ events, for -4, 0, 4, 8, and 12 dB respectively, based on 25 000 stimuli.

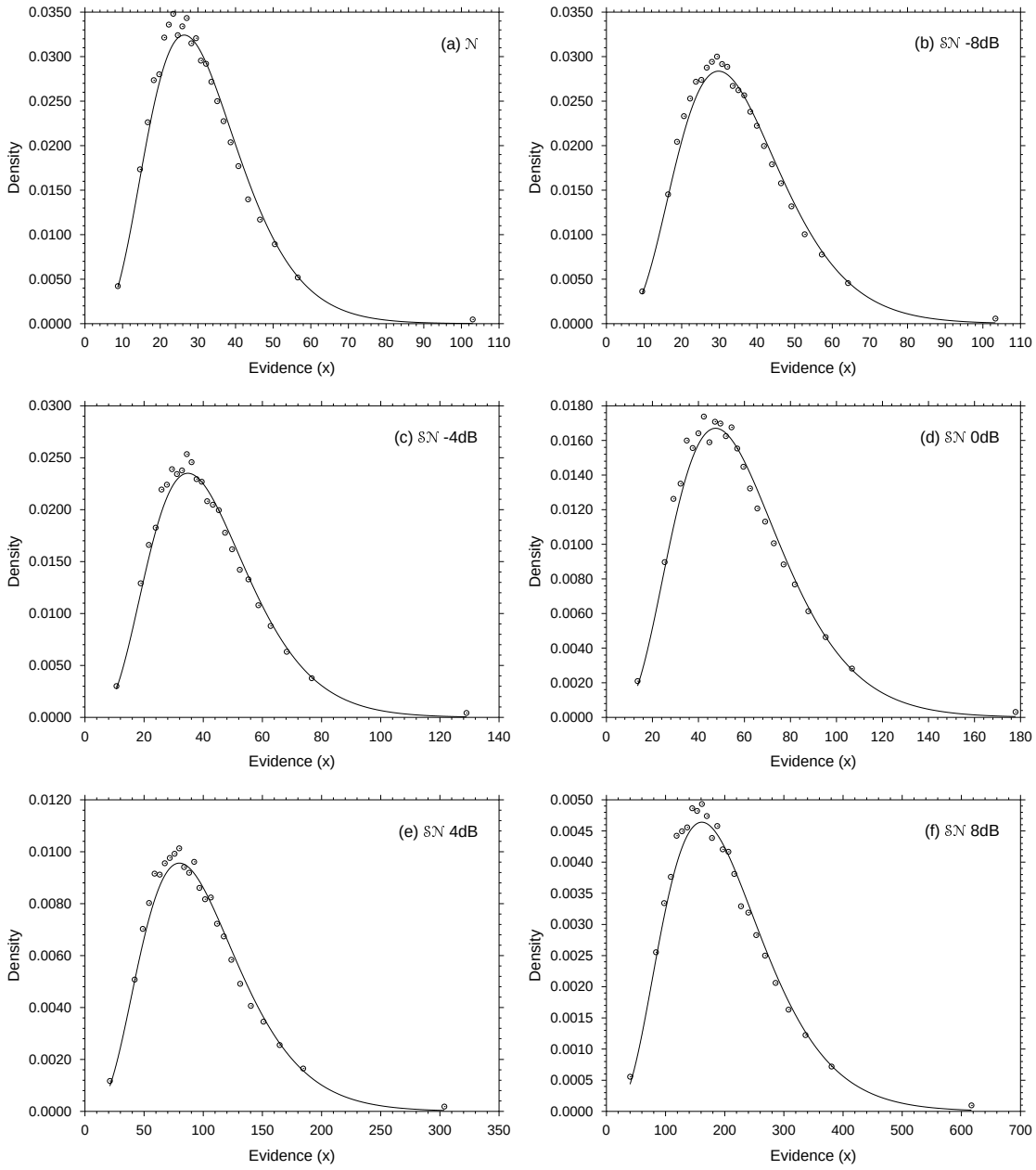

Figure 4.7: Histograms with unequal bin–widths (represented with circles instead of bars for clarity) of the output of energy detector simulations for the {50ms, 80Hz} VLS waveforms compared with gamma probability density functions, fitted using MLE estimation. (a) is the $\mathcal{N}$ event based on 25 000 stimuli and (b–f) are the $\mathcal{SN}$ events, for -8, -4, 0, 4, and 8 dB respectively, based on 25 000 stimuli.

is also bigger than that estimated from the signal analysis. The estimates of signal–to–noise ratio are also slightly larger.

Buffer energy detector simulations

Despite having only 500 samples per event for the buffer simulations, parameter estimation using the gamma distribution was done so that the estimators could be directly compared to the signal analysis (Tables 4.1, 4.2, and 4.3). Because the distributions are asymmetrical, unequal bin widths were used. Each bin contained at least 50 tallies.

Tables 4.5, 4.6, and 4.7 show:

- That the estimates of $\mathcal{WT}$ and signal–to–noise ratio appear to be similar to the original signal parameters.
- The estimated $\mathcal{WT}$ tends to decrease with increasing signal–to–noise ratio. This is the same phenomenon that was noticed in the signal analysis of the buffers, however, the actual values are different.
- The estimated signal–to–noise ratio appears to be similar to, but not the same as, both the theoretical signal–to–noise ratio and that estimated from the signal analysis of the buffers.
- $\mathcal{WT}$ and signal–to–noise ratio estimators tend to be slightly larger than the values from the signal analysis.
- The size of the χ^2 test statistic is relatively the same across signal level for the experimental condition $\mathcal{WT}$=1, but diverges for $\mathcal{WT}$=2 and $\mathcal{WT}$=4. For $\mathcal{WT}$=2, the $\mathcal{N}$ test statistic is 2–3 times larger than for the largest $\mathcal{SN}$ event indicating that as signal–to–noise ratio decreases the distribution of the evidence becomes less like a gamma distribution. For $\mathcal{WT}$=4 the χ^2 test statistic is huge for small signal–to–noise ratios.

The results were very similar to the gamma estimation based on the VLS signals but more variable. The MLE estimators for the $\mathcal{N}$ events do not describe the histograms particularly well. The fit improves for the $\mathcal{SN}$ event as signal–to–noise ratio increases, but the results of χ^2 tests indicate that there is still a difference for many cases.

Summary

All the comparisons of histograms and estimated parameters indicate that the gamma probability density function is not appropriate for modelling detection of these types of waveforms—especially for $\mathcal{N}$ and $\mathcal{SN}$ signals with small signal–to–noise ratio. The histograms tend to be more peaked than the gamma *p.d.f.* with the same statistics. This does *not* mean that the energy detector is inappropriate, merely that the detectability of the small–$\mathcal{WT}$ signals is not based on a gamma distribution. This is because the models of energy detection make a number of assumptions about the form of the input waveforms: (a) the spectrum is rectangular, (b) the filtering is ideal, and (c) the waveform is continuous in the time domain—or that a finite duration waveform does not cause spread in frequency domain. None of these assumptions are practically possible so it is no surprise that the distribution of the evidence is not the same.

Even though the gamma is not an ideal fit, is it still a useful descriptor of the evidence? One way of assessing this is by seeing if the ROC curves produced by the gamma densities are good fits to the ROC curves of the simulations.

Table 4.5: The gamma estimators of the evidence from ideal energy detector simulations on $\mathcal{WT}$=1 buffers and results of the χ^2 tests.

Signal Set	SNR (dB)	$\widehat{\mathcal{WT}}$	$\hat{\alpha}$	$\widehat{\mathrm{SNR}}$ (dB)	χ^2	*df*	Significance
400ms, 2.5Hz	$-\infty$	1.4817	0.7687	$-\infty$	63.0695	47	ns
	0	1.2169	1.6790	0.7337	71.3826	47	$p < 0.05$
	4	1.1646	3.0731	4.7678	42.0652	47	ns
	8	1.1094	6.7141	8.8840	40.0437	47	ns
	12	1.0971	13.9318	12.3358	45.9152	47	ns
	16	1.0987	32.0624	16.0968	42.3613	47	ns
200ms, 5Hz	$-\infty$	1.5221	0.5825	$-\infty$	43.9237	47	ns
	0	1.4293	1.0049	-1.3955	33.3610	47	ns
	4	1.1962	2.1120	4.1925	51.5921	47	ns
	8	1.0821	4.4916	8.2678	67.2670	47	$p < 0.05$
	12	1.1025	10.2512	12.2008	41.7298	47	ns
	16	1.0444	26.6114	16.5016	59.2353	47	ns
100ms, 10Hz	$-\infty$	1.6851	0.5118	$-\infty$	50.8509	47	ns
	0	1.4025	0.9680	-0.5005	62.6991	47	ns
	4	1.2371	2.0419	4.7557	47.7270	47	ns
	8	1.1744	4.4165	8.8244	61.4472	47	ns
	12	1.0483	10.4447	12.8793	51.6183	47	ns
	16	1.0244	26.9175	17.1255	63.1983	47	ns
50ms, 20Hz	$-\infty$	1.5427	0.5913	$-\infty$	53.4345	47	ns
	0	1.3832	1.1228	-0.4626	44.3256	47	ns
	4	1.1832	1.8831	3.3942	50.9732	47	ns
	8	1.1186	4.4755	8.1753	50.6593	47	ns
	12	1.0918	9.8057	11.9269	55.3164	47	ns
	16	1.0681	26.1276	16.3537	54.4244	47	ns
25ms, 40Hz	$-\infty$	1.5776	0.5269	$-\infty$	53.7034	47	ns
	0	1.3772	1.1148	0.4752	57.6628	47	ns
	4	1.2132	2.0190	4.5202	49.4631	47	ns
	8	1.1478	4.3471	8.6033	43.9225	47	ns
	12	1.0933	10.8708	12.9293	47.9336	47	ns
	16	1.0093	26.9241	16.9980	54.6320	47	ns
12.5ms, 80Hz	$-\infty$	1.4615	0.6028	$-\infty$	61.3070	47	ns
	0	1.3842	1.1114	-0.7377	52.8520	47	ns
	4	1.1402	2.1240	4.0203	70.0764	47	$p < 0.05$
	8	1.0842	4.2627	7.8331	51.9751	47	ns
	12	1.0662	10.6827	12.2329	42.9531	47	ns
	16	1.0720	25.4171	16.1454	49.3413	47	ns
6.25ms, 160Hz	$-\infty$	1.5539	0.5566	$-\infty$	47.8500	47	ns
	0	1.3035	1.2676	1.0636	53.6667	47	ns
	4	1.2024	2.0322	4.2344	81.5485	47	$p < 0.05$
	8	1.1068	4.4971	8.5002	30.5063	47	ns
	12	1.1636	9.4501	12.0354	64.3799	47	$p < 0.05$
	16	1.0463	25.7964	16.5656	55.6755	47	ns

Table 4.6: The gamma estimators of the evidence from ideal energy detector simulations on $\mathcal{WT}$=2 buffers and results of the χ^2 tests.

Signal Set	SNR (dB)	$\widehat{\mathcal{WT}}$	$\hat{\alpha}$	$\widehat{\mathrm{SNR}}$ (dB)	χ^2	*df*	Significance
400ms, 5Hz	$-\infty$	2.7333	0.7974	$-\infty$	80.5658	47	$p < 0.05$
	-4	2.3480	1.2196	-2.7616	82.5057	47	$p < 0.05$
	0	2.4129	1.5668	-0.1555	46.3998	47	ns
	4	2.2552	3.0399	4.4905	104.8504	47	$p < 0.05$
	8	2.4956	5.0613	7.2812	64.8280	47	$p < 0.05$
	12	2.3195	12.9688	11.8366	46.9882	47	ns
200ms, 10Hz	$-\infty$	2.8520	0.7676	$-\infty$	132.0424	47	$p < 0.05$
	-4	2.6319	1.0849	-3.8369	65.1422	47	$p < 0.05$
	0	2.8215	1.3894	-0.9148	73.1057	47	$p < 0.05$
	4	2.6600	2.5095	3.5589	45.9307	47	ns
	8	2.2071	5.8471	8.2069	74.2254	47	$p < 0.05$
	12	2.2492	13.8711	12.3226	49.1052	47	ns
100ms, 20Hz	$-\infty$	2.6679	0.8183	$-\infty$	135.1572	47	$p < 0.05$
	-4	2.7103	1.1215	-4.3112	89.2775	47	$p < 0.05$
	0	2.3623	1.7801	0.7021	56.5323	47	ns
	4	2.3736	2.7915	3.8228	74.3970	47	$p < 0.05$
	8	2.4789	5.6214	7.6864	53.9394	47	ns
	12	2.1989	14.0464	12.0861	53.9453	47	ns
50ms, 40Hz	$-\infty$	2.8997	0.7389	$-\infty$	78.0733	47	$p < 0.05$
	-4	2.5775	1.1619	-2.4223	52.5189	47	ns
	0	2.4153	1.7581	1.3970	65.3078	47	$p < 0.05$
	4	2.3987	2.6716	4.1758	56.0055	47	ns
	8	2.3264	5.8732	8.4190	36.0063	47	ns
	12	2.2391	13.4471	12.3550	54.8343	47	ns
25ms, 80Hz	$-\infty$	2.6431	0.8591	$-\infty$	120.6787	47	$p < 0.05$
	-4	2.6693	1.0906	-5.6956	86.6553	47	$p < 0.05$
	0	2.5276	1.6061	-0.6076	82.0287	47	$p < 0.05$
	4	2.4520	2.8319	3.6102	65.6740	47	$p < 0.05$
	8	2.2015	5.8811	7.6682	56.3617	47	ns
	12	2.1758	13.7001	11.7454	32.4579	47	ns
12.5ms, 160Hz	$-\infty$	2.9155	0.7498	$-\infty$	92.7545	47	$p < 0.05$
	-4	2.5444	1.1235	-3.0230	54.7852	47	ns
	0	2.5873	1.4796	-0.1172	64.0813	47	$p < 0.05$
	4	2.3021	3.0209	4.8132	50.8469	47	ns
	8	2.2050	5.7969	8.2812	53.6605	47	ns
	12	2.1004	13.9691	12.4629	46.0187	47	ns

Table 4.7: The gamma estimators of the evidence from ideal energy detector simulations on $\mathcal{WT}$=4 buffers and results of the χ^2 tests.

Signal Set	SNR (dB)	$\widehat{\mathcal{WT}}$	$\hat{\alpha}$	$\widehat{\mathrm{SNR}}$ (dB)	χ^2	*df*	Significance
400ms, 10Hz	$-\infty$	5.6078	0.9305	$-\infty$	66824.2977	47	$p < 0.05$
	-8	5.7778	1.0373	-9.4038	3610.3830	47	$p < 0.05$
	-4	5.7923	1.2329	-4.8810	830.4632	47	$p < 0.05$
	0	4.9339	1.9336	0.3260	261.6715	47	$p < 0.05$
	4	4.9543	3.2170	3.9045	107.8578	47	$p < 0.05$
	8	4.6170	7.1088	8.2214	68.9207	47	$p < 0.05$
200ms, 20Hz	$-\infty$	5.3325	0.9764	$-\infty$	667.4780	47	$p < 0.05$
	-8	5.1629	1.1263	-8.1404	495.0537	47	$p < 0.05$
	-4	4.9387	1.3865	-3.7681	231.5173	47	$p < 0.05$
	0	5.1050	1.8322	-0.5730	119.4464	47	$p < 0.05$
	4	4.5949	3.5708	4.2440	87.1131	47	$p < 0.05$
	8	4.6382	7.1024	7.9754	47.5130	47	ns
100ms, 40Hz	$-\infty$	5.6389	0.9096	$-\infty$	2951.8289	47	$p < 0.05$
	-8	6.1070	0.9679	-11.9341	1151.3025	47	$p < 0.05$
	-4	5.6834	1.1669	-5.4846	511.7167	47	$p < 0.05$
	0	4.9563	1.8912	0.3305	172.9706	47	$p < 0.05$
	4	4.7460	3.2888	4.1756	83.8762	47	$p < 0.05$
	8	4.5850	7.0094	8.2645	55.2484	47	ns
50ms, 80Hz	$-\infty$	5.6692	0.8897	$-\infty$	1743.7155	47	$p < 0.05$
	-8	5.0160	1.1544	-5.2641	1165.1101	47	$p < 0.05$
	-4	5.5780	1.1871	-4.7585	627.7478	47	$p < 0.05$
	0	5.1376	1.8130	0.1612	175.4348	47	$p < 0.05$
	4	4.9757	3.1113	3.9745	119.8667	47	$p < 0.05$
	8	4.6676	6.9505	8.3331	67.9883	47	$p < 0.05$
25ms, 160Hz	$-\infty$	5.3195	0.9422	$-\infty$	353.5518	47	$p < 0.05$
	-8	5.9728	0.9823	-13.7119	326.4431	47	$p < 0.05$
	-4	5.2457	1.2892	-4.3394	195.0536	47	$p < 0.05$
	0	4.9610	1.9069	0.1023	101.1715	47	$p < 0.05$
	4	5.2635	3.0579	3.5128	86.1814	47	$p < 0.05$
	8	4.8297	6.7072	7.8663	63.5610	47	ns

4.4.6 Assessing the gamma estimators in the ROC space

Assessing the gamma parameter estimators in the ROC space is appropriate because ROC analysis provides the basic data for many other analyses. The problem is that the $\mathcal{SN}$ events cannot be assessed independently from the $\mathcal{N}$ event, because the $\mathcal{N}$ evidence is used for the false–alarm rate for *all* the ROC curves. This is a particular problem here, because the indications from the other analyses suggest that the evidence from the $\mathcal{N}$ event is not distributed as a gamma.

Figure 4.8 compares the ROC curves from (a) the energy detection simulations, (b) the gamma MLE estimations, and (c) the ideal energy detector (χ^2 mathematical model) assuming the theoretical $\mathcal{WT}$ and signal–to–noise ratios—not those estimated from the signal analyses. For all $\mathcal{WT}$, the estimated gamma ROC curves and the theoretical χ^2 ROC curves are reasonably similar. The simulation ROC curves, however, are more symmetrical and fall slightly below the other ROC curves, on the left hand side of the ROC space, and cross over on the right-hand side. The estimated gamma ROC curve is closer to the χ^2 curve than the simulation from which the gamma was estimated. All three sets of ROC curves are different, but are probably not different enough to reliably test them in an experiment.

Figures 4.9, 4.10, and 4.11, on pages 136–138 at the end of this section, make the same comparisons for the buffer simulations. In general, for $\mathcal{WT}$=1, the fitted gamma ROC curves are a better fit to the simulation ROC curve than the χ^2 energy detector model with theoretical parameters (note that the gamma model using the theoretical parameters would result in the *same* ROC curve as the χ^2 model.) The χ^2 ROC curves tend to fall above the gamma and simulated ROC curves, indicating that the buffers did not result in the desired signal–to–noise ratio; in particular, levels {50ms, 20Hz} and {12.5ms, 80Hz}. Much of this difference, however, can be accounted for by sampling variability. There are also a number of excellent fits where the three sets of ROC curves essentially fall on top of one another. The three sets of $\mathcal{WT}$=2 and $\mathcal{WT}$=4 ROC curves were also very close for all experimental levels, and there were no signal levels with curves as disparate from the χ^2 model as the few rogue levels for $\mathcal{WT}$=1.

In sum, the majority of the χ^2 and gamma ROC curves fitted the simulated energy detector ROC curves surprisingly well. The gamma tended to give a better fit than the χ^2 model, because the gamma parameters were estimated from the data, rather than the theory, thereby reducing the effects of sampling variability. The results seem to indicate that although the distributions may be different to that of the simulation evidence, their properties are such that they still provide a reasonable model of the simulation's performance. It could be the case, for instance, that the simulation evidence is a subtle transform of a gamma distribution.

4.4.7 Summary

It is possible that another density could be more appropriate than the gamma. It is preferable, however, to only use densities that have some basis in theory (even if not fully developed as with the gamma). There does not appear to be any other distribution suggested in the literature, other than the χ^2 and gamma, which could be used as a model of energy detection. Drga (1998, personal communication) has mentioned some distributions, for which the gamma is a special case, but the difference could be modelled as the form of the rectifier. Because the rectifier in an energy detector is square–law, the other distributions could not be justified on this basis. They may, however, be useful in modelling a linear detector.

Investigation in the ROC space indicates that the χ^2 and gamma models are very nearly indis-

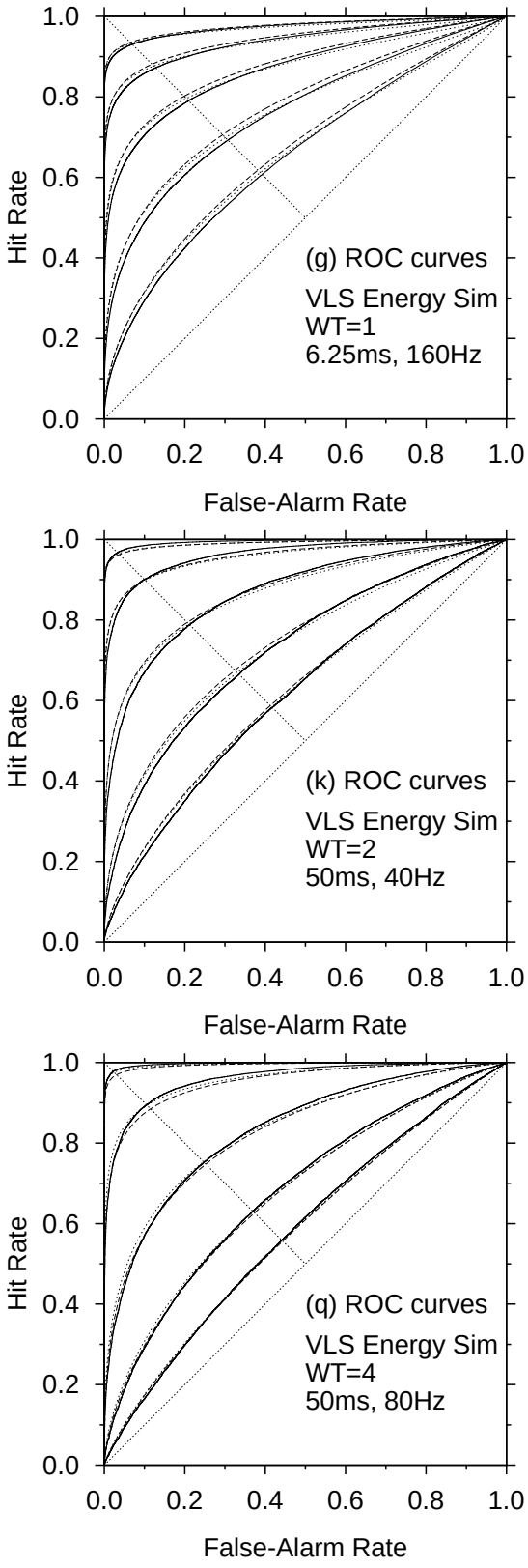

Figure 4.8: The simulated energy detector ROC curves for the VLS signal transients (solid curves), the corresponding fitted gamma ROC curves estimated from the simulation evidence (dotted curves), and the ideal χ^2 energy detector ROC curves (dashed curves).

tinguishable, and, therefore, practically equivalent. It is unlikely experimental data could make a distinction between these models in the ROC space. The simulated ROC curve, however, was slightly different to both models indicating that neither are totally appropriate. The gamma model may, however, prove more useful in estimating parameters when $\mathcal{WT}$ is not the same for each event.

Little mention has been made so far of the linear detectors. This is because mathematical theory has not been fully developed for them. It is assumed that if the energy detector does not produce evidence distributed as χ^2 for these types of signals, then it is likely the linear detectors would do the same (but for a different distribution such as the χ). This is because the signals and the filters are the same for all these detectors.

The next section explores in more detail the differences between these different detectors by considering differences in their psychometric functions.

4.5 Comparisons among detectors

There is some doubt as to whether the mathematical energy detector accurately predicts performance for real signals, because of simplifying assumptions in the mathematical derivation. There is also some doubt as to whether the energy, envelope, and full–linear detectors are equivalent for all $\mathcal{WT}$. These problems are not resolved, primarily because the mathematics involved is difficult. Simulations, however, provide an alternative approach to evaluating these models.

Firstly, the psychometric functions from the simulated energy detector are compared to the mathematical energy detector. Then, various comparisons are made among the energy, full–linear, and envelope detectors.

4.5.1 Energy detector psychometric functions

The psychometric function can be used to compare the performance of the mathematical energy detector, from Equation (1.15), to the simulated energy detector. Figure 4.12 shows the performance of the simulated energy detector (square symbols) relative to the theoretical psychometric function for (a) the buffer simulations and (b) the transient simulations. The most salient finding is that the empirical points for each experimental level are very similar, within each experimental condition, give or take some sampling variability. For the values considered, $\mathcal{W}$ and $\mathcal{T}$ may be traded off for the same performance. This also implies that the results may be averaged across signal–to–noise ratio for each type of simulation within an experimental condition.

Comparing the points for each experimental condition with the theory shows that

- for the seven $\mathcal{WT}$=1 levels, the fit with respect to shape is very good but there is some attenuation of the data points.
- for the six $\mathcal{WT}$=2 levels, the $\mathcal{WT}$=2 mathematical model does not fit the simulations' psychometric function; instead the mathematical model for $\mathcal{WT}$=3, with some linear attenuation, provides a more satisfactory fit.
- for the five $\mathcal{WT}$=4 levels, the mathematical model for $\mathcal{WT}$=5, with some linear attenuation, provides a better fit than $\mathcal{WT}$=4.

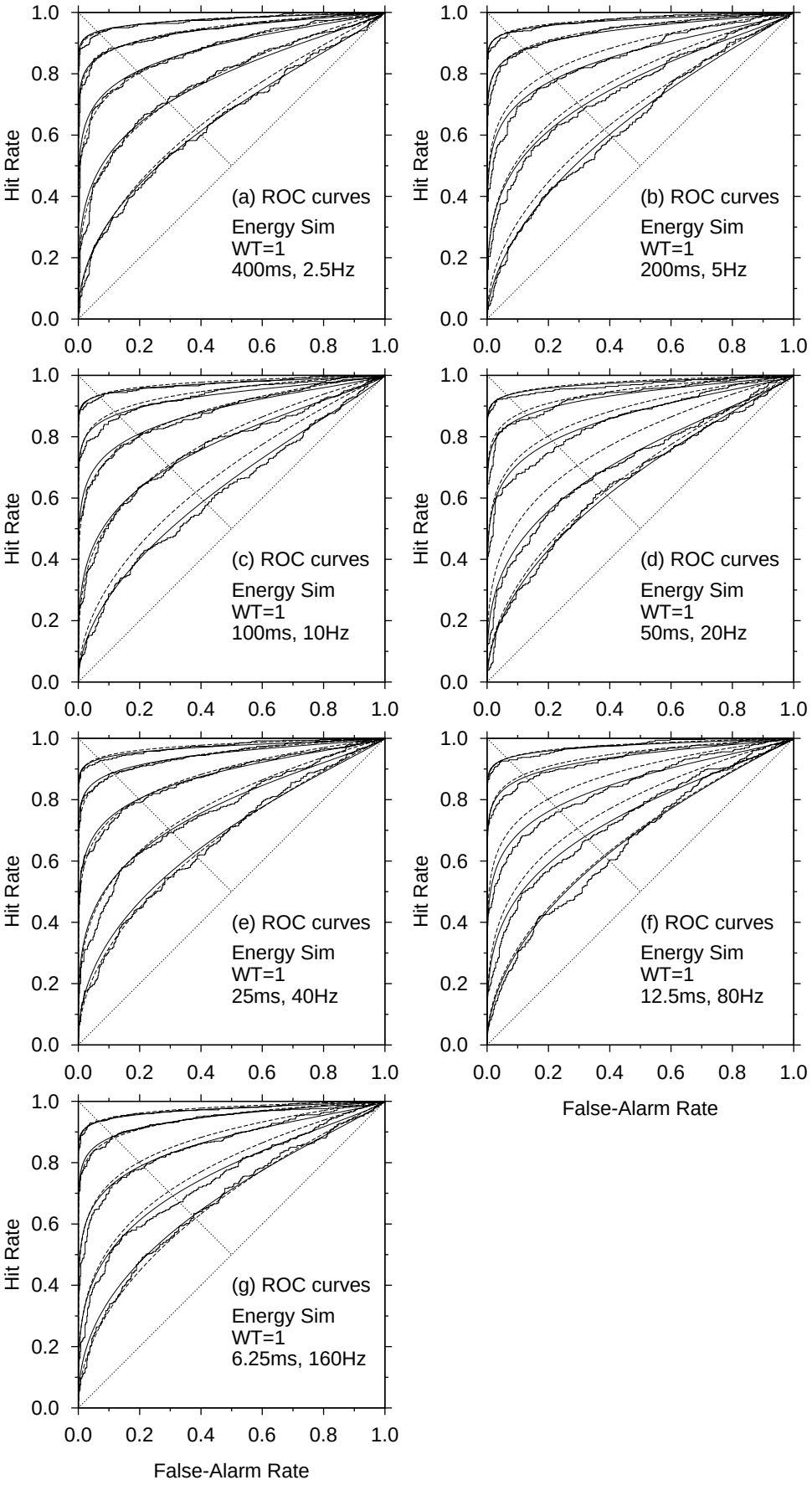

Figure 4.9: The simulated energy detector ROC curves for the signals used in condition $\mathcal{WT}=1$ (jagged solid curves), the corresponding fitted gamma ROC curves estimated from the simulation evidence (smooth solid curves), and the ideal χ^2 energy detector ROC curves (smooth dashed curves).

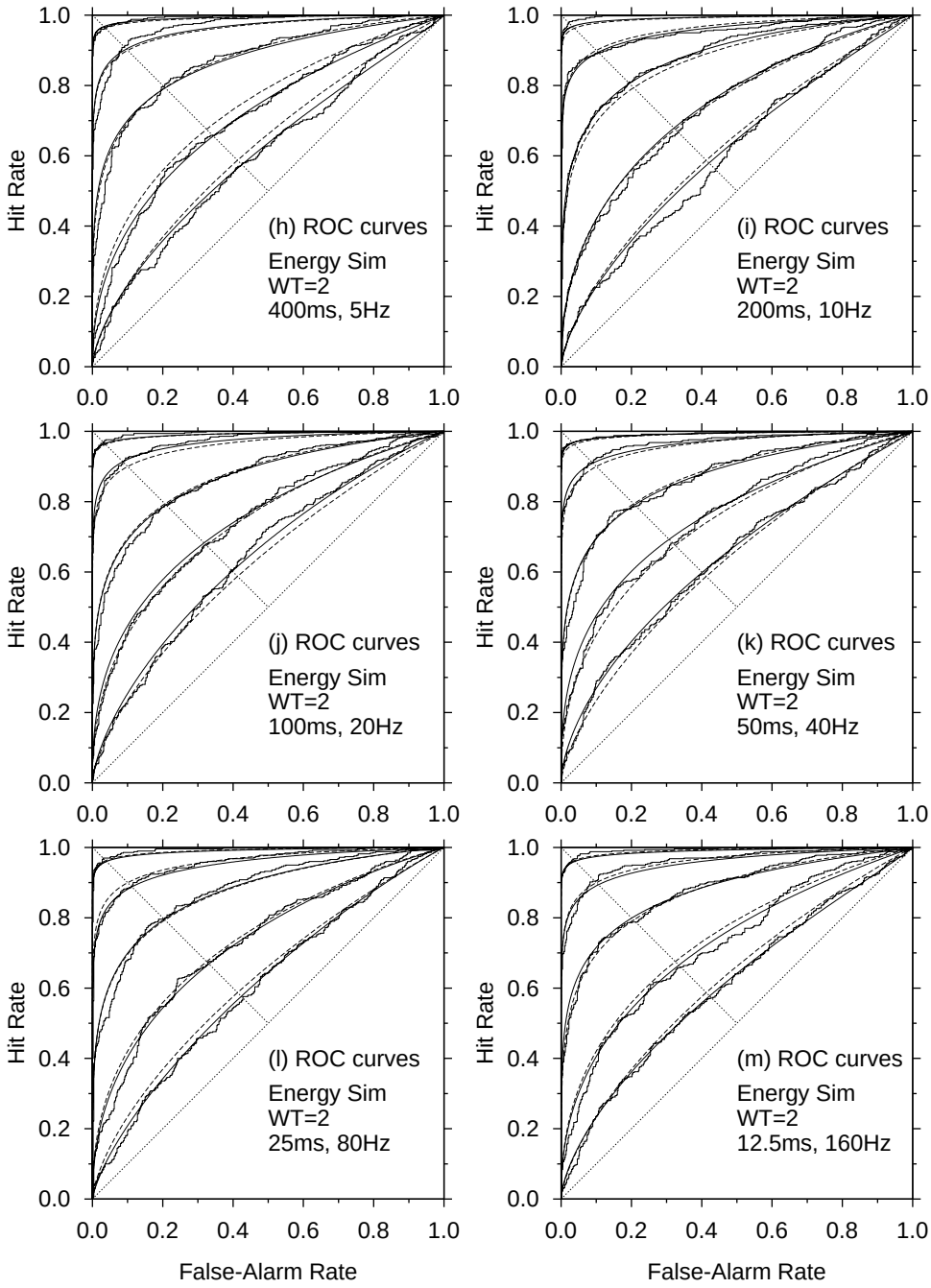

Figure 4.10: The simulated energy detector ROC curves for the signals used in condition $\mathcal{WT}=2$ (jagged solid curves), the corresponding fitted gamma ROC curves estimated from the simulation evidence (smooth solid curves), and the ideal χ^2 energy detector ROC curves (smooth dashed curves).

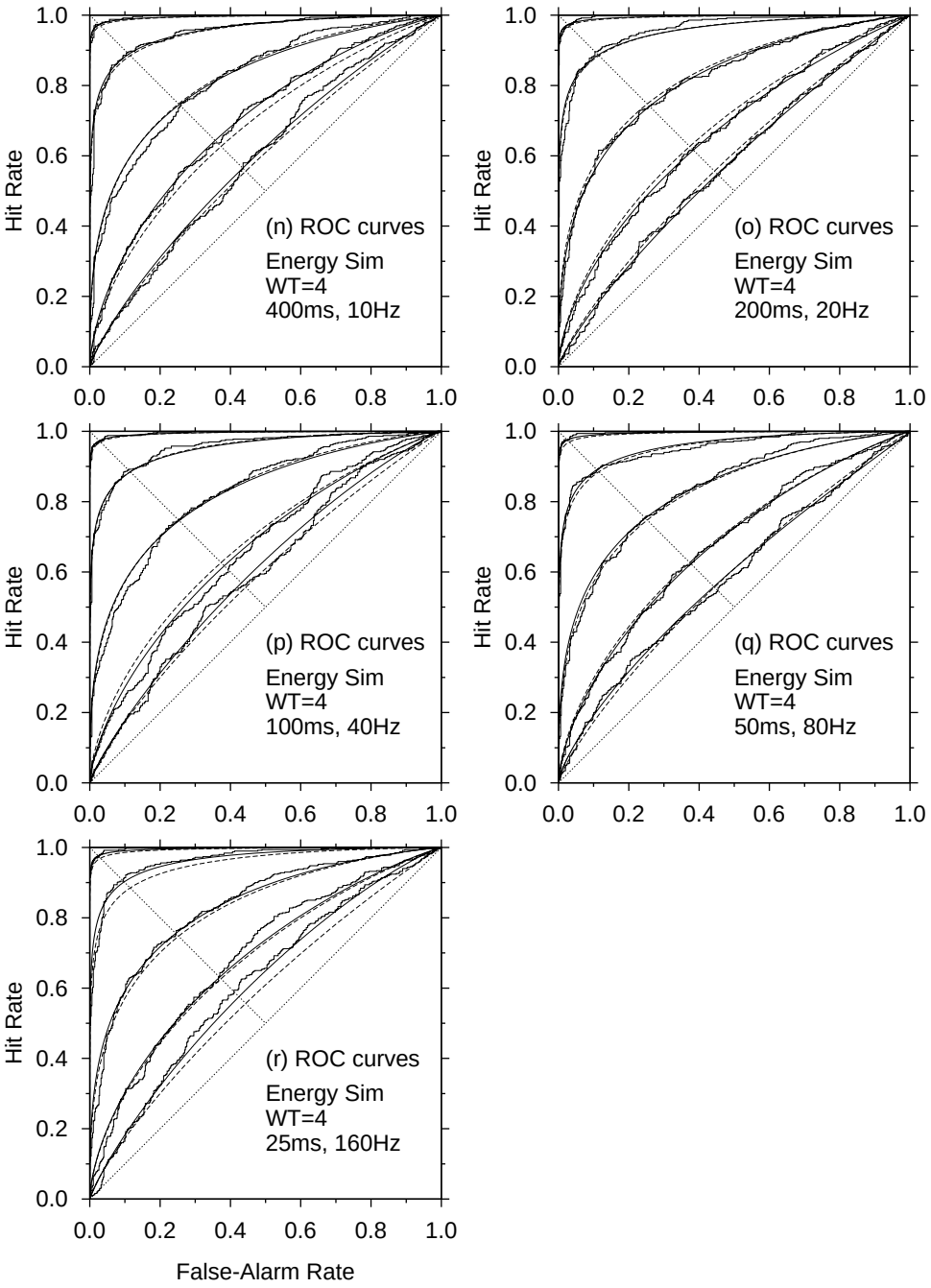

Figure 4.11: The simulated energy detector ROC curves for the signals used in condition $\mathcal{WT}$=4 (jagged solid curves), the corresponding fitted gamma ROC curves estimated from the simulation evidence (smooth solid curves), and the ideal χ^2 energy detector ROC curves (smooth dashed curves).

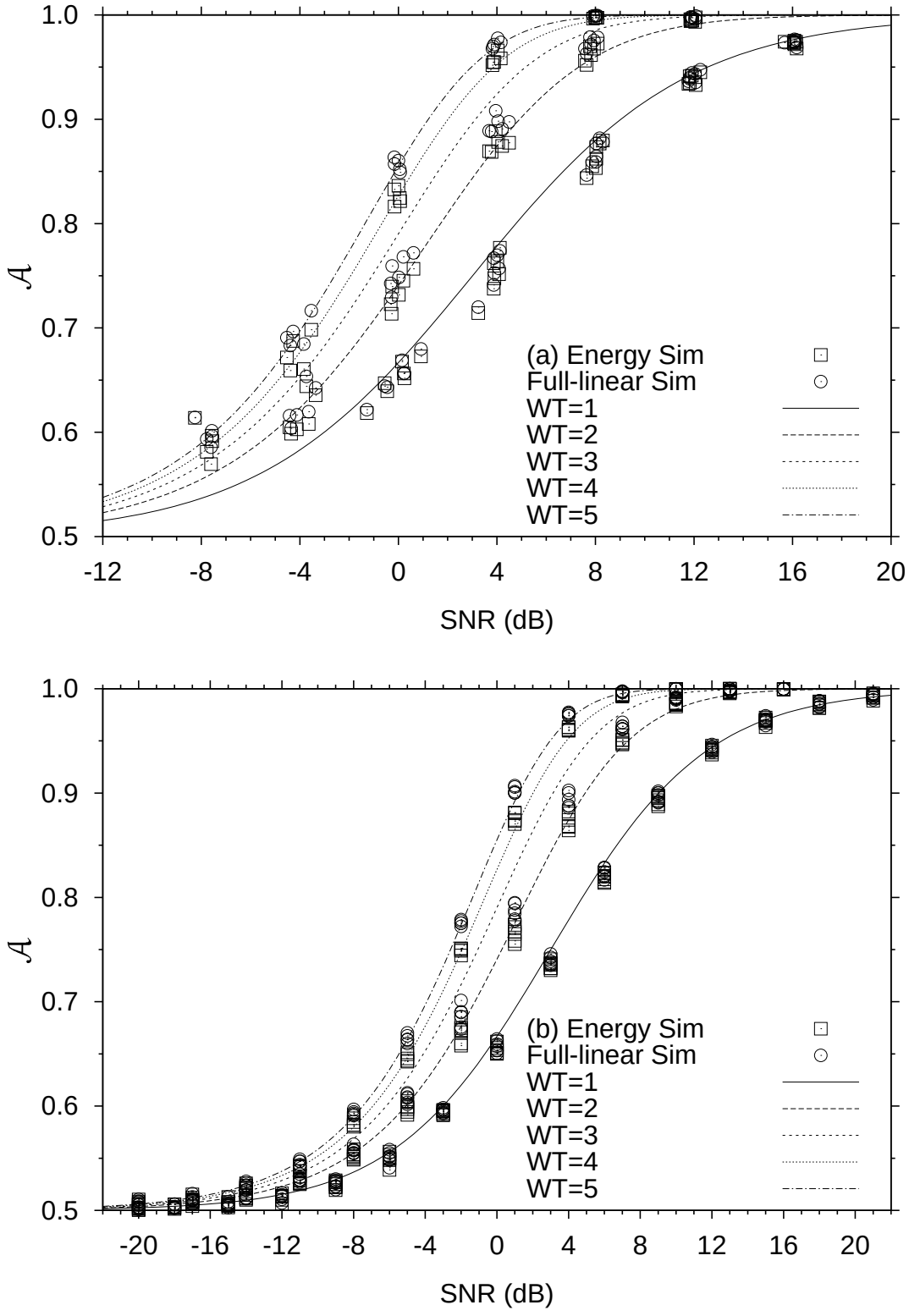

Figure 4.12: Comparison of the energy and full–linear detectors for the (a) the 18 buffer simulations and (b) the 18 transient simulations. Both are compared with the energy detector theoretical psychometric functions.

On closer inspection, the psychometric functions of the simulated energy detector, for the $\mathcal{WT}$=2 condition, fall between the $\mathcal{WT}$=2 and $\mathcal{WT}$=3 mathematical functions. Similarly, for the $\mathcal{WT}$=4 condition, the psychometric functions of the simulated energy detector, fall between the $\mathcal{WT}$=4 and $\mathcal{WT}$=5 mathematical functions. This phenomenon is supported by the findings in the previous sections that $\mathcal{WT}$ changes with signal–to–noise ratio. The $\mathcal{N}$ event contributes to every point in the psychometric function so the relationship of $\mathcal{WT}$ to the shape of the function is likely to be complicated. This is because each point is made up of two different $\mathcal{WT}$ values—one value is common across signal–to–noise ratio and the other varies.

The attenuation of data points may be because of the way signal–to–noise ratio is defined, but this does not change the shape of the psychometric function, merely the position.

4.5.2 Energy and full–linear detector comparisons

Some researchers suggest that the energy and envelope detectors should result in virtually the same performance for the same stimuli. The envelope detector, however, differs from the energy detector in two respects: the rectifier and integrator; and has an additional parameter: the time constant. To establish whether any difference between the energy and envelope detector was due to the rectifier or the integrator, a third detector was tested: the full–linear detector. This detector was the same as the energy detector, except it used a full–wave rectifier. According to Marcum (1960), the output of the two rectifiers should be the same when $\mathcal{WT}$=1 and $\mathcal{WT}$=70. In between these values, the linear detector gives a higher effective signal–to–noise ratio and for values greater than 70, the square–law detector is superior.

Figure 4.12 shows the psychometric functions for the energy and full–linear simulations (top graph for buffers and bottom graph for transients), and compares them with the mathematical (χ^2) psychometric function, for the energy detector, for $\mathcal{WT}$=1 to 5. The full–linear detector is generally superior to the simulated energy detector, especially for the $\mathcal{WT}$=2 and 4 waveforms. The full–linear detector is also superior than the mathematical energy detector for the $\mathcal{WT}$=2 and 4 waveforms, but is worse than the mathematical detector for $\mathcal{WT}$=1 (as is the simulated energy detector).

Comparisons using $\mathcal{D}_2$

To compare detectors, differences in $\mathcal{D}_2$ were computed and plotted as a function of signal–to–noise ratio. $\mathcal{D}_2$ was chosen as the measure of detectability for this analysis, because its properties are particularly suited to making comparisons.

Figure 4.13(a) shows the $\mathcal{D}_2$ psychometric functions for the energy and full–linear transient simulations, averaged across experimental level within an experimental condition (the fitted lines are a data model, introduced in the next section). This figure shows that the full–linear is a superior detector for these small–$\mathcal{WT}$ signals. Figure 4.13(b) plots the differences in $\mathcal{D}_2$, in bits, for each $\mathcal{WT}$. As $\mathcal{WT}$ increases, so does the difference in performance of the two detectors. For $\mathcal{WT}$=1, the difference is very tiny; presumably the waveforms have an effective $\mathcal{WT}$ that is just slightly greater than unity, assuming the two detectors are meant to perform identically at $\mathcal{WT}$=1 as Marcum (1960) would suggest. For larger $\mathcal{WT}$ the difference is greater, reaching a maximum of approximately 0.085 bits at 0 dB for $\mathcal{WT}$=4. The difference across signal–to–noise ratio is a result of the shape of the psychometric function. As the signal–to–noise ratio gets very small or very

large both functions asymptotically approach zero or one, so the functions become shallower and the difference becomes smaller.

4.5.3 Polynomial psychometric functions

The analyses in the previous section have shown a difference between the energy detector mathematical model and the simulations. There are no mathematical models for the full–linear detector, so polynomial psychometric functions were fitted to the transient simulations' results, to build data models for the full–linear and energy detector. Because performance in detecting the small–$\mathcal{WT}$ signals was similar across combinations, three models were fitted for each detector: one for each $\mathcal{WT}$.

The fitting procedure used was the Marquardt–Levenberg non-linear least–squares algorithm implemented in GNUPLOT (Williams & Kelley, 1997). The form and order of the fitted polynomial was based on the energy detector polynomial psychometric functions, in Equation (1.15), derived by Green and McGill (1970) and McGill and Teich (1991). The free parameters were the coefficients of the polynomial (the exponents were fixed) and a translation factor. The initial values were the ideal energy detector coefficients and a translation factor of 0.1.

The results of the transient simulations were used, because they had at least 12 points on the psychometric function. The $\mathcal{A}$'s for each signal–to–noise ratio were averaged across each level for a given $\mathcal{WT}$ condition. The fitting procedure weighted each point on the psychometric function by the sampling standard deviation (from Equation (2.7) with the correction for empirical data).

The psychometric function with the best fitting coefficients for the $\mathcal{WT}$=1 energy detector is:

$$\mathcal{A}(x) \quad = \quad \frac{a(S(x-b)+1)}{(S(x-b)+2)} \tag{4.16}$$

where $a = 1.00176$, $b = 0.49549$, x is the signal–to–noise ratio in decibels, and the function $S(x)$ converts decibels back into power (assuming the decibels are relative to unity power) as $S(y) = 10^{y/10}$. For $\mathcal{WT}$=2 the best fitting solution is:

$$\mathcal{A}(x) \quad = \quad \frac{a(S(x-b)+1)^5 + c(S(x-b)+1)^4 + d(S(x-b)+1)^3}{(S(x-b)+2)^5} \tag{4.17}$$

where $a = 1.00231$, $b = 1.65215$, $c = 4.92657$, and $d = 10.15526$, and for $\mathcal{WT}$=4 the best fitting solution is:

$$\mathcal{A}(x) \quad = \quad \frac{a(S(x-b)+1)^9 + c(S(x-b)+1)^8 + d(S(x-b)+1)^7 + e(S(x-b)+1)^6 + f(S(x-b)+1)^5}{(S(x-b)+2)^9} \tag{4.18}$$

where $a = 0.99357$, $b = 0.67053$, $c = 9.21547$, $d = 33.76075$, $e = 92.62537$, and $f = 119.77846$.

The best fitting solutions for the full–linear detector simulations, for each $\mathcal{WT}$, are the same as Equations (4.16), (4.17), and (4.18) for the energy detector but with different coefficients:

$\mathcal{WT}$=**1:** $a = 1.00208$ and $b = 0.37701$.

$\mathcal{WT}$=**2:** $a = 1.00065$, $b = 1.06690$, $c = 4.97944$, and $d = 10.07645$.

$\mathcal{WT}$=**4:** $a = 0.99330$, $b = 0.11746$, $c = 9.26275$, $d = 32.71755$, $e = 99.33391$, and $f = 113.77148$.

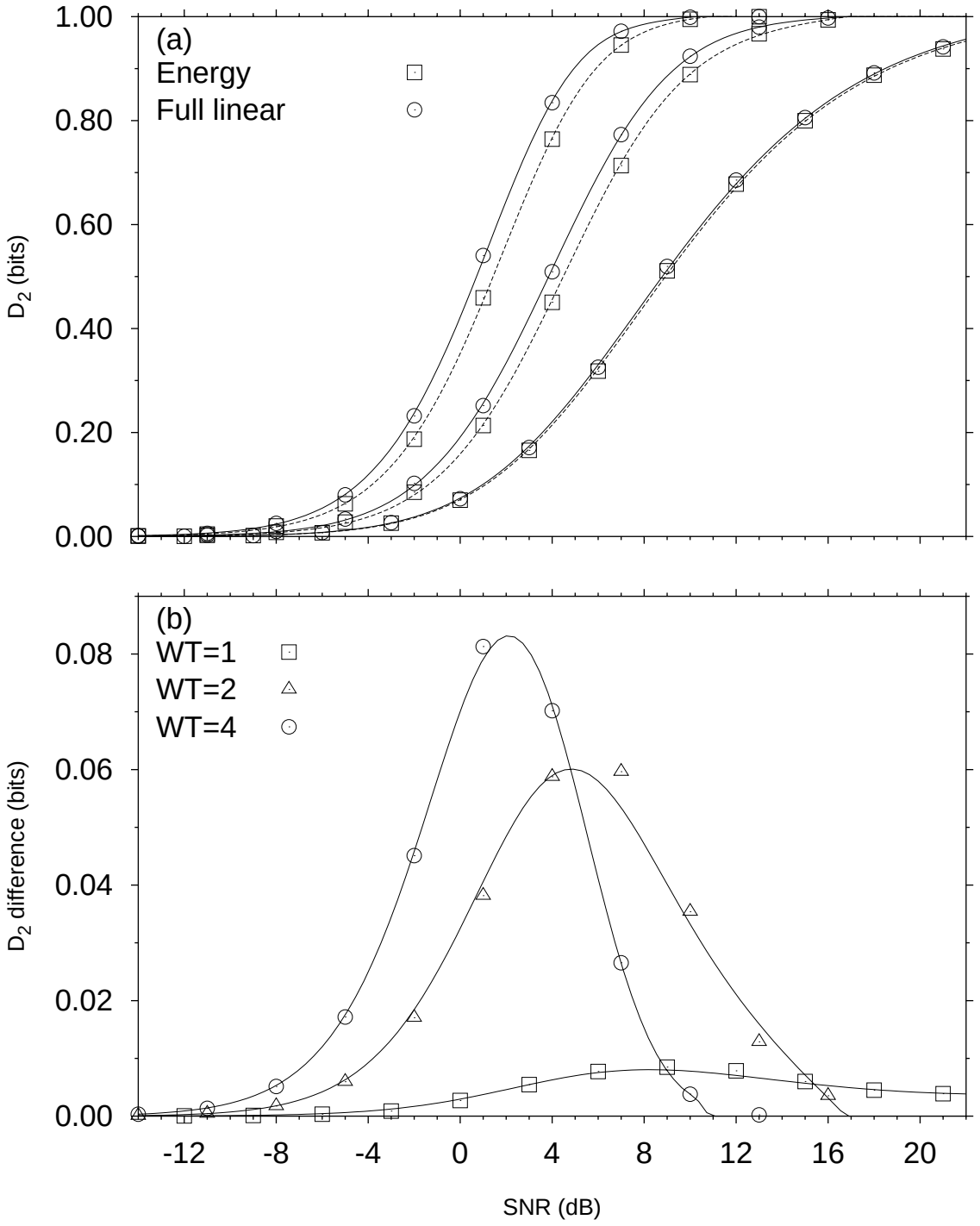

Figure 4.13: Comparison of the energy and full–linear detectors for the transient waveforms used in the experiments. (a) The fitted polynomial psychometric functions converted to $\mathcal{D}_2$ (lines) and the empirical data points for the simulations averaged over signal–to–noise ratio for each experimental level at each $\mathcal{WT}$, (b) the difference between the energy and full–linear fitted psychometric functions (lines) and the difference between the full–linear and energy detectors, averaged over each waveform set, for each signal–to–noise ratio and $\mathcal{WT}$ (points).

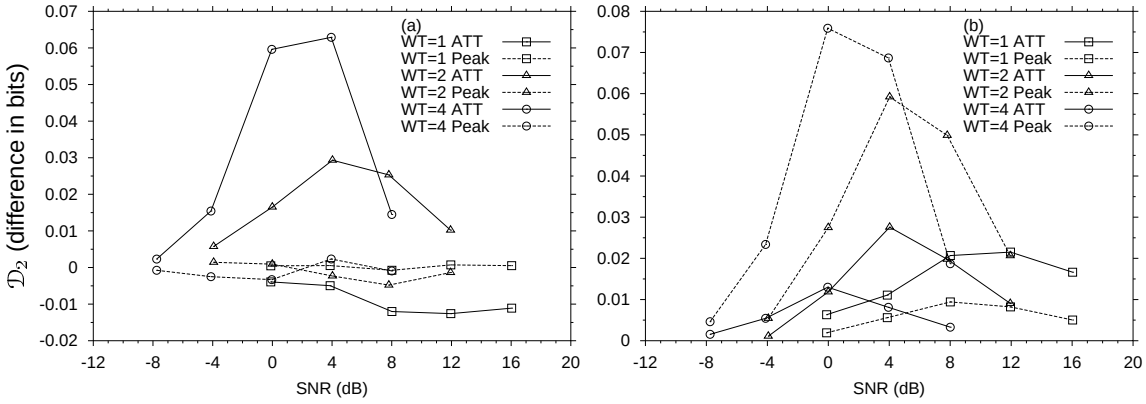

Figure 4.14: Comparison of the arithmetic difference of $\mathcal{D}_2$ between (a) the envelope and energy detectors, and (b) the full–linear and envelope detectors, for the buffer waveforms used in the experiments. Plotted is the difference in $\mathcal{D}_2$ (bits), averaged over each buffer waveform set for each (averaged) signal–to–noise ratio and $\mathcal{WT}$, for the two envelope sampling strategies: at this time (ATT) and peak.

The fitted energy and full–linear psychometric functions are plotted in Figure 4.13. They clearly show that the full–linear detector is superior for small $\mathcal{WT}$ Gaussian noise, for $\mathcal{WT} > 1$.

4.5.4 Envelope detector simulations

As described in Chapter 1, the ideal mathematical envelope detector has only been derived for tone–in–noise signals and for $\mathcal{WT}$=1 noise–in–noise signals. Therefore, the aims of the ideal envelope simulations were to:

- To establish whether performance is the same as for the energy detector or the full–linear detector, especially for higher–order $\mathcal{WT}$ where no mathematical theory exists.
- To compare sampling strategies: peak detection versus sampling at a specific time.

Comparison using $\mathcal{D}_2$

Buffer simulations of the envelope, energy, and full–linear detectors were compared. The envelope detector used an equal–bandwidth Butterworth filter, a half–wave rectifier, a leaky integrator with a time constant equal to the ERD of the input signals, and two sampling plans: peak detection and sampling at the time equal to the absolute duration of the waveform.

Comparison of the envelope detector with the energy and full–linear detectors revealed some surprises:

- The difference between the energy detector and the envelope detector with a peak sampler is virtually nonexistent. Previous suggestions have implied that the two would be equal if the envelope detector sampling was at a particular time, rather than at the peak. This may be partly because the shape of the waveform (which was windowed over the entire duration) forced the peak to occur, on average, at a particular time instead of any time over the duration of the waveform.

- The ATT envelope detector is better than the energy detector for $\mathcal{WT}$=2 and $\mathcal{WT}$=4, and slightly worse than the energy detector for $\mathcal{WT}$=1.
- Compared to the full–linear detector, the envelope detector (for both sampling plans) did not perform as well. As $\mathcal{WT}$ increased, the difference in performance became greater. The ATT envelope detector was, however, better than the peak envelope detector.

The reasons the full–linear detector performed better than the envelope detectors could be:

- Because the sampling time was not optimally chosen (this should be investigated further).
- Because the integrator time constant was not optimal (this should also be investigated further).

4.6 Comparisons with human data

Section 4.5 showed that the full–linear detector is generally better than the energy detector for the types of signals used here. Chapter 1 also pointed out other research that supported the idea that the linear rectifier is more appropriate for human detectors. Therefore, in lieu of a theoretical model for linear detection of noise signals, the fitted polynomial functions were compared to the human data from Chapter 3 by comparing psychometric functions and differences in the attenuation functions.

Like the theoretical psychometric functions, the fitted models can be used to compare performance by considering the shape and the attenuation of data. In particular, the attenuation may be estimated by solving the roots of the polynomial to obtain a function in terms of signal–to–noise ratio, calculated by numerical estimation using a bisection method.

4.6.1 Psychometric functions

Appendix F shows the human psychometric functions compared to the mathematical energy detector and the simulated full–linear model. The results for the comparison with the energy detector were reported in Section 3.4.5 and 3.4.6.

Figures F.2, F.4, F.6, F.8, F.10, and F.12, show the psychometric functions from the FORCE, GOC, and mROC analyses, calculated from the six replications completed by Observer 1 and Observer 2 for each experimental condition and level. Each figure shows:

1. a psychometric function fitted to the appropriate full–linear detector from Equations (4.16), (4.17), or (4.18).
2. the empirical psychometric points for each signal–to–noise ratio from the FORCE, GOC, and mROC analyses for each observer, and
3. the same full–linear psychometric function *attenuated* to fit the human psychometric functions. The attenuation is the linear average of attenuation in decibels for each five signal–to–noise ratios relative to the model being tested. For the FORCE analysis, a point was only included in the attenuation calculation if the point had converged sensibly.

The functions are plotted on the opposite page to the energy detector analysis to allow comparison.

Results are very similar to the mathematical energy detector analysis except for the degree of attenuation. This is partly because the full–linear model was fitted with an attenuation parameter. The goodness–of–fit between the human data and psychometric functions for $\mathcal{WT}$=1 is essentially the same for each type of detector, because the models have virtually the same shape. For $\mathcal{WT}$=2 and $\mathcal{WT}$=4, the human data appear to be slightly better fitted by the full–linear detector.

4.6.2 Attenuation analysis

The attenuation analysis summarises the information from the psychometric functions by plotting the average linear attenuation of the data from the model as a function of bandwidth and as a function of duration. A signal level only contributed to the attenuation of the FORCE function if it converged successfully. Rogue data points from the FORCE analysis for Observer 1 in levels {400ms, 5Hz: -4,12dB}, {200ms, 10Hz: -4dB}, and {400ms, 10Hz}, were dropped, because they biased the attenuations. This is not unreasonable as it was obvious that these levels converged with difficulty. Attenuations were calculated, using a bisection method, from the fitted polynomials from Equations (4.16), (4.17), or (4.18).

Figure 3.16 (on page 103) shows the *average attenuation* in decibels from two types of psychometric function: the mathematical energy detector (dashed lines and open symbols) and the data model for the full–linear detector (solid lines and solid symbols). For the two models, attenuation functions were of similar shape but the actual attenuations were different. It is not surprising that the shape is similar, because the differences between the shape of the psychometric functions for the energy and full–linear detector is subtle. The attenuation from the full–linear detector compared to the energy detector was less for $\mathcal{WT}$=1 and $\mathcal{WT}$=2 but greater for $\mathcal{WT}$=4.

An obvious question is whether the attenuation functions from the different conditions give the same pattern when matched for bandwidth or when they are matched for duration (or neither). If they match for bandwidth, then that would indicate that the ear was adjusting its detection bandwidth and/or its integration time based on the bandwidth of the signal. If duration is matched instead, then the ear was adjusting its parameters based only on the duration of the signal. Figure 4.15 shows, for each observer, the three attenuation functions (after GOC analysis) matched for duration then for bandwidth. They are matched for GOC and not FORCE analysis because of Observer 1's rogue points, which changed the *shape* of the function.

Observer 1's attenuation functions are more similar when they are matched for duration than for bandwidth, but Observer 2's attenuation functions match better for bandwidth. In neither case is the relationship between bandwidth and duration–matching clear cut. It looks like both bandwidth and duration information is used by the hearing system but that bandwidth and duration are being traded off in a suboptimal way.

Because the attenuation functions of the human performance are not constant with experimental level, the human hearing system cannot be trading–off $\mathcal{W}$ and $\mathcal{T}$ optimally. The next chapter considers more realistic models of human performance by taking the energy, full–linear, and envelope detectors and systematically manipulating their parameters until they start performing like humans.

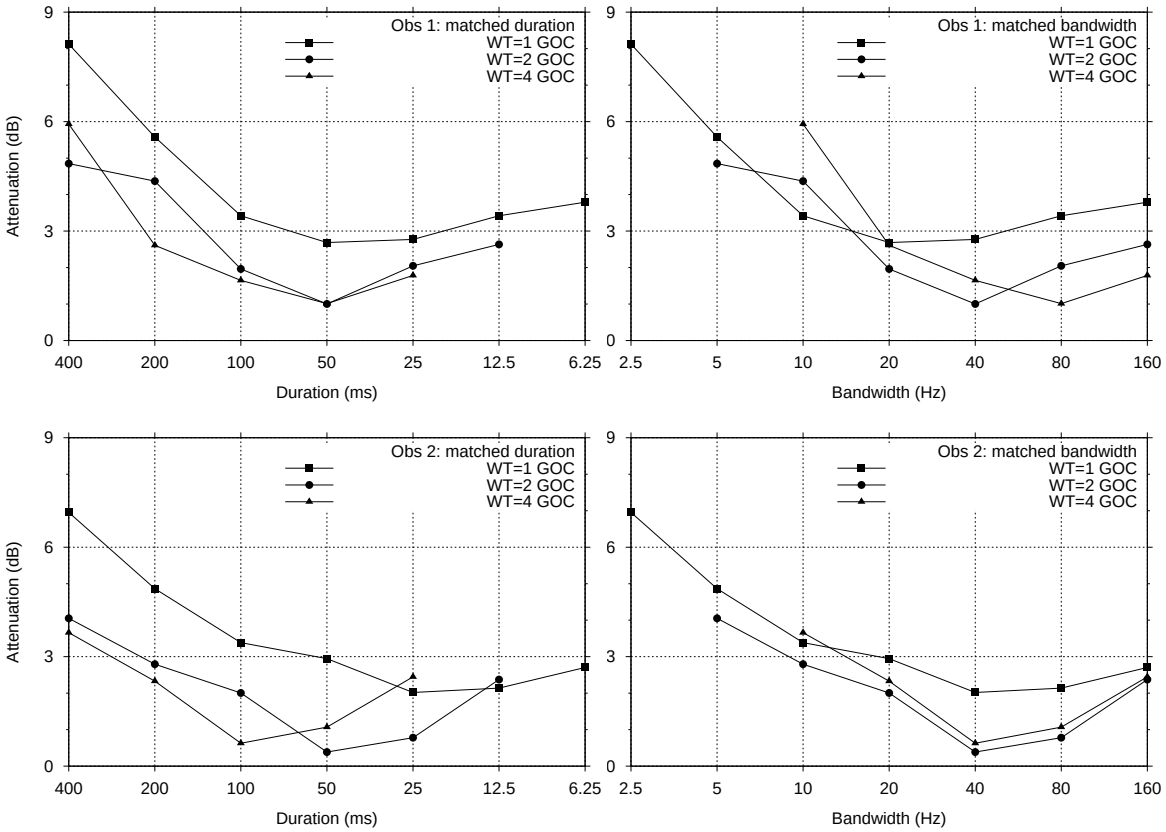

Figure 4.15: Attenuation functions, relative to the full–linear detector, after GOC analyses, for each observer in all conditions. Left functions are matched for duration; right functions are matched for bandwidth.

Chapter 5

Correlations between humans and simulations

"How little you mortals understand time.
Must you be so linear, Jean–Luc?"

Q (Star Trek: The Next Generation).

.....

Stimulus–level, or molecular, analysis has been successfully employed by both Ahumada (1967) and Gilkey (1981) to study the psychophysics of human hearing. Their general method was to measure the association between human observers' ratings for each stimulus and the output from simulated electronic and computational detectors for the *same* stimulus, to estimate properties of the human hearing system (see Section 1.4.1 of Chapter 1).

Like GOC analysis (see Chapter 2), their approach reduces the effects of observer inconsistency in decisions (ratings). Experiments are replicated multiple times, with the same stimuli, then ratings are averaged to remove error. The averaged ratings are then correlated with the output of a simulated detector, which has been presented the same waveforms.[1]

The ratings from the human observers can be viewed as implying an ordering, or sorting, of the stimulus set. On each replication, the stimuli are sorted into a particular order, using the rating scale. The order is slightly different on each replication due to unique noise. If the ratings are averaged for each stimulus, then the unique noise is removed, and as more replications are added, the order of the stimuli tends to a fixed order. The evidence output from a computational detector also implies a sorting of the stimuli. The correlation between the two observers is then a measure of the extent the two observers sorted the stimuli into the same order. Thus, like ROC analysis, only an ordinal scale is required for the analysis.

By manipulating the properties of the detector, the pattern of correlation, as a function of the manipulated property, can indicate the nature of an underlying process.

[1]Gilkey's method required converting the sum–of–ratings and evidence values to hit rates and false–alarm rates first. This step is unnecessary.

5.1 The Bester correlation method

The *Bester correlation method*[2] was developed from the methods of Ahumada (1967) and Gilkey (1981), and was used to evaluate the hearing systems of the human observers from Chapter 3:

- To evaluate if the square–law or the linear (full or half–wave) rectifier was a better model for human hearing.
- To estimate the detection bandwidth for each observer, and each experimental level, by varying the bandwidth of the simulation's filter.
- To establish the form and parameters of the temporal integrator and sampler.
- To determine the detector that best described the human observers' performances in all conditions and levels, and compare its performance directly to human psychometric functions.

The reliability of the Bester correlation method was also tested by analysing data sets from another experiment, where the number of replications was larger than in the current experiment.

The Bester correlation method can be summarised as follows:

- Energy, full–linear, and envelope detector simulations were run, using SIM IO (Chapter 4; Lapsley Miller, 1998c), where either a particular parameter was systematically varied (e.g., the filter bandwidth), while holding all other parameters and processes constant, or different forms of a process (e.g., square–law versus linear rectification) were compared.
- The simulated detectors were presented the same digital waveforms that were used in the human experiments.
- The evidence output from SIM IO's sampler was then correlated with the sum–of–ratings from the human experiments, using Spearman's rank–order correlation coefficient (using the program BESTER (Lapsley Miller, 1998b)).
- The correlation was plotted, as a function of the varied parameter or process, and the maximum correlation point or maximum correlated process was determined.
- If a parameter was varied, the value of the parameter associated with the maxima was used to estimate the parameter of the underlying process. If processes were compared, then the process resulting in the highest correlation was assumed to indicate the best model (out of those tested) for the underlying human hearing process.

The Bester correlation method measures to what extent two observers (humans and / or computer simulations) have sorted the stimuli into the same order. The closer the two orderings match, the stronger the inferences that may be drawn from the data. The actual ratings or evidence values are irrelevant, only the rank–order is important. This is why Spearman's rank–order correlation coefficient was used, rather than Pearson's product–moment correlation coefficient, because it measured the correlation between the orderings. Kendall's rank–order correlation coefficient was also considered, but gave very similar results to Spearman's coefficient. Because Spearman's is better known than Kendall's coefficient, it was chosen in preference, although either would be appropriate.

[2] *Bester* stands for *best–estimated–ranking*.

5.1.1 Estimating the form of the rectifier

The results from Section 4.5.2 indicated that there was a difference between the energy and full–linear detector for $\mathcal{WT} > 1$, and that this difference was due to the rectifier (square–law or linear). The envelope detector was also found to be better than the energy detector for $\mathcal{WT} > 1$. If the main difference between these models is the rectifier, then the first analysis should be to see which form of rectification is more likely in the human hearing system. This may be done by comparing the correlations for the energy and full–linear detectors, because all their other parameters and processes are equal. If there is a difference in correlation then it can only be due to the rectifier. Section 5.3 reports this analysis, in conjunction with estimating the detection bandwidths for the human observers in the current experiment.

5.1.2 Estimating critical and detection bandwidths

The simulated detector's filter is meant to represent the critical, or detection, bandwidth of the auditory filter. If simulations are run with the same signals, and the bandwidth of the simulation's filter is systematically varied, then it is possible to estimate the detection bandwidths of the human observers. This is achieved by calculating the correlation between the simulated detector and the human observer's ratings for the same stimuli, and plotting the resulting correlations as a function of the simulation's filter bandwidth.

The correlation–bandwidth function shows the relationship between the bandwidth of the simulation's filter (the independent variable) and the correlation between the human's and simulation's stimulus orderings (the dependent variable). The shape of this function differs depending on the signal–to–noise ratio of the waveforms. Large signal–to–noise ratios produce the same, high, correlation for many bandwidths. This is because the large signal swamps the noise, such that the ordering of the stimuli is similar, regardless of the amount of masker passed by the filter (with unequal bandwidth). The $\mathcal{N}$ event, on the other hand, tends to result in a correlation–bandwidth function with a single peak. This phenomenon is used to measure the detection bandwidth most correlated with human performance, because the estimate is not confounded with the signal bandwidth. Because there is no signal to match bandwidths in the $\mathcal{N}$ event trials, the estimated bandwidth should reflect that of the detection bandwidth. This approach is also justified from the signal analysis of the simulations in Chapter 4, where it was shown that the signal bandwidth changed after filtering as a function of signal–to–noise ratio. Only the bandwidth of the $\mathcal{N}$ signals indicated the actual bandwidth of the filter, despite the same filter being used for all the signals. Section 5.3 reports the estimated detection bandwidth for the human observers in the current experiment, using the energy and full–linear detectors.

"Detection" or "critical" bandwidth?

Bernstein and Raab (1990) suggested that the term *detection bandwidth* should be used instead of *critical bandwidth*, and that the latter should be reserved for the smallest estimate of the detection bandwidth. This is because detection bandwidths are estimates of the effective bandwidth of the auditory system, and may arise from a higher–order process that combines information across critical bands. It is premature to say whether the bandwidth estimates from the Bester correlation method are estimates of the critical band, therefore, the more neutral term, detection bandwidth, is used at this stage.

5.1.3 Estimating the form and parameters of the integrator

Once the detection bandwidth has been estimated, and the form of the rectifier determined, the different forms of integration may be tested with little confounding from the other processes. The integrators that were evaluated included the true integrator and the leaky integrator with various time constants. This analysis was done in conjunction with testing the best sampling time, because the two processes are interrelated.

5.1.4 Estimating the form and parameters of the sampler

To estimate how the human observers sampled the output of the integrator, two sampling strategies were simulated, and the correlations resulting from the Bester correlation method analysis were compared. Sampling at a particular time was investigated by sampling at approximately 30 different times, then calculating the correlation with the human observers. The results were plotted as correlation–duration functions, where the duration represented the sampling time, or duration from the beginning of the integrator output. There was a separate correlation–duration function for every time constant and integrator. By comparing these functions, it was possible to estimate the best integrator time constant and sampling time combination.

The other sampling strategy tested was a peak detector. The strategy was to use the peak of the output of the integrator for each individual signal, as the sample or evidence. This peak could occur at any time, but tended towards the same time, because of the Kaiser window. Thus, the difference between sampling at a particular time, and peak sampling, was not necessarily that great. This strategy produces only one correlation coefficient, rather than a function.

5.2 Analysis of the $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{\mathrm{2IFC}}$ project with the Bester correlation method

The current experimental data are based on only six replications of the experiment. If there is observer inconsistency in the data, then six replications of data may still be too variable to draw conclusions from the correlation analysis. One way of assessing the effect of this variability is by considering data from a different experiment where more than six replications were run. Kindly the other authors of the "$\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{\mathrm{2IFC}}$" project (Lapsley Miller et al., 1999) have let me use the data sets and signals to help assess how the estimates from the Bester correlation method vary when individual replications, or groups of six are analysed.

The experiment was an SIFC amplitude discrimination experiment. The observer's task was to indicate their confidence that a $\mathcal{SN}$ event occurred by making a rating on a continuous rating scale. The ratings were subsequently binned into a 64 point rating scale for analysis.

The stimuli were narrow–band Gaussian noise transients with a duration of 8.2 ms (ERD), windowed with a Kaiser window ($\kappa = 9$), and a bandwidth of 92 Hz (ERB), centred at 250 Hz, and masked by wider band Gaussian noise (1.5 kHz) with a signal–to–noise ratio of 7.5 dB. These transients had an essential $\mathcal{WT}$ of approximately unity, and an equivalent $\mathcal{WT}$ of approximately 0.75. The background masker (4 kHz, low–pass) was continuous. Three observers ran 32 replications and one observer ran 48 replications[3] of the experiment with 500 stimuli per event. Each

[3] When comparing across observers, only the first 32 replications were used unless otherwise stated.

observer used a different set of stimuli. Observers 1 and 3 in the $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{\mathrm{2IFC}}$ project are also Observer 1 and Observer 3 in this project.

The data analysed here are the sum–of–ratings for each observer, for various combinations of replications: single replications, consecutive groups of six replications, and the total number of replications. The Bester correlation method was used to calculate the correlation between the human observers' ratings and the evidence from a simulated energy detector, presented with the same digital signals.

Method

Energy detector simulations were run using the same digital signals that were used in the experiments, where the filter bandwidth was systematically varied.[4] The aim was to find the maximum of the correlation–bandwidth function for the $\mathcal{N}$ event, for sum–of–ratings over all the replications, by iteratively varying the filter bandwidth until the maximum was found. The bandwidth associated with the maximum correlation was used as an estimate of the detection bandwidth (resolution was $\pm$ 10–20 Hz).

Results

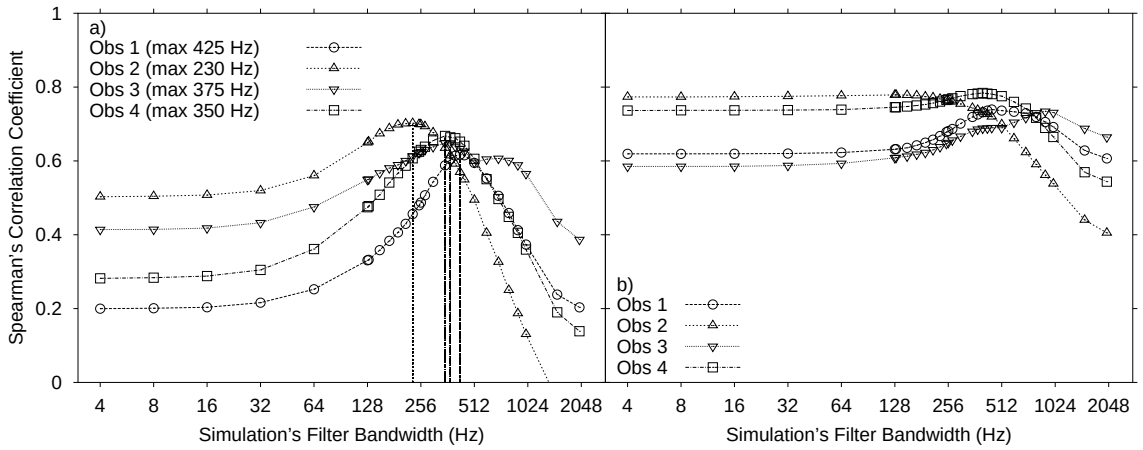

Figure 5.1: Correlation–bandwidth functions based on the sum–of–ratings from 32 replications for each observer. (a) $\mathcal{N}$ event (b) $\mathcal{SN}$ event. The vertical bars indicate the bandwidth corresponding to the maximum correlation for the $\mathcal{N}$ event.

Figure 5.1 shows the correlation–bandwidth functions for each observer. Also shown in Figure 5.1(a), with vertical bars, are the filter bandwidths of the energy detector best correlated with the human sum–of–ratings for the $\mathcal{N}$ event. Observer 2 had the narrowest filter so it is no surprise that his performance in the ROC space was greater than the other observers. Observer 3's correlation–bandwidth plot shows two maxima; indicating that perhaps he changed strategy, or perhaps treated larger signals differently to smaller signals.

To investigate this phenomenon further, correlations were calculated over sets of six consecutive replications (six replications were chosen for comparison with the current experiment). The

[4]These analyses were done before it was found that the full–linear detector better described human performances. For $\mathcal{WT}$=1, however, the difference is very slight.

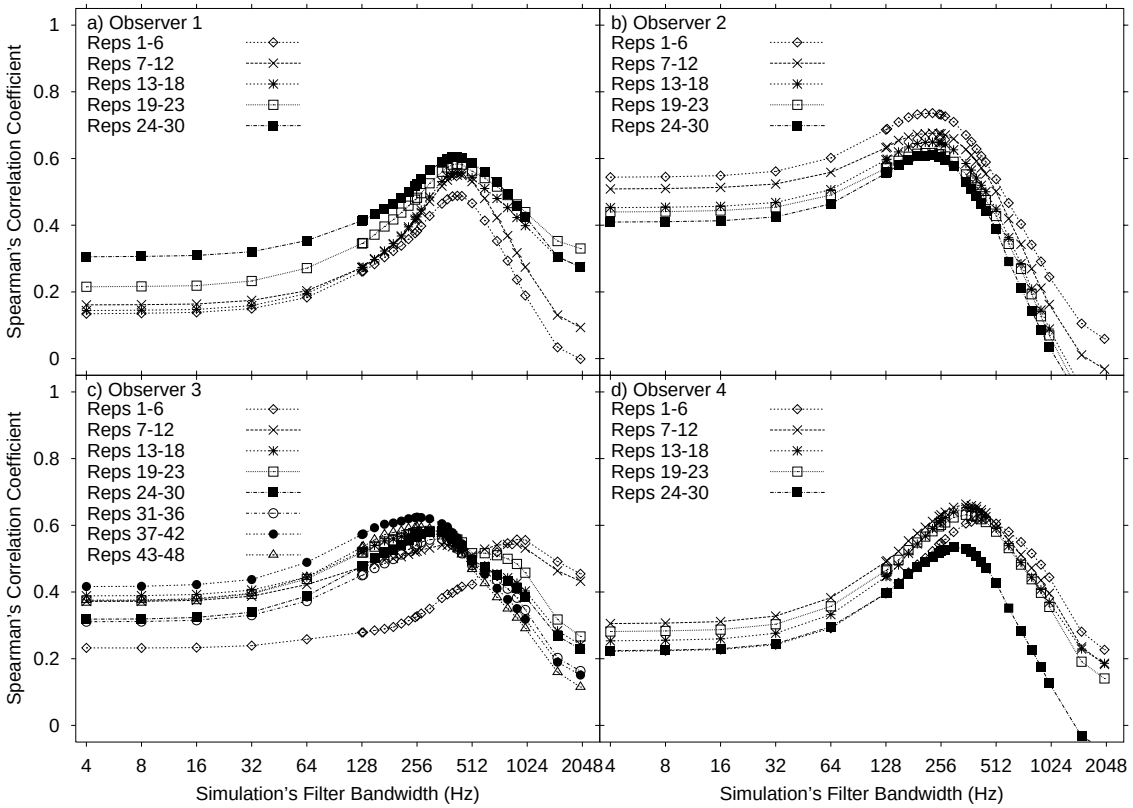

Figure 5.2: Correlation–bandwidth functions for sets of six consecutive replications for each observer.

results, shown in the top four graphs in Figure 5.2 indicate that the estimated detection bandwidths of Observers 1, 2, and 4 were relatively stable over time but Observer 3's bandwidth changed dramatically.

Correlations were calculated over individual replications, to further assess variability of the $\mathcal{N}$ event. The bandwidth associated with the maximum of the correlation–bandwidth function, for each replication, is illustrated in Figure 5.3 for each observer. It appears that the change in Observer 3's bandwidth occurred in separate replications, not within a replication. This indicated a change in the detection bandwidth with time, rather than the adoption of a different bandwidth for different types of signals. It should be noted that no additional simulations were run to isolate each maximum, so the resolution of the detection bandwidth estimates, for estimates greater than 1 kHz, are only accurate to 500 Hz. Observer 2 was very stable over time. Observer 4 stabilised over the first few replications. Observer 1 was stable except for three replications in the middle. Observer 3, however, was very unstable with widely varying bandwidths over the first 24 replications. The second 24 replications were far more consistent.

Reliability of the Bester correlation method

Variability of the estimated detection bandwidth across replications indicates that some unique noise in amplitude discrimination experiments comes from observers not holding a constant detection bandwidth. What effect does this have on analysis using the sum–of–ratings data? Using sum–of–ratings data does not appear to average the maximum correlated bandwidths from the

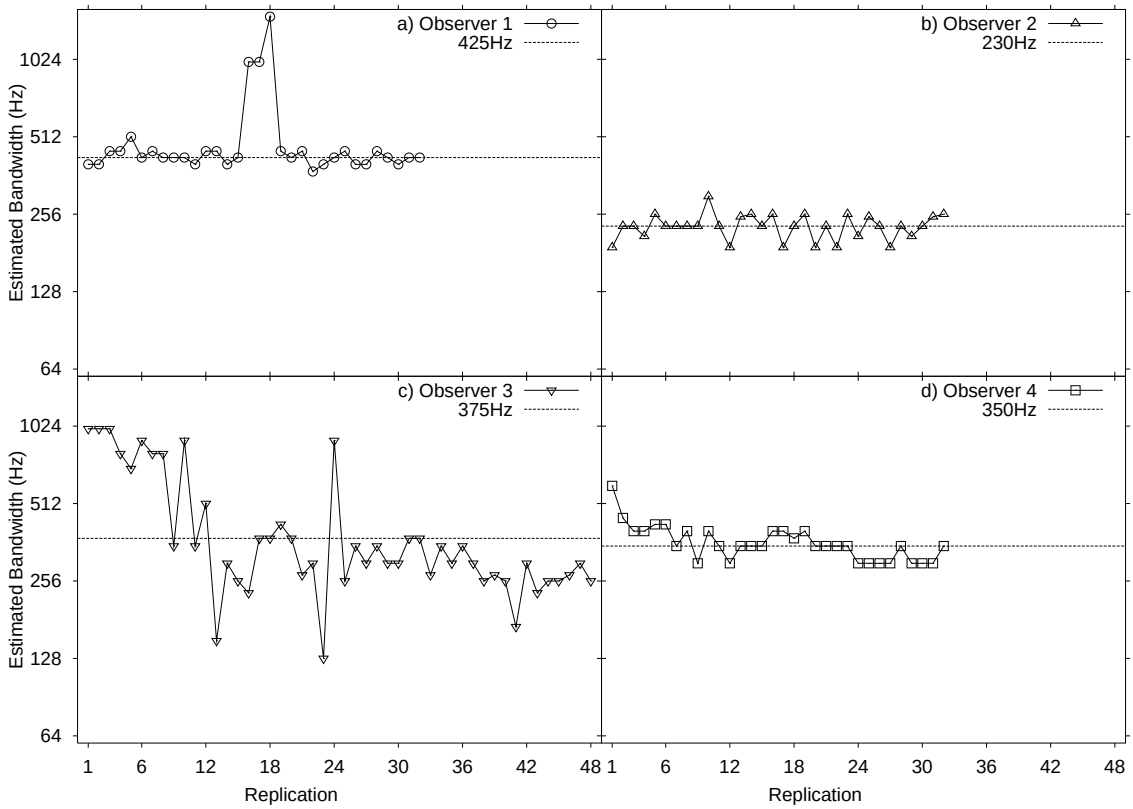

Figure 5.3: Detection bandwidths for individual replications, plotted in order of completion, for each observer. The horizontal line indicates the estimated detection bandwidth calculated over all replications. (a) Observer 1, (b) Observer 2, (c) Observer 3, and (d) Observer 4.

individual replications. Instead, the correlation–bandwidth functions relationship may have multiple maxima. In general, however, the analyses also indicate that the detection bandwidths estimated from the sum–of–ratings data may not be that different, on average, from single replication estimates.

In the current experiment, the methodology was different from the $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{\mathrm{2IFC}}$ project in that replications were not completed in one session (or consecutive sessions), but were spread out over weeks in a randomised session design. Thus, analysis on single replications may not show whether the observer is using different detection bandwidths across experimental sessions.

In sum, the only effects on correlations when using six replications are (a) for the correlation coefficients to be smaller than for larger replication sizes, and (b) for the estimates to be more variable. The correlation–bandwidth function, however, tends to be similar. Further analysis, however, may be indicated if the correlation–bandwidth function has multiple maxima.

5.3 Energy and full–linear detectors

One of the simplest models for detection of small–$\mathcal{WT}$ Gaussian noise is the energy detector, because it has only one free parameter—the bandwidth of the filter. Similar to the energy detector is the full–linear detector, which uses a linear instead of square–law rectifier. It was shown in Chapter 4 that the ideal full–linear detector was better than the ideal energy detector at detecting the signals used in the experiments (Chapter 3).

Preliminary analyses indicated that for simulated observers, one of the biggest contributors to the ordering of the stimuli was the form of the rectifier, rather than the form of the integration or sampling. Because the envelope detector differs from the energy detector, in both the form of the rectifier and the integrator, the effect of the rectifier cannot be isolated. By using a full–linear detector, however, the effect of the rectifier can be isolated without confounding from the integrator. The best correlated detection bandwidth can then be used as a constant, so the temporal characteristics of the envelope detector may be manipulated more easily.

The aims of the correlation analysis with the energy and full–linear detectors, were to (a) establish the rectifier best correlated with the human sum–of–ratings, (b) use the detector associated with the best correlated rectifier to estimate the detection bandwidths for the human observers, in each experimental level, and (c) use the detector with the best correlated bandwidth and rectifier to estimate the best sampling time, assuming a true integrator (leaky integration is evaluated in the next section).

5.3.1 Method

Energy detector simulations, similar to those from Section 4.2 in Chapter 4, were run repeatedly on all the buffer signal sets, with each repeat using a different 3 dB bandwidth for the Butterworth filter. All other parameters and processes were held constant. Because the filter bandwidth was not equal to the $\mathcal{S}$ signal bandwidth, the simulated observers were not optimal observers.

Simulations were run in sets (of various sizes), where the filter bandwidth was systematically varied. After each set the correlation–bandwidth function was plotted, and the maximum found. The resolution of the maximum was increased by running further sets of simulations, where the filter bandwidth was varied over a smaller range, encompassing the maximum. Thus, sets of simulations were run iteratively until the maximum was estimated to an accuracy of about 1 Hz. The number of simulations run, and the bandwidths used, depended on where, and how quickly, the maximum was found. This analysis was done individually for each observer so there were, in fact, three maxima to be estimated for each buffer set.

The full–linear detector simulations were run repeatedly on all the buffer signal sets, with each repeat using a different 3 dB bandwidth for the Butterworth filter. Bandwidths were selected from 10 Hz to 4096 Hz in 50 equal steps on an octave scale. All other parameters of the detector were held constant (full–wave rectification, true integration) and the integrator was sampled at the end of the output. Thus, the resolution of the bandwidth estimate was not quite as high as for the energy detector, but the process was faster because it was automated.

Once the best correlated bandwidth was found, the simulation for that bandwidth was repeated, but this time the output of the integrator was sampled at thirty different points spaced equally between 10 ms, and then end of the integrator (on a linear scale) to estimate the best correlated sampling time.

5.3.2 Results

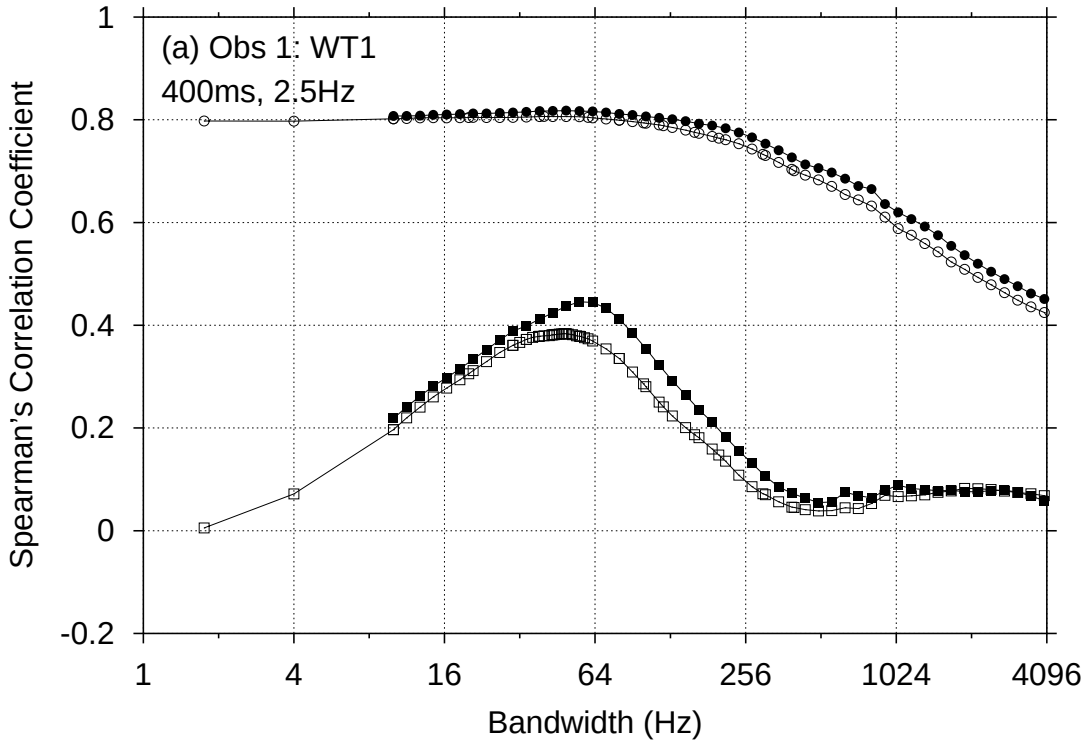

Figure 5.4: Correlation–bandwidth functions of Observer 1 based on the energy detector simulation (open symbols), and the full–linear detector simulation (closed symbols), in condition $\mathcal{WT}$=1, level {400ms, 2.5Hz}, for the $\mathcal{N}$ event (square symbols) and largest $\mathcal{SN}$ event (circle symbols).

The correlation was calculated between the sampled output for each stimulus, from each simulation, and the human observer sum–of–ratings for the same stimulus. Figures G.1–G.6 in Appendix G show the correlations coefficients plotted as a function of filter bandwidth for the $\mathcal{N}$ and largest $\mathcal{SN}$ event for when the integrator was sampled at the end of output, for both the energy and full–linear simulations. An example is shown in Figure 5.4.

The rationale for estimating the detection bandwidth from the correlation–bandwidth function was described in Section 5.1.2. As expected, the shape of the correlation–bandwidth function differed depending on the signal–to–noise ratio of the waveforms. Large signal–to–noise ratios produced the same, high, correlation for many bandwidths. The $\mathcal{N}$ event, on the other hand, tended to result in a correlation–bandwidth function with a single peak. Except for very wide band signals the final estimate of the detection bandwidth is within 1 Hz of the actual peak for the energy detector.[5]

The correlation–bandwidth functions for the full–linear detector tended to be a similar shape, but the correlation coefficients were consistently higher than for the energy detector, indicating that the full–linear detector described the human's detectability better than the energy detector. The exception was for Observer 1 in the experiment level {6.25ms, 160Hz}, where the difference

[5]The small blip at 1000 Hz in the correlation–bandwidth function is due to the changeover from a band–pass to a low–pass filter.

between the energy and full–linear detectors was negligible. The estimated detection bandwidths from the full–linear simulations tended to slightly larger than for the energy detector simulations.

The correlation–duration function, where each point is a different sampling time of the output of the integrator, showed an increase in correlation as a function of duration, reaching a plateau just before the point where the integrator was full (see Figure 5.10 on p. 163). This time was, coincidently, about the same time as the absolute duration of the waveform.

Table 5.1: Comparison of the estimated TdB and Ess$\mathcal{W}_{92.4\%}$ detection bandwidth (Hz) for each observer, for each experimental condition and level. The detection bandwidth was estimated from the full–linear simulations.

Experimental	Detection Bandwidth (TdB)			Detection Bandwidth (Ess$\mathcal{W}_{92.4\%}$)		
Level	Observer 1	Observer 2	Observer 3	Observer 1	Observer 2	Observer 3
400ms, 2.5Hz	55	49	—	58	51	—
200ms, 5Hz	63	43	—	66	45	—
100ms, 10Hz	71	55	—	74	58	—
50ms, 20Hz	71	71	—	74	74	—
25ms, 40Hz	147	90	188	154	95	197
12.5ms, 80Hz	722	102	—	755	107	—
6.25ms, 160Hz	566	450	650	591	471	677
400ms, 5Hz	55	43	55	58	45	58
200ms, 10Hz	71	49	—	74	51	—
100ms, 20Hz	90	49	—	95	51	—
50ms, 40Hz	115	55	115	121	58	121
25ms, 80Hz	240	80	—	251	84	—
12.5ms, 160Hz	347	102	566	363	107	591
400ms, 10Hz	55	49	63	58	51	66
200ms, 20Hz	71	49	—	74	51	—
100ms, 40Hz	90	63	80	95	66	84
50ms, 80Hz	147	71	—	154	74	—
25ms, 160Hz	213	102	639	222	107	668

Plotting the detection bandwidth as a function of original signal bandwidth and duration indicates that both the bandwidth and the duration of the original signal plays a large role in setting the bandwidth of the auditory filter (see Figure 5.5). It looks as though the detection bandwidth is flexible, and that it *does not* match the bandwidth of the signal. The detection bandwidth also appears to have a minimum of around 55 Hz (TdB), or 65 Hz (Ess$\mathcal{W}_{92.4\%}$), for Observer 1, and 43 Hz (TdB), or 51 Hz (Ess$\mathcal{W}_{92.4\%}$), for Observer 2. Table 5.1 shows the detection bandwidths for all levels and all observers, including Observer 3 (although these figures are based only on three replications). This minimum could, perhaps, be considered an estimate of the critical bandwidth.

Also plotted on these figures are the bandwidths and durations associated with the ideal full–linear detector. The ideal bandwidth is defined to be equal to the Ess$\mathcal{W}_{92.4\%}$ bandwidth of the δ transients, and the ideal duration is equal to the Ess$\mathcal{T}_{92.4\%}$ duration of the Kaiser window, because, according to Green and McGill (1970), an ideal noise–in–noise detector should match the bandwidth and duration of the signal. These definitions were used to determine the parameters of the experimental signals in Chapter 3. Observer 2 crosses the *equal–bandwidth* line, so that for wide–band short–duration signals, his detection bandwidth is narrower than that of the sig-

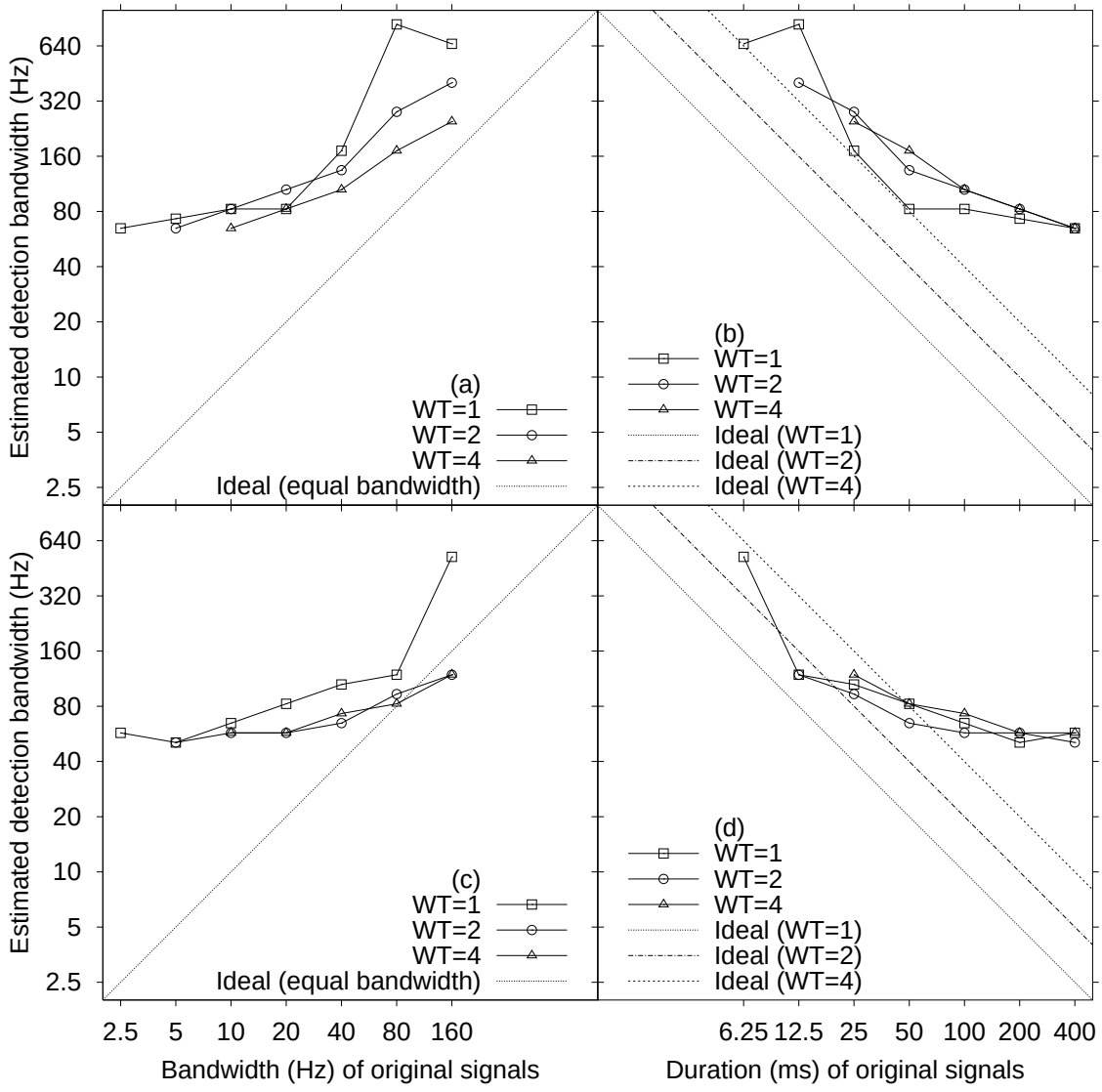

Figure 5.5: Comparison of the estimated Ess$\mathcal{W}_{92.4\%}$ detection bandwidth of Observer 1 and Observer 2 for each experimental condition and level, plotted against the Ess$\mathcal{W}_{92.4\%}$ bandwidth of the original signals ((a) and (c), respectively) and the Ess$\mathcal{T}_{92.4\%}$ duration of the original signals ((b) and (d), respectively). The detection bandwidth was estimated from the full–linear simulations.

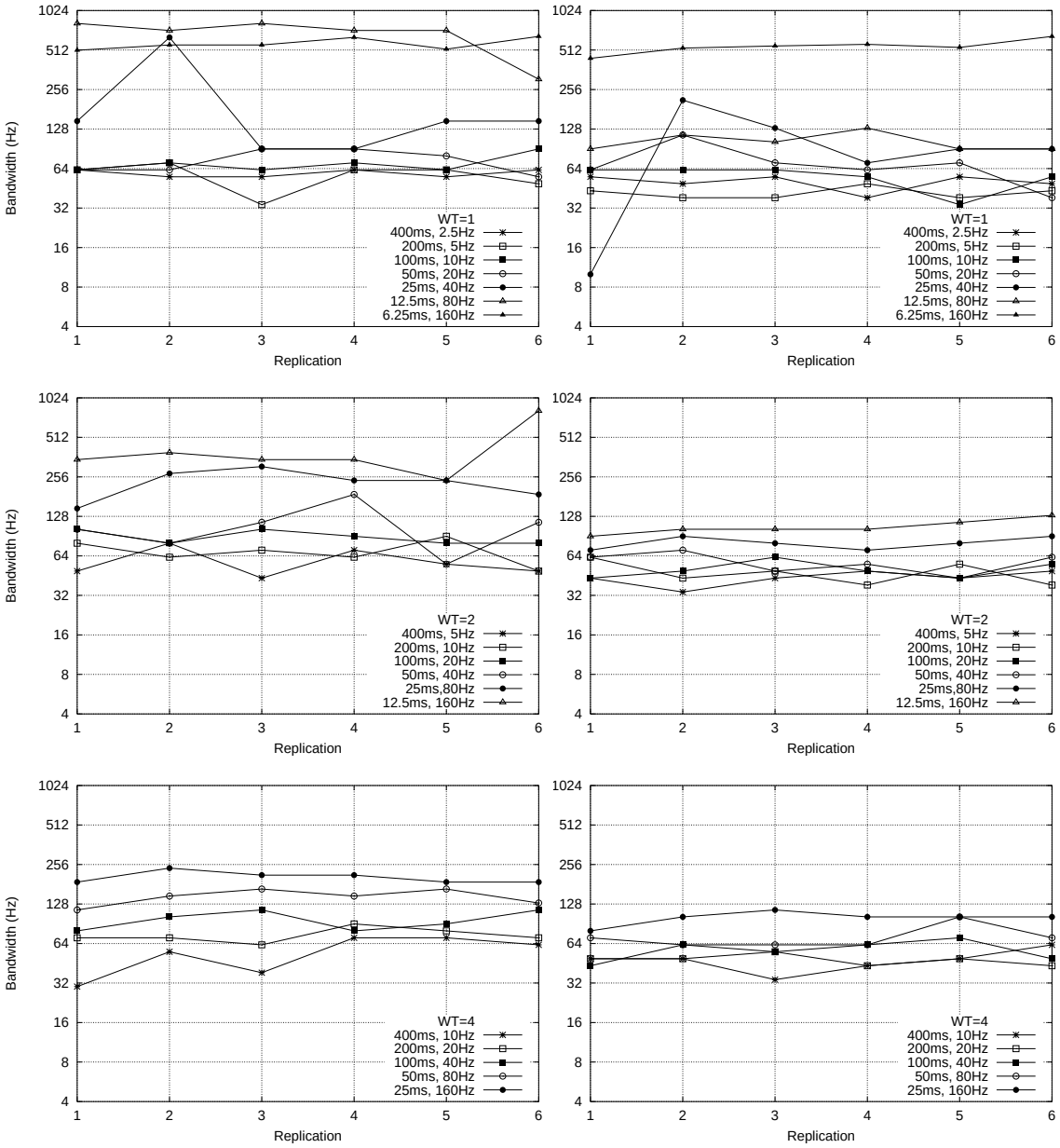

Figure 5.6: Detection bandwidths for Observer 1 and Observer 2 estimated from individual replication ratings and the full–linear detector simulations.

nal. Observer 1's detection bandwidth is always wider than the signal. This would imply that Observer 1 always performs worse than the full–linear detector, but that Observer 2 is able to perform as well as the full–linear detector at some levels once unique noise in the sum–of–ratings was removed. In fact, one level does appear to fall on the ideal line for Observer 2: {25ms, 80Hz}. These functions are very similar for the energy detector simulations but the bandwidths tended to be slightly smaller. They also resulted in lower correlations.

Variability of the correlation–bandwidth functions over replications was also considered. A succinct way of displaying this information is to plot the maximum correlated bandwidth as a function of replication. These plots, for the full–linear detector simulations, are shown in Figure 5.6. Observer 1 is reasonably stable for longer durations but for the short durations the estimated bandwidth varies considerably—especially for $\mathcal{WT}$=1. It should be noted that this analysis is on single replication data so the bandwidth estimates are confounded with other unique noise sources to a greater degree than the six replications combined. Both Observer 1 and Observer 2 are more stable for $\mathcal{WT}$=2 and 4, compared with $\mathcal{WT}$=1. This is perhaps due to the difficulty of minimising the filter bandwidth, because of the effect of the lower bound imposed by the acoustical uncertainty principle as $\mathcal{WT}$ gets small.

For the current experiment, the correlation analysis would not easily pick up if an observer kept a constant bandwidth within an experimental session, but varied it across sessions. This is because each replication was completed over a period of weeks in 10–15 sessions. The results reported in Section 5.2, from the $\mathcal{A}_{\mathrm{SIFC}} = P(\mathrm{C})_{\mathrm{2IFC}}$ project, were from an experiment where the replications were completed in 1–2 sessions within the same day. Thus, a variable detection bandwidth explanation can be put forward with more confidence.

5.4 Envelope detectors

A variety of envelope detector simulations were run on all the buffer signal sets where the filter bandwidths used were the best correlated bandwidths for each observer from the full–linear simulation of the previous section.

A preliminary simulation indicated that there was no difference between half–wave and full–wave rectification on subsequent correlations with human observers (i.e., the two forms of rectification resulted in the same stimulus orderings) so half–wave rectification was chosen, to be true to Jeffress' notion of the envelope detector. There was, however, a substantial difference in the stimulus orderings by a square–law rectifier compared to both the full–wave and half–wave rectifiers.

5.4.1 Method

Preliminary tests indicated that the best correlations occurred for leaky integrators with long time constants. To investigate this finding, ten envelope detector simulations were run for each bandwidth—each with a different integrator time constant (10, 120, 230, 340, 450, 560, 670, 780, 890, and 1000 ms).

Two sampling strategies were used (a) peak detection of the output of the integrator for each waveform, and (b) sampling the output of the integrator at thirty different points spaced equally between 10 ms and the end of the integrator (on a linear scale).

5.4.2 Results

Figure 5.7 shows an example of the correlation between the sum–of–ratings for Observer 2 in level {400ms, 2.5Hz} and the output of each envelope detector simulation (for each integrator time constant), as a function of thirty sample times. The results from other simulations tended to have higher correlations, but the difference between short and long time constants was not as extreme.

Figure 5.8 shows that for envelope detector simulations, the correlation as a function of sampling time increases with time, but only very slightly after about 200ms. Figure 5.9 shows that the best correlated sampling time is *relatively* independent from the integrator time constant. In log coordinates these sets of lines are roughly equally spaced as a function of signal duration, especially for Observer 1.

These sets of figures indicate that (a) the correlation increases as the integrator time constant increases and that this increase appears to tend towards an asymptote, and (b) the correlation–duration function reaches a plateau near the peak of the integrator output. If the integrator was allowed to leak further then this correlation would eventually drop as the signal decayed away. The overall maximum correlation does not always occur with the longest time constant, but once the plateau has been reached, differences are tiny and are presumably just statistical fluctuations.

The difference in the shape of the correlation–duration function between the $\mathcal{N}$ and $\mathcal{SN}$ events is minimal; that is, the best correlations are for a similar decay time and sampling time. The actual correlations for the $\mathcal{SN}$ events tend to much larger (often over 0.9). There is less cause for concern in using the $\mathcal{SN}$ event data for testing duration parameters, because the shape of the signal window was identical for all events. The shape in the frequency domain for each event, however, was not identical for filters with bandwidths different to the signal bandwidth, thus leading to the difference in the shape of the correlation–bandwidth functions. In the end, the $\mathcal{SN}$ event data were not used for estimating duration properties, mainly due to the sheer amount of data analysis and presentation that would entail.

5.5 Assessing the best detector model

Although there is no one detector that is always better correlated with human performance than any other, the full–linear detector stands out as generally being better correlated. The difference, however, between the full–linear detector and the envelope detector with a long time constant is tiny. The latter may be more justifiable physiologically, but the full–linear detector is simpler to work with, so will be used to compare with the human psychometric functions.

5.5.1 Differences in correlation

Figure 5.10 shows some examples of the correlation between the sum–of–ratings for Observer 1 and Observer 2 and the output of (a) the energy detector with the best correlated bandwidth, (b) the envelope detector for the 230 ms and 1000 ms time constants, and (c) the full–linear detector with the best correlated bandwidth; all sampled at thirty different times. Except for the envelope detector with the 230 ms time constant, all the detectors tended to be the best in their class. The additional envelope detector was included, because its time constant was closest to Green's posited 200 ms (Green, 1973, 1985). Also included in these graphs are the standardised temporal–averaged output of the four integrators. The biggest difference between detectors occurs for the

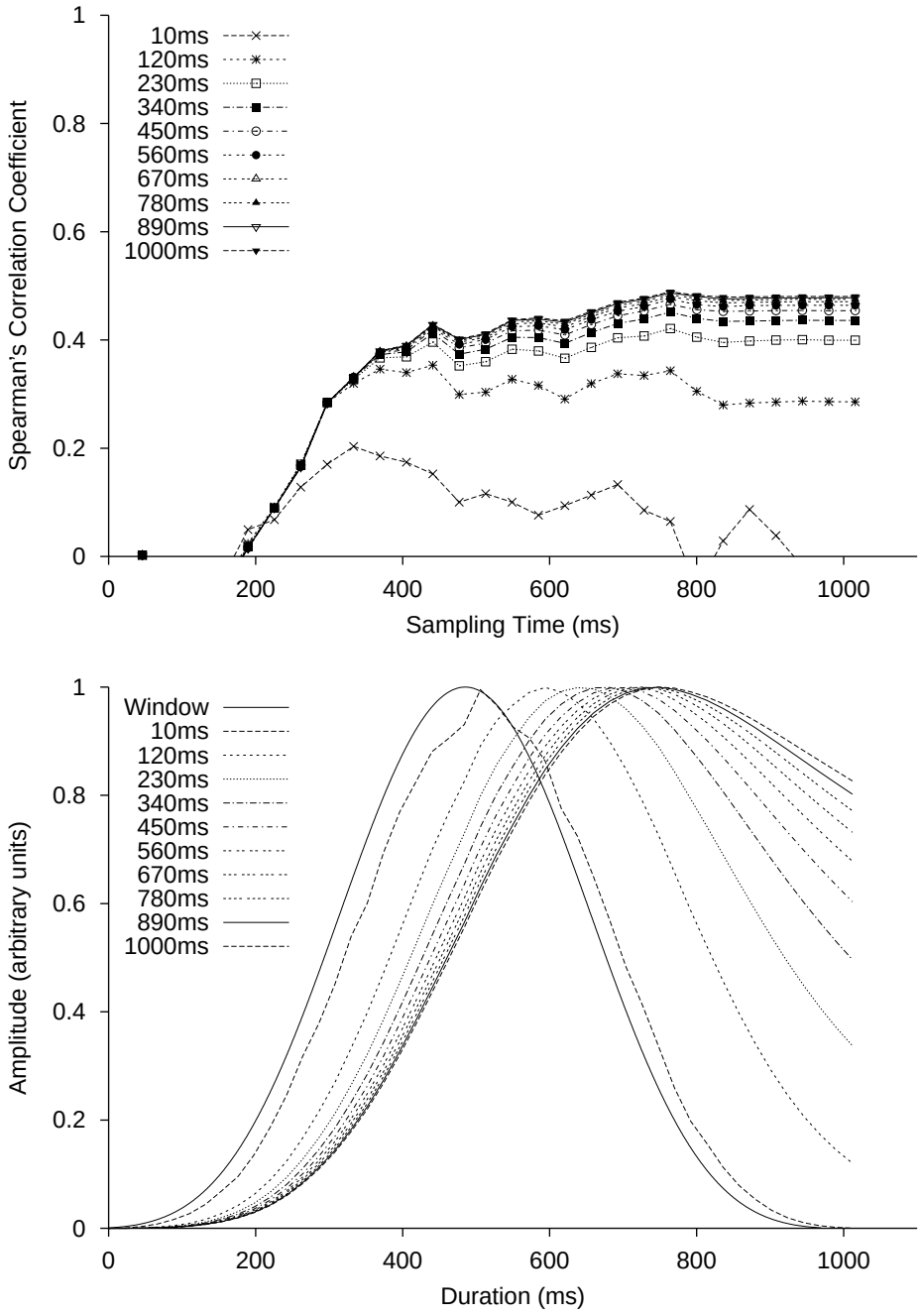

Figure 5.7: Correlations between the sum–of–ratings from Observer 2 and ten envelope detectors, each using a different integrator time constant, and sampled at thirty different times, for the {400ms, 2.5Hz} $\mathcal{N}$ stimuli. (a) shows the correlation–duration functions from each integrator time constant and sampling time, and (b) shows the temporal–averaged output of the envelope detector compared with the window of the original waveforms. The two graphs have the same time base so the correlations can be compared directly to the integrator output.

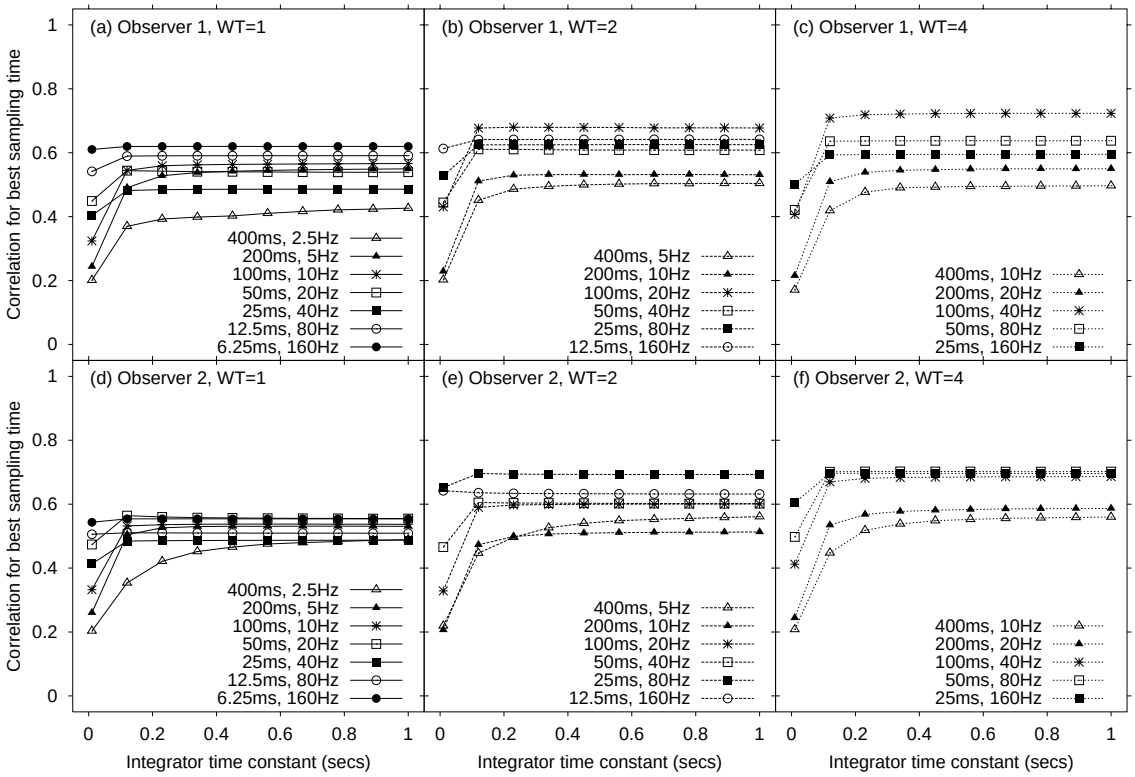

Figure 5.8: The correlation coefficient for the best sampling time tends to asymptotically increase as the integrator time constant increases, for each observer and experimental condition.

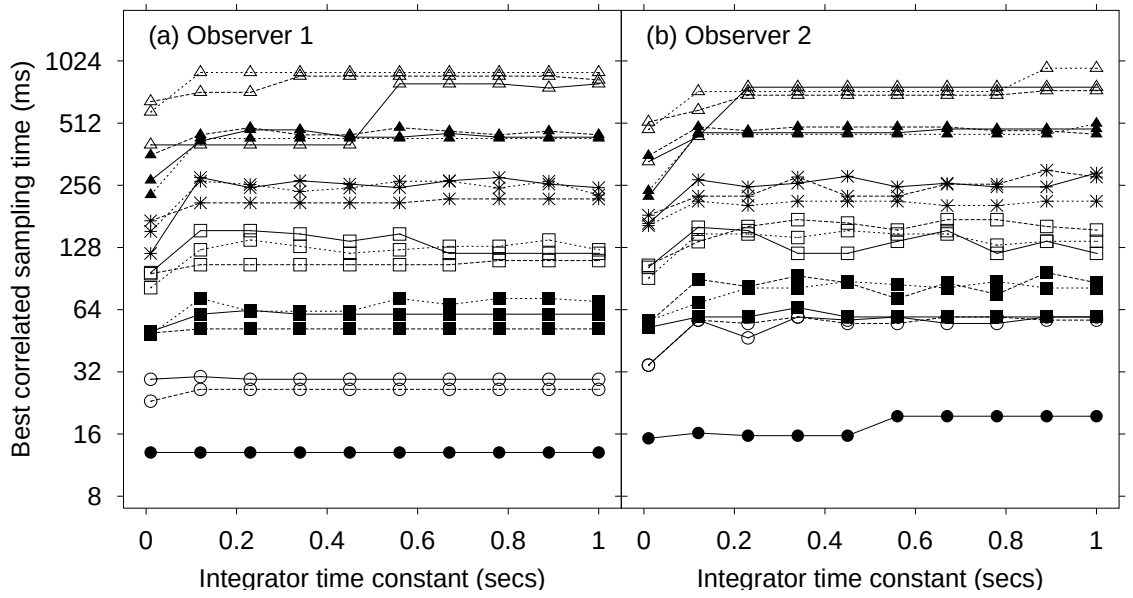

Figure 5.9: The best correlated sampling time for each integrator time constant (key is the same as for Figure 5.8) for all conditions and levels for each observer. Note that the sampling time is in $\log_2$ coordinates.

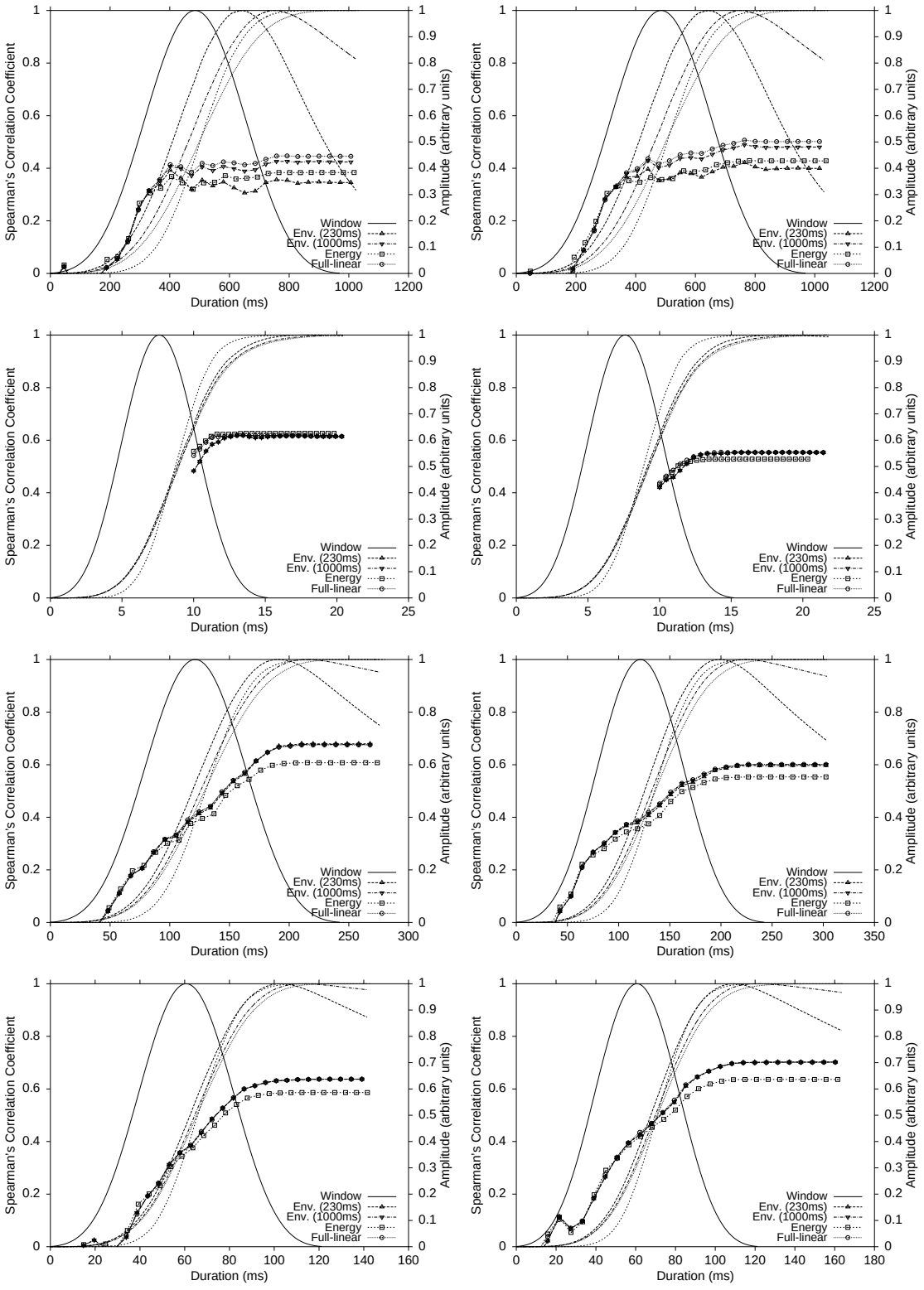

Figure 5.10: Correlation–duration functions based on the output of four different detectors (energy, full–linear, and envelope [time constant=230 ms and 1000 ms]), sampled at thirty different times, and the sum–of–ratings from Observer 1 (left) and Observer 2 (right). Overlayed are the temporal–averaged output of the four integrators, and the Kaiser window. Stimulus sets from top to bottom are {400ms, 2.5Hz}, {6.25ms, 160Hz}, {100ms, 20Hz}, and {50ms, 80Hz}.

long duration signals. This is not surprising, because the effect of the integrator time constant should be relative to the signal duration.

The energy detector, except for level {6.25ms, 160Hz} for Observer 1, is consistently worse than the full–linear detector and the envelope detector with a time constant of 1000 ms. For the longer duration signals, the envelope detector with a time constant of 230 ms tended to be the worst. Looking at the temporal–averaged integrator outputs, the correlation–duration functions are more similar if the output of the integrators are similar. In particular, there is not much difference between full–linear with a true integrator and the envelope detector with a very long time constant. If the integrator output was allowed to continue then the leaky integrator would eventually dissipate, whereas the full integrator would hold its value. It seems to be the case that for the envelope detector, that correlations would continue to improve if the time constant was increased further, but the most parsimonious model is the full–linear with an unequal bandwidth, full–wave rectifier, true integrator, and sampling strategy at the end of the signal.

The best sampling times were related to the time the temporal–averaged integrator output reached its peak. This is not the same sampling strategy as sampling the peak of individual waveforms (that detection scheme resulted in lower correlations in general).

The simulations that tended to give high correlations were compared to see if there was one detector that best described human hearing (bar graphs of the correlations are shown in Figures G.7 and G.8 in Appendix G). The candidates included (a) the energy detector with the best correlated bandwidth, sampled at the end of the integrator [Energy–EOI], (b) the same energy detector but the best correlated sampling time—this is usually the same as the end of the integrator, but small statistical fluctuations may favour another time [Energy–ATT], (c) the full–linear detector with the best correlated bandwidth, sampled at the end of the integrator [Full–EOI], (d) the same full–linear detector, but the best correlated sampling time [Full–ATT], (e) the envelope detector with the best correlated time constant (usually 1000 ms), sampled with a peak detector [Env.–Peak], and (f) the envelope detector with the best correlated time constant *and* sample time [Env.–ATT]. The actual values for the various parameters do not matter, because this comparison is among different *classes* of simulated observers.

These comparisons indicated that both energy detectors (Energy–ATT and Energy–EOI), and the peak envelope detector (Env.–Peak), gave consistently lower correlations, compared with both the full–linear detectors and the ATT envelope detector. There was essentially no difference among the latter detectors. This is probably because a leaky integrator with a long time constant tends towards a true integrator, making the difference between the two models negligible.

5.5.2 Psychometric functions

A test of the Bester correlation method is to compare the psychometric functions of the best correlated simulation with the human observers. Figures G.9 and G.10 show the psychometric function of the best–correlated full–linear detector, the human functions after GOC and FORCE analysis, and the ideal full–linear detector polynomial psychometric function estimated in Chapter 4. The data points have all been translated by their attenuation, such that they fall onto the ideal full–linear model so their shapes may be easily compared. Labels on each graph report the degree of attenuation in decibels. When considering the comparison, remember that the best fit was chosen solely from the $\mathcal{N}$ event ratings. Many of the functions are extremely well fitted by the full–linear detector simulations, especially for non-extreme signal sets.

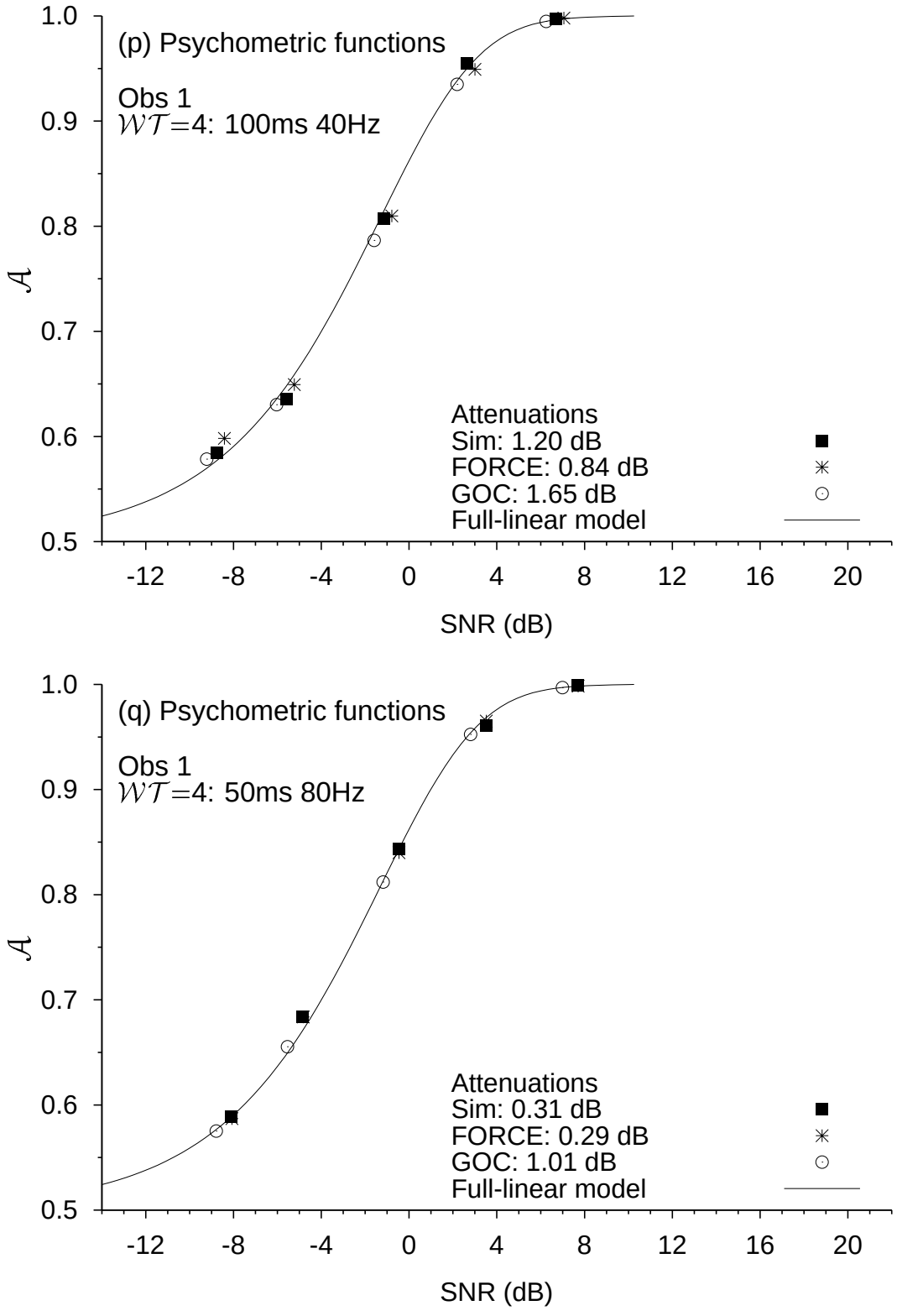

Figure 5.11: Examples of the overall best correlated full–linear detector psychometric functions compared with FORCE and GOC analyses, for Observer 1. The data points from the GOC, FORCE, and best correlated detector have been translated by their average attenuation so that they fall onto the ideal full–linear detector function.

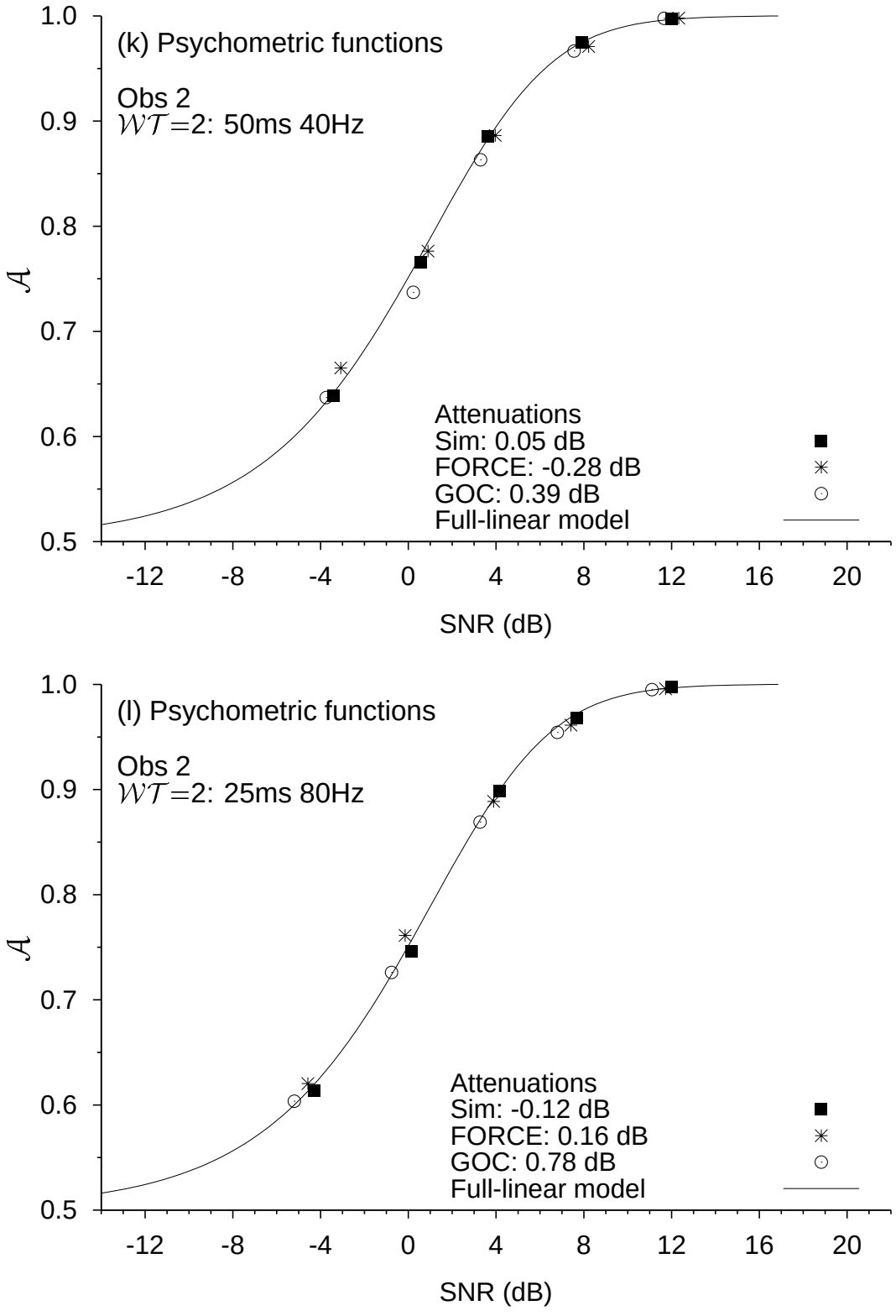

Figure 5.12: Examples of the overall best correlated full–linear detector psychometric functions compared with FORCE and GOC analyses, for Observer 2. The data points from the GOC, FORCE, and best correlated detector have been translated by their average attenuation so that they fall onto the ideal full–linear detector function.

The important findings are that:

- the shape of the ideal full–linear psychometric function accurately describes the shape of most of the empirical data points (human and simulation) except for some of the long–duration levels. The data points, as a function of signal–to–noise ratio, steepen in a similar way as the ideal model and change in the same way as $\mathcal{WT}$ of the signals change.

- the human data points tend to follow the same pattern where the best–correlated data points fall away from the ideal function, for instance, the point associated with the smallest signal–to–noise ratio in Figure G.9(i) and Figure G.10(i) is well under the fitted function for all observers.

- in some cases there is essentially no attenuation between the human data and the best–correlated full–linear simulation. This can be seen where the human and the best–correlated data points fall on top of one another—if there was just differing levels of attenuation then the points may fall on the same function but at different places. This indicates that even though performance may not be ideal, the best–correlated full–linear detector still predicts performance extremely well (e.g., Figures G.9 (b), (d), (j), (l), and (q) for Observer 1 and Figures G.10(c) and (l) for Observer 2). What is unclear is why sometimes it is the GOC data that fit well and other times the FORCE data.

- in some cases there is essentially no attenuation from the ideal full–linear detector indicating that humans, once error has been removed, can perform ideally (e.g., Figures G.9(q) and (r) for Observer 1 and Figures G.10(k) and (l) for Observer 2).

Figures G.9(p) and (q) and G.10(k) and (l) are enlarged and reproduced as Figures 5.11 and 5.12, for comparison.

5.5.3 Signal analysis

Tables G.1–G.6 in Appendix G show the results of signal analysis on the best–correlated full–linear detector simulations. The interaction between estimated signal power (converted from estimated signal–to–noise ratio) and estimated $\mathcal{WT}$ are illustrated in Figures 5.13 and 5.14. These figures show that as signal power increases, estimated $\mathcal{WT}$ decreases—in some cases quite markedly. Note that Observer 2 crosses the ideal $\mathcal{WT}$ line. This does not mean that $\mathcal{W}$ and $\mathcal{T}$ were optimally matched. The analysis was similar to that for the ideal energy detector in Section 4.3 except that the duration measures were for a linear rectifier, so they did not need the additional transformations required for the square–law rectifier. The tables and figures show that

- The signal–to–noise ratio, estimated using Equation (3.5), is barely changed by the detector, despite the bandwidth of the filter being unequal to the signal. This is due to the method of calculating the signal–to–noise ratio, which was estimated from the pass band of the spectrum to the masker floor (this method is described in Section 3.3.5). It is possible that for the cases where the observer bandwidth is narrower than the signal bandwidth, the signal–to–noise ratio would be underestimated slightly as the average ERB measure requires averaging through the passband of the signal. The cutoffs used were for the unfiltered signal. This does not seem to have made much difference to the final estimates.

- The duration of the waveform out of the rectifier is barely changed from the input duration, because the filter bandwidths are all rather wide so there is little ringing. There is a small tendency for durations to decrease as signal–to–noise ratio increases.

- The ERD duration estimates are smaller than the $\mathrm{Ess}\mathcal{T}_{92.4\%}$ estimates in the cases where the observer's bandwidth is wider than the signal bandwidth. For the input signals these measures were essentially the same so presumably this difference comes about from a change in the shape of the waveform envelope after filtering.

- For the cases where the observer's bandwidth is wider than the signal bandwidth (which is most cases), the estimated bandwidth (for all bandwidth measures) decreases markedly as signal–to–noise ratio increases. The degree of decrease is dependent on the bandwidth measure, because the TdB bandwidth only considers the peak of the waveform (and is therefore sensitive to the shape of the spectrum), whereas the $\mathrm{Ess}\mathcal{W}_{\%}$ and AERB are calculated over the entire spectrum. This phenomenon was described and illustrated in Section 4.3.1.

- Because the bandwidth changes with signal–to–noise ratio, the estimates of $\mathcal{WT}$ also decrease with signal–to–noise ratio. The bigger the difference between observer and signal bandwidth, the greater the change in $\mathcal{WT}$ as a function of signal–to–noise ratio. Again the degree of decrease also depends on which definition of $\mathcal{WT}$ is used. Because the duration measure is stable, this difference in $\mathcal{WT}$ is primarily due to the change in bandwidth.

The information in these tables goes some way to answering the question "What is the rôle of the bandwidth–duration product $\mathcal{WT}$ in the detectability of diotic signals?". They show that for a set of signals with the same $\mathcal{WT}$, but different bandwidths, durations, and signal–to–noise ratios, that the human hearing system does not use a constant $\mathcal{WT}$ to detect those signals, even though such a strategy would be optimal. In other words, the parameters of the human hearing system cannot be modelled with a constant $\mathcal{WT}$ theory (such as the χ^2 energy detector model). But neither does the hearing system use fixed parameters. It appears that although the hearing system is able to integrate over the full duration of the waveform, with apparently little error, that its ability to filter with a bandwidth equal to that of the signal, is limited to a minimum bandwidth of around 40 Hz.

This chapter has explored the quasi–molecular, or stimulus–by–stimulus, approach to studying human hearing. The other extreme is to use an measure that encompasses all stimuli *and* all events. Such a measure is Scurfield's (1995, 1996) multi–event information theory measure of detectability, $\mathcal{D}$. The next chapter looks at this analysis more closely by calculating $\mathcal{D}_6$ for both the human data (after GOC and FORCE analysis) and the simulations (ideal and best–correlated).

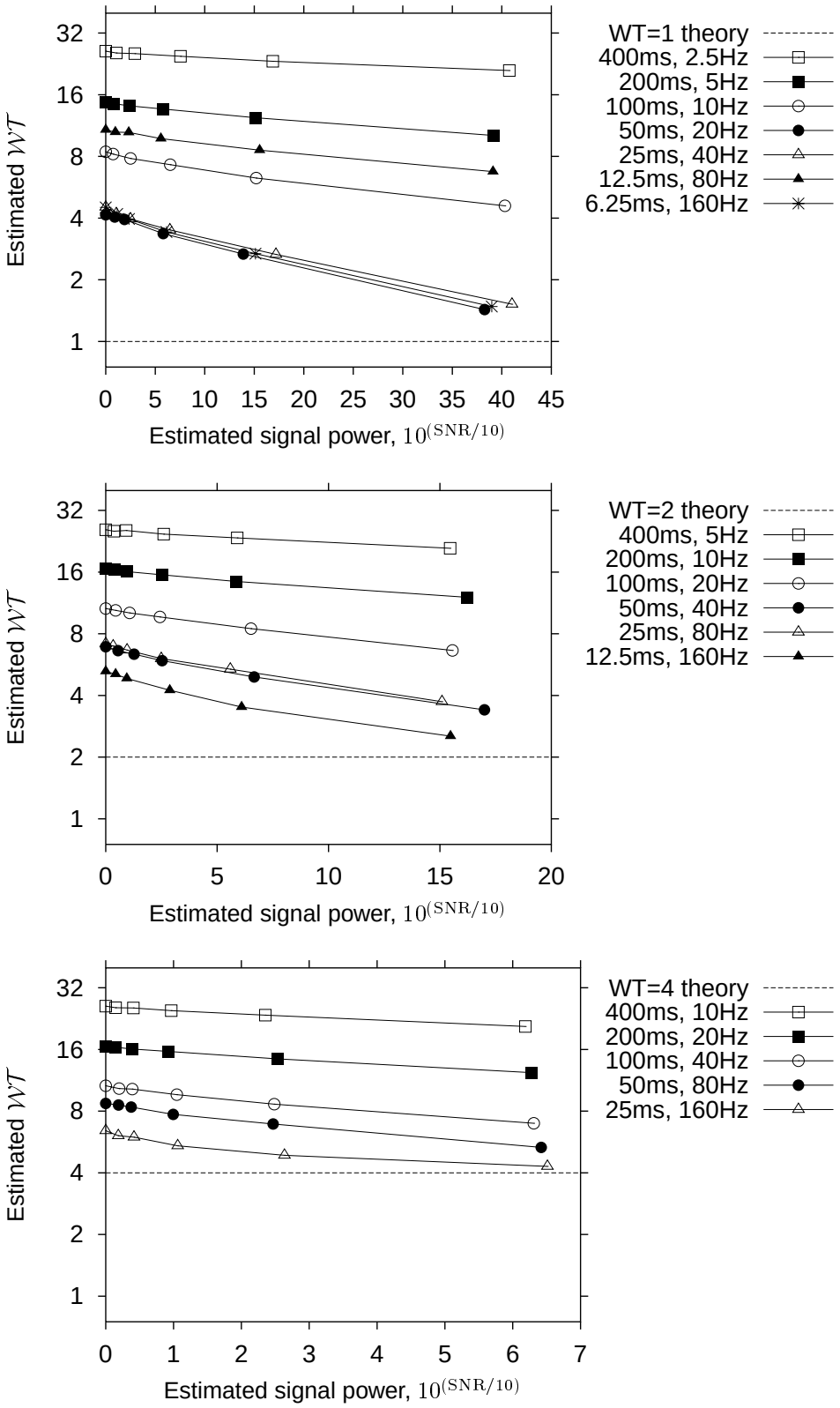

Figure 5.13: Estimated $\mathcal{WT}$ as a function of estimated signal power (converted from the estimated signal–to–noise ratio, by assuming unity noise power), for Observer 1, in each experimental condition and level. (a) $\mathcal{WT}$=1, (b) $\mathcal{WT}$=2, and (c) $\mathcal{WT}$=4. Note that estimated $\mathcal{WT}$ is plotted in $\log_2$ coordinates.

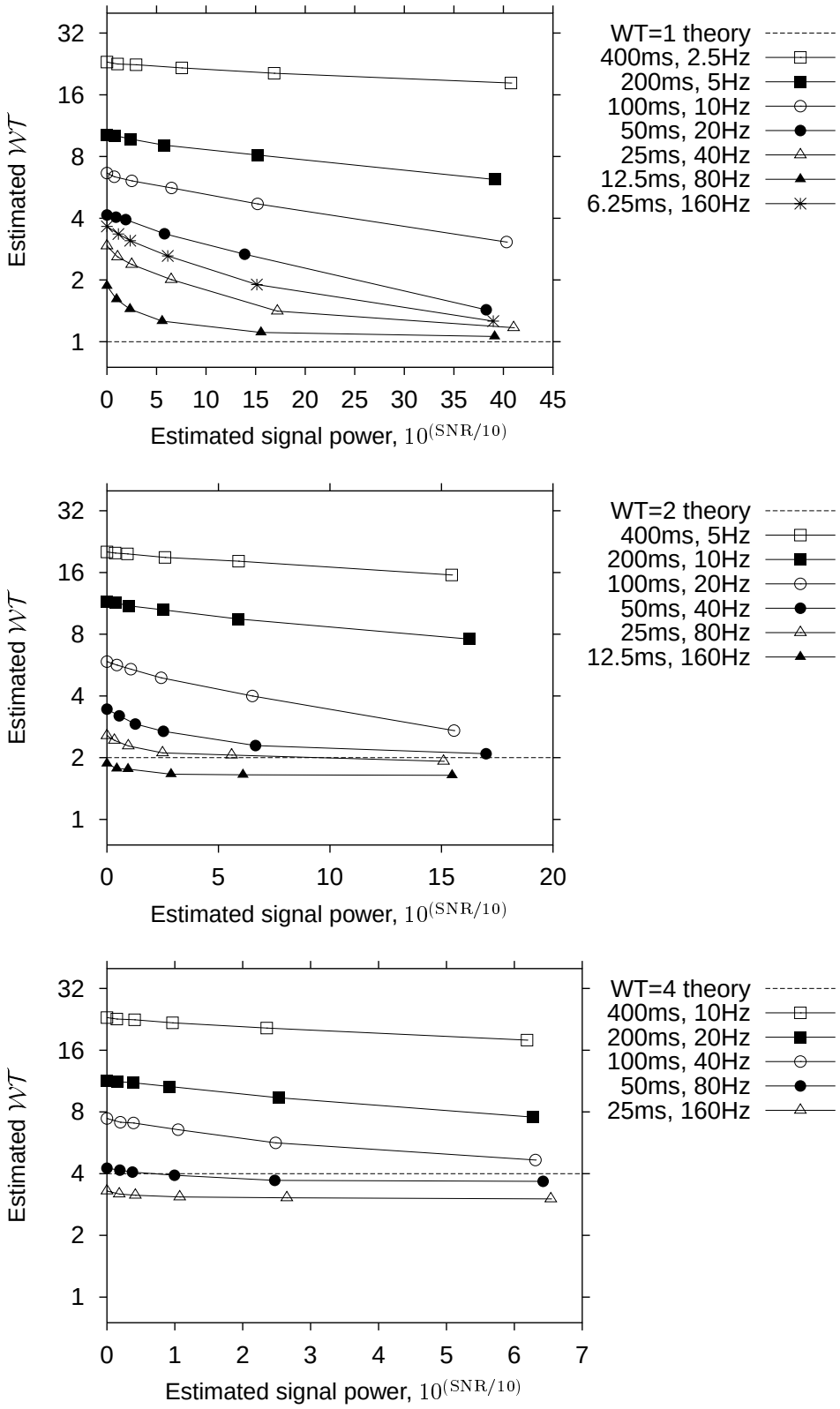

Figure 5.14: Estimated $\mathcal{WT}$ as a function of estimated signal power (converted from the estimated signal–to–noise ratio, by assuming unity noise power), for Observer 2, in each experimental condition and level. (a) $\mathcal{WT}$=1, (b) $\mathcal{WT}$=2, and (c) $\mathcal{WT}$=4. Note that estimated $\mathcal{WT}$ is plotted in $\log_2$ coordinates.

Chapter 6

Multiple–event discriminability

"Reality is nothing but a collective hunch."

Lily Tomlin.

.....

The previous chapter considered a stimulus–by–stimulus approach to studying human hearing. This chapter explores the other extreme of analysing all stimuli *and* all events together. By using Scurfield's (1995, 1996) multi–event information theory measure of detectability, $\mathcal{D}_n$, on both human and simulated ratings, overall performance can be assessed by measuring human discriminability relative to simulated discriminability. The term "discriminability" is preferred over "detectability" here, because it better reflects the task of distinguishing among n events rather than two events.

An abiding problem in psychophysics, and especially ROC analysis, has been the inability to measure performance over multiple events; most theory applies only to two–event tasks. Recently Brian Scurfield generalised ROC analysis to multiple–events and derived a *discriminability* measure, $\mathcal{D}_n$, (where the subscript stands for the number of events) which overcame the problems of previous methods (Scurfield, 1995, 1996, 1998; Lapsley Miller, Scurfield, Drga, Galvin, & Whitmore, 1999). This measure was derived using both ROC analysis and information theory. Scurfield (1995, 1996) was primarily concerned with unidimensional decision axes, but recently he generalised his results to multi–dimensional multiple–event ROC analysis (Scurfield, 1998).

6.1 Scurfield's discriminability measure $\mathcal{D}_n$

Scurfield (1995, 1996) showed that an observer's performance in discriminating n independent events may be represented as $n!$ ROC hypersurfaces, in n–dimensional ROC hyperspaces. For instance, in the three event case, performance is described by six ROC surfaces in three dimensions. A transformation of the hypervolumes under each of the ROC hypersurfaces is then used to measure discriminability. This measure is called $\mathcal{D}_n$.

Scurfield (1996) defined $\mathcal{D}_n$, for the n–event single–interval forced–choice task, as

$$\mathcal{D}_n \quad = \quad \log(n!) - \mathcal{H}_n \tag{6.1}$$

where

$$\mathcal{H}_n = -\sum_{\alpha} \mathcal{V}_{\alpha(12\ldots n)} \log\left(\mathcal{V}_{\alpha(12\ldots n)}\right) \tag{6.2}$$

$$= -\sum_{\alpha} P(X_{\alpha(12\ldots n)}) \log\left(P(X_{\alpha(12\ldots n)})\right) . \tag{6.3}$$

The notation $\alpha(12\ldots n)$ is the index for a possible permutation of the ordering of n events. To sum over α means to sum over all possible permutations. $\mathcal{V}$ is the hypervolume under the ROC hypersurface, and X is the univariate random variable conditional on each event. $\mathcal{H}_n$ is both a measure of the *prior uncertainty* of the ordering of the n events and a measure of the *average information* from a particular ordering of the n events. Therefore, $\mathcal{D}_n$ is a measure of the *average certainty* about the ordering of the n events.[1] Scurfield also showed, with particular assumptions, that $\mathcal{D}_n$ is equal to the channel capacity of the corresponding n–event n–interval forced–choice task.

Scurfield derived $\mathcal{D}_n$ to have the nine properties he believed were essential for a psychophysical measure:

1. $\mathcal{D}_n$ is non–parametric, allowing it to be applied to any detection problem regardless of the underlying distributions, and allowing comparisons across detection tasks regardless of the decision axis.

2. $\mathcal{D}_n$ is independent from the criterion adopted by an observer.

3. It is independent from the prior probabilities of each event occurring.

4. It has a finite value when events are perfectly discriminable.

5. The discriminability of n events is no less than the discriminability of a subset of these events—that is the measure of the task of discriminating more events reflects the additional difficulty.

6. $\mathcal{D}_n$ has a true zero, so that the inability to discriminate among the events has an appropriate value.

7. Ratios of $\mathcal{D}_n$ are relative measures of discriminability, allowing comparisons among observers and tasks. In particular, Scurfield (1996) showed that $\mathcal{D}_n$ may be transformed to a measure of relative discriminability, $\mathcal{D}'_n$, by dividing through by the maximum value of $\mathcal{D}_n$, which is $\log(n!)$.

8. It is invariant to scaling of the decision axis.

9. It is invariant to the *labelling* of the events, because the important factor in discrimination is the correspondence between decisions and events—not the identity of the events.

Implicit in these properties is that $\mathcal{D}_n$ is applicable to more than two events.

$\mathcal{D}_n$ may also be calculated using empirical data. If a rating scale experiment is used, then the calculation of the ROC hypersurfaces may be bypassed, and a generalisation of the Mann–Whitney $\mathcal{U}$ statistic used instead. Miller (1998a) wrote an algorithm to calculate $\mathcal{D}_n$ using this

[1] For discrete random variables this is not entirely true due to equalities of the evidence, but this does not negate the use of $\mathcal{D}_n$.

method. This algorithm is, unfortunately, factorially, exponentially, complex in the number of events. Although it has huge complexity, it is currently practical up to about $\mathcal{D}_9$.

According to Miller (1999, personal communication) it is extremely difficult to describe how to calculate $\mathcal{D}_n$ for arbitrary n. This is because in the discrete case (empirical data are invariably discrete) there is the problem of determining all possible combinations of ties, and their weightings, when evidence from different events is given the same rating. The calculation of $\mathcal{D}_n$ is relatively straightforward when there are fewer than four events, but it is more complex for four or more events. Scurfield (1995) put forward a conjecture for the n–event case, but according to Miller (1998, personal communication) Scurfield did not correctly express the weightings on the different possible ties. Miller's algorithm addresses the problem of determining all the ties and weightings for any given n. As an example, Equation (6.4) shows the formula for one of the twenty–four hypervolumes used to calculate $\mathcal{D}_4$, where N_i is the number of stimuli for event i, and R_i is one of the set of ratings for event i. The operator $\#(x)$ is the number of permutations of ratings for which the argument x is true (Scurfield, 1996).

$$\begin{aligned}
\mathcal{V}_{\alpha(1234)} \quad = \quad & \frac{1}{(N_1N_2N_3N_4)} \times \big[\#(R_{\alpha(1)} < R_{\alpha(2)} < R_{\alpha(3)} < R_{\alpha(4)}) \\
& + \frac{1}{2}(\#(R_{\alpha(1)} < R_{\alpha(2)} < R_{\alpha(3)} = R_{\alpha(4)}) \\
& \quad + \#(R_{\alpha(1)} < R_{\alpha(2)} = R_{\alpha(3)} < R_{\alpha(4)}) \\
& \quad + \#(R_{\alpha(1)} = R_{\alpha(2)} < R_{\alpha(3)} < R_{\alpha(4)})) \\
& + \frac{1}{4}(\#(R_{\alpha(1)} = R_{\alpha(2)} < R_{\alpha(3)} = R_{\alpha(4)})) \\
& + \frac{1}{6}(\#(R_{\alpha(1)} = R_{\alpha(2)} = R_{\alpha(3)} < R_{\alpha(4)}) + \#(R_{\alpha(1)} < R_{\alpha(2)} = R_{\alpha(3)} = R_{\alpha(4)})) \\
& + \frac{1}{24}\,(\#(R_{\alpha(1)} = R_{\alpha(2)} = R_{\alpha(3)} = R_{\alpha(4)}))\big]\,. \qquad (6.4)
\end{aligned}$$

The last four parts of this equation deal with the problem of establishing all the possible combinations of ties, and their weightings.

This algorithm is different to that suggested by Scurfield (1995, p. 95). Scurfield's conjecture was only applicable to three events or less. Scurfield constructed the equations by generalising the two and three event cases, and overlooked the problem shown in Equation (6.4), where the weight of the third term is 1/4, not 1/6. According to Miller (1999, personal communication), the weighting of each term relates to the number of ways the term can be constructed, remembering that equality is symmetric: $R_{\alpha(1)} = R_{\alpha(2)}$ is the same as $R_{\alpha(2)} = R_{\alpha(1)}$. In the term $R_{\alpha(1)} = R_{\alpha(2)} = R_{\alpha(3)} < R_{\alpha(4)}$ there are six possible ways of arranging $R_{\alpha(1)}$, $R_{\alpha(2)}$, and $R_{\alpha(3)}$ across the equalities, hence the weighting for this term is 1/6. In the term $R_{\alpha(1)} = R_{\alpha(2)} < R_{\alpha(3)} = R_{\alpha(4)}$, however, $R_{\alpha(1)}$ and $R_{\alpha(2)}$ are interchangeable, and similarly $R_{\alpha(3)}$ and $R_{\alpha(4)}$, but $R_{\alpha(1)}$ and $R_{\alpha(2)}$ cannot be permuted with $R_{\alpha(3)}$ and $R_{\alpha(4)}$, because the inequality in between is not symmetric. Thus, there are only four possible arrangements of $R_{\alpha(i)}$ in this term, so the weighting is only 1/4.

Miller also showed that the overall weighting should be $1/(N_1N_2N_3 \ldots N_n)$, rather than the size of the rating scale to the power of the number of events as Scurfield (1996) suggested.

6.2 Multiple–event analysis of multiple–signal–level tasks

Scurfield's generalised ROC analysis is relevant to the experiments from Chapter 3, which used a multiple–signal–level task (Tucker et al., 1967). This is because the $\mathcal{N}$ event and five $\mathcal{SN}$ events may be viewed as the six events in a single–interval, six–event, unidimensional, forced–choice experiment. In each experimental session, all six signal levels were interleaved in random order, but the ROC analysis considered only the pairing of each $\mathcal{SN}$ event with the $\mathcal{N}$ event.[2] Results from a six–event analysis, however, should be interpreted cautiously, because the observers were not told to treat the experiment as a six–event task. They were all aware, however, that six events were presented within a session, and that the analysis would be two–event. A similar degree of caution should also be used in evaluating the results of the two–event analyses. There is, however, some precedent for two–event analysis of a multiple–signal–level experiment.

Although the observers were not explicitly asked to make ratings about the ordering of the events, their ratings may be still be analysed as if they were participating in a multiple–event experiment. This is because the observer only had access to the evidence, rather than the events, therefore, the ratings about the *evidence* are being ordered on the rating scale, rather than the events. In making ratings, the observer maps the (expected) range of the evidence onto the entire range of the rating scale.

Therefore, it is suggested that it is appropriate to analyse the current experiment as a six–event task. If there is a problem, then the six–event analysis would presumably return results inconsistent with the findings from the other analyses. Regardless of whether the analysis is totally appropriate for the current experiment, the testing and evaluation of Scurfield's analysis is important, because it is new and untried in human psychophysical experiments.

6.3 Multiple–event analysis of experiments and simulations

In the context of this experiment, $\mathcal{D}_6$ may be used for global comparisons between the simulations and the human performances. This is in contrast to the previous chapter where the focus was on how observers order individual stimuli. There are some problems with using $\mathcal{D}_6$ for this type of experiment:

- Each signal level is being treated like a separate discrete event when they are points from a continuous event set (that is, signal–to–noise ratio is a continuous quantity). Therefore, unlike psychometric functions and attenuation analysis, results from different conditions cannot be compared if the signal–to–noise ratios are different. In the current experiment, relative comparisons can still be made amongst experimental levels, because the signal–to–noise ratios are the same within an experimental level.

- $\mathcal{D}_6$ has not been derived for the theoretical energy, full–linear, or envelope detector (ideal or non-ideal), so there is no baseline for comparison. It is possible though, in this case, to

[2] Traditionally, stimuli from one $\mathcal{SN}$ event and one $\mathcal{N}$ event would be presented in random order, but if the aim is to test multiple $\mathcal{SN}$ events, then this requires the unnecessary presentation of as many $\mathcal{N}$ events as $\mathcal{SN}$ events. The current method is more efficient. There are some other practical benefits for running a multiple–signal–level experiment with interleaved events, rather than separate sessions of each level. In the latter case, each session would require the observer to recalibrate their use of the rating scale, so that the entire scale was used. On some sessions, this would require rating very small signals at the extreme, and on others very large signals at the the same extreme, possibly leading to confusion and an increase in observer inconsistency. It is probably better, therefore, to mix the levels within a session, so that the overall expected range of sound pressure level is the same over all sessions.

estimate $\mathcal{D}_6$ using computer simulation. Given that there were differences shown between the ideal energy detector theory and the simulated observers (in Chapter 4), it is likely that the theoretical multi–event observer, based on the same distributions of energy, would not be appropriate for the current type of signals anyway.

Each empirical $\mathcal{D}_6$ was calculated from the GOC hit rates and false–alarm rates, using the program DELEN (Miller, 1998a). Each $\mathcal{D}_6$ for the ideal full–linear detector simulations was calculated with the same algorithm by sorting the evidence output of the simulation, and calculating hit rates and false–alarm rates.

Asymptotic $\mathcal{D}_6$ were also estimated using FORCE analysis. The method was the same as described in Chapter 2, with average $\mathcal{D}_6$ being calculated for each combination of replications added, a FORA being fitted, then the asymptotic value estimated. The program FORCED (Miller, 1998b) was used for this analysis. Miller (1998b) used Drga's FORCE analysis algorithm (Drga, 1997a, 1999), but it was implemented using efficient dynamic programming methods, and incorporated the program DELEN.

The results of the FORCE analyses are presented at the end of Appendix E. Tables E.19 and E.20 show the FORCE results, including the parameters k, m, and $\mathcal{D}_{6,1}$ used to fit the FORA, the asymptotic measure of detectability, and the correlation. Figures E.13–E.18 show the empirical and fitted FORA with asymptotic values. The figures are scaled over the entire range of $\mathcal{D}_6$ and the error bars are ± 1 standard deviation. Finally, Figure E.19 shows the FORCE results in log–increment coordinates.

Only two conditions did not sensibly converge: {400ms, 5Hz} and {400ms, 10Hz} for Observer 1. These experimental levels converged to impossible values. The other long duration–narrow bandwidth levels, {400ms, 2.5Hz}, {200ms, 10Hz} and {200ms, 20Hz} also indicated problems, because the empirical FORA was still far from the asymptotic estimate for Observer 1. This indicates there is still a lot of unique noise in the data. Figure E.19 also highlights the problem with these conditions: the slopes of the fitted and empirical FORA in log–increment coordinates is more shallow compared to the other levels, indicating the rate of convergence is slower. All the other levels have very similar slopes. All experimental levels converged for Observer 2.

Figure 6.1 shows graphs of $\mathcal{D}'_6$ as a function of experimental level for each experimental condition. There are two x–axes showing bandwidth and duration respectively. The data are (a) $\mathcal{D}'_6$ for each human observer after GOC analysis, (b) asymptotic $\mathcal{D}'_6$ for each observer after FORCE analysis, and (c) $\mathcal{D}'_6$ for the ideal full–linear detector simulation.

As with the previous FORCE analyses in Chapter 3, the FORCE results follow those of the GOC results but with a larger measure of discriminability. The exception is for the long duration–narrow bandwidth levels of Observer 1 mentioned above, which tended to indicate that asymptotic performance would be better than expected given the difference between FORCE and GOC results in the other levels.

The ideal full–linear detector performs at approximately the same level across experimental levels within an experimental condition. The average is 2.651955 bits for $\mathcal{WT}$=1, 3.620842 bits for $\mathcal{WT}$=2, and 3.87244 bits for $\mathcal{WT}$=4 where the maximum (perfect performance) is

$$\mathcal{D}_{6(\max)} \quad = \quad \log(6!)/\log(2) \approx 9.5 \text{ bits.} \tag{6.5}$$

In all cases, the ideal full–linear detector discriminates better than the human observers after GOC analysis but the degree of difference varies with condition and level. The FORCE analyses indi-

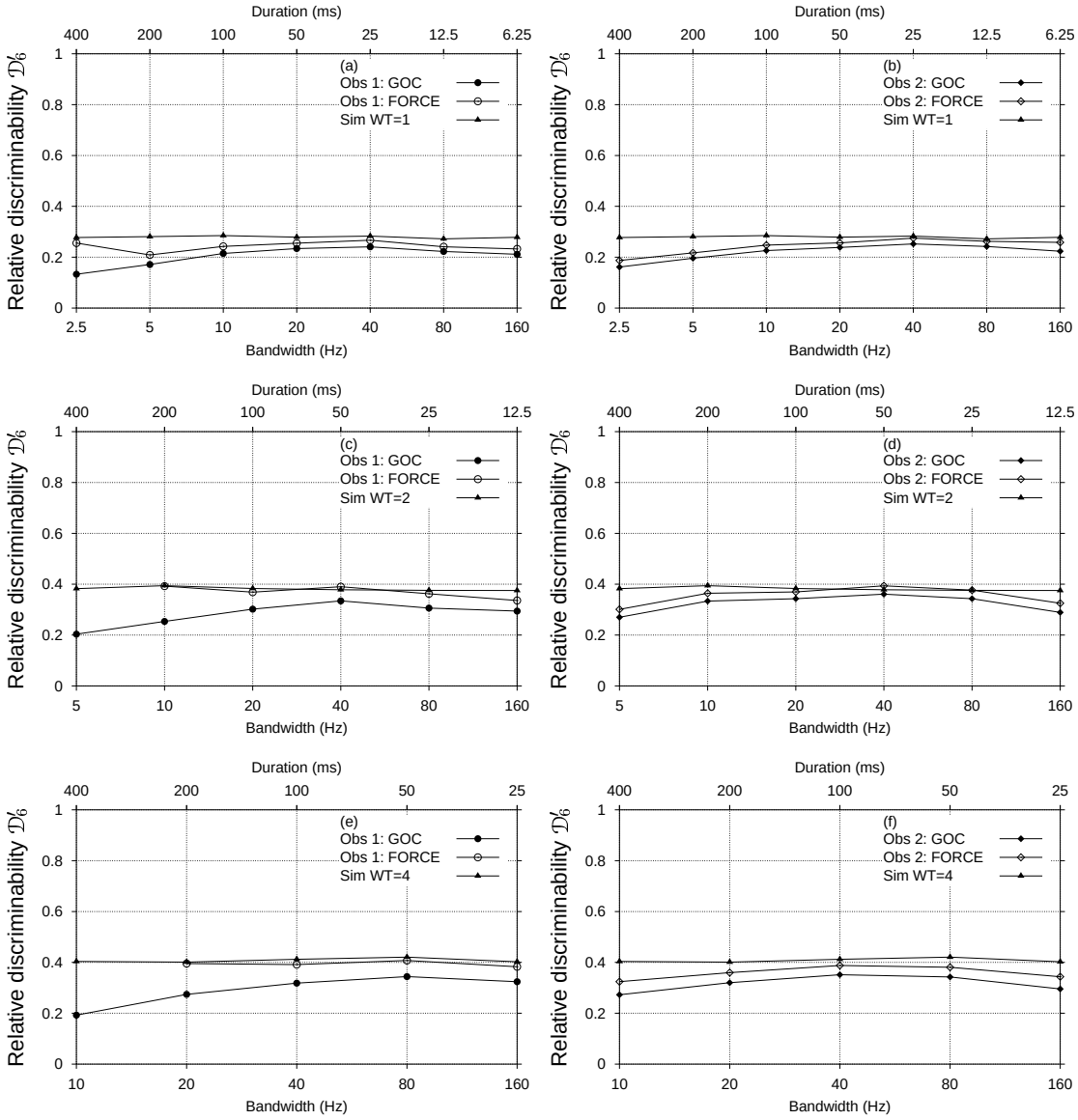

Figure 6.1: Relative discriminability $\mathcal{D}'_6$, after GOC and FORCE analyses, for Observer 1 and Observer 2 compared with the ideal full–linear detector simulation.

cate, however, that in some experimental levels, the human observers are capable of performing the task as well as the simulated full–linear detector if all unique noise is removed.

According to Scurfield (1996), the relative discriminability is a measure of the ordering of the events relative to perfect discrimination. If the ideal full–linear detector simulation data are viewed as approximations or estimations of ideal performance, for these signals, then the relative discriminability of the signals is 28% for $\mathcal{WT}$=1, 38% for $\mathcal{WT}$=2, and 41% for $\mathcal{WT}$=4. In comparison, the relative discriminability of the ordering of these events for the energy detector is 28% for $\mathcal{WT}$=1, 35% for $\mathcal{WT}$=2, and 36% for $\mathcal{WT}$=4.

Discriminability may also be calculated relative to another observer. For instance, the relative discriminability of the event orderings for the energy detector, compared with the full–linear detector, are 98.6% for $\mathcal{WT}$=1, 91.2% for $\mathcal{WT}$=2, and 89.3% for $\mathcal{WT}$=4, indicating again that the full–linear detector is a better detector.

Figure 6.2 shows graphs of relative detectability for each observer relative to the ideal full–linear simulation using the same stimuli. This way of presenting the data highlights how $\mathcal{D}_n$ may be used to make relative comparisons. Comparing the difference between the graphs of $\mathcal{D}_6$ versus those of the proportion of $\mathcal{D}_6$ relative to the ideal full–linear detector shows roughly the same pattern within an experimental level. The proportion measure, however, accentuates the differences among each level and is possibly more useful. The scale of the other graphs compresses the differences due to the scaling being over the entire range of possible performance.

For $\mathcal{WT}$=1, the human observers after GOC analysis tend to do best at levels {50ms, 20Hz} through {12.5ms, 80Hz}, where they perform at least 80% as well as the ideal full–linear detector simulation with a maximum of just over 90% for Observer 2 in level {25ms, 40Hz}. For the narrower bandwidths, performance drops off with the full–linear detector discriminating over twice as well as Observer 1 in level {400ms, 2.5Hz}. In the condition $\mathcal{WT}$=2, Observer 2 nearly reaches ideal performance in level {50ms, 40Hz}. His performance is over 80% in each level except for the two extremes. Observer 1, on the other hand, has a more peaked function (at {50ms, 40Hz}) with performance dropping either side especially for the narrower bandwidths. For the condition $\mathcal{WT}$=4, Observer 1 function slightly peaks at {50ms, 80Hz} whereas Observer 2 peaks at {100ms, 40Hz}.

After FORCE analysis, the relative discriminability of the humans to the full–linear detector reaches 100±5% for the best conditions. The fact that the human observers reach over 100% is probably just due to sampling variability, or possibly the FORCE analysis overestimated the asymptotic value.

Comparing Observer 1 and Observer 2 indicates that Observer 1 does not discriminate among the event orderings as well as Observer 2 except for the short duration—wideband signals (especially {25ms, 160Hz}). The degree of improvement is greater for Observer 1 compared with Observer 2 but this may be because her performance was more noise–degraded. Her asymptotic performance is, in many cases, higher than Observer 2 but because only six replications were completed it may be that the FORCE analysis is biased (Drga, 1998; personal communication).

6.3.1 Comparison of $\mathcal{D}_6$ with psychometric functions

The psychometric function is a powerful descriptor of an observer's performance regardless of whether there is a theoretical function to use as comparison. This is because it directly relates performance to changes in the independent variable. But there are problems with psychometric

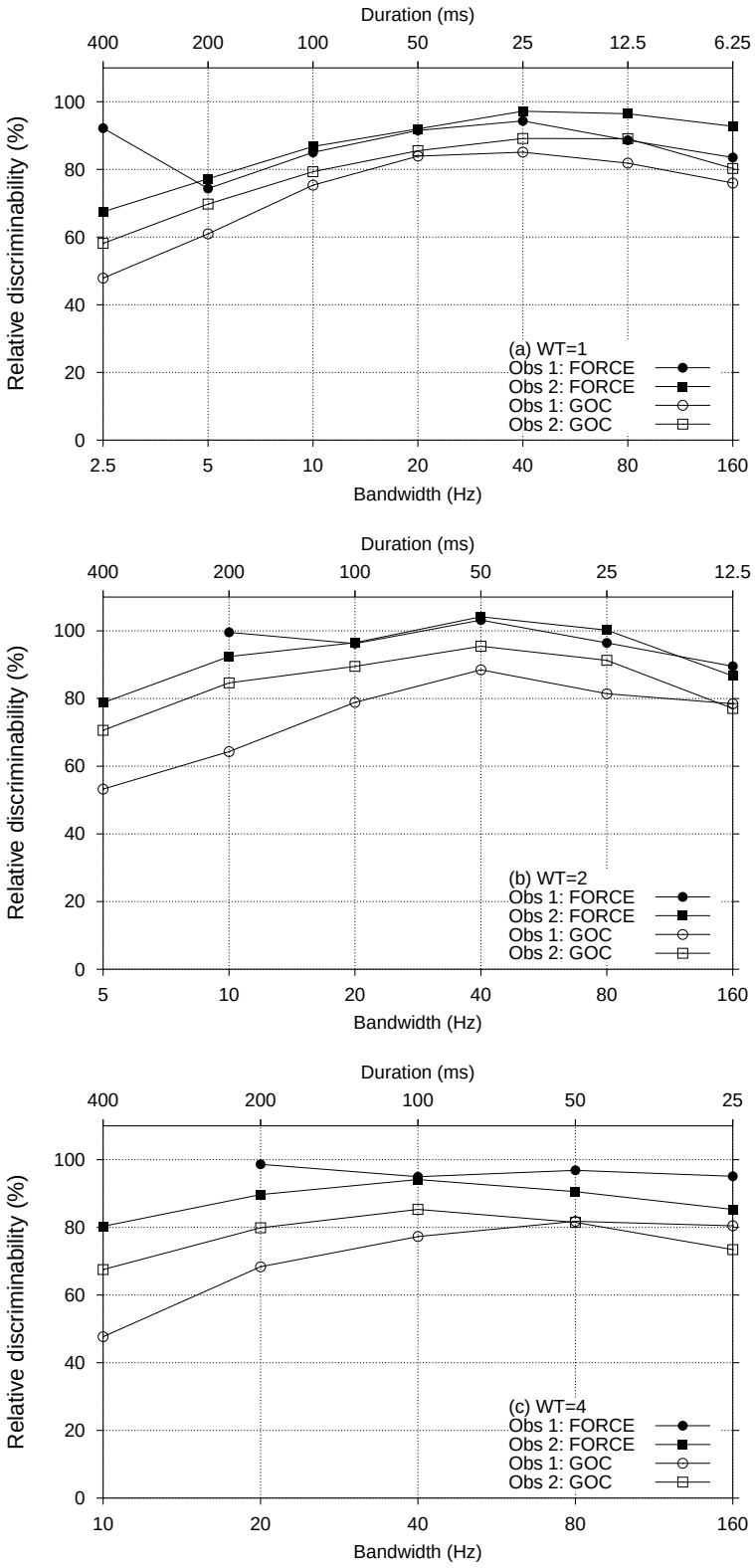

Figure 6.2: Discriminability $\mathcal{D}_6$, for Observer 1 and Observer 2, relative to the ideal full–linear detector simulation, after GOC and FORCE analyses, for each experimental condition.

functions: (a) it can be hard to assess the degree of difference between data and theory if the data are not simply linearly attenuated from the theory, (b) the $\mathcal{N}$ event contributes to all points thereby increasing its contribution to the analysis over the other events, and (c) it is difficult to compare performance across different tasks.

In the current analysis, one of the more noticeable differences between analysis with psychometric functions and that of $\mathcal{D}_6$, is that there is little difference between $\mathcal{D}_6$ for the $\mathcal{WT}$=2 and $\mathcal{WT}$=4 events—only about 0.25 bits, whereas the $\mathcal{WT}$=1 events were less discriminable by about 1 bit. In comparison the $\mathcal{WT}$=4 psychometric functions shows greater detectability than $\mathcal{WT}$=2. The difference is that with $\mathcal{D}_6$, the measure is discriminability among all events, not just the $\mathcal{N}$ event relative to each $\mathcal{SN}$ event.

It is possible that $\mathcal{D}_6$ is related to the slope, or shape, of the psychometric function—as $\mathcal{WT}$ increases so does the slope of the function such that each function looks similar once $\mathcal{WT} > 4$. If this is the case then the set of five signal–to–noise ratios, which result in the same $\mathcal{A}$ on their psychometric function, will result in the same $\mathcal{D}_6$. Consider again, the hypothetical psychometric functions in Figure 3.1. The experimental signal–to–noise ratios are spread further on $\mathcal{WT}$=2, and 4 compared with $\mathcal{WT}$=1. This would imply that the events may be more discriminable.

The only other appropriate measure we have that summarises across all signal levels is the attenuation from the psychometric function. The pros of attenuation analysis are that it is with respect to a particular theory, returns a number with a useful unit (dB), and specifically deals with the continuous nature of signal–to–noise ratio even though we only have six levels. The con is that it is with respect to a particular theory! If we're unsure of the theory, for whatever reason, then interpreting the attenuation factor is difficult and may be misleading.

$\mathcal{D}_6$ on the other hand is independent of any theory or ideal observer, summarises performance across all signal levels, and allows for relative comparisons within an experimental level. If $\mathcal{D}_6$ is also calculated on an appropriate theory (mathematical or computational) then further comparisons may be made.

Attenuation analysis across signal level biases the results by the $\mathcal{N}$ event, because the false–alarm rate is used in the calculation of each $\mathcal{A}$. $\mathcal{D}_6$, on the other hand, treats each event equally so the global measure of detectability is weighted evenly over event.

Comparing the empirical attenuation functions (in Figure 3.16) with the empirical $\mathcal{D}_6$ graphs (in Figure 6.2) indicates that the *shape* of the function relating $\mathcal{D}_6$ to bandwidth or duration is very similar to the attenuation function within an experimental level (except of course that the attenuation gets smaller as performance improves whereas $\mathcal{D}_6$ gets larger).

Both analyses show that humans can act analogously to an ideal full–linear detector if $\mathcal{WT}$=2 or $\mathcal{WT}$=4 and the bandwidth is about 40–80 Hz. But $\mathcal{D}_6$ can do this without having to appeal to a specific theory about how signal–to–noise ratio should be measured or how the measure of detectability relates to the signal–to–noise ratio.

It is illustrative to consider the limitations of using a two–event measure of detectability, such as $\mathcal{A}$, in summarising overall discriminability in the current experiment. With $\mathcal{A}$ there is (a) no method of combining information across all signal levels, (b) no easy way of enumerating how performance differs within and among observers, and (c) differences in attenuation cannot be measured without a specific theory.

For example, Observer 1 obtained $\mathcal{A}$s of 0.52, 0.56, 0.68, 0.79, and 0.89 in level {400ms, 2.5Hz} after GOC analysis. The ideal full–linear detector obtained $\mathcal{A}$s of 0.64, 0.77, 0.88, 0.94, and 0.97. The ideal full–linear detector did better—that is obvious. But the numerical difference of $\mathcal{A}$ does

not mean anything, nor can relative comparisons be made, because $\mathcal{A}$ is not a metric (there is no zero). Comparisons can be made, however, to the fitted ideal polynomial psychometric function, and the average attenuation measured. The best–correlated full–linear detector was attenuated from the fitted model by 0.15 dB, but the data points fell close to the attenuated psychometric function. Observer 1's data were attenuated by 8.1 dB. Thus, Observer 1 would need the signal–to–noise ratios increased by about 8 dB to perform as well as the ideal full–linear detector. But Observer 1's data points do not fall on the attenuated psychometric function (as can be seen in Figure F.2(a); some points are above and others below the curve) so average attenuation is not an appropriate descriptor of performance. In comparison, the corresponding $\mathcal{D}_6$ measures are 1.26 bits for Observer 1 and 2.63 bits for the ideal full–linear detector. Thus, the ideal full–linear detector can discriminate among the event orderings twice as well as Observer 1.

The limitations of $\mathcal{D}_6$ in the current analysis reflect ignorance about how $\mathcal{D}_6$ relates to the underlying evidence distributions for the task of noise–in–noise discrimination. Scurfield (1995, 1996) shows how the theory could be derived, but the math is forbidding.

6.3.2 Detectability as a function of $\mathcal{W}$ and $\mathcal{T}$

As with the attenuation analysis, plotting each $\mathcal{D}_6$ as a function of bandwidth or duration may indicate how the human observers deal with these parameters as $\mathcal{WT}$ changes. Figure 6.3 shows, for each observer, the proportion of $\mathcal{D}_6$ for each experimental level relative to the full–linear detector simulation, but plotted either as a function of signal duration (left hand side) or signal bandwidth (right hand side). The GOC data are used, because there are valid points for all experimental levels.

It is fair to say that performance is dependent on both the bandwidth and duration of the signals but the importance of each parameter is not necessarily the same. It appears that performance for Observer 1 is best described by duration; that is the hearing system is adjusting how it detects signals by using the duration of the signal more than the bandwidth. For Observer 2 the opposite is indicated, that the hearing system is using frequency information more so than duration.

A different way of viewing the same data is by plotting a three dimensional surface of detectability with $\mathcal{W}$ and $\mathcal{T}$ forming the two independent variable axes and some measure of detectability forming the third. In Figure 6.4 the proportion of $\mathcal{D}_6$ relative to the ideal full–linear detector simulation was used, because ideal detection would give a flat surface. The four figures show contours for each observer after GOC and FORCE analysis. Attenuation could also be used, but gives saddle–shaped surfaces that do not show the different regions of equal detectability as clearly. The figures show the contour map of human performance relative to the simulation. High–order cubic spline approximations were used to fit the contours, using GNUPLOT (Williams & Kelley, 1997). Linear interpolation gave similar results but tended to overemphasise the discrete grid of values tested.

It should be noted that because only three $\mathcal{WT}$ values were tested, the contours are restricted to a narrow diagonal band. Thus, the contours are artificially bounded, especially to the top right of the diagram where $\mathcal{WT}$ would normally increases without bound. The bottom left is a more accurate depiction because the uncertainty principle limits how small $\mathcal{WT}$ can get—for real signals it is difficult to generate $\mathcal{WT} < 1$. Obviously more data need to be collected for larger $\mathcal{WT}$ to help map out this surface.

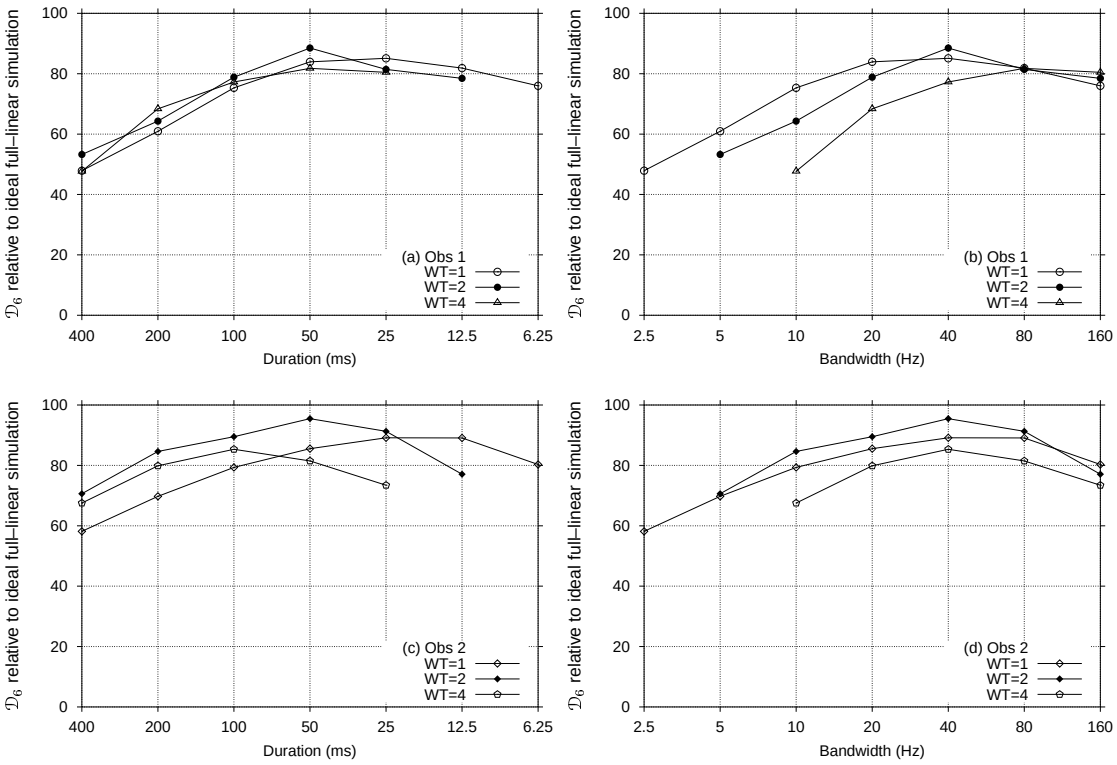

Figure 6.3: Proportion of $\mathcal{D}_6$, relative to the ideal full–linear detector simulation, after GOC analyses, for each observer and each condition, plotted as a function of signal duration (a) and (c) or signal bandwidth (b) and (d).

6.3.3 The rôle of $\mathcal{WT}$

This new method of analysis gives insight into the question of how $\mathcal{WT}$ affects detectability. The contour maps in Figure 6.4 show relatively similar contours for both GOC and FORCE analyses. Because the data are so sparse the contours must only be viewed as indicative. They suggest that for the types of signals tested, that the human hearing system is designed to best detect signals of durations concentrated within 25–50 ms, bandwidths concentrated between 40–80 Hz, and $\mathcal{WT} \approx 2$. Detectability drops off around this area in what appears to be elongated concentric rings, relative to the simulated ideal full–linear detector. This indicates that the detection bandwidth, and the temporal integrator's time constant and sampler, are neither fixed nor ideally flexible. If these parameters were fixed then the rings would not be as elongated; if they were ideally flexible then there would be ridges running along equal $\mathcal{WT}$ coordinates rather than rings. It would seem that the human hearing system is ideally suited to detect a particular bandwidth and duration, but, within a small range, can trade bandwidth for duration for the price of a drop in detectability.

Scurfield's (1995, 1996) multiple–event ROC analysis and associated measure of discriminability gives another way of viewing performance that complements psychometric function and attenuation analysis. $\mathcal{D}_6$ does not really help in determining whether a theory is appropriate or not but does help in enumerating how well observers are performing at a task as well as allowing relative comparisons among experimental levels and among observers (human and computational).

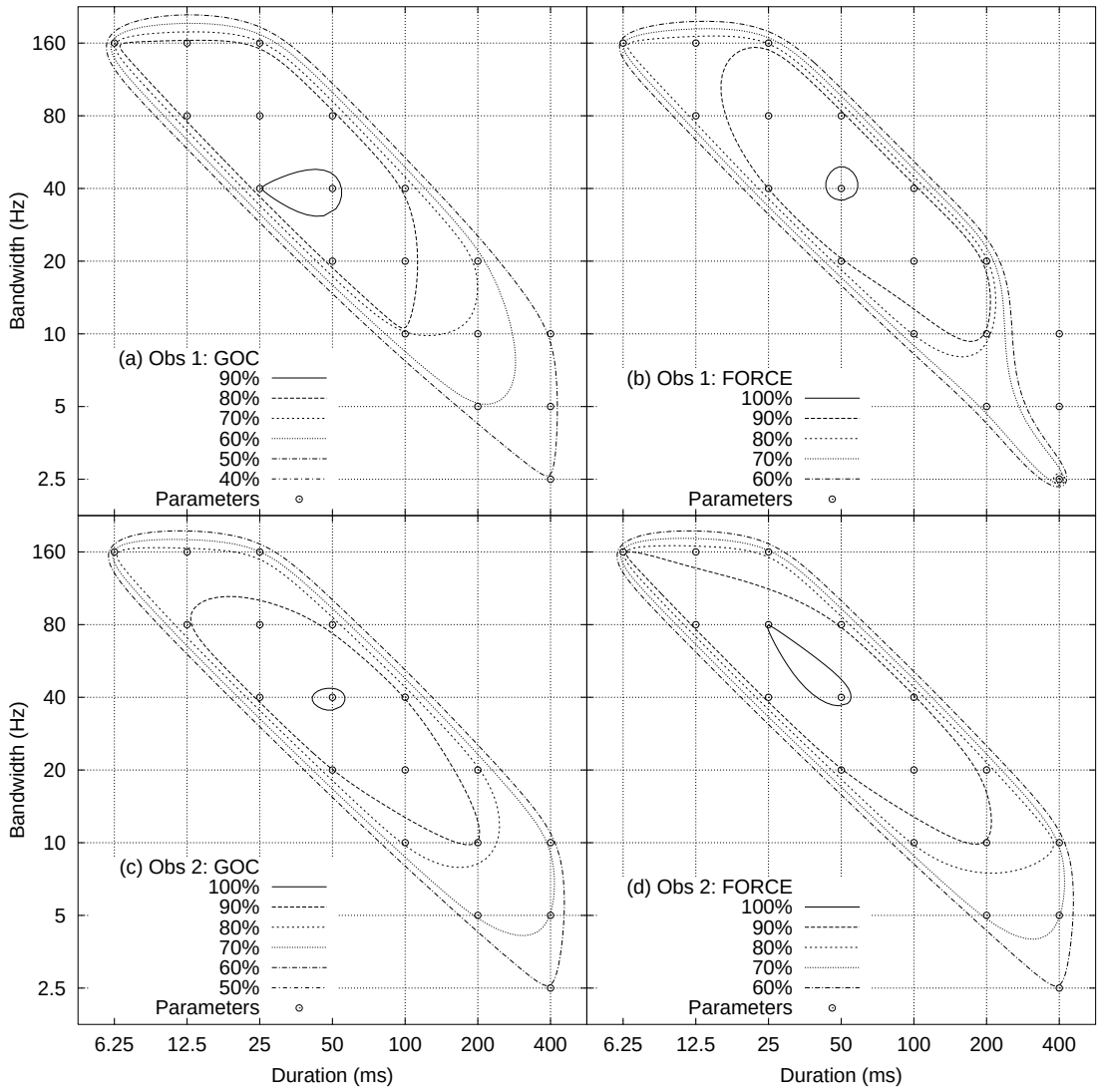

Figure 6.4: Contour maps in $\mathcal{WT}$–space of the proportion of $\mathcal{D}_6$ (in percent), relative to ideal full–linear detector simulation, after GOC analyses, for Observer 1 (top panel) and Observer 2 (bottom panel). Left hand panel is GOC performance and right hand panel is asymptotic FORCE performance. Each contour represents a proportion of $\mathcal{D}_6$. Parameters are the original signal bandwidths and durations.

Chapter 7

Discussion

"...luckily Owl kept his head and told us that the Opposite to an Introduction, my dear Pooh, was a Contradiction; and, as he is very good at long words, I am sure that that's what it is."

A. A. Milne.

.....

Chapter 1 showed that the models derived from TSD all incorporate the bandwidth–duration product $\mathcal{WT}$ as a fundamental parameter. Detectability of a waveform is, therefore, dependent on the product, but not on the individual values of $\mathcal{W}$ and $\mathcal{T}$. This result is inconsistent with human experience and the human data in the literature. The data reported in the literature, however, are also ambiguous and contradictory. There is some evidence that human hearing processes are fixed, and other evidence that they are flexible. Also, the mathematical models, which are used to compare and model human hearing, have unrealistic assumptions, flaws, and simplifications.

A crucial problem with psychophysical experiments on human hearing is the level of observer inconsistency. This error can change the *apparent* form of the underlying evidence distributions, which can lead to inappropriate conclusions. Chapter 2 discussed the problem of observer inconsistency and stated that this variability could be partitioned into unique and common noise. Two related methods of *removing* the effects of unique noise, GOC and FORCE analysis, were described, as were methods of estimating the amount of sampling variability in the common noise.

Chapter 3 described the experimental method and results. Briefly, two observers participated in an experiment where $\mathcal{W}$, $\mathcal{T}$, and $\mathcal{WT}$ were systematically varied. The experiment was replicated six times to allow for mROC, GOC, and FORCE analyses. Results from these analyses were compared with the attenuated χ^2 energy detector models, and it was found that humans were not optimally, or even consistently, matching $\mathcal{WT}$ as predicted by the χ^2 model. Data from a third observer, for three replications of eight (out of eighteen) experimental levels, supported the results from the other observers.

Chapter 4 focused on comparisons among computational simulations and mathematical models. There were differences found between the simulated and mathematical energy detector, because the signals and simulation components did not have ideal (impossible) properties. In comparing computer simulations (energy, full–linear, and envelope) it was found that the full–linear detector did perform at a higher level than the energy detector for $\mathcal{WT} > 1$—at least for

the types of signals used in the experiments. Because there was no mathematical theory for the full–linear detector, polynomial psychometric functions were fitted to the full–linear simulation data to be used as a data model. The human psychometric functions, with an attenuation factor, were slightly better fitted by the full–linear model, but the degree of attenuation was not constant across experimental level or condition.

The pattern of the attenuation of human performance from theory was partially explained, in Chapter 5, after running multitudes of suboptimal energy, full–linear, and envelope detector simulations, then correlating the simulation evidence with the sum–of–ratings from the human experiments. Firstly, it was found that a linear rectifier (full or half–wave) described the human data better than a square–law rectifier. By varying the filter bandwidths of simulated energy and full–linear detectors, it was found that not only did the detection bandwidth not match the signal bandwidth it was also not fixed. There was, however, a minimum value of around 40–50 Hz that was only obtainable with long duration signals.

Compared to the energy detector, the simulated envelope detectors were found to be slightly more correlated with human performance, when the time constant was very long, and the sampling time was approximately at the end of the input waveform—or just past the peak of the integrator. The full–linear detector, however, explained the data just as well as the envelope detector, but with fewer parameters. Comparisons of the best–correlated full–linear detector with human data and the fitted polynomial model showed striking similarities for many of the experimental levels, indicating that humans can perform as well as an ideal observer, after reducing unique noise.

Chapter 6 described a multiple–event analysis, run on both the human data and simulated full–linear detector using Scurfield's information theory measure of discriminability: $\mathcal{D}_n$. By comparing the relative discriminability of the human performance to the simulated performance, contour plots of discriminability were drawn for each observer. These plots showed that the hearing system detected signals of around 40 Hz and 50 ms the best, and performance was attenuated around these values. Therefore, the hearing system appears to be trading bandwidth and duration, but not entirely succeeding.

It could be argued, however, that the differences in the attenuation functions were due to differences in unique noise. That is, the observers were able to trade $\mathcal{W}$ and $\mathcal{T}$, but were more noisy in some levels, than others. Ignoring the rogue levels, FORCE analysis indicated that the differences within the attenuation functions were not due to more unique noise in some levels than others, because the analysis estimated performance without unique noise. It showed that the change in attenuation of the attenuation functions, after FORCE analysis, was relatively constant within a condition. Therefore, the overall shape of the attenuation function was not greatly changed.

7.1 Comparison with previous research

Comparing the current results with previous research, described in Chapter 1, shows support for the findings that the human hearing system was flexible, but not optimal.

Critical band estimates

Even if the auditory filter is flexible, the concept of the critical band still has meaning as the minimum bandwidth the filter can adopt. The results from the correlation analysis in Chapter 5 indicated that the detection bandwidths, for the two observers, were asymptotically approaching about 40–50 Hz (depending which bandwidth definition was used) as signal duration increased.

Patterson (1976) estimated the critical band (for a 500 Hz tone, with a 600 ms duration), to be 69.2 Hz (TdB). This condition was most similar to {400ms, 2.5Hz}, where the estimated TdB bandwidths were 55 Hz for Observer 1 and 49 Hz for Observer 2. These figures are very similar given the different type of signal, different observers, and different method of estimation.

The critical band estimates, for the current observers, are close to the predictions of 77 Hz, from the analytical formulae in Equations 1.8–1.10 by Moore and Glasberg (1983, 1987) and Moore et al. (1990). Zwicker and Terhardt's (1980) analytical formula for critical bandwidth, on the other hand, predicted 117 Hz (from Equation 1.7).

The current estimates are also roughly consistent with the smaller estimates of the critical bandwidth from Swets et al. (1962) and Kollmeier and Holube (1992).

Evidence for an adjustable auditory filter

The results from the correlation analysis in Chapter 5 indicated that the detection bandwidths for the two observers varied, depending on the bandwidth, duration, and $\mathcal{WT}$ of the signal. It appeared as if the hearing system was attempting to match the bandwidth of the signal, but tended to overestimate the optimal size. This is consistent with the research of Hamilton (1957), Majerník et al. (1978), Northern (1967) and van den Brink (1964), because they also found wider bandwidth estimates for short duration signals.

de Boer (1966) suggested that for wide–band signals that the critical band is 'turned–off' and that for short duration signals that the integrator is switched off too. Similarly, Scharf (1970) suggested that for short duration signals, detectability was independent of bandwidth, implying there was no filtering. The current data indicate that for wide–band signals, the detection bandwidth widens considerably, but it is still more highly correlated with a filtered waveform than with no filter at all.

Schacknow and Raab (1976) extended Green's (1960a) noise–in–noise theory to account for the case where the critical band was wider than the signal bandwidth, but only for large $\mathcal{WT}$, thus the current results cannot be analysed with Schacknow and Raab's (1976) model. They did not have enough data points to allow for accurate estimation of the width of the critical band, but it appeared to be wider than other estimates. For a constant duration of 100 ms, and various wide bandwidths, the critical band was estimated to be somewhere between 100 and 1000 Hz. This is consistent with the very wide detection bandwidths found in the wide bandwidth conditions (160 Hz) in the current experiment. It is inconsistent with the results from the 100 ms duration levels, however, because those levels resulted in reasonably narrow detection bandwidth estimates. This perhaps indicated that Schacknow and Raab's (1976) observers were influenced more by bandwidth, than by duration, when detecting the wide band signals.

Bernstein and Raab (1990) also found evidence that the critical band was not fixed, and that it was often wider than the signal bandwidth. Their signal parameters were closest to {100ms, 10Hz}. This level, in the current experiment, resulted in detection bandwidth estimates of around 55 and 71 Hz (TdB). Bernstein and Raab (1990) estimates, however, were around 700 Hz. The

centre frequency of the tone was 1000 Hz so direct comparison may not be justified.

Another similar condition from Bernstein and Raab (1990) was for $\mathcal{WT} \approx 5$, where the bandwidth was 50 Hz and the duration around 100 ms. The centre frequency was much higher, however, at 1500 Hz. Their detection bandwidth estimates were 314 and 659 Hz, which can be compared to 63, 80, and 90 Hz (TdB) from the experimental level {100ms, 40Hz}. One possible reason, aside from the differences between the signals, is that their estimates were inflated due to the presence of unique noise. It is difficult to make comparisons because as the current research shows, different observers can have quite different detection bandwidths.

The results are also consistent with Wright and Dai (1994a, 1994b). They found that the detection bandwidth varied, depending on when and where observers *expected* time and frequency information. They also showed that different observers sometimes solved detection tasks differently.

Detection bandwidth versus the critical band

The current study was unable to distinguish whether observers were adjusting the bandwidth of the auditory filter, or summing across critical bands (Spiegel, 1979), because both theories predict the same results for this type of experiment. The results were consistent, however, with the concept of a detection bandwidth (Bernstein & Raab, 1990), or an attention bandwidth (Dai, 1991; in Moore et al., 1996). These detection bandwidths may be the result of the observer combining information across critical bandwidths or adjusting the bandwidth for a filter at the centre frequency. The term "detection bandwidth" is preferable to "attention bandwidth", because the latter implies other cognitive processes.

Rectification

The simulations in Chapter 4 showed that a linear rectifier outputs a higher signal–to–noise ratio than a square–law rectifier, for the values of $\mathcal{WT}$ tested. The simulations in Chapter 5 clearly indicated that the human observers were also behaving as if they were using a linear, rather than a square–law, rectifier. These results are consistent with Gilkey (1981) and Gilkey and Robinson (1986), and with the physiological studies reported by Jeffress (1964) and Laming (1986).

Pfafflin and Mathews (1962) thought that any differences due to the rectifier would be too small to be shown experimentally. What the current study has shown is if unique noise is removed, and $\mathcal{WT} > 1$, then the full–linear detector fits the human psychometric functions better than the square–law detector.

Integration time and time constant

The correlation analysis (Chapter 5) results showed higher correlations between human ratings and simulated observers, when a true integrator, or a leaky integrator with a very long time constant, was used. This is consistent with the findings of Makita and Miyatani (1950) and Henning and Psotka (1969), but not consistent with studies supporting leaky integration with a short time constant (Jeffress, 1967, 1968; Robinson & Pollack, 1973; Robinson, 1974; Ronken, 1973).

Although the observers' performance was attenuated for long durations, the correlation analysis did not indicate that the hearing system was integrating over too short a time, or with too short a time constant. In most cases, the correlations with the leaky integrator simulations showed

increasing correlations as the time constant increased for all experimental levels, and the best–correlated sampling time coincided with approximately the end of the waveform. This point coincided with either the (nearly) saturated value of the true integrator, or just past the peak output of a leaky integrator.

Previous researchers have found performance was attenuated when using very long stimuli (Green et al., 1957). It was difficult to tell, however, whether this drop was caused by observer fatigue, or an inability to maintain integration. The indications are that observers can integrate for very long times, but that unique noise may also increase, because it is more difficult to concentrate on the task.

The current results are somewhat consistent with Gerken et al. (1990), who found evidence against the exponential temporal integrator. They also suggested that there is quite possibly cognitive control over temporal integration, and thus methodology may interact with detectability. The results are also consistent with the 'multi–look' model of Viemeister and Wakefield (1991), but like Dai and Wright (1995), it is difficult to distinguish among different models, because they all predict similar outcomes for the type of experiment used in the current study.

Sampling strategies

The results from the correlation analysis (Chapter 5) suggested that humans sampled the output of the integrator at a particular time, rather than sampling the peak output. The best–correlated sampling time tended to coincide with a time just past the *average* peak output of a leaky integrator, or a time just before the true integrator was saturated. For the types of signals used in the current study, this time also coincided with approximately the end of the absolute duration of the original signal. This finding is consistent with Gilkey (1981) and Gilkey and Robinson (1986).

Nichols and Jeffress (1966) also found that peak detection did not describe human performance. Instead they found average voltage a better predictor. Gaston and Jeffress (1974), on the other hand, found that the peak envelope detector was the best predictor of human performance for long duration signals. They used only one decay time (50 ms), however, for their envelope detector. The correlation analysis indicated that that humans do not use such a short time constant, and the performance is much better described by time constants an order of magnitude larger. The difference between peak and ATT sampling is then minimal.

In the current study, the Kaiser window was used, which forced the peak to occur in the middle of the waveform, on average. This process possibly changes the statistics of peak detection. Because the mathematics of peak detection are difficult, simulations were used instead. They showed that for $\mathcal{WT} > 1$, the peak–envelope detector behaved similarly to the energy detector, rather than the full–linear or envelope (with ATT sampling) detectors.

Attenuation from theory

Asymptotic performance was predicted, using GOC and FORCE analysis, then compared to ideal performance. One comparison was to measure the attenuation of the human psychometric functions from the ideal functions (both mathematical and simulated). This analysis showed that GOC and FORCE analyses reduced attenuation. In some levels, the attenuation was so small as to be essentially due to sampling variability; in some cases it was slightly negative.

Comparing the $\mathcal{WT}=1$ data to Ronken (1969), the current mROC attenuation for the comparable condition is 6 dB, which is better than Ronken's result by about 1.5 dB. After FORCE analysis, the attenuation is about 2.8 dB from the theoretical. It is difficult to compare the shapes of the psychometric function because Ronken chose to present his data as functions of decibel increments.

Hautus and Irwin (1992) replicated and compared Ronken (1969) and Whitmore et al. (1968) and found the attenuation to be around 4 dB from the theoretical, and that the attenuation was about the same for each experiment, within an observer. The reason Ronken and Whitmore et al. found a difference may, therefore, have been due to individual differences in the observers. As was seen in the correlation analysis in Chapter 5, there could be large differences in the estimated critical band—particularly for short duration signals. It should be noted that Whitmore et al. used very short duration signals (10 ms) and Ronken used reasonably long duration signals (100 ms). It may have been the case that the short duration signals resulted in an extremely wide bandwidth that would have passed more noise and therefore increased attenuation. For Ronken's longer duration signals, the observers may have found it easier to more closely match the bandwidth. Finally it should be noted that Whitmore et al.'s study used the same observer as in the present study (Observer 3). This observer's detection bandwidth has been shown here, and in a number of other experiments, to be particularly wide and rather variable over replications.

Simulated detectors best correlated with human observers

Ahumada found that correlations for $\mathcal{SN}$ signals were higher for narrower filters than the correlations for $\mathcal{N}$ signals (Ahumada, 1967; Ahumada & Lovell, 1969, 1971; Ahumada et al., 1975). Ahumada interpreted this as suggesting that the observers monitored the output of a filter–bank. In $\mathcal{SN}$ trials, the signal would always occur in one filter, therefore, the correlations would match a narrow filter. In $\mathcal{N}$ trials, the energy in the signal could be distributed over a wider bandwidth; therefore, the best correlations would be for a filter covering the width of the filter bank.

The correlations, with respect to filter width, in the current study indicated a similar phenomenon, and were not inconsistent with a filter bank model. The number of filter widths tested, however, were greater than Ahumada, and indicated than the largest $\mathcal{SN}$ signals showed similar correlations for a wide range of filter bandwidths. As signal–to–noise ratio decreased, the correlation–bandwidth function tended to converge to a function with a single peak.

The process of making ratings, or decisions, about the existence of a signal masked by noise can be seen as a *sorting* or *ordering* exercise. The correlation analysis measures the extent to which two observers *sort* the stimuli into the same order. If the observers sort the stimuli into exactly the same order, then it is impossible to tell the difference between them, because the same ROC curves will result.

The $\mathcal{SN}$ signals, with large signal–to–noise ratios, result in very similar correlations across a wide range of filter bandwidths because filtering does not change the order in which they are sorted. The signals are so big, compared to the noise, that even if the filter is much wider than the signal, relatively little noise is passed, therefore, the sorting order is not greatly affected. As signal–to–noise ratio decreases, the more the wide–band random noise affects the effective signal–to–noise ratio, as the filter bandwidth is varied. The result is that large $\mathcal{SN}$ trials cannot be reliably used to estimate the detection bandwidth. Instead, it is only the $\mathcal{N}$ trials that will show the detection bandwidth. This does not rule out Ahumada's filter bank theory, but merely shows that it cannot be assessed with this particular method. Ahumada has based his theory on data that can

be explained a simpler way, that is, a fixed bandwidth model.

The findings from the correlation analysis in Chapter 5 are similar to that of Gilkey (1981) and Gilkey and Robinson (1986) even though they studied tone–in–noise detection. Firstly, he found that half–wave rectification resulted in higher correlations with the human data than square–law rectification. Gilkey found two detectors that gave better results for some observers than any of the other standard detectors. The full–linear detector is similar to Gilkey's best detector, which computed the average value of the integrator output (the time constant of the integrator was very short making the output essentially the same as the rectifier) over the signal duration of the half–wave rectified waveform. The difference is that the full–linear detector summed the integrator output of a full–wave rectifier, whereas Gilkey essentially implements a true integrator by using an averager as a sampling strategy, thereby making the two models virtually identical.

Evidence for matching bandwidth and duration

The results showed that, in general, the human observers did not match the signal parameters. The exception was for Observer 2, who crossed the equal bandwidth lines in Figure 5.5. One level fell on this line, {100ms, 40Hz}, and levels {25ms, 80Hz} and {25ms, 160Hz} were very close. This implies there are particular combinations of $\mathcal{W}$ and $\mathcal{T}$ that could result in ideal detection by humans, once unique noise was removed. Observer 1 also neared the equal bandwidth line for level {25ms, 160Hz}.

It is difficult to compare the current results with Green (1960a), because his signals were all for larger $\mathcal{WT}$ noises. He showed that the human hearing system appeared to be able to match bandwidth and duration, and perform consistently with his energy detection model, except for an attenuation factor. Most of this attenuation was removed in the current study, through GOC and FORCE analyses, showing that humans were better modelled by a full–linear detector.

Green's (1960a) finding was typical of most human experiments: human performance never matched that of the ideal observer, because data were always degraded by unique noise. This has caused many to question the usefulness of the concept of the ideal observer. The justification given in defence was that an ideal observer gave the limits of detectability, and served as a baseline for comparison. The finding that, under some circumstances, observers can match the parameters of the stimuli, is interesting, because it suggests that the concept of the ideal observer is not just good as a baseline, but is also useful as a model of actual human hearing.

Bandwidth–duration reciprocity

The energy and envelope detectors of Green (1960a), Green and McGill (1970), Jeffress (1968), McGill (1967, 1968b), McGill and Teich (1991) and Whitmore (1969) all predict bandwidth–duration reciprocity. The results of the correlation analysis, and the multiple–event analysis, indicate that human observers do not show ideal reciprocity. Instead, the contour plots in Chapter 6 suggest there are regions of equal detectability, indicating reciprocity, but it is attenuated relative to the simulated full–linear detector. Both bandwidth and duration information is used, but the two observers solved the trade–off differently, perhaps with different weightings for the information in each domain. It seems the hearing system tries to trade bandwidth and duration, to optimise detection, but does not, in general, succeed. Performance falls between a detector with fixed parameters and a detector that matches parameters.

Michaels (1961) found that detectability decreased with decreasing duration for $\mathcal{WT}$=1 and 4

whereas the current results showed the opposite effect. It was not clear in Michael's experiment, however, that the signals had the desired parameters, and observer inconsistency was also high.

The data reported by Creelman (1961) give some support for the current results for the condition $\mathcal{WT}=1$ and $\mathcal{WT}=2$, because he too found that performance dropped at longer durations. Performance also decreased with increasing $\mathcal{WT}$ that may or may not be paradoxical depending on whether his signals can be considered more like tones, or more like noise.

Unlike Whitmore (1969), the shape of the psychometric functions did change with varying duration and fixed bandwidth. The condition of $\mathcal{WT}=1/2$ was not tested but $\mathcal{WT}=1$ and 2 were in the current experiment. Whitmore (1998, personal communication) said that no frequency domain analysis was done to test that the signal generation method produced signals with the required parameters. It was assumed that the bandwidth was the same for each condition, and that the duration did not change after filtering. Simulations reported in Chapter 1 indicated that Whitmore probably could not have generated signals with the stated, nominal bandwidths and durations.

Unlike the current study, which showed sub-optimal bandwidth–duration reciprocity, Raab and Goldberg (1975) found no bandwidth–duration reciprocity for the conditions they studied (although only two conditions allowed this comparison). Raab and Goldberg (1975) also found that bandwidth and duration were not treated symmetrically by the auditory system, with results being largely independent of bandwidth, but dependent on duration.

Hanna (1984) found that performance increased as duration increased up to 25 ms and that increasing bandwidth increased performance to a lesser extent. It is hard to compare his results, however, because there were problems with his definitions of bandwidth and duration.

van den Brink and Houtgast (1990) showed in their first experiment that as bandwidth or duration increased, detectability decreased. Their signals were quite different to the current signals, they used a different experimental procedure, and they tested at a higher frequency, but their results are roughly similar to the current findings. The exception is for their second experiment, where they showed increasing thresholds for short–duration, wide–band, signals, whereas the attenuation functions in Appendix F show the biggest attenuations for long–duration, narrow–band signals.

van den Brink and Houtgast (1990) tested shorter durations, and wider bandwidths, than the current study. Their results indicated that for wide band signals, the critical masking interval equated to about 80 ms. Converting their estimates of maximum efficiency to the 500 Hz region of the current experiment implied that the optimum bandwidth should be 9 Hz, and the optimum duration to be about 80 ms. The product is approximately $\mathcal{WT}=1$ and is similar to the condition $\{100\text{ms}, 10\text{Hz}\}$. This condition does not produce optimum detectability for $\mathcal{WT}=1$, nor is $\mathcal{WT}=1$ better than larger $\mathcal{WT}$.

Formby et al. (1994) conducted a study of $\mathcal{WT}$, similar to the current experiment, but at a higher test frequency (2500 Hz). Their study was flawed, however, because they did not provide important details, such as how they defined bandwidth, or $\mathcal{WT}$. It is likely that they measured bandwidth prior to gating. Their results were partly inconsistent with the current results. They found that performance was attenuated as duration increased, but they also found that performance was attenuated as bandwidth increased. In comparison, the current study indicated that observers were able to accommodate a widening bandwidth, albeit sub-optimally. The difference between the two studies probably hinges on how bandwidth was defined and measured.

7.2 Analysis of the new analyses

This project involved a number of analyses, including some new ones. GOC analysis, FORCE analysis, attenuation analysis, Bester correlation method analysis, and multiple–event analysis all provided different ways of viewing the data, but also, they all had limitations.

GOC analysis

GOC analysis proved its worth by removing unique noise to reveal the shape of the error–reduced experimental psychometric functions and ROC curves.

The mROC curves in Appendix D show a similar phenomenon to Whitmore et al. (1968), where they look less asymmetrical than would be expected from the χ^2 energy detector model. The psychometric functions, however, are better fitted by the model—also like Whitmore et al. (1968). After GOC analysis, however, the GOC curves are more asymmetrical than the mROC curves, and are better fitted by the theory. The difference in the psychometric functions, however, is mainly a linear attenuation, rather than a change in the shape of the function.

The improvement in average attenuation from mROC to GOC analysis (see Tables F.1–F.6) ranged from 1.8–3 dB, after six replications, with slightly more improvement at longer durations. This should not be seen as a general prediction, because the improvement is dependent on the level of unique noise.

FORCE analysis

FORCE analysis provided estimations of asymptotic performance, for both the usual two–event measures of detectability and the new multiple–event measure $\mathcal{D}_n$ for six events. When FORCE analysis converged quickly, it provided reasonable estimates of asymptotic performance. In some experimental levels, however, there was still too much unique noise in the data for sensible estimates to emerge.

The problem of the signal levels being too low in some experimental levels was considered before the experiments were run. At the time, it was considered more important to have the same signal–to–noise ratios for each level, in each condition, so direct comparisons could be made. It would have also been presupposing the outcome to have run higher signal–to–noise ratios for the levels in which the observers were doing badly. As a result, there appears to be more unique noise for the long duration signals, making the six replication FORCE analysis less reliable. It is unclear how many more replications would be needed to reduce the unique noise in these levels.

Another reason why the long duration conditions could be problematic is due to the common noise. Drga (1998, personal communication) has suggested that long durations may result in an observer using information from different parts of the waveform, on successive presentations of the same stimulus. Therefore, there may be no common component across replications. This, of course, would not only be a problem in the time domain; a wide bandwidth signal may produce the same phenomenon in the frequency domain. There is little evidence, however, to suggest that there was no common noise in these conditions.

FORCE analysis, in general, resulted in a 0.5–2 dB improvement over the GOC results, indicating that the GOC analysis had removed most of the unique noise even after only six replications. The improvement was greater, however, for the lower signal–to–noise ratios, indicating that the amount of unique noise is dependent on signal level. It may also be an artifact due to mea-

surement error in the shallow regions of the psychometric functions. FORCE analysis showed, however, that observers were able to perform ideally under some circumstances.

Bester correlation method analysis

The Bester correlation method analysis played an integral part in this study, because it helped reveal the nature of the different processes making up human hearing. By carefully reducing the possibility of confounding, the energy detector was shown to not be a good model of human hearing, compared with the full–linear or envelope detector. This analysis also showed that some observer inconsistency may come from a detection bandwidth that varies over time.

Multiple–event analysis

Multiple–event analysis was shown to provide a useful measure of overall performance, independent of any theory about the detectability of the signals. As noted, however, it does not generalise to arbitrary signal–to–noise ratio in the same way as attenuation analysis. Therefore, it is recommended that in this type of experiment, both analyses are used, because they make complementary contributions.

The multiple–event measure of discriminability, $\mathcal{D}_n$, for six events, was also shown to work with FORCE analysis to give good estimates of asymptotic performance. The amount of number–crunching involved with this particular analysis is impressive, because both procedures require calculating large combinatorics. It took only a few minutes on a Dec Alpha (Personal Workstation, 333 MHz clock speed), however, to run the analysis on the entire data set.

7.3 The rôle of $\mathcal{WT}$

The human hearing system cannot be modelled as a constant $\mathcal{WT}$ system. The results from the experiments and simulations indicate that for a set of signals with the same $\mathcal{WT}$, but different bandwidths and durations, the human hearing system does not use a constant $\mathcal{WT}$ to detect those signals, even though such a strategy would be optimal. The hearing system cannot be modelled as a fixed parameter system either, because (a) it appears to integrate over duration with apparently little error, and (b) it is able to vary the detection bandwidth. Instead, the hearing system appears to use both time and frequency information to fit the detection bandwidth to the signal. It does not succeed, in general, in making an optimal match, and it is also limited to a minimum bandwidth of around 40–50 Hz. The reason the hearing system does not match bandwidth and duration exactly, even though it is flexible, maybe because it errs on the side of caution due to uncertainty about the edges of the signal in time and frequency.

The evidence from the current experiments quite clearly indicates that the human hearing system uses linear rectification. This finding eliminates the energy detector as a model for the detectability of small $\mathcal{WT}$ Gaussian noise in human hearing. It was impossible to tell, however, which form of linear rectification was more appropriate (full–wave or half–wave), because both forms of rectification led to essentially the same sorting of the stimuli.

$\mathcal{WT}$ was also shown to decrease as signal–to–noise ratio increased, for physical and computational signals with the same bandwidth and duration. The human data were also consistent with this phenomenon. Any new mathematical model of detectability of small $\mathcal{WT}$ signals must take this into account.

7.4 Future directions

On reflection, this experiment did not treat the time domain equally with the frequency domain. Either the gated masker should have been continuous—or at least longer than the signal duration—or the maskers should have had the same bandwidth as the signals. If the signal was embedded in a masker, which was wider in time and in frequency, then the uncertainty as to the position of the signal would be equivalent in both domains. A better understanding of temporal integration may result by studying how the hearing system attempts to detect a non–existent signal in a $\mathcal{N}$ trial.

Future experiments could include:

- extending the current parameters to larger $\mathcal{WT}$,
- isolating the parameters that result in the best detectability to find out if the hearing system is trading $\mathcal{W}$ and $\mathcal{T}$ ideally, but over a much smaller range than was considered, and
- mapping out $\mathcal{WT}$ space for other centre frequencies.

Additional simulations could include:

- implementing some of the recent, more complex, models of human hearing, such as those suggested by Dau, Puschel, and Kohlrausch (1996a, 1996b) and Richards (1992),
- implementing more realistic filters, such as the gamma–chirp, gamma–tone, and roex (Irino & Patterson, 1997; Patterson et al., 1982),
- extending the simulations to do correlation analysis in a similar direction to Ahumada et al. (1975), but perhaps using more sophisticated analyses, such as wavelet decomposition or instantaneous time–frequency analysis.

Further theoretical work could include:

- deriving the detection statistics for the full–linear detector,
- deriving energy, full–linear, and envelope detectors for multiple–events, and
- incorporating a $\mathcal{WT}$ parameter that varies with signal–to–noise ratio.

References

Ahumada, A., Jr., Marken, R., & Sandusky, A. (1975). Time and frequency analyses of auditory signal detection. *Journal of the Acoustical Society of America*, *57*(2), 385-390.

Ahumada, A. J., Jr. (1967). *Detection of tones masked by noise: A comparison of human observers with digital–computer–simulated energy detectors of varying bandwidths.* Unpublished doctoral dissertation, University of California, Los Angeles.

Ahumada, A. J., Jr., & Lovell, J. (1969). Bandwidth estimates in human signal detection. *Journal of the Acoustical Society of America*, *45*, 327(A).

Ahumada, A. J., Jr., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, *49*(6(2)), 1751-1756.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, *12*, 387-415.

Barbosa, L. C. (1986). A maximum–energy–concentration spectral window. *IBM Journal of Research and Development*, *30*(3), 321-325.

Bendat, J. S., & Piersol, A. G. (1986). *Random data: Analysis and measurement procedures* (Second ed.). NY: John Wiley and Sons.

Bernstein, R. S., & Raab, D. H. (1990). The effects of bandwidth on the detectability of narrow and wideband signals. *Journal of the Acoustical Society of America*, *88*(5), 2115-2125.

Birdsall, T. G. (1960). Detection of a signal specified exactly with a noisy stored reference signal. *Journal of the Acoustical Society of America*, *32*(8), 1038-1045.

Boring, E. G. (1957). *A history of experimental psychology* (Second ed.). NY: Appleton–Century–Crofts.

Borowski, E. J., & Borwein, J. M. (1989). *Dictionary of mathematics.* Glasgow: Collins.

Bourret, R. (1958). A note on an information theoretic form of the uncertainty principle. *Information and Control*, *1*(4), 398-401.

Boven, R. (1976). *The use of multiple observers in signal detection theory: A method to remove the effect of unique noise from experimental data.* Unpublished master's thesis, Victoria University of Wellington, New Zealand.

Bowman, K. O., & Shenton, L. R. (1988). *Properties of estimators for the gamma distribution.* NY: Marcel Dekker.

Brigham, E. O. (1974). *The fast Fourier transform.* Prentice–Hall.

Brignell, J., & Rhodes, G. (1975). *Laboratory on–line computing.* International Textbook Company Limited.

Brillouin, L. (1962). *Science and information theory* (Second ed.). Academic Press.

Chalk, J. H. H. (1950). The optimum pulse–shape for pulse communication. *Proceedings of the IEE*, *87*, 88-92.

Cohen, L. (1989). Time–frequency distributions—a review. *Proceedings of the IEEE, 77*(7), 941-981.

Copeland, A. B., & Mowry, H. J., III. (1971). Real–ear attenuation characteristics of selected noise–excluding audiometric receiver enclosures. *Journal of the Acoustical Society of America, 49*(6(2)), 1757-1761.

Corliss, E. L. R. (1963). Resolution limits of analyzers and oscillatory systems. *Journal of Research of the National Bureau of Standards–A. Physics and Chemistry, 67A*(5), 461-474.

Corliss, E. L. R. (1967). Mechanistic aspects of hearing. *Journal of the Acoustical Society of America, 41*(6), 1500-1516.

Creelman, C. D. (1961). Detection of complex signals as a function of signal bandwidth and duration. *Journal of the Acoustical Society of America, 33*(1), 89-94.

Creelman, C. D. (1965). Discriminability and scaling of linear extent. *Journal of Experimental Psychology, 70*(2), 192-200.

Dai, H., & Green, D. M. (1993). Discrimination of spectral shape as a function of stimulus duration. *Journal of the Acoustical Society of America, 93*(2), 957-965.

Dai, H., & Wright, B. A. (1995). Detecting signals of unexpected or uncertain durations. *Journal of the Acoustical Society of America, 98*, 798-806.

Dai, H., & Wright, B. A. (1996). The lack of frequency dependence of thresholds for short tones in continuous broadband noise. *The Journal of the Acoustical Society of America, 100*(1), 467-472.

Dau, T., Puschel, D., & Kohlrausch, A. (1996a). A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements. *The Journal of the Acoustical Society of America, 99*(6), 3623-3631.

Dau, T., Puschel, D., & Kohlrausch, A. (1996b). A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. *The Journal of the Acoustical Society of America, 99*(6), 3615-3622.

Davenport, W. B., Jr., & Root, W. L. (1958). *An introduction to the theory of random signals and noise.* NY: McGraw–Hill Book Company.

de Boer, E. (1962). Note on the critical bandwidth. *Journal of the Acoustical Society of America, 34*, 985-986.

de Boer, E. (1966). Intensity discrimination of fluctuating signals. *Journal of the Acoustical Society of America, 40*(3), 552-560.

de Boer, E. (1985). Auditory time constants: A paradox? In A. Michelsen (Ed.), *Time resolution in auditory systems. Proceedings of the 11th Danavox Symposium on hearing. Gamle Avernaes, Denmark, August 28–31, 1984* (p. 141-158). Springer-Verlag.

de Bruijn, N. G. (1967). Uncertainty principles in Fourier analysis. In O. Shisha (Ed.), *Inequalities* (p. 57-71). NY: Academic Press.

de Lano, R. H. (1949). Signal–to–noise ratios of linear detectors. *Proceedings of the IRE, 37*(10), 1120-1126.

Dorfman, D. D., & Berbaum, K. S. (1986). RSCORE–J: Pooled rating–method data: A computer program for analyzing pooled ROC curves. *Behavior Research Methods, Instruments, and Computers, 18*(5), 452-462.

Drga, V. (1988). *2IFC ROC analysis of narrow–band noise detection.* Unpublished honours thesis, Victoria University of Wellington, New Zealand.

Drga, V. (1993a, August). *Theory of group operating characteristic (GOC) analysis.* Paper presented at the Psychophysics Symposium, New Zealand Psychological Society Annual Conference, Wellington, New Zealand.

Drga, V. (1993b, August). *Transform-averaged group operating characteristic (GOC) analysis.* Paper presented at the Psychophysics Symposium, New Zealand Psychological Society Annual Conference, Wellington, New Zealand.

Drga, V. (1995). *Estimating errorless performance in discrimination tasks.* Department of Psychology Research Seminar, Victoria University of Wellington, New Zealand, 27 Oct.

Drga, V. (1997a). ACA*: A FORCE analysis program* [Computer software in BORLAND PASCAL 7.0, for MS–DOS.].

Drga, V. (1997b). SIGGEN*: signal generation program, using the inverse FFT, for bandlimited Gaussian noise* [Computer software in BORLAND PASCAL 7.0, for MS–DOS.].

Drga, V. (1998). SIGGEN*: signal generation program, using the inverse FFT, for bandlimited Gaussian noise* (Tech. Rep. No. 3). Victoria University of Wellington, Wellington, New Zealand: Psychophysics Laboratory.

Drga, V. (1999). *Theory of group operating characteristic analysis in discrimination tasks.* Unpublished doctoral dissertation, Victoria University of Wellington, Wellington, New Zealand.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, *46*, 372-383.

Egan, J. P. (1975). *Signal detection theory and ROC analysis.* NY: Academic Press.

Egan, J. P., & Clarke, F. R. (1966). Psychophysics and signal detection. In J. B. Sidowski (Ed.), *Experimental methods and instrumentation in psychology.* McGraw Hill Book Company.

Egan, J. P., Schulman, A. I., & Greenberg, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: contemporary readings* (1964 ed.). NY: John Wiley & Sons.

Elliot, P. B. (1964). Tables of d'. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: contemporary readings* (1964 ed.). NY: John Wiley & Sons.

Emmerich, D. S. (1968). Receiver–operating characteristics determined under several interaural conditions of listening. *Journal of the Acoustical Society of America*, *43*(2), 298-307.

Falmagne, J. C. (1985). *Elements of psychophysical theory.* Oxford: Oxford University Press.

Fisher, A. J. (1996). *Interactive filter design: Mkfilter (version 2).* World Wide Web page at http://dcpu1.cs.york.ac.uk:6666/fisher/mkfilter.

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, *12*, 47-65.

Fletcher, H. (1953). *Speech and hearing in communication.* Princeton, NJ: D. van Nostrand. (Reprinted 1958.)

Formby, C., Heinz, M. G., Luna, C. E., & Shaheen, M. K. (1994). Masked detection thresholds and temporal integration for noise band signals. *Journal of the Acoustical Society of America*, *96*(1), 102-114.

Franks, L. E. (1969). *Signal theory.* Englewood Cliffs, NJ: Prentice–Hall.

Friedman, M. P., & Carterette, E. C. (1964). Detection of Markovian sequences of signals. *Journal of the Acoustical Society of America*, *36*(12), 2334-2339.

Gabor, D. (1946). Theory of communication. *Journal of the Institution of Electrical Engineers*, *93*, 429-457.

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (1999). *Extending the theory of signal detectability to discrimination between correct and incorrect decisions.* (Manuscript submitted for publication.)

Galvin, S. J., & Whitmore, J. K. (1986). *The equivalent rectangular duration of the Kaiser data window.* (Psychophysics Lab, Victoria University of Wellington, New Zealand.)

Gambardella, G., & Trautteur, G. (1966). Time–frequency analysis in the hearing process. *The Journal of the Acoustical Society of America*, *40*(5), 1187-1189.

Garner, W. R. (1947). The effect of frequency spectrum on temporal integration of energy in the ear. *Journal of the Acoustical Society of America*, *19*(5), 808-815.

Garner, W. R., & Miller, G. A. (1947). The masked threshold of pure tones as a function of duration. *Journal of Experimental Psychology*, *37*, 293-305.

Gaston, A. D., & Jeffress, L. A. (1974). Effect of sampling procedure upon the performance of an electrical model of auditory detection. *Journal of the Acoustical Society of America*, *56*(6), 1815-1817.

Geckinli, N. C., & Yavuz, D. (1978). Some novel windows and a concise tutorial comparison of window families. *IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-26*(6), 501-507.

Gerken, G. M., Bhat, V. K. H., & Hutchison-Clutter, M. (1990). Auditory temporal integration and the power function model. *Journal of the Acoustical Society of America*, *88*(2), 767-778.

Gilkey, R. H. (1981). *Molecular psychophysics and models of auditory signal detectability.* Unpublished doctoral dissertation, Indiana University, Bloomington, IN.

Gilkey, R. H., & Robinson, D. E. (1986). Models of auditory masking: A molecular psychophysical approach. *Journal of the Acoustical Society of America*, *79*(5), 1499-1510.

Gilkey, R. H., Robinson, D. E., & Hanna, T. E. (1985). Effects of masker waveform and signal–to–masker relation on diotic and dichotic masking by reproducible noise. *Journal of the Acoustical Society of America*, *78*, 1207-1219.

Goldman, S. (1953). *Information theory.* NY: Prentice–Hall.

Green, D. M. (1960a). Auditory detection of a noise signal. *Journal of the Acoustical Society of America*, *32*(1), 121-131.

Green, D. M. (1960b). Psychoacoustics and detection theory. *Journal of the Acoustical Society of America*, *32*(10), 1189-1203.

Green, D. M. (1964a). Consistency of auditory detection judgments. *Psychological Review*, *71*(5), 392-407.

Green, D. M. (1964b). General prediction relating yes–no and forced–choice results. *Journal of the Acoustical Society of America*, *36*, 1042(A).

Green, D. M. (1969). Masking with continuous and pulsed sinusoids. *Journal of the Acoustical Society of America*, *46*(4(2)), 939-946.

Green, D. M. (1973). Minimum integration time. In A. R. Moller & P. Boston (Eds.), *Basic mechanisms in hearing. Proceedings of the First Royal Swedish Academy of Sciences Symposium held October 30–November 1, 1972 at the Academy, Stockholm, Sweden* (p. 829-846). NY and London: Academic Press.

Green, D. M. (1985). Temporal factors in psychoacoustics. In A. Michelson (Ed.), *Time resolution in auditory systems. Proceedings of the 11th Danavox Symposium on hearing. Gamle Avernaes, Denmark, August 28–31, 1984* (p. 122-140). Berlin: Springer-Verlag.

Green, D. M. (1988). *Profile analysis: Auditory intensity discrimination.* NY: Oxford University Press.

Green, D. M., Birdsall, T. G., & Tanner, W. P., Jr. (1957). Signal detection as a function of signal intensity and duration. *Journal of the Acoustical Society of America*, *29*(4), 523-531.

Green, D. M., & Luce, R. D. (1974). Counting and timing mechanisms in auditory discrimination and reaction time. In D. H. Krantz, R. O. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary development in mathematical psychology* (p. 372-415). San Francisco: Freeman.

Green, D. M., & McGill, W. J. (1970). On the equivalence of detection probabilities and well-known statistical quantities. *Psychological Review*, *77*(4), 294-301.

Green, D. M., & Moses, F. L. (1966). On the equivalence of two recognition measures of short–term memory. *Psychological Bulletin*, *66*(3), 228-234.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* NY: Wiley. (Reprinted by Krieger, Huntingdon, NY, 1974.)

Greenwood, D. D. (1961). Auditory masking and the critical band. *Journal of the Acoustical Society of America*, *33*(4), 484-502.

Grenander, U., Pollak, H. O., & Slepian, D. (1959). The distribution of quadratic forms in normal variates: A small sample theory with applications to spectral analysis. *Journal of the Society for Industrial and Applied Mathematics*, *7*(4), 374-401.

Hamilton, P. M. (1957). Noise masked thresholds as a function of tonal duration and masking noise band width. *Journal of the Acoustical Society of America*, *29*(4), 506-511.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Diagnostic Radiology*, *143*(1), 29-36.

Hanna, T. E. (1984). Discrimination of reproducible noise as a function of bandwidth and duration. *Perception & Psychophysics*, *36*(5), 409-416.

Harmuth, H. F. (1968). A generalized concept of frequency and some applications. *IEEE Transactions on Information Theory*, *IT-14*(3), 375-382.

Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, *66*(1), 51-83.

Hautus, M. J., & Irwin, R. J. (1992). Amplitude discrimination of sinusoids and narrow-band noise with Rayleigh properties. *Perception & Psychophysics*, *52*(1), 53-62.

Helstrom, C. W. (1960). *Statistical theory of signal detection.* NY: Pergamon Press.

Henning, G. B. (1967). A model for auditory discrimination and detection. *Journal of the Acoustical Society of America*, *42*(6), 1325-1334.

Henning, G. B., & Psotka, J. (1969). Effect of duration on amplitude discrimination in noise. *Journal of the Acoustical Society of America*, *45*(4), 1008-1013.

Hilberg, W., & Rothe, P. G. (1971). The general uncertainty relation for real signals in communication theory. *Information and Control*, *18*, 103-125.

Hodos, W. (1970). Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin*, *74*(5), 351-354.

Houtgast, T. (1974). *Lateral suppression in hearing: a psychophysical study on the ear's capability to preserve and enhance spectral contrasts.* Unpublished doctoral dissertation, Vrije Universiteit te Amsterdam, Amsterdam, Netherlands.

Hsueh, K. D., & Hamernik, R. P. (1990). A generalized approach to random noise synthesis: Theory and computer simulation. *Journal of the Acoustical Society of America*, *87*(3), 1207-1217.

Indlin, A., Y. (1979a). Loudness of narrowband noise pulses. *Soviet Physics–Acoustics*, *25*(4), 301-402.

Indlin, A., Y. (1979b). Model of the auditory detection of intensity increments of narrowband noise signals. *Soviet Physics–Acoustics*, *25*(5), 400-402.

Irino, T., & Patterson, R. D. (1997). A time–domain, level–dependent auditory filter: The gammachirp. *The Journal of the Acoustical Society of America*, *101*(1), 412–419.

Irwin, R. J. (1989). Psychometric functions for the discrimination of differences in intensity of Gaussian noise. *The Quarterly Journal of Experimental Psychology*, *41A*(4), 655-674.

Jeffress, L. A. (1964). A stimulus–oriented approach to detection theory. *Journal of the Acoustical Society of America*, *36*, 766-774.

Jeffress, L. A. (1967). Stimulus–oriented approach to detection re-examined. *Journal of the Acoustical Society of America*, *41*(2), 480-488.

Jeffress, L. A. (1968). Mathematical and electrical models of auditory detection. *Journal of the Acoustical Society of America*, *44*(1), 563-569.

Jeffress, L. A. (1970). Masking. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 1, p. 87-114). NY: Academic Press.

Jeffress, L. A., Blodgett, H. C., Sandel, T. T., & Wood, C. L., III. (1956). Masking of tonal signals. *Journal of the Acoustical Society of America*, *28*(3), 416-426.

Jeffress, L. A., & Gaston, A. D., Jr. (1967). Electrical model for binaural detection. *Journal of the Acoustical Society of America*, *41*, 1612(A).

Jeruchim, M. C., Balaban, P., & Shanmugan, K. S. (1992). *Simulation of communication systems.* NY: Plenum Press.

Kaiser, J. F. (1966). Digital filters. In F. F. Kuo & J. F. Kaiser (Eds.), *System analysis by digital computer* (p. 218-285). NY: John Wiley & Sons.

Kaplan, H. L. (1975). The five distractors experiment: exploring the critical band with contaminated white noise. *Journal of the Acoustical Society of America*, *58*(2), 404-411.

Kay, I., & Silverman, R. A. (1957). On the uncertainty relation for real signals. *Information and Control*, *1*(1), 64-75.

Kay, I., & Silverman, R. A. (1959). On the uncertainty relation for real signals: Postscript. *Information and Control*, *2*, 396-397.

Kharkevich, A. A. (1960). *Spectra and analysis.* NY: Consultants Bureau Enterprises.

Kock, W. E. (1935). On the principle of uncertainty in sound. *Journal of the Acoustical Society of America*, *7*, 56-58.

Kollmeier, B., & Holube, I. (1992). Auditory filter bandwidths in binaural and monaural listening conditions. *Journal of the Acoustical Society of America*, *92*(4 Pt.1.), 1889-1901.

Laming, D. (1986). *Sensory analysis.* London: Academic Press.

Lampard, D. G. (1956). Definitions of “bandwidth” and “time duration” of signals which are connected by an identity. *IRE Transactions on Circuit Theory*, *CT-3*, 286-288.

Landau, H. J., & Pollak, H. O. (1961). Prolate spheroidal wave functions, Fourier analysis and uncertainty—II. *The Bell System Technical Journal*, *40*(1), 65-84.

Landau, H. J., & Pollak, H. O. (1962). Prolate spheroidal wave functions, Fourier analysis and uncertainty—III: The dimension of the space of essentially time–and band–limited signals. *The Bell System Technical Journal*, *41*(4), 1295-1336.

Langhans, A., & Kohlrausch, A. (1992). Spectral integration of broadband signals in diotic and dichotic masking experiments. *Journal of the Acoustical Society of America*, *91*(1), 317-326.

Lapsley, J. A., Scurfield, B. K., Drga, V. F., Galvin, S. J., & Whitmore, J. K. (1993, August). *Measures of sensitivity in single–interval and two–interval forced–choice tasks.* Paper presented at the Psychophysics Symposium, New Zealand Psychological Society Annual Conference, Wellington, New Zealand.

Lapsley Miller, J. A. (1995a). *Common noise sampling variability in psychophysical experiments.* Department of Psychology Research Seminar, Victoria University of Wellington, New Zealand, 25 Aug.

Lapsley Miller, J. A. (1995b). RAYSIML*: A program to estimate sampling variability of measures of detectability.* [Computer software in BORLAND PASCAL 7.0, for MS–DOS.].

Lapsley Miller, J. A. (1997). RANDOMBOT*: an SIFC and 2IFC experiment counter–balancing and randomisation program* [Computer software in BORLAND PASCAL 7.0, for MS–DOS.].

Lapsley Miller, J. A. (1998a). ANALBOT*: ROC, mROC, and GOC analysis program* [Computer software in BORLAND PASCAL 7.0, for MS–DOS.].

Lapsley Miller, J. A. (1998b). BESTER*: a program to calculate correlations among the ratings of human and simulated observers.* [Computer software in BORLAND PASCAL 7.0, for MS–DOS.].

Lapsley Miller, J. A. (1998c). SIM IO*: Simulated ideal observer for amplitude discrimination* [Computer software in BORLAND PASCAL 7.0, for MS–DOS.].

Lapsley Miller, J. A., & Miller, L. (1997). RUNITBOT*: experimental control program* [Computer software in HP BASIC 5.1.].

Lapsley Miller, J. A., & Miller, L. (1998). GAMMAIT*: an implementation of Bowman and Shenton's (1988) maximum likelihood estimation algorithm for gamma distributions* [Computer software in BORLAND PASCAL 7.0.].

Lapsley Miller, J. A., Scurfield, B. K., Drga, V., Galvin, S. J., & Whitmore, J. (1999). *Non-parametric relationships between single-interval and two-interval forced-choice tasks in the theory of signal detectability.* (Manuscript submitted for publication.)

Larson, H. J. (1982). *Introduction to probability theory and statistical inference.* NY: John Wiley & Sons.

Lawson, J. L., & Uhlenbeck, G. E. (1950). *Threshold signals.* London: Academic Press.

Leipnik, R. (1959). Entropy and the uncertainty principle. *Information and Control*, *2*, 64-79.

Leshowitz, B. (1969). Comparison of ROC curves from one– and two–interval rating–scale procedures. *Journal of the Acoustical Society of America*, *46*(2(2)), 399-402.

Licklider, J. C. R. (1963). Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (p. 985-1039). NY: Wiley.

Luce, R. D. (1997). Some unresolved conceptual problems in mathematical psychology. *Journal of Mathematical Psychology*, *41*, 79-87.

MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: a user's guide.* Cambridge: Cambridge University Press.

MacMillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false–alarm rates. *Psychological Bulletin*, *98*(1), 185-199.

Majerník, V., Konečný, M., & Nehnevaj, D. (1978). Intensity discrimination of band–noise signals as a function of their durations and bandwidths. *Acustica*, *39*, 117-122.

Makita, Y., & Miyatani, S. (1950). The time in which equilibrium state of audition is established part III. A new method of measuring the damping constant of the mechanical system of the ear. *Memoirs of the Research Institute of Acoustical Science Japan*, *1*, 38-46.

Marcum, J. I. (1960). A statistical theory of target detection by pulsed radar (mathematical appendix). *IRE Transactions on Information Theory (Special Monograph Issue on Studies of Target Detection by Pulsed Radar), IT-6*(2), 145-268.

Marill, T. (1956). *Detection theory and psychophysics* (Tech. Rep. No. MIT Technical Report 319). MIT Research Laboratory of Electronics.

Markowitz, J., & Swets, J. A. (1967). Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. *Perception & Psychophysics, 2*, 91-97.

Martel, H. C., & Mathews, M. V. (1961). Further results on the detectability of known signals in Gaussian noise. *The Bell System Technical Journal, 40*, 423-447.

McAulay, K. (1978). *A temporal model of aural frequency discrimination.* Unpublished doctoral dissertation, Victoria University of Wellington, Wellington, New Zealand.

McFadden, D. (1970). Three computational versions of proportion correct for use in forced–choice experiments. *Perception & Psychophysics, 8*(5B), 336-342.

McGill, W. J. (1967). Neural counting mechanisms and energy detection in audition. *Journal of Mathematical Psychology, 4*, 351-376.

McGill, W. J. (1968a). Polynomial psychometric functions in audition. *Journal of Mathematical Psychology, 5*, 369-376.

McGill, W. J. (1968b). Variations on Marill's detection formula. *Journal of the Acoustical Society of America, 43*(1), 70-73.

McGill, W. J. (1971). Poisson counting and detection in sensory systems. In E. F. Beckenbach & C. B. Tompkins (Eds.), *Concepts of communication: Interpersonal, intrapersonal, and mathematical* (p. 257-281). NY: Wiley.

McGill, W. J., & Teich, M. C. (1991). Auditory signal detection and amplification in a neural transmission network. In M. C. Commons, J. A. Nevin, & M. C. Davison (Eds.), *Signal detection: mechanisms, models and applications.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Metz, C. E., & Shen, J.-H. (1992). Gains in accuracy from replicated readings of diagnostic images: prediction and assessment in terms of ROC analysis. *Medical Decision Making, 12*, 60-75.

Michaels, R. M. (1961). Intensity discrimination for narrow bandwidths of noise at various pulse lengths. *Journal of Auditory Research, 1*, 88-110.

Miller, G. A. (1947). Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *The Journal of the Acoustical Society of America, 19*, 609-619.

Miller, L. (1997a). BIGARRAYS*: memory allocation program* [Computer software in BORLAND PASCAL 7.0 and assembly code, for MS–DOS.].

Miller, L. (1997b). FILTERIT*: a* BORLAND PASCAL 7.0 *implementation of Fisher's* GENCODE *and* MKFILTER C *programs, for* SIM IO [Computer software in BORLAND PASCAL 7.0.].

Miller, L. (1997c). SIGNALS*: signal output program* [Computer software in BORLAND PASCAL 7.0 and Motorola 65001 code, for MS–DOS.].

Miller, L. (1998a). DELEN*: a program to calculate Scurfield's multi–event measure of discriminability:* $\mathcal{D}_n$ [Computer software in C.].

Miller, L. (1998b). FORCED*: a program for FORCE analysis with Scurfield's multi–event measure of discriminability:* $\mathcal{D}_n$ [Computer software in C.].

Moller, A. R. (1983). Use of pseudorandom noise in studies of frequency selectivity: the periphery of the auditory system. *Biological Cybernetics, 47*, 95-102.

Moller, A. R. (1989). Analysis and modeling of the auditory system dynamics. In *Modelling and control in biomedical systems. Selected papers from the IFAC Symposium, Venice, Italy. 6-8 april 1988* (p. 507-518). Oxford, UK: Pergamon.

Moore, B. C. J. (1975). Mechanisms of masking. *Journal of the Acoustical Society of America, 57*(2), 391-399.

Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory–filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America, 74*(3), 750-753.

Moore, B. C. J., & Glasberg, B. R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hearing Research, 28*(2-3), 209-225.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., & Biswas, A. K. (1988). The shape of the ear's temporal window. *Journal of the Acoustical Society of America, 83*(3), 1102-1116.

Moore, B. C. J., Hafter, E. R., & Glasberg, B. R. (1996). The probe–signal method and auditory–filter shape: Results from normal– and hearing–impaired subjects. *The Journal of the Acoustical Society of America, 99*(1), 542–552.

Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1990). Auditory filter shapes at low center frequencies. *Journal of the Acoustical Society of America, 88*(1), 132-140.

Mulligan, B. E., & Elrod, M. (1970). Monaural detection and filtering. *Journal of the Acoustical Society of America, 47*, 1548-1556.

Munson, W. A., & Karlin, J. E. (1954). The measurement of human channel transmission characteristics. *Journal of the Acoustical Society of America, 26*(4), 542-553.

Nachmias, J. (1968). Effects of presentation probability and number of response alternatives on simple visual detection. *Perception & Psychophysics, 3*(2B), 151-155.

Nichols, T. L., & Jeffress, L. A. (1966). Detection performance and two parameters of the auditory stimulus. *Journal of the Acoustical Society of America, 40*, 1250(A).

North, D. O. (1963). An analysis of the factors which determine signal/noise discrimination in pulsed–carrier systems. *Proceedings of the IEEE, 51*(7), 1016-1027. (Original work published 1943; *Tech. Rept. No. PTR-6C, RCA Laboratories, Princeton, N.J., June 23, 1943.*)

Northern, J. L. (1967). Temporal summation for critical bandwidth signals. *Journal of the Acoustical Society of America, 42*(2), 456-461.

Nuttall, A. H. (1981). Some windows with very good sidelobe behavior. *IEEE Transactions on Acoustics, Speech and Signal Processing, 29*(1), 84-91.

Papoulis, A. (1962). *The Fourier integral and its applications.* McGraw–Hill Book Company.

Parzen, E. (1960). *Modern probability theory and its applications.* NY: John Wiley & Sons.

Patterson, J., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. *Journal of the Acoustical Society of America, 72*, 1788-1803.

Patterson, R. D. (1974). Auditory filter shape. *Journal of the Acoustical Society of America, 55*(4), 802-809.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America, 59*(3), 640-654.

Patterson, R. D., & Henning, G. B. (1977). Stimulus variability and auditory filter shape. *Journal of the Acoustical Society of America, 62*(3), 649-664.

Penner, M. J. (1972). Neural of energy summation in a Poisson counting model. *Journal of Mathematical Psychology*, *9*, 286-293.

Perniske, P. (1987). *The application of group–operating–characteristic analysis to two–interval forced–choice experiment.* Unpublished honours thesis, Victoria University of Wellington, New Zealand.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Trans I.R.E., PGIT-4*, 171-212.

Pfafflin, S. M., & Mathews, M. V. (1962). Energy–detection model for monaural auditory detection. *Journal of the Acoustical Society of America*, *34*(12), 1842-1853.

Pfafflin, S. M., & Mathews, M. V. (1966). Detection of auditory signals in reproducible noise. *Journal of the Acoustical Society of America*, *39*(2), 340-345.

Pollack, I., & Hsieh, R. (1969). Sampling variability of the area under the ROC–curve and of d'_e. *Psychological Bulletin*, *71*(3), 161-173.

Press, W. H. (1988). *Numerical recipes in C: The art of scientific computing.* Cambridge [Cambridgeshire], NY: Cambridge University Press.

Pridham, R. G. (1968). Comment on "Energy detection of unknown deterministic signals". *Proceedings of the IEEE*, *56*, 1379-1380.

Raab, D. H., & Goldberg, I. A. (1975). Auditory intensity discrimination with bursts of reproducible noise. *Journal of the Acoustical Society of America*, *57*(2), 437-447.

Raab, D. H., & Leshowitz, B. (1968). Use of an average response computer to provide reproducible noise bursts. *Journal of the Acoustical Society of America*, *44*, 282-283.

Rabiner, L. R., & Gold, B. (1975). *Theory and application of digital signal processing.* Englewood Cliffs, NJ: Prentice–Hall.

Reed, C. M., & Bilger, R. C. (1973). A comparative study of S/N_0 and E/N_0. *Journal of the Acoustical Society of America*, *53*(4), 1039-1043.

Rice, S. O. (1954). Mathematical analysis of random noise. In N. Wax (Ed.), *Selected papers on noise and stochastic processes* (p. 133-294). NY: Dover.

Rice, S. O. (1982). Envelopes of narrow–band signals. *Proceedings of the IEEE, 70*(7), 692-699.

Richards, V. M. (1992). The detectability of a tone added to narrow bands of equal–energy noise. *Journal of the Acoustical Society of America*, *91*(6), 3424-3435.

Richards, V. M., Heller, L. M., & Green, D. M. (1991). The detection of a tone added to a narrow band of noise: The energy model revisited. *The Quarterly Journal of Experimental Psychology*, *43A*(3), 481-501.

Robert, C. P. (1994). *The Bayesian choice: A decision–theoretic motivation.* NY: Springer–Verlag.

Robinson, C. E. (1974). Simple form of the auditory running–average hypothesis: application to the temporal summation of loudness and to the delayed perception of the offset of brief stimuli. *Journal of the Acoustical Society of America*, *55*(3), 645-648.

Robinson, C. E., & Pollack, I. (1973). Interaction between forward and backward masking: a measure of the integrating period of the auditory system. *Journal of the Acoustical Society of America*, *53*(5), 1313-1316.

Rockette, H. E., Gur, D., & Metz, C. E. (1992). The use of continuous and discrete judgements in ROC studies of diagnostic imaging techniques. *Investigative Radiology*, *27*, 169-172.

Ronken, D. A. (1969). Intensity discrimination of Rayleigh noise. *Journal of the Acoustical Society of America*, *45*(1), 54-57.

Ronken, D. A. (1970a). Reciprocal spreading effects on frequency discrimination. *Journal of the Acoustical Society of America*, *47*, 119(A).

Ronken, D. A. (1970b). Some effects of bandwidth–duration constraints on frequency discrimination. *Journal of the Acoustical Society of America*, *49*(4(2)), 1232-1242.

Ronken, D. A. (1973). Masking produced by sinusoids of slowly changing frequency. *Journal of the Acoustical Society of America*, *54*(4), 905-915.

Schacknow, P. N., & Raab, D. H. (1976). Noise–intensity discrimination: Effects of bandwidth conditions and mode of masker presentation. *Journal of the Acoustical Society of America*, *60*(4), 893-905.

Schafer, T. H., Gales, R. S., Shewmaker, C. A., & Thompson, P. O. (1950). The frequency selectivity of the ear as determined by masking experiments. *Journal of the Acoustical Society of America*, *22*(4), 490-496.

Scharf, B. (1970). Critical bands. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 1, p. 157-202). NY: Academic Press.

Scharf, B. (1978). Loudness. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. IV). NY: Academic Press.

Schulman, A. I., & Mitchell, R. R. (1963). Operating characteristics from yes–no and forced–choice procedures. *Journal of the Acoustical Society of America*, *35*, 785(A).

Schulman, A. I., & Mitchell, R. R. (1966). Operating characteristics from yes–no and forced–choice procedures. *Journal of the Acoustical Society of America*, *40*(2), 473-477.

Scurfield, B. K. (1994). `RandUnit`*: a program to implement the linear congruent pseudo–random–number–generator, described as the "minimal standard random number generator" in Park (1988)* [Computer software in BORLAND PASCAL 7.0.].

Scurfield, B. K. (1995). *Discrimination among events by neural networks.* Unpublished doctoral dissertation, Victoria University of Wellington, Wellington, New Zealand.

Scurfield, B. K. (1996). Multiple–event forced–choice tasks in the theory of signal detectability. *Journal of Mathematical Psychology*, *40*(3), 253-269.

Scurfield, B. K. (1998). Generalization of the theory of signal detectability to m–dimensional n–event forced–choice tasks. *Journal of Mathematical Psychology*, *42*(1), 5-31.

Shannon, C. E. (1949a). Communication in the presence of noise. *Proceedings of the IRE*, *37*(1), 10-21.

Shannon, C. E. (1949b). The mathematical theory of communication. In C. E. Shannon & W. Weaver (Eds.), *The mathematical theory of communication* (p. 3-99). Urbana, IL: The University of Illinois Press.

Sherwin, C. W., Kodman, F. J., Kovaly, J. J., Prothe, W. C., & Melrose, J. (1956). Detection of signals in noise: A comparison between the human detector and an electronic detector. *Journal of the Acoustical Society of America*, *28*(4), 617-622.

Simpson, A. J., & Fitter, M. J. (1973). What is the best index of discriminability. *Psychological Bulletin*, *80*(6), 481-488.

Sinclair, I. R. (1988). *Dictionary of electronics.* Glasgow: Collins.

Skilling, H. H. (1965). *Electrical engineering circuits* (second ed.). NY: John Wiley & Sons.

Slepian, D. (1954). Estimation of signal parameters in the presence of noise. *IRE Transactions Professional Group on Information Theory (IRE Trans PGIT)*, *3*, 68-89.

Slepian, D. (1976). On bandwidth. *Proceedings of the IEEE, 64*(3), 292-300.

Slepian, D. (1983). Some comments on Fourier analysis, uncertainty and modeling. *SIAM Review, 25*(3), 379-393.

Slepian, D., & Pollak, H. O. (1961). Prolate spheroidal wave functions, Fourier analysis and uncertainty—I. *The Bell System Technical Journal, 40*(1), 43-63.

Smith, M., & Wilson, E. (1953). A model of the auditory threshold and its application to the problem of the multiple observer. *Psychological Monographs: General and Applied, 67*(9), 1-35.

Spiegel, M. F. (1979). The range of spectral integration. *Journal of the Acoustical Society of America, 66*(5), 1356-1363.

Stearns, S. D. (1975). *Digital signal analysis.* Rochelle Park, NJ: Hayden Book Company.

Steenson, B. O., & Stirling, N. C. (1965). The amplitude distribution and false–alarm rate of filtered noise. *Proceedings of the IEEE, 53*, 42-55.

Stewart, G. W. (1931). Problems suggested by an uncertainty principle in acoustics. *Journal of the Acoustical Society of America, 2*(3), 325-329.

Swets, J. A. (1959). Indices of signal detectability obtained with various psychophysical procedures. *Journal of the Acoustical Society of America, 31*(4), 511-513.

Swets, J. A. (1963). Central factors in auditory frequency selectivity. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: contemporary readings* (1964 ed.). NY: John Wiley & Sons.

Swets, J. A., & Green, D. M. (1961). Sequential observations by human observers of signals in noise. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: contemporary readings* (1964 ed.). NY: John Wiley & Sons.

Swets, J. A., Green, D. M., & Tanner, W. P., Jr. (1962). On the width of critical bands. *Journal of the Acoustical Society of America, 34*(1), 108-113.

Tanner, W. P., Jr. (1960a). Theory of signal detectability as an interpretive tool for psychophysical data. *Journal of the Acoustical Society of America, 32*(9), 1140-1147.

Tanner, W. P., Jr. (1960b). Use of the Fourier series bandlimitation assumption in the analysis of psychophysical data. *Journal of the Acoustical Society of America, 32*, 931(A).

Tanner, W. P., Jr. (1961). Application of the theory of signal detectability to amplitude discrimination. *Journal of the Acoustical Society of America, 33*(9), 1233-1244.

Tanner, W. P., Jr., & Birdsall, T. G. (1958). Definitions of d' and η as psychophysical measures. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: contemporary readings* (1964 ed.). NY: John Wiley & Sons.

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision–making theory of visual detection. *Psychological Review, 61*(6), 401-409.

Taylor, A., Boven, R., & Whitmore, J. (1991). Reduction of unique noise in the psychophysics of hearing by group operating characteristic analysis. *Psychological Bulletin, 109*(1), 133-146.

Taylor, A. J. (1984). *Auditory psychophysics in birds: The effects of unique noise on sensitivity.* Unpublished doctoral dissertation, Victoria University of Wellington, New Zealand.

Tucker, A., Evans, R. B., & Jeffress, L. A. (1967). ROC curves for multiple–signal levels in a detection task. *Journal of the Acoustical Society of America, 41*, 1611(A).

Urkowitz, H. (1967). Energy detection of unknown deterministic signals. *Proceedings of the IEEE, 55*(4), 523-531.

Vakman, D. E. (1968). *Sophisticated signals and the uncertainty principle in radar.* NY: Springer–Verlag.

van den Brink, G. (1964). Detection of tone pulses of various durations in noise of various bandwidths. *Journal of the Acoustical Society of America*, *36*, 1206-1211.

van den Brink, W. A. C., & Houtgast, T. (1990). Spectro–temporal integration in signal detection. *Journal of the Acoustical Society of America*, *88*(4), 1703-1708.

Viemeister, N. F., & Wakefield, G. H. (1991). Temporal integration and multiple looks. *Journal of the Acoustical Society of America*, *90*(2(1)), 858-865.

Watson, C. S. (1963). Signal detection and certain physical characteristics of the stimulus during the observation interval (Doctoral dissertation, Indiana University, 1962). *Dissertation Abstracts International*, *24*, 2995.

Watson, C. S., Kellogg, S. C., Kawanishi, D. T., & Lucas, P. A. (1973). The *uncertain* response in detection–oriented psychophysics. *Journal of Experimental Psychology*, *99*(2), 180-185.

Watson, C. S., Rilling, M. E., & Bourbon, W. T. (1964). Receiver–operating characteristics determined by a mechanical analog to the rating scale. *Journal of the Acoustical Society of America*, *36*(2), 283-288.

Weyl, H. (1931). *The theory of groups and quantum mechanics.* Methuen and Co.

Whitfield, I. C. (1967). *The auditory pathway.* Baltimore: Williams and Wilkins.

Whitmore, J. K. (1969). Auditory detection of short–duration narrow–band noise signals. *Journal of the Acoustical Society of America*, *45*, 327(A).

Whitmore, J. K., & Bowden, C. J. (1985). *Electronic simulation of human binaural detection.* Paper presented to the 9th New Zealand Electronics Conference. Oct 16-18, Massey University, Palmerston North.

Whitmore, J. K., Drga, V., & Taylor, A. (1993, August). *A diotic amplitude discrimination experiment replicated 100 times.* Paper presented at the Psychophysics Symposium, New Zealand Psychological Society Annual Conference, Wellington, New Zealand.

Whitmore, J. K., & Galvin, S. J. (1993, August). *Group operating characteristic (GOC) analysis of Type II decisions.* Paper presented at the Psychophysics Symposium, New Zealand Psychological Society Annual Conference, Wellington, New Zealand.

Whitmore, J. K., Williams, P. I., & Ermey, H. L. (1968). Psychometric function from Rayleigh–Rayleigh ROC curves. *Journal of the Acoustical Society of America*, *44*(1), 370-371(A).

Wilbanks, W. A., & Whitmore, J. K. (1967). Detection of a narrow–band noise as a function of its interaural correlation. *Psychonomic Science*, *8*(6), 235-236.

Wilcox, G. W. (1967). Measures of performance sensitivity. *Journal of the Acoustical Society of America*, *42*, 1195(A).

Wilcox, G. W. (1968). *Inter-observer agreement and models of monaural auditory processing in detection tasks.* Unpublished doctoral dissertation, University of Michigan.

Williams, P. I., & Jeffress, L. A. (1967). Agreement in detection: Observers and electrical model. *Journal of the Acoustical Society of America*, *42*(5), 1194(A).

Williams, T., & Kelley, C. (1997). GNUPLOT*: version 3.0. A freeware graphics and analysis program* [Computer software.].

Wright, B. A., & Dai, H. (1994a). Detection of unexpected tones in gated and continuous maskers. *Journal of the Acoustical Society of America*, *95*(2), 939-948.

Wright, B. A., & Dai, H. (1994b). Detection of unexpected tones with short and long durations. *Journal of the Acoustical Society of America*, *95*(2), 931-938.

Zakai, M. (1960). A class of definitions of "duration" (or "uncertainty") and the associated uncertainty relations. *Information and Control*, *3*, 101-115.

Zwicker, E., & Fastl, H. (1972). On the development of the critical band. *Journal of the Acoustical Society of America*, *52*(2(2)), 699-702.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical–band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, *68*(5), 1523-1525.

Appendix A

The problem of waveform representation

The problem of waveform representation is the cause of most of the criticisms about ideal observers (Green & Swets, 1966; Tanner, 1960b). This appendix expands on some of these problems, which were introduced in Section 1.1.3 of Chapter 1.

The acoustical uncertainty principle shows that a waveform cannot be both band and time–limited. This is a problem because, paradoxically, most psychophysical tasks, using human or animal observers, appear to involve detecting a transient waveform, with a limited bandwidth. Conversely, many psychophysical theories make the assumption that the waveform is infinitely long (or, less commonly, an infinite bandwidth). The theories that attempt to deal with the detectability of transient waveforms, of limited bandwidth, generally require more sophisticated mathematics, because they need to deal with the effects of the uncertainty principle on the waveform representation. The result is that waveform representation of essentially band and time–limited waveforms merely approximates the original waveform of interest. This approximation is not necessarily accurate as the bandwidth–duration product nears the limit imposed by the uncertainty principle. Because many psychophysical theories are interested in small $\mathcal{WT}$ signals, the accuracy of this representation becomes paramount.

Waveform representation with the Fourier Integral

If $f(t)$ is the representation of a waveform in the time domain and $F(\omega)$ is the representation of a waveform in the angular frequency domain, then the two are related via the Fourier integral equation

$$f(t) \quad = \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}d\omega \tag{A.1}$$

and the inverse Fourier integral equation

$$F(\omega) \quad = \quad \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt, \tag{A.2}$$

to form what is known as a *Fourier pair*. The Fourier integral *analyses* or *decomposes* a waveform into a linear combination of weighted sinusoids (Franks, 1969; Papoulis, 1962; Vakman, 1968).

$F(\omega)$ provides information about both amplitude and phase as a function of frequency. If $A(\omega)$ is the amplitude spectrum (or Fourier spectrum), and $\phi(\omega)$ is the phase spectrum, then the relationship is $F(\omega) = A(w)e^{j\phi(\omega)}$. Another representation of the amplitude frequency domain information is the *energy spectrum*, $A^2(\omega)$, which is the square of the Fourier spectrum. The concept of energy here is a mathematical concept, and should not be confused with the energy in a physical signal.

An important result, which links the energy content in the two domains, is Parseval's theorem

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt \quad = \quad \frac{1}{2\pi} \int_{-\infty}^{\infty} A^2(\omega) \, d\omega \tag{A.3}$$

which shows that the energy of the waveform in the time domain is equal to the energy of the waveform in the frequency domain.

Fourier analysis and acoustical uncertainty principles

It is impossible for a waveform to be limited in both time and frequency. If a waveform is limited in time, then it must have an infinitely wide bandwidth, and conversely, if a waveform has a finite bandwidth, then it must have an infinite duration. This result is known as the *acoustical uncertainty principle*. There are many forms of the acoustical uncertainty principle because $\mathcal{W}$ and $\mathcal{T}$ can be defined in many ways. All acoustical uncertainty principles, however, specify a lower bound to the bandwidth–duration product, $\mathcal{WT}$, indicating that a waveform cannot have finite spread in both domains.

The acoustical uncertainty principle contradicts what seems to be a defining property of the waveforms to which the human hearing system is sensitive: that waveforms are both band–limited and time–limited. How can this contradiction be resolved? If a small amount of energy is allowed to exist outside the bounds of bandwidth or duration then it is meaningful to refer to band–limited and time–limited waveforms, because these waveforms have essentially all of their energy between two well-defined bounds. The energy outside these bounds is so small as to be virtually nonexistent. The problem is that the energy is there, and this makes defining the bandwidth and duration arbitrary. As a result there are many definitions of $\mathcal{W}$ or $\mathcal{T}$, depending on the nature of the problem or the context, and there are also many different forms of the uncertainty principle (see Bourret, 1958; Brillouin, 1962; de Bruijn, 1967; Cohen, 1989; Franks, 1969; Gabor, 1946; Hilberg & Rothe, 1971; Kay & Silverman, 1957, 1959; Lampard, 1956; Landau & Pollak, 1961, 1962; Leipnik, 1959; Slepian & Pollak, 1961; Weyl, 1931; Zakai, 1960).

The acoustical uncertainty principle was traditionally based on the analogous uncertainty principle of quantum physics. This has lead to problems of interpretation when applied to acoustics. Landau, Pollak and Slepian have provided the best alternative approach to the classical acoustical uncertainty principle. In a series of papers entitled "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty" they have proposed new definitions of the concepts of bandwidth and duration, and derived the corresponding uncertainty principle (Landau & Pollak, 1961, 1962; Slepian & Pollak, 1961). An excellent overview of this research is in Slepian (1983).

Landau and Pollak (1961) suggested that a good definition of $\mathcal{W}$ and $\mathcal{T}$ would describe the behaviour of $f(t)$ in a given finite time interval and likewise the behaviour of $F(\omega)$ in a given finite frequency band. They argued that although waveforms cannot be bounded in both domains, bounds that essentially constrain the waveforms can still be specified. One way to specify these bounds is to look at the *energy* content and *energy* spread of a waveform in both the time and frequency domains. Bandwidth and duration can be specified by calculating the proportion of energy constrained between two bounds relative to the total energy of the waveform.

The essential bandwidth and duration, defined in Equations (1.3) and (1.4), are reproduced here as Equations (A.4) and (A.5). The proportion of energy β^2, constrained by the *essential bandwidth*, Ω, is defined as

$$\beta^2 \quad = \quad \frac{\int_{-\Omega}^{\Omega} |A(\omega)|^2 \, d\omega}{\int_{-\infty}^{\infty} |A(\omega)|^2 \, d\omega}. \tag{A.4}$$

Similarly, the proportion of energy, α^2, constrained by the essential duration,[1] $\mathcal{T}$, is

$$\alpha^2 = \frac{\int_{-\mathcal{T}/2}^{\mathcal{T}/2} |f(t)|^2 \, dt}{\int_{-\infty}^{\infty} |f(t)|^2 \, dt} \tag{A.5}$$

where α^2 and β^2 range between 0 and 1, $f(t)$ is an acoustic waveform, and $A^2(\omega)$ is its energy spectrum.

Landau and Pollak (1961) derived an acoustical uncertainty principle from these definitions of bandwidth and duration. They showed that if either α^2 or β^2 is specified, the other must remain below a certain maximum, which depended on the bandwidth–duration product $\Omega\mathcal{T}$. This acoustical uncertainty principle does not specify a single number as a lower bound, like in some definitions, but a *function* of the energy constrained in each domain. From this uncertainty principle, the waveform which best constrained energy in both domains was shown to be a prolate spheroidal wave function. This function concentrates the most energy, in both domains, out of all the finite energy functions. Its one parameter is proportional to the bandwidth–duration product.[2]

It may be asked what is *the* minimum bandwidth–duration product for this new definition. The answer is no longer quite as simple because it depends on what is considered to be a suitable amount of energy to be left out of one, or the other, domain. This may be seen by some as a failing of the new definition, but it is not, because (a) it shows that there is no such thing as a truly time–limited and band–limited waveform, and (b) it highlights the arbitrary nature of the definitions of bandwidth and duration. The classical approach avoids these issues by defining a measure (variance) which does not explicitly indicate how well the waveform is bounded in one domain or the other.

The dimensionality of waveforms

Shannon's sampling theorem is the best known example of waveform representation using a sampling space (Shannon, 1949a; Goldman, 1953). This theorem shows how a band–limited signal may be represented by taking samples from the time series every $1/2\mathcal{W}$ seconds. These samples are Fourier coefficients. The waveform is reconstructed by interpolating between the samples with sinc functions. Due to Fourier symmetry, a similar result holds for time–limited signals where samples are taken in the frequency domain every $1/2\mathcal{T}$ Hz. The sampling theorem assumes that sampling is done over infinite time or infinite frequency respectively. Practically, this type of sampling can never be achieved, however, excellent approximations can be achieved by taking $2\mathcal{WT}$ samples of the waveform. This approximation is good for large $\mathcal{WT}$ and for signals concentrated in both time and frequency.

This theorem has an intuitive appeal, but the formulation lacks mathematical rigour, especially with respect to taking $2\mathcal{WT}$ samples. It is easier to understand why the sampling theorem works, and why there are limitations, when it is seen as a special case of signal representation using a sampling space. This representation uses the concept of approximate dimensionality of sampling spaces and relies on a generalised view of frequency.

The set of all finite energy waveforms may be defined as the space, $\mathcal{L}_\infty^2$, where all complex valued waveforms, $f(t)$, are defined on the real line and are absolute square integrable. Any arbitrary real waveform, $s(t)$, in $\mathcal{L}_\infty^2$ space can be expanded into a *generalised Fourier Series* as a complete system of orthogonal functions

$$s(t) \quad = \quad \sum_{k=1}^{m} s_k \varphi_k(t) \tag{A.6}$$

[1] Kharkevich (1960) has also suggested this definition. He also states that the integral is a constant proportion of the total energy. Kharkevich goes on to examine the resulting $\mathcal{WT}=$ for a 90% proportion of energy constrained in both domains for a variety of data windows. Additionally, Chalk (1950) uses a similar definition to derive the optimum pulse–shape for pulse communication.

[2] For the proofs of the mathematical properties of prolate spheroidal wave functions see Landau and Pollak (1961, 1962), Papoulis (1962), Slepian and Pollak (1961).

where the s_k are the Fourier coefficients. Because the basis functions, $\varphi_k(t)$, are orthogonal, these coefficients are given by

$$s_k = \int_{-\infty}^{\infty} s(t)\varphi_k(t)\,dt. \tag{A.7}$$

This generalised Fourier Series tends to an integral when $m \to \infty$.

The process of sampling can be seen as a mapping from $\mathcal{L}^2_{\mathcal{T}} \to \mathrm{C}^m$ where m is chosen as a compromise between accuracy and economy (Franks, 1969) and C is the complex number plane. It is a many–to–one mapping in the sense that there is no unique set of numbers for each waveform in $\mathcal{L}^2$. The sampling space is said to have m–dimensions and m degrees of freedom. As $m \to \infty$, the degrees of freedom increase without bound. The waveform, $s(t)$, is also m–dimensional, thus $s(t)$ is a point in m–dimensional space, or in other words, $s(t)$ can be represented as a multi–dimensional radius vector from the origin. The waveform energy is represented as the squared magnitude of the sampling[3]:

$$\int_{-\infty}^{\infty} s^2(t)\,dt \simeq \sum_{k=1}^{m} s_k^2. \tag{A.8}$$

The nature of the orthogonal basis functions used depends on the goal. Shannon's sampling theorem, for instance, uses sinc functions. However, Landau and Pollak (1962) demonstrate convincingly that sinc functions are not optimal; instead prolate spheroidal wave functions are the best basis functions. Their argument can be summarised as follows:

If a waveform $s(t)$ is band–limited then it cannot also be time–limited. From Equation (A.5), however, $s(t)$ can be thought of as approximately time–limited if most of its energy is contained in the interval $|t| \leqslant \mathcal{T}/2$:

$$\frac{\int_{-\mathcal{T}/2}^{\mathcal{T}/2} |s(t)|^2\,dt}{\int_{-\infty}^{\infty} |s(t)|^2\,dt} = \alpha^2 = 1 - \epsilon_T^2 \tag{A.9}$$

where ϵ_T is a measure of the degree to which $s(t)$ fails to be time–limited. E(ϵ_T) is the set of all band–limited waveforms satisfying Equation (A.9).

How can the dimensionality of E(ϵ_T) be measured? E(ϵ_T) is not finite–dimensional for there is no finite set of functions whose linear combinations exactly represent each waveform in E(ϵ_T). However, E(ϵ_T) is *approximately* m–dimensional, if there exists m linearly independent functions $\varphi_0, \ldots, \varphi_{m-1}$, whose linear combinations approximate each $s(t)$ in E(ϵ_T) to within a small fraction of the energy in $s(t)$:

$$\min_{\{a_i\}} \int_{-\infty}^{\infty} \left| s(t) - \sum_{0}^{m-1} a_i\varphi_i(t) \right|^2 dt < \delta_m^2 \tag{A.10}$$

where again, δ_m is small.

The form of the orthogonal basis functions $\varphi_0, \ldots, \varphi_{m-1}$ that are used is arbitrary. Landau and Pollak (1962) compare the *best* function with the more traditional sinc function. They state that for real understanding of the dimension of E(ϵ_T) the φ_i that gives the best approximation, by making δ_m as small as possible over all E(ϵ_T), should be used, but for many practical applications choosing the simplest function is better (as is perhaps the case with Shannon's sampling theorem). They show that the best basis, for any m, should be $(m+1)$ linearly independent *most concentrated* band–limited functions. These functions are proven by Landau and Pollak (1962) to be the prolate spheroidal wave functions.

Landau and Pollak (1962) go on to prove that $\lfloor 2\mathcal{W}\mathcal{T} \rfloor + 1$ of the best basis functions, ψ_i, approximate a concentrated function to a degree proportional to the energy ϵ_T^2 outside of the interval. With sinc functions, $\lfloor 2\mathcal{W}\mathcal{T} \rfloor + 1$ sinc functions will approximate $s(t)$ in energy roughly to within a constant times ϵ_T, that is, within a constant times the square root of the unconcentrated energy.

[3] Also known as the generalised Pythagorean theorem.

A sampling series approximation with $2\mathcal{WT}$ + C terms will not approximate every concentrated function to a degree proportional to the unconcentrated energy. This is in direct contrast to the prolate spheroidal wave functions.

Similar results can be derived for waveforms that are not totally band–limited. Landau and Pollak (1962) conclude that "the degree of approximation achievable by sampling functions is in a very real sense poorer than the degree achievable by the *best* basis functions" (p. 1297).

Band–pass waveform representation

In the time domain, the representation of a band–pass signal, $f(t)$, is achieved by representing the waveform in terms of its odd and even components such that the odd components are in phase and the even components are in quadrature. This is possible through the properties of the Hilbert transform (Goldman, 1953).

$$f(t) \quad = \quad a(t)\cos\omega_0 t - b(t)\sin\omega_0 t \tag{A.11}$$

where $a(t)$ and $b(t)$ are low–pass signals with upper cutoffs $\omega_c = \mathcal{W}/2$ and carrier frequency $\omega_0 > \mathcal{W}/2$.

Gaussian noise waveform representation

The concept of the sampling space simplifies the representation of stochastic waveforms. Assuming stationary, white Gaussian noise, the probability density of a given noise level, n, at any arbitrary time, t, is

$$p(n) \quad = \quad \frac{1}{\sqrt{2\pi}\sigma} e^{-n^2/2\sigma^2} \tag{A.12}$$

where σ is the rms noise level. Using the concept of the sampling space, noise can be represented as the weighted linear combination of $2\mathcal{WT}$ basis functions:

$$n(t) \quad = \quad \sum_{k=1}^{2\mathcal{WT}} n_k \varphi_k(t) \tag{A.13}$$

where the basis functions are prolate spheroidal wave functions (Slepian & Pollak, 1961) and the coefficients are given by

$$n_k \quad = \quad \int_{-\infty}^{\infty} n(t)\varphi_k(t)\,dt. \tag{A.14}$$

The coefficients, n_k, are the Gaussian random variables, $p(n)$, from Equation (A.12), with mean $\bar{n}_k = 0$, and variance $\sigma^2 = \mathrm{N}_0$ (uniform power spectrum), $\mathcal{N}(0, \mathrm{N}_0)$. They are independently, identically, distributed (*i.i.d.*) for all k:

$$p(n_k) \quad = \quad \frac{1}{\sqrt{2\pi\mathrm{N}_0}} e^{-n_k^2/2\mathrm{N}_0} \tag{A.15}$$

The Karhunen–Loève expansion shows how band–limited white noise can be represented in m–dimensions for the interval $\mathcal{T}$, by using an orthogonal basis in $\mathcal{L}^2_{\mathcal{T}}$ (Davenport & Root, 1958; Franks, 1969; Urkowitz, 1967)

$$n(t) \quad \simeq \quad \sum_{i=1}^{m} \alpha_i \varphi_i(t), \qquad\qquad |t| \leqslant \mathcal{T}. \tag{A.16}$$

If the noise is second–order stationary, band–limited, white Gaussian noise, Slepian shows that the basis functions, $\varphi_i(t)$, are prolate spheroidal wave functions, that $m = 2\mathcal{WT}$, and the α_i coefficients are independent Gaussian variates n_k from Equation (A.14) (Slepian, 1954; Slepian &

Pollak, 1961). Because the φ_i are orthogonal, the energy of $n(t)$ in the interval $[0, \mathcal{T}]$ is

$$\int_0^{\mathcal{T}} n^2(t)\,dt \quad \simeq \quad \sum_{k=1}^{2\mathcal{W}\mathcal{T}} n_k^2. \tag{A.17}$$

If the noise is band–pass instead of low–pass, the representation changes but the result is the same (see Davenport & Root, 1958; Franks, 1969; Goldman, 1953; Urkowitz, 1967). The noise is split into two modulation components—one in phase (cosine component) and one in quadrature (sine component)

$$n(t) \quad = \quad n_{\mathrm{c}}(t)\cos\omega_{\mathrm{c}}t - n_{\mathrm{s}}(t)\sin\omega_{\mathrm{s}}t \tag{A.18}$$

where ω is the reference angular frequency and $n_{\mathrm{c}}(t)$ and $n_{\mathrm{s}}(t)$ are low–pass functions with bandwidth $|f| < \mathcal{W}/2$. If $n(t)$ has a flat power spectrum, N_0, then $n_{\mathrm{c}}(t)$ and $n_{\mathrm{s}}(t)$ also have flat power spectra each equal to $2\mathrm{N}_0$ over $|f| < \mathcal{W}/2$. Thus, $n_{\mathrm{c}}(t)$ and $n_{\mathrm{s}}(t)$ each have $\mathcal{W}\mathcal{T}$ degrees of freedom and variance $2\mathrm{N}_0\mathcal{W}$.

According to Urkowitz (1967), the energy in $n(t)$ can be approximated as

$$\int_0^{\mathcal{T}} n^2(t)\,dt \quad \simeq \quad \frac{1}{2}\int_0^{\mathcal{T}} \left[n_{\mathrm{c}}^2(t) + n_{\mathrm{s}}^2(t)\right]\,dt \tag{A.19}$$

which improves as $\mathcal{T}$ increases. The series expansion for the energy in $n_{\mathrm{c}}(t)$ and $n_{\mathrm{s}}(t)$ is

$$\int_0^{\mathcal{T}} n_{\mathrm{c}}^2(t)\,dt \quad = \quad \frac{1}{\mathcal{W}}\sum_{i=1}^{\mathcal{W}\mathcal{T}} a_{\mathrm{c}i}^2 \tag{A.20}$$

$$\int_0^{\mathcal{T}} n_{\mathrm{s}}^2(t)\,dt \quad = \quad \frac{1}{\mathcal{W}}\sum_{i=1}^{\mathcal{W}\mathcal{T}} a_{\mathrm{s}i}^2 \tag{A.21}$$

where $a = n(\frac{i}{\mathcal{W}})$.

Summary

By using generalised Fourier analysis, with prolate spheroidal wave basis functions, Gaussian noise may be represented, in an approximate form, by $2\mathcal{W}\mathcal{T}$ samples. This representation is better than that achievable with Shannon's sampling theorem.

More generally, because the definitions of $\mathcal{W}$, $\mathcal{T}$, and $\mathcal{W}\mathcal{T}$ are somewhat arbitrary, and because the representation of band and time–limited waveforms involves approximations, it is important to make clear which definitions are being used.

As a final comment, about the nature of band and time–limited waveforms, Slepian (1983) provides a useful insight. He argued that it is "senseless to ask if real signals are bandlimited, or timelimited" because "continuity is not a verifiable notion in the real world" (p. 389). Band–limiting and time–limiting are *models*. It may be the case that real waveforms may be both band and time–limited, but because they cannot be measured at their limit, it is impossible to know. Slepian warns that it may be unwise to assume that the limits imposed by the extremes of a mathematical model in fact exist in real waveforms. As a result, Slepian (1976, 1983) gave an alternative interpretation to the notion of band and time limited signals by redefining band and time limiting so that they did not depend on the detailed behaviour of waveforms, or their spectra, at infinity. He suggested that a waveform is band–limited and time–limited if, at a given level, there is energy lying outside $\mathcal{W}$ and $\mathcal{T}$ which is not measurable in the real world.

Appendix B

An algorithm for calculating the essential bandwidth and duration

This appendix presents an heuristic algorithm for calculating the essential bandwidth $\mathrm{Ess}\mathcal{W}_{\beta^2\%}$ and duration $\mathrm{Ess}\mathcal{T}_{\alpha^2\%}$ of digital or digitised signals. The code for the time domain and frequency domain is effectively the same, but for clarity they have both been presented. The main difference is that the energy spectrum is already in units of energy, whereas each element in the time series must first be squared. Also, depending on the application, the DC component may, or may not be, important. The algorithm, as presented, includes the DC component by looping from zero, but this may be changed to one if necessary.

The algorithm in pseudo–code

input

waveform : an array of real values representing a time series.

wavespectrum : an array of real values representing a power spectrum.

wavesize : the number of elements in the waveform.

spectrumsize : the number of elements in the spectrum.

timebase : the resolution of the time series in seconds.

freqbase : the resolution of the spectrum in hertz.

tolerance : an arbitrary cutoff to determine how close the energy is to the desired bound.

α^2 : proportion of energy constrained in the time domain.

β^2 : proportion of energy constrained in the frequency domain.

output

essential_duration: a real value representing duration in seconds.

essential_bandwidth: a real value representing bandwidth in Hz.

energy_time: a real value representing the total energy in the time domain.

energy_freq: a real value representing the total energy in the frequency domain.

```
function essential_bandwidth (wavespectrum, β², spectrumsize, freqbase, tolerance)
(* Essential bandwidth for the discrete spectrum, wavespectrum, with spectrumsize elements, and with
resolution of freqbase Hz, for a given proportion of energy constrained (β²). Based on the definition for
essential bandwidth for continuous spectra by Landau and Pollak (1961). *)
  energy ← 0; total_energy ← 0; p ← 0; max_e_elt ← 0; max_e_idx ← 0
  call find_maximum_energy_freq (wavespectrum, spectrumsize, max_e_elt, max_e_idx)
  total_energy ← energy_freq (wavespectrum, spectrumsize)
  lowerindex ← max_e_idx; upperindex ← max_e_idx
  p ← wavespectrum[max_e_idx] / total_energy
  energy ← energy + p
  if (energy + tolerance) ⩾ β² then
    (* Required proportion of energy is contained in max_e_idx, so interpolate to get bandwidth. *)
    bandwidth ← β² / energy × freqbase
    return bandwidth
  else
    while (energy + tolerance) < β² do
      if (lowerindex > 0) and (upperindex < spectrumsize) then
        (* Sum bin with most energy from either side of max_e_idx. *)
        if wavespectrum[lowerindex-1] > wavespectrum[upperindex+1] then
          lowerindex ← lowerindex - 1
          p ← wavespectrum[lowerindex] / total_energy
          energy ← energy + p
        else
          upperindex ← upperindex + 1
          p ← wavespectrum[upperindex] / total_energy
          energy ← energy + p
        fi
      else if (lowerindex = 0) and (upperindex < spectrumsize) then
        (* If the lowest index is reached then sum only from the upper index. *)
        upperindex ← upperindex +1
        p ← wavespectrum[upperindex] / total_energy
        energy ← energy + p
      else if (lowerindex > 0) and (upperindex = spectrumsize) then
        (* If the highest index is reached then sum only from the lower index. *)
        lowerindex ← lowerindex - 1
        p ← wavespectrum[lowerindex] / total_energy
        energy ← energy + p
      else if (lowerindex = 0) and (upperindex = spectrumsize) then
        bandwidth ← spectrumsize × freqbase
        return bandwidth
      fi
    endwhile
  fi
  (* Use linear interpolation to calculate the bandwidth. *)
  bandwidth ← ((β² - (energy-p)) / p + (upperindex-lowerindex)) × freqbase
  return bandwidth
end
```

```
function essential_duration (waveform, α², wavesize, timebase, tolerance)
(* Essential duration for the discrete timeseries, waveform, with wavesize elements, and with resolution
of timebase seconds, for a given proportion of energy constrained (α²). Based on the definition for essential
duration for continuous timeseries by Landau and Pollak (1961). *)
  energy ← 0; total_energy ← 0; p ← 0; max_e_elt ← 0; max_e_idx ← 0
  call find_maximum_energy_time (waveform, wavesize, max_e_elt, max_e_idx)
  total_energy ← energy_time (waveform, wavesize)
  lowerindex ← max_e_idx; upperindex ← max_e_idx
  p ← sqr(waveform[max_e_idx]) / total_energy
  energy ← energy + p
  if (energy + tolerance) ⩾ α² then
    (* Required proportion of energy is contained in max_e_idx, so interpolate to get duration. *)
    duration ← α² / energy × timebase
    return duration
  else
    while (energy + tolerance) < α² do
      if (lowerindex > 1) and (upperindex < wavesize) then
        (* Sum bin with most energy from either side of max_e_idx. *)
        if sqr(waveform[lowerindex - 1]) > sqr(waveform[upperindex +1]) then
          lowerindex ← lowerindex +1
          p ← sqr(waveform[lowerindex]) / total_energy
          energy ← energy + p
        else
          upperindex ← upperindex + 1
          p ← sqr(waveform[upperindex]) / total_energy
          energy ← energy + p
        fi
      else if (lowerindex = 1) and (upperindex < wavesize) then
        (* If the lowest index is reached then sum only from the upper index. *)
        upperindex ← upperindex +1
        p ← sqr(waveform[upperindex]) / total_energy
        energy ← energy + p
      else if (lowerindex > 1) and (upperindex = wavesize) then
        (* If the highest index is reached then sum only from the lower index. *)
        lowerindex ← lowerindex+1
        p ← sqr(waveform[lowerindex]) / total_energy
        energy ← energy + p
      else if (lowerindex = 1) and (upperindex = wavesize) then
        duration ← wavesize × timebase
        return duration
      fi
    endwhile
  fi
  (* Use linear interpolation to calculate the duration. *)
  duration ← ((α² - (energy-p)) / p + (upperindex-lowerindex)) × timebase
  return duration
end
```

```
function energy_frequency (wavespectrum, spectrumsize)
(* Assumes input is a discrete energy spectrum, wavespectrum, which contains spectrumsize elements.
Energy at 0 Hz (DC) is included. *)
  energy ← 0
  for i ← 0 to spectrumsize do
    energy ← energy + wavespectrum[i]
  return energy
end

function energy_time (waveform, wavesize)
(* Assumes input is a discrete time series, waveform, of wavesize elements. *)
  energy ← 0
  for i ← 1 to wavesize do
    energy ← energy + sqr(waveform[i])
  return energy
end

procedure find_maximum_energy_freq (wavespectrum, spectrumsize, max_e_elt, max_e_idx)
(* Assumes input is a discrete energy spectrum, wavespectrum, of spectrumsize elements. *)
  max_e_elt ← wavespectrum[0]
  max_e_idx ← 0
  for i ← 1 to spectrumsize do
    if wavespectrum[i] > max_e_elt then
      max_e_elt ← wavespectrum[i]
      max_e_idx ← i
    fi
end

procedure find_maximum_energy_time (waveform, wavesize, max_e_elt, max_e_idx)
(* Assumes input is a discrete timeseries, waveform, of wavesize elements. *)
  max_e_elt ← sqr(waveform[1])
  max_e_idx ← 1
  for i ← 2 to wavesize do
    if sqr(waveform[i]) > max_e_elt then
      max_e_elt ← sqr(waveform[i])
      max_e_idx ← i
    fi
end
```

Appendix C

Signal analysis

This appendix describes the properties of the waveforms generated for the experiments and simulations. The analyses are described in more detail in Chapter 3.

Properties of the Kaiser window: The standardised $\mathrm{Ess}\mathcal{T}_{\%}$, as a function of κ, is tabulated in tables C.1 and C.2 for α^2=5–100%, and is shown graphically in Figure 3.11.

Signal generation parameters: Tables C.3 and C.4 show the necessary parameters required to generate signals using SIGGEN. All other parameters are derivable from these values.

Transient bandwidths, durations, and $\mathcal{WT}$: Table C.5 shows the measured bandwidths, durations, and bandwidth–duration products of the transient waveforms for a variety of definitions. See Section 3.3.3 for more detail.

Power spectra: Figures C.1 and C.2 show the spectral–averaged power spectra for the premixed $\mathcal{S}$ and $\mathcal{N}$ transients. The spectra are displayed in dB coordinates which were scaled by the maximum power within a signal set.

Roll–off of $\mathcal{S}$ and $\mathcal{N}$ transition bands: Table C.6 shows the roll–off, in decibels per octave, of the transition band of the spectral–averaged transient power spectra. Roll–offs are calculated for both the $\mathcal{S}$ and $\mathcal{N}$ signal sets between -3 dB and -60 dB. The upper transition band was used in both cases.

Descriptive statistics of transients: Tables C.7 and C.8 present the descriptive statistics (mean, standard deviation, skewness, and kurtosis) of each transient signal set, before and after scaling.

Correlation tests on transients: Tables C.9 and C.10 show the results of the correlation tests described in Section 3.3.5.

Signal–to–noise ratio estimates: Tables C.11, C.12, and C.13 give the signal–to–noise ratio estimates of the $\mathcal{SN}$ transients based on the bandwidths of the $\mathcal{S}$ and $\mathcal{N}$ transients, for a variety of bandwidth definitions. Tables C.14, C.15, and C.16 give the signal–to–noise ratio estimates of the $\mathcal{SN}$ buffers based on the measurements made on the spectral–averaged $\mathcal{SN}$ power spectra.

Table C.1: The $\mathrm{Ess}\mathcal{T}_{\%}$ of the Kaiser window, as a function of κ, and as a proportion of absolute duration. Each column represents a different proportion of energy constrained for the essential duration from $\mathrm{Ess}\mathcal{T}_{5\%}$, increasing in steps of 5% to $\mathrm{Ess}\mathcal{T}_{50\%}$.

κ	$\mathrm{Ess}\mathcal{T}_{5\%}$	$\mathrm{Ess}\mathcal{T}_{10\%}$	$\mathrm{Ess}\mathcal{T}_{15\%}$	$\mathrm{Ess}\mathcal{T}_{20\%}$	$\mathrm{Ess}\mathcal{T}_{25\%}$	$\mathrm{Ess}\mathcal{T}_{30\%}$	$\mathrm{Ess}\mathcal{T}_{35\%}$	$\mathrm{Ess}\mathcal{T}_{40\%}$	$\mathrm{Ess}\mathcal{T}_{45\%}$	$\mathrm{Ess}\mathcal{T}_{50\%}$
0	0.05000	0.10000	0.15000	0.20000	0.25000	0.30000	0.35000	0.40000	0.45000	0.50000
1	0.04329	0.08666	0.13017	0.17390	0.21793	0.26235	0.30724	0.35269	0.39881	0.44571
2	0.03315	0.06641	0.09987	0.13365	0.16786	0.20263	0.23811	0.27446	0.31188	0.35059
3	0.02662	0.05333	0.08022	0.10740	0.13498	0.16307	0.19180	0.22134	0.25186	0.28361
4	0.02273	0.04554	0.06851	0.09174	0.11532	0.13935	0.16396	0.18928	0.21548	0.24277
5	0.02018	0.04043	0.06084	0.08147	0.10242	0.12378	0.14566	0.16819	0.19153	0.21585
6	0.01835	0.03676	0.05532	0.07408	0.09314	0.11258	0.13249	0.15301	0.17426	0.19643
7	0.01695	0.03395	0.05109	0.06842	0.08603	0.10399	0.12240	0.14136	0.16102	0.18153
8	0.01582	0.03171	0.04771	0.06390	0.08035	0.09712	0.11432	0.13205	0.15042	0.16960
9	0.01490	0.02986	0.04493	0.06018	0.07566	0.09147	0.10767	0.12437	0.14169	0.15977
10	0.01412	0.02830	0.04258	0.05704	0.07172	0.08670	0.10206	0.11790	0.13432	0.15147
11	0.01346	0.02696	0.04057	0.05434	0.06833	0.08261	0.09725	0.11235	0.12800	0.14435
12	0.01287	0.02580	0.03882	0.05200	0.06539	0.07905	0.09307	0.10751	0.12250	0.13816
13	0.01236	0.02477	0.03728	0.04994	0.06279	0.07592	0.08938	0.10326	0.11765	0.13269
14	0.01191	0.02386	0.03591	0.04810	0.06049	0.07313	0.08610	0.09947	0.11334	0.12783
15	0.01150	0.02305	0.03468	0.04645	0.05842	0.07063	0.08315	0.09607	0.10947	0.12347

Table C.2: The proportion of absolute duration for the essential duration of the Kaiser window as a function of κ. Each column represents a different proportion of energy constrained for the essential duration from $\mathrm{Ess}\mathcal{T}_{55\%}$, increasing in steps of 5% to $\mathrm{Ess}\mathcal{T}_{100\%}$.

κ	$\mathrm{Ess}\mathcal{T}_{55\%}$	$\mathrm{Ess}\mathcal{T}_{60\%}$	$\mathrm{Ess}\mathcal{T}_{65\%}$	$\mathrm{Ess}\mathcal{T}_{70\%}$	$\mathrm{Ess}\mathcal{T}_{75\%}$	$\mathrm{Ess}\mathcal{T}_{80\%}$	$\mathrm{Ess}\mathcal{T}_{85\%}$	$\mathrm{Ess}\mathcal{T}_{90\%}$	$\mathrm{Ess}\mathcal{T}_{95\%}$	$\mathrm{Ess}\mathcal{T}_{100\%}$
0	0.55000	0.60000	0.65000	0.70000	0.75000	0.80000	0.85000	0.90000	0.95000	1.00000
1	0.49353	0.54240	0.59250	0.64401	0.69718	0.75226	0.80961	0.86963	0.93286	1.00000
2	0.39088	0.43311	0.47775	0.52542	0.57701	0.63381	0.69790	0.77293	0.86648	1.00000
3	0.31685	0.35196	0.38944	0.42998	0.47460	0.52490	0.58371	0.65675	0.75951	1.00000
4	0.27141	0.30173	0.33419	0.36942	0.40840	0.45263	0.50486	0.57079	0.66688	1.00000
5	0.24139	0.26847	0.29750	0.32908	0.36409	0.40394	0.45120	0.51126	0.59990	1.00000
6	0.21973	0.24445	0.27097	0.29986	0.33193	0.36853	0.41203	0.46754	0.55007	1.00000
7	0.20309	0.22599	0.25058	0.27737	0.30716	0.34119	0.38172	0.43359	0.51110	1.00000
8	0.18978	0.21120	0.23423	0.25933	0.28727	0.31921	0.35732	0.40618	0.47946	0.99999
9	0.17879	0.19900	0.22073	0.24443	0.27082	0.30102	0.33709	0.38341	0.45309	0.99965
10	0.16952	0.18870	0.20933	0.23185	0.25692	0.28564	0.31996	0.36411	0.43064	0.99544
11	0.16157	0.17986	0.19954	0.22103	0.24497	0.27240	0.30522	0.34746	0.41124	0.98305
12	0.15464	0.17216	0.19101	0.21160	0.23455	0.26086	0.29234	0.33291	0.39425	0.96627
13	0.14853	0.16537	0.18349	0.20329	0.22535	0.25067	0.28097	0.32004	0.37920	0.94811
14	0.14310	0.15933	0.17679	0.19588	0.21717	0.24158	0.27083	0.30857	0.36576	0.92964
15	0.13822	0.15390	0.17078	0.18923	0.20981	0.23343	0.26172	0.29824	0.35365	0.91131

Table C.3: Input to SIGGEN IFFT and resulting SIGGEN parameters for all $\mathcal{S}$ transient sets.

Signal set	Bandwidth Input (Hz)			IFFT			Points[a]	Comp.[b]	Chopping Index		Points between signals		
$\mathcal{S}$ transients	Low	High	Width	Nr.	Size	Sigs / IFFT			First	Last	Min	Ave	Max
400ms, 2.5Hz	499.75	500.25	0.50	278	2^{22}	9	42787	48	357013	4194303	383390	426177	468964
200ms, 5Hz	499.50	500.50	1.00	132	2^{22}	19	21393	96	176594	4194303	178043	199436	220829
100ms, 10Hz	499.00	501.00	2.00	66	2^{22}	38	10697	192	87835	4194303	89013	99710	110407
50ms, 20Hz	499.00	501.00	2.00	60	2^{22}	42	5348	192	87835	4194303	89200	94548	99896
25ms, 40Hz	498.00	502.00	4.00	30	2^{22}	84	2674	382	44035	4194303	44590	47264	49938
12.5ms, 80Hz	496.00	504.00	8.00	15	2^{22}	169	1337	762	22046	4194303	22144	23481	24818
6.25ms, 160Hz	496.00	504.00	16.00	14	2^{22}	179	669	762	22046	4194303	22094	22763	23432
400ms, 5Hz	497.75	502.25	4.50	76	2^{22}	33	42787	429	39199	4194303	41598	84385	127172
200ms, 10Hz	495.60	504.40	8.80	38	2^{22}	66	21393	838	20044	4194303	20775	42168	63561
100ms, 20Hz	490.99	509.00	18.01	19	2^{22}	134	10697	1713	9800	4194303	9907	20604	31301
50ms, 40Hz	482.01	518.00	35.99	10	2^{22}	268	5348	3424	4901	4194303	4954	10302	15650
25ms, 80Hz	464.00	536.00	72.00	5	2^{22}	537	2674	6848	2450	4194303	2462	5136	7810
12.5ms, 160Hz	429.00	571.00	142.00	3	2^{22}	1337	1337	13506	1242	4194303	1242	2579	3916
400ms, 10Hz	494.75	505.25	11.00	63	2^{22}	40	42787	1000	16794	4194303	19347	62134	104921
200ms, 20Hz	489.50	510.50	21.00	31	2^{22}	81	21393	1998	8401	4194303	9002	30395	51788
100ms, 40Hz	479.00	521.00	42.00	16	2^{22}	163	10697	3996	4200	4194303	4338	15035	25732
50ms, 80Hz	458.00	542.00	84.00	8	2^{22}	327	5348	7990	2100	4194303	2130	7478	12826
25ms, 160Hz	416.00	584.00	168.00	4	2^{22}	655	2674	15980	1050	4194303	1055	3729	6403

[a] Per transient.
[b] Total sinusoidal components included in IFFT.

Table C.4: Input to SIGGEN IFFT and resulting SIGGEN parameters for all $\mathcal{N}$ transient sets.

Signal sets	Bandwidth Input (Hz)			IFFT			Points[a]	Comp.[b]	Chopping Index		Points between signals		
$\mathcal{N}$ transients	Low	High	Width	Nr.	Size	Sigs / IFFT			First	Last	Min	Ave	Max
400ms, 4000Hz	0	4327.44	4327.44	63	2^{22}	48	42787	411579	510	4194303	1834	44621	87408
200ms, 4000Hz	0	4325.88	4325.88	32	2^{22}	96	21393	411431	510	4194303	908	22301	43694
100ms, 4000Hz	0	4322.75	4322.75	16	2^{22}	191	10697	411133	510	4194303	566	11263	21960
50ms, 4000Hz	0	4316.49	4316.49	9	2^{22}	374	5348	410538	511	4194303	518	5866	11214
25ms, 4000Hz	0	4329.00	4329.00	5	2^{22}	716	2674	411727	509	4194303	509	3183	5857
12.5ms, 4000Hz	0	4329.00	4329.00	3	2^{22}	1317	1337	411728	509	4194303	510	1847	3184
6.25ms, 4000Hz	0	4329.00	4329.00	2	2^{22}	2270	669	411728	509	4194303	509	1178	1847
400ms, 4000Hz	0	4329.00	4329.00	63	2^{22}	48	42787	411728	509	4194303	1834	44621	87408
200ms, 4000Hz	0	4329.00	4329.00	32	2^{22}	96	21393	411728	509	4194303	908	22301	43694
100ms, 4000Hz	0	4329.00	4329.00	16	2^{22}	191	10697	411728	509	4194303	566	11263	21960
50ms, 4000Hz	0	4329.00	4329.00	9	2^{22}	374	5348	411728	509	4194303	518	5866	11214
25ms, 4000Hz	0	4329.00	4329.00	5	2^{22}	716	2674	411728	509	4194303	509	3183	5857
12.5ms, 4000Hz	0	4329.00	4329.00	3	2^{22}	1317	1337	411728	509	4194303	510	1847	3184
400ms, 4000Hz	0	4329.00	4329.00	63	2^{22}	48	42787	411728	509	4194303	1834	44621	87408
200ms, 4000Hz	0	4329.00	4329.00	32	2^{22}	96	21393	411728	509	4194303	908	22301	43694
100ms, 4000Hz	0	4329.00	4329.00	16	2^{22}	191	10697	411728	509	4194303	566	11263	21960
50ms, 4000Hz	0	4329.00	4329.00	9	2^{22}	374	5348	411728	509	4194303	518	5866	11214
25ms, 4000Hz	0	4329.00	4329.00	5	2^{22}	716	2674	411728	509	4194303	509	3183	5857

[a] Per transient.
[b] Total sinusoidal components included in IFFT.

Table C.5: Estimates of $\mathcal{WT}$ for the $\mathcal{S}$ transients, based on various definitions of bandwidth and duration.

Signal Set	Bandwidth $\mathcal{W}$ (Hz), duration $\mathcal{T}$ (ms), and $\mathcal{WT}$ estimates														
	$\text{Ess}_{92.4\%}$			ERB–ERD			AERB–ERD			$\text{Ess}_{95\%}$			TdB–ERD		
	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 2.5Hz	2.56	399.99	1.02	1.85	399.38	0.74	1.86	399.38	0.74	2.82	439.59	1.24	1.75	399.38	0.70
200ms, 5Hz	5.12	199.99	1.02	3.69	199.68	0.74	3.82	199.68	0.76	5.63	219.78	1.24	3.50	199.68	0.70
100ms, 10Hz	10.25	99.99	1.02	7.39	99.84	0.74	7.39	99.84	0.74	11.27	109.89	1.24	6.98	99.84	0.70
50ms, 20Hz	20.20	49.99	1.01	14.57	49.91	0.73	14.57	49.91	0.73	22.17	54.94	1.22	13.81	49.91	0.69
25ms, 40Hz	40.36	24.99	1.01	29.13	24.95	0.73	29.13	24.95	0.73	44.36	27.46	1.22	27.63	24.95	0.69
12.5ms, 80Hz	80.74	12.49	1.01	58.29	12.47	0.73	58.32	12.47	0.73	88.74	13.73	1.22	55.29	12.47	0.69
6.25ms, 160Hz	160.85	6.24	1.00	116.17	6.24	0.72	116.22	6.24	0.72	176.79	6.86	1.21	110.22	6.24	0.69
400ms, 5Hz	5.02	399.99	2.01	4.39	399.38	1.75	4.42	399.38	1.76	5.40	439.59	2.37	4.40	399.38	1.76
200ms, 10Hz	9.94	199.99	1.99	8.98	199.68	1.79	9.06	199.68	1.81	10.68	219.78	2.35	8.93	199.68	1.78
100ms, 20Hz	20.14	99.99	2.01	18.21	99.84	1.82	18.55	99.84	1.85	21.65	109.89	2.38	18.13	99.84	1.81
50ms, 40Hz	40.19	49.99	2.01	35.57	49.91	1.78	35.80	49.91	1.79	43.22	54.94	2.37	35.67	49.91	1.78
25ms, 80Hz	80.69	24.99	2.02	72.91	24.95	1.82	73.18	24.95	1.83	86.71	27.46	2.38	72.64	24.95	1.81
12.5ms, 160Hz	159.46	12.49	1.99	140.69	12.47	1.75	141.45	12.47	1.76	171.54	13.73	2.35	141.09	12.47	1.76
400ms, 10Hz	10.09	399.99	4.04	10.21	399.38	4.08	10.48	399.38	4.19	10.59	439.59	4.66	10.43	399.38	4.17
200ms, 20Hz	20.16	199.99	4.03	20.08	199.68	4.01	20.90	199.68	4.17	21.17	219.78	4.65	20.78	199.68	4.15
100ms, 40Hz	40.24	99.99	4.02	40.22	99.84	4.02	41.42	99.84	4.13	42.27	109.89	4.65	41.48	99.84	4.14
50ms, 80Hz	80.48	49.99	4.02	80.28	49.91	4.01	83.81	49.91	4.18	84.53	54.94	4.64	82.95	49.91	4.14
25ms, 160Hz	160.97	24.99	4.02	162.95	24.95	4.07	166.69	24.95	4.16	169.06	27.46	4.64	166.38	24.95	4.15

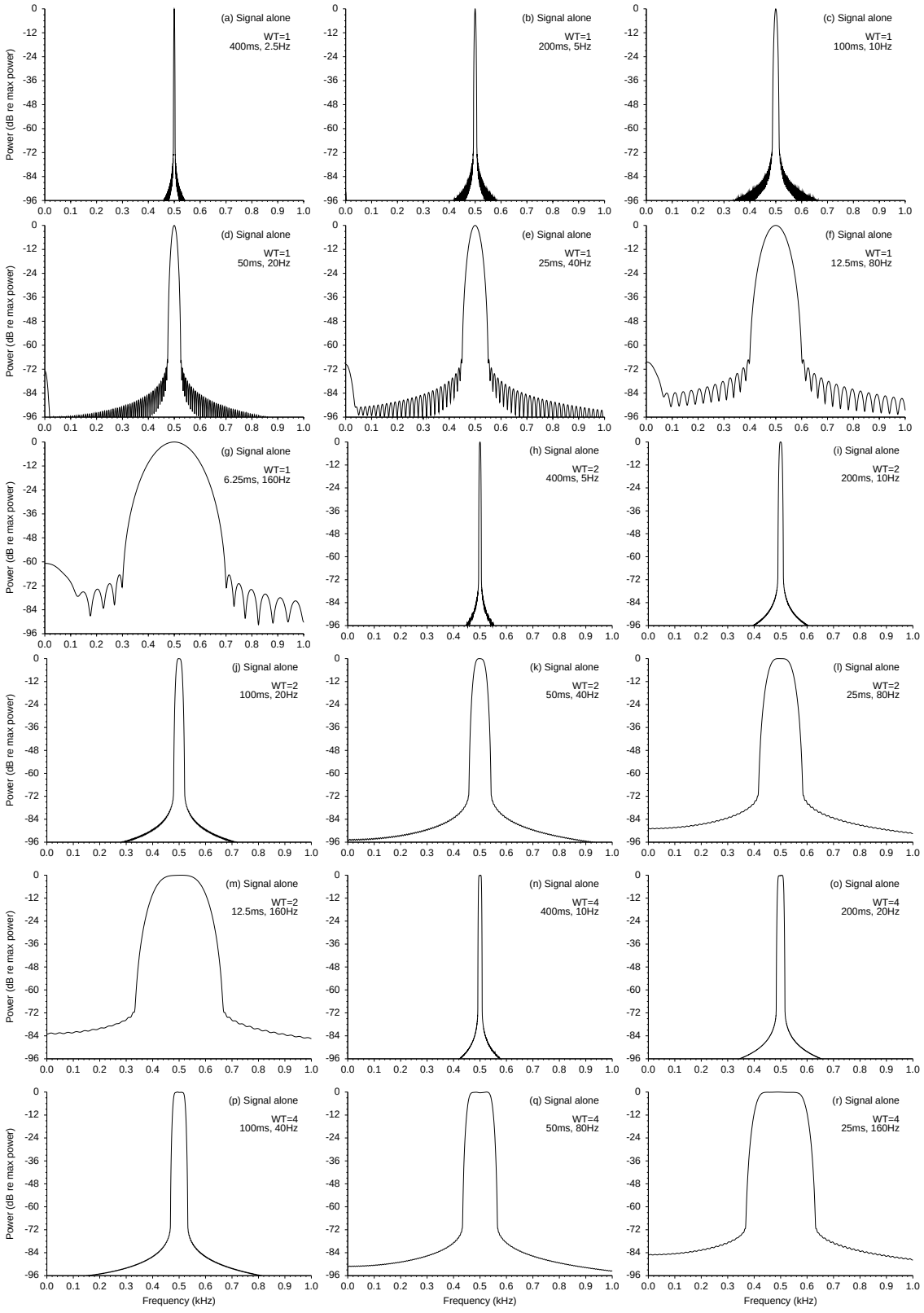

Figure C.1: S power spectra in dB (relative to maximum power) for each experimental level.

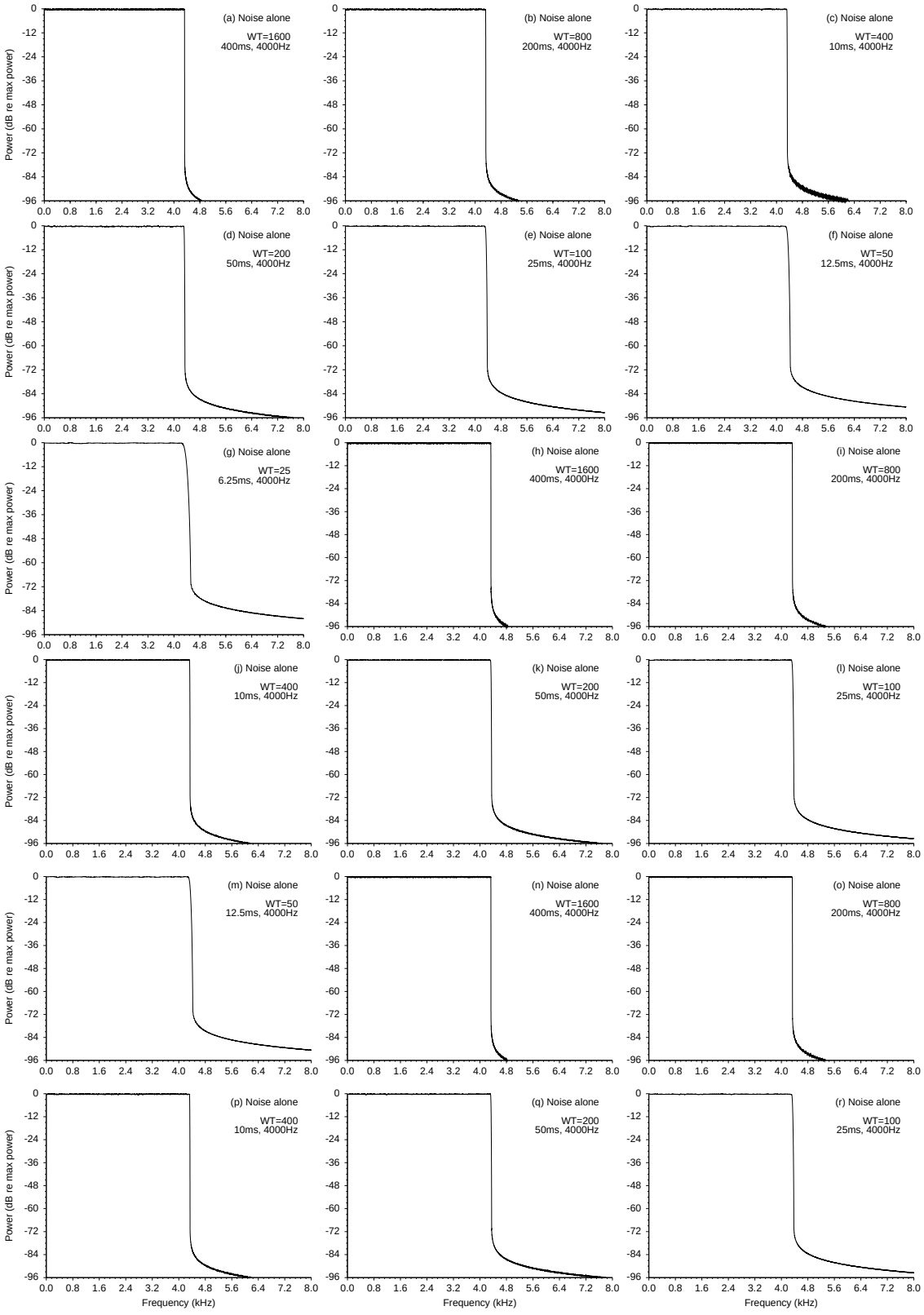

Figure C.2: $\mathcal{N}$ power spectra in dB (relative to maximum power) for each experimental level.

Table C.6: Transition band roll–off, in dB / octave, of $\mathcal{S}$ and $\mathcal{N}$ spectral–averaged transient power spectra.

Signal Set	-3dB freq. (Hz)	-60dB freq. (Hz)	Roll–off (dB / octave)
400ms, 2.5Hz	500.87	503.17	8637.99
200ms, 5Hz	501.76	506.00	4690.37
100ms, 10Hz	503.48	512.67	2183.55
50ms, 20Hz	506.90	524.57	1152.86
25ms, 40Hz	513.81	549.17	593.71
12.5ms, 80Hz	527.61	598.34	314.04
6.25ms, 160Hz	555.18	694.64	176.30
400ms, 5Hz	502.22	505.09	6937.10
200ms, 10Hz	504.48	510.09	3569.66
100ms, 20Hz	509.09	520.37	1803.50
50ms, 40Hz	517.69	540.73	907.23
25ms, 80Hz	536.15	581.55	486.12
12.5ms, 160Hz	570.58	662.09	265.61
400ms, 10Hz	505.23	508.09	6990.15
200ms, 20Hz	510.40	516.19	3501.15
100ms, 40Hz	520.76	532.36	1793.68
50ms, 80Hz	541.64	564.74	946.18
25ms, 160Hz	583.03	629.52	514.94
400ms, 4000Hz	4327.36	4330.28	58598.09
200ms, 4000Hz	4325.39	4331.23	29262.67
100ms, 4000Hz	4322.52	4334.09	14784.26
50ms, 4000Hz	4316.05	4339.21	7382.65
25ms, 4000Hz	4327.65	4374.50	3669.40
12.5ms, 4000Hz	4326.57	4420.08	1847.71
6.25ms, 4000Hz	4327.04	4511.19	948.03
400ms, 4000Hz	4328.96	4331.85	59163.10
200ms, 4000Hz	4328.90	4334.69	29572.11
100ms, 4000Hz	4328.70	4340.37	14670.20
50ms, 4000Hz	4328.57	4351.74	7399.81
25ms, 4000Hz	4327.82	4374.47	3685.04
12.5ms, 4000Hz	4327.15	4419.84	1864.24
400ms, 4000Hz	4328.95	4331.85	59059.69
200ms, 4000Hz	4328.88	4334.69	29445.51
100ms, 4000Hz	4328.79	4340.36	14801.86
50ms, 4000Hz	4328.74	4351.74	7453.42
25ms, 4000Hz	4328.38	4374.45	3732.04

Table C.7: Descriptive statistics of the $\mathcal{S}$ transients.

	Original		Rescaled				
$\mathcal{S}$ transients	Mean	SD	Mean	SD	Skewness	Kurtosis	N
400ms, 2.5Hz	−0.00000	0.00277	0.00000	1.00000	0.00000	3.03247	106967500
200ms, 5Hz	−0.00000	0.00396	0.00000	1.00000	−0.00000	2.99468	53482500
100ms, 10Hz	0.00000	0.00558	0.00000	1.00000	0.00011	3.03102	26742500
50ms, 20Hz	−0.00000	0.00548	0.00000	1.00000	−0.00021	2.99095	13370000
25ms, 40Hz	−0.00000	0.00789	0.00000	1.00000	−0.00009	2.95012	6685000
12.5ms, 80Hz	0.00000	0.01107	0.00000	1.00000	−0.00067	2.94173	3342500
6.25ms, 160Hz	0.00000	0.01097	0.00000	1.00000	−0.00265	3.01035	1672500
400ms, 5Hz	0.00000	0.00831	0.00000	1.00000	0.00000	2.99832	106967500
200ms, 10Hz	0.00000	0.01158	0.00000	1.00000	0.00001	2.99543	53482500
100ms, 20Hz	0.00000	0.01646	0.00000	1.00000	0.00014	2.99807	26742500
50ms, 40Hz	0.00000	0.02326	0.00000	1.00000	0.00010	2.98556	16044000
25ms, 80Hz	0.00000	0.03309	0.00000	1.00000	0.00015	2.99650	6685000
12.5ms, 160Hz	0.00001	0.04626	0.00000	1.00000	0.00029	2.98804	3342500
400ms, 5Hz	0.00000	0.01264	0.00000	1.00000	0.00001	2.96158	106967500
200ms, 10Hz	—	—	0.00000	1.00000	0.00000	2.99519	53482500
100ms, 20Hz	0.00000	0.02522	0.00000	1.00000	−0.00003	2.97760	26742500
50ms, 40Hz	0.00000	0.03562	0.00000	1.00000	−0.00001	3.00250	13370000
25ms, 80Hz	0.00000	0.05057	0.00000	1.00000	0.00014	3.00517	6685000

Table C.8: Descriptive statistics of the $\mathcal{N}$ transients.

	Original		Rescaled				
$\mathcal{N}$ transients	Mean	SD	Mean	SD	Skewness	Kurtosis	N
400ms, 4000Hz	−0.00004	0.25577	0.00000	1.00000	−0.00088	3.00032	128361000
200ms, 4000Hz	−0.00000	0.25577	0.00000	1.00000	0.00038	3.00028	64179000
100ms, 4000Hz	0.00015	0.25569	0.00000	1.00000	−0.00132	2.99821	32091000
50ms, 4000Hz	−0.00012	0.25545	0.00000	1.00000	0.00088	2.99743	16044000
25ms, 4000Hz	−0.00020	0.25597	0.00000	1.00000	0.00092	3.00551	8022000
12.5ms, 4000Hz	−0.00051	0.25587	0.00000	1.00000	0.00051	2.99708	4011000
6.25ms, 4000Hz	0.00003	0.25570	0.00000	1.00000	0.00408	2.99602	2007000
400ms, 4000Hz	−0.00002	0.25581	0.00000	1.00000	−0.00034	2.99933	128361000
200ms, 4000Hz	0.00007	0.25586	0.00000	1.00000	0.00034	2.99913	64179000
100ms, 4000Hz	−0.00010	0.25593	0.00000	1.00000	−0.00023	2.99684	32091000
50ms, 4000Hz	0.00002	0.25572	0.00000	1.00000	−0.00011	2.99887	16044000
25ms, 4000Hz	−0.00028	0.25569	0.00000	1.00000	0.00337	2.99507	8022000
12.5ms, 4000Hz	0.00021	0.25568	0.00000	1.00000	−0.00154	2.99930	4011000
400ms, 4000Hz	−0.00004	0.25583	0.00000	1.00000	0.00055	3.00035	128361000
200ms, 4000Hz	0.00003	0.25585	0.00000	1.00000	−0.00035	3.00051	64179000
100ms, 4000Hz	0.00000	0.25587	0.00000	1.00000	−0.00008	3.00292	32091000
50ms, 4000Hz	−0.00015	0.25578	0.00000	1.00000	−0.00089	2.99942	16044000
25ms, 4000Hz	−0.00001	0.25577	0.00000	1.00000	0.00147	3.00333	8022000

Table C.9: Correlation tests on the $\mathbb{S}$ transient sets.

Correlation Tests: overlapping, ordered pairs					
$\mathbb{S}$–transients	global-r	mean-r	z	Pairs	Total Points
400ms, 2.5Hz	0.009041	0.006410	93.488862	2499	106924713
200ms, 5Hz	0.002365	0.003787	17.290944	2499	53461107
100ms, 10Hz	−0.011461	−0.017752	−59.258716	2499	26731803
50ms, 20Hz	−0.018665	−0.016006	−68.243216	2499	13364652
25ms, 40Hz	0.002810	−0.005217	7.263397	2499	6682326
12.5ms, 80Hz	−0.006210	−0.004286	−11.351398	2499	3341163
6.25ms, 160Hz	0.012534	0.009464	16.207579	2499	1671831
400ms, 5Hz	0.004018	0.004681	41.552021	2499	106924713
200ms, 10Hz	−0.009435	−0.009483	−68.991676	2499	53461107
100ms, 20Hz	0.007681	0.006646	39.714298	2499	26731803
50ms, 40Hz	−0.003022	−0.000184	−11.048327	2499	13364652
25ms, 80Hz	−0.004181	−0.007164	−10.808000	2499	6682326
12.5ms, 160Hz	0.006619	0.010813	12.098112	2499	3341163
400ms, 10Hz	0.000013	−0.000417	0.129883	2499	106924713
200ms, 20Hz	−0.006851	−0.006885	−50.094882	2499	53461107
100ms, 40Hz	−0.002584	−0.003514	−13.359034	2499	26731803
50ms, 80Hz	0.001925	0.003144	7.037321	2499	13364652
25ms, 160Hz	0.001540	0.001881	3.981042	2499	6682326
Correlation Tests: independent, random pairs					
$\mathbb{S}$–transients	global-r	mean-r	z	Pairs	Total Points
400ms, 2.5Hz	0.001886	−0.001182	13.796454	1250	53483750
200ms, 5Hz	0.001499	0.012477	7.750095	1250	26741250
100ms, 10Hz	0.008447	0.011354	30.888166	1250	13371250
50ms, 20Hz	−0.014088	−0.007312	−36.427075	1250	6685000
25ms, 40Hz	−0.015057	0.000369	−27.529466	1250	3342500
12.5ms, 80Hz	0.015791	0.001881	20.416230	1250	1671250
6.25ms, 160Hz	0.019031	0.006033	17.405548	1250	836250
400ms, 5Hz	0.009165	0.003151	67.029573	1250	53483750
200ms, 10Hz	0.003133	0.006375	16.199555	1250	26741250
100ms, 20Hz	0.004059	0.006529	14.840827	1250	13371250
50ms, 40Hz	−0.003192	−0.006472	−8.253628	1250	6685000
25ms, 80Hz	−0.005369	−0.009551	−9.815069	1250	3342500
12.5ms, 160Hz	0.016167	0.014601	20.902063	1250	1671250
400ms, 10Hz	−0.002087	−0.002028	−15.262937	1250	53483750
200ms, 20Hz	0.000689	0.002161	3.561387	1250	26741250
100ms, 40Hz	−0.005292	−0.006188	−19.351298	1250	13371250
50ms, 80Hz	0.004382	0.006183	11.329296	1250	6685000
25ms, 160Hz	0.010269	0.009506	18.774795	1250	3342500

Table C.10: Correlation tests on the $\mathcal{N}$ transient sets.

Correlation Tests: overlapping, ordered pairs					
$\mathcal{N}$–transients	global-r	mean-r	z	Pairs	Total Points
400ms, 4000Hz	−0.000178	−0.000182	−2.013098	2999	128318213
200ms, 4000Hz	0.000392	0.000383	3.137236	2999	64157607
100ms, 4000Hz	−0.000522	−0.000510	−2.953992	2999	32080303
50ms, 4000Hz	−0.000338	−0.000377	−1.353773	2999	16038652
25ms, 4000Hz	−0.000300	−0.000276	−0.849448	2999	8019326
12.5ms, 4000Hz	0.000556	0.000594	1.112982	2999	4009663
6.25ms, 4000Hz	−0.002237	−0.002154	−3.169221	2999	2006331
400ms, 4000Hz	−0.000865	−0.000864	−9.797942	2999	128318213
200ms, 4000Hz	0.000051	0.000050	0.408733	2999	64157607
100ms, 4000Hz	0.000561	0.000567	3.178341	2999	32080303
50ms, 4000Hz	−0.000246	−0.000217	−0.984548	2999	16038652
25ms, 4000Hz	−0.000628	−0.000670	−1.777344	2999	8019326
12.5ms, 4000Hz	−0.000142	−0.000163	−0.284549	2999	4009663
400ms, 4000Hz	−0.000101	−0.000103	−1.139714	2999	128318213
200ms, 4000Hz	−0.000404	−0.000401	−3.234474	2999	64157607
100ms, 4000Hz	0.000216	0.000195	1.224154	2999	32080303
50ms, 4000Hz	0.001046	0.001059	4.187269	2999	16038652
25ms, 4000Hz	−0.000206	−0.000193	−0.582270	2999	8019326
Correlation Tests: independent, random pairs					
$\mathcal{N}$–transients	global-r	mean-r	z	Pairs	Total Points
400ms, 4000Hz	−0.000509	−0.000506	−4.078443	1500	64180500
200ms, 4000Hz	0.000056	0.000061	0.315154	1500	32089500
100ms, 4000Hz	−0.000213	−0.000210	−0.853712	1500	16045500
50ms, 4000Hz	−0.000143	−0.000178	−0.405307	1500	8022000
25ms, 4000Hz	−0.000201	−0.000259	−0.401600	1500	4011000
12.5ms, 4000Hz	0.003686	0.003716	5.220354	1500	2005500
6.25ms, 4000Hz	−0.001506	−0.001378	−1.508233	1500	1003500
400ms, 4000Hz	−0.000134	−0.000131	−1.075034	1500	64180500
200ms, 4000Hz	0.000154	0.000147	0.870715	1500	32089500
100ms, 4000Hz	−0.000369	−0.000378	−1.478273	1500	16045500
50ms, 4000Hz	0.000553	0.000533	1.567580	1500	8022000
25ms, 4000Hz	0.001269	0.001226	2.542123	1500	4011000
12.5ms, 4000Hz	−0.001891	−0.001850	−2.678134	1500	2005500
400ms, 4000Hz	0.000197	0.000191	1.577698	1500	64180500
200ms, 4000Hz	0.000209	0.000214	1.181257	1500	32089500
100ms, 4000Hz	0.000448	0.000458	1.793088	1500	16045500
50ms, 4000Hz	0.000618	0.000689	1.749532	1500	8022000
25ms, 4000Hz	−0.000393	−0.000427	−0.786912	1500	4011000

Table C.11: Estimates of signal–to–noise ratio (dB) calculated from the $\mathcal{WT}$=1 transient spectra, using a variety of bandwidth estimates. Difference (Diff) is relative to ERB.

Transient Set	ERB	AERB	Diff	$\text{Ess}_{92.4\%}$	Diff	$\text{Ess}_{95\%}$	Diff	TdB	Diff
400ms, 2.5Hz	0.00000	0.00000	−0.00000	−1.76031	1.76031	−2.05311	2.05311	0.24343	−0.24343
	4.00000	4.00000	−0.00000	2.23969	1.76031	1.94689	2.05311	4.24343	−0.24343
	8.00000	8.00000	−0.00000	6.23969	1.76031	5.94689	2.05311	8.24343	−0.24343
	12.00000	12.00000	0.00000	10.23969	1.76031	9.94689	2.05311	12.24343	−0.24343
	16.00000	16.00000	−0.00000	14.23969	1.76031	13.94689	2.05311	16.24343	−0.24343
200ms, 5Hz	0.00000	−0.08760	0.08760	−1.76249	1.76249	−2.05360	2.05360	0.23592	−0.23592
	4.00000	3.91240	0.08760	2.23751	1.76249	1.94640	2.05360	4.23592	−0.23592
	8.00000	7.91240	0.08760	6.23751	1.76249	5.94640	2.05360	8.23592	−0.23592
	12.00000	11.91240	0.08760	10.23751	1.76249	9.94640	2.05360	12.23592	−0.23592
	16.00000	15.91240	0.08760	14.23751	1.76249	13.94640	2.05360	16.23592	−0.23592
100ms, 10Hz	0.00000	0.00000	0.00000	−1.76277	1.76277	−2.05663	2.05663	0.24544	−0.24544
	4.00000	4.00000	0.00000	2.23723	1.76277	1.94337	2.05663	4.24544	−0.24544
	8.00000	8.00000	−0.00000	6.23723	1.76277	5.94337	2.05663	8.24544	−0.24544
	12.00000	12.00000	0.00000	10.23723	1.76277	9.94337	2.05663	12.24544	−0.24544
	16.00000	16.00000	−0.00000	14.23723	1.76277	13.94337	2.05663	16.24544	−0.24544
50ms, 20Hz	0.00000	−0.00000	0.00000	−1.76301	1.76301	−2.04711	2.04711	0.23112	−0.23112
	4.00000	4.00000	−0.00000	2.23699	1.76301	1.95289	2.04711	4.23112	−0.23112
	8.00000	8.00000	−0.00000	6.23699	1.76301	5.95289	2.04711	8.23112	−0.23112
	12.00000	12.00000	0.00000	10.23699	1.76301	9.95289	2.04711	12.23112	−0.23112
	16.00000	16.00000	−0.00000	14.23699	1.76301	13.95289	2.04711	16.23112	−0.23112
25ms, 40Hz	0.00000	0.00000	0.00000	−1.76215	1.76215	−2.05024	2.05024	0.22763	−0.22763
	4.00000	4.00000	0.00000	2.23785	1.76215	1.94976	2.05024	4.22763	−0.22763
	8.00000	8.00000	−0.00000	6.23785	1.76215	5.94976	2.05024	8.22763	−0.22763
	12.00000	12.00000	0.00000	10.23785	1.76215	9.94976	2.05024	12.22763	−0.22763
	16.00000	16.00000	−0.00000	14.23785	1.76215	13.94976	2.05024	16.22763	−0.22763
12.5ms, 80Hz	0.00000	−0.00265	0.00265	−1.76543	1.76543	−2.05269	2.05269	0.22762	−0.22762
	4.00000	3.99735	0.00265	2.23457	1.76543	1.94731	2.05269	4.22762	−0.22762
	8.00000	7.99735	0.00265	6.23457	1.76543	5.94731	2.05269	8.22762	−0.22762
	12.00000	11.99735	0.00265	10.23457	1.76543	9.94731	2.05269	12.22762	−0.22762
	16.00000	15.99735	0.00265	14.23457	1.76543	13.94731	2.05269	16.22762	−0.22762
6.25ms, 160Hz	0.00000	−0.00176	0.00176	−1.75729	1.75729	−2.04678	2.04678	0.22380	−0.22380
	4.00000	3.99824	0.00176	2.24271	1.75729	1.95322	2.04678	4.22380	−0.22380
	8.00000	7.99824	0.00176	6.24271	1.75729	5.95322	2.04678	8.22380	−0.22380
	12.00000	11.99824	0.00176	10.24271	1.75729	9.95322	2.04678	12.22380	−0.22380
	16.00000	15.99824	0.00176	14.24271	1.75729	13.95322	2.04678	16.22380	−0.22380

Table C.12: Estimates of signal–to–noise ratio (dB) calculated from the $\mathcal{WT}$=2 transient spectra, using a variety of bandwidth estimates. Difference (Diff) is relative to ERB.

Transient Set	ERB	AERB	Diff	$\text{Ess}_{92.4\%}$	Diff	$\text{Ess}_{95\%}$	Diff	TdB	Diff
400ms, 5Hz	−4.00000	−4.02425	0.02425	−4.92202	0.92202	−5.11946	1.11946	−4.01203	0.01203
	0.00000	−0.02425	0.02425	−0.92202	0.92202	−1.11946	1.11946	−0.01203	0.01203
	4.00000	3.97575	0.02425	3.07798	0.92202	2.88054	1.11946	3.98797	0.01203
	8.00000	7.97575	0.02425	7.07798	0.92202	6.88054	1.11946	7.98797	0.01203
	12.00000	11.97575	0.02425	11.07798	0.92202	10.88054	1.11946	11.98797	0.01203
200ms, 10Hz	−4.00000	−4.03878	0.03878	−4.78110	0.78110	−4.97612	0.97612	−3.97419	−0.02581
	0.00000	−0.03878	0.03878	−0.78110	0.78110	−0.97612	0.97612	0.02581	−0.02581
	4.00000	3.96122	0.03878	3.21890	0.78110	3.02388	0.97612	4.02581	−0.02581
	8.00000	7.96122	0.03878	7.21890	0.78110	7.02388	0.97612	8.02581	−0.02581
	12.00000	11.96122	0.03878	11.21890	0.78110	11.02388	0.97612	12.02581	−0.02581
100ms, 20Hz	−4.00000	−4.08050	0.08050	−4.78325	0.78325	−4.97577	0.97577	−3.98313	−0.01687
	0.00000	−0.08050	0.08050	−0.78325	0.78325	−0.97577	0.97577	0.01687	−0.01687
	4.00000	3.91950	0.08050	3.21675	0.78325	3.02423	0.97577	4.01687	−0.01687
	8.00000	7.91950	0.08050	7.21675	0.78325	7.02423	0.97577	8.01687	−0.01687
	12.00000	11.91950	0.08050	11.21675	0.78325	11.02423	0.97577	12.01687	−0.01687
50ms, 40Hz	−4.00000	−4.02727	0.02727	−4.87342	0.87342	−5.06880	1.06880	−4.01072	0.01072
	0.00000	−0.02727	0.02727	−0.87342	0.87342	−1.06880	1.06880	−0.01072	0.01072
	4.00000	3.97273	0.02727	3.12658	0.87342	2.93120	1.06880	3.98928	0.01072
	8.00000	7.97273	0.02727	7.12658	0.87342	6.93120	1.06880	7.98928	0.01072
	12.00000	11.97273	0.02727	11.12658	0.87342	10.93120	1.06880	11.98928	0.01072
25ms, 80Hz	−4.00000	−4.01628	0.01628	−4.78075	0.78075	−4.97474	0.97474	−3.98481	−0.01519
	0.00000	−0.01628	0.01628	−0.78075	0.78075	−0.97474	0.97474	0.01519	−0.01519
	4.00000	3.98372	0.01628	3.21925	0.78075	3.02526	0.97474	4.01519	−0.01519
	8.00000	7.98372	0.01628	7.21925	0.78075	7.02526	0.97474	8.01519	−0.01519
	12.00000	11.98372	0.01628	11.21925	0.78075	11.02526	0.97474	12.01519	−0.01519
12.5ms, 160Hz	−4.00000	−4.02355	0.02355	−4.88751	0.88751	−5.08463	1.08463	−4.01818	0.01818
	0.00000	−0.02355	0.02355	−0.88751	0.88751	−1.08463	1.08463	−0.01818	0.01818
	4.00000	3.97645	0.02355	3.11249	0.88751	2.91537	1.08463	3.98182	0.01818
	8.00000	7.97645	0.02355	7.11249	0.88751	6.91537	1.08463	7.98182	0.01818
	12.00000	11.97645	0.02355	11.11249	0.88751	10.91537	1.08463	11.98182	0.01818

Table C.13: Estimates of signal–to–noise ratio (dB) calculated from the $\mathcal{WT}$=4 transient spectra, using a variety of bandwidth estimates. Difference (Diff) is relative to ERB.

Transient Set	ERB	AERB	Diff	$\text{Ess}_{92.4\%}$	Diff	$\text{Ess}_{95\%}$	Diff	TdB	Diff
400ms, 10Hz	−8.00000	−8.11368	0.11368	−8.29249	0.29249	−8.38335	0.38335	−8.09454	0.09454
	−4.00000	−4.11368	0.11368	−4.29249	0.29249	−4.38335	0.38335	−4.09454	0.09454
	0.00000	−0.11368	0.11368	−0.29249	0.29249	−0.38335	0.38335	−0.09454	0.09454
	4.00000	3.88632	0.11368	3.70751	0.29249	3.61665	0.38335	3.90546	0.09454
	8.00000	7.88632	0.11368	7.70751	0.29249	7.61665	0.38335	7.90546	0.09454
200ms, 20Hz	−8.00000	−8.17270	0.17270	−8.35954	0.35954	−8.45186	0.45186	−8.14876	0.14876
	−4.00000	−4.17270	0.17270	−4.35954	0.35954	−4.45186	0.45186	−4.14876	0.14876
	0.00000	−0.17270	0.17270	−0.35954	0.35954	−0.45186	0.45186	−0.14876	0.14876
	4.00000	3.82730	0.17270	3.64046	0.35954	3.54814	0.45186	3.85124	0.14876
	8.00000	7.82730	0.17270	7.64046	0.35954	7.54814	0.45186	7.85124	0.14876
100ms, 40Hz	−8.00000	−8.12766	0.12766	−8.34633	0.34633	−8.43888	0.43888	−8.13290	0.13290
	−4.00000	−4.12766	0.12766	−4.34633	0.34633	−4.43888	0.43888	−4.13290	0.13290
	0.00000	−0.12766	0.12766	−0.34633	0.34633	−0.43888	0.43888	−0.13290	0.13290
	4.00000	3.87234	0.12766	3.65367	0.34633	3.56112	0.43888	3.86710	0.13290
	8.00000	7.87234	0.12766	7.65367	0.34633	7.56112	0.43888	7.86710	0.13290
50ms, 80Hz	−8.00000	−8.18692	0.18692	−8.35481	0.35481	−8.44753	0.44753	−8.14381	0.14381
	−4.00000	−4.18692	0.18692	−4.35481	0.35481	−4.44753	0.44753	−4.14381	0.14381
	0.00000	−0.18692	0.18692	−0.35481	0.35481	−0.44753	0.44753	−0.14381	0.14381
	4.00000	3.81308	0.18692	3.64519	0.35481	3.55247	0.44753	3.85619	0.14381
	8.00000	7.81308	0.18692	7.64519	0.35481	7.55247	0.44753	7.85619	0.14381
25ms, 160Hz	−8.00000	−8.09863	0.09863	−8.28949	0.28949	−8.38229	0.38229	−8.09341	0.09341
	−4.00000	−4.09863	0.09863	−4.28949	0.28949	−4.38229	0.38229	−4.09341	0.09341
	0.00000	−0.09863	0.09863	−0.28949	0.28949	−0.38229	0.38229	−0.09341	0.09341
	4.00000	3.90137	0.09863	3.71051	0.28949	3.61771	0.38229	3.90659	0.09341
	8.00000	7.90137	0.09863	7.71051	0.28949	7.61771	0.38229	7.90659	0.09341

Table C.14: Estimates of signal–to–noise ratio (dB) calculated from the $\mathcal{WT}$=1 buffer power spectra.

Signal Set	Desired SNR	Maximum SNR	Average SNR
400ms, 2.5Hz	0.00000	−0.55626	−0.55626
	4.00000	4.11807	4.11807
	8.00000	8.30068	8.30068
	12.00000	11.82199	11.82199
	16.00000	15.67593	15.67593
200ms, 5Hz	0.00000	−0.45463	−0.45463
	4.00000	4.08816	4.08816
	8.00000	7.87108	7.87108
	12.00000	12.03830	12.03830
	16.00000	16.14839	16.14839
100ms, 10Hz	0.00000	−1.28081	−1.28081
	4.00000	4.01578	4.01578
	8.00000	8.17273	8.17273
	12.00000	11.83220	11.83220
	16.00000	16.07561	16.07561
50ms, 20Hz	0.00000	0.24428	0.24189
	4.00000	3.23998	3.23998
	8.00000	8.01368	8.01368
	12.00000	11.77771	11.77771
	16.00000	16.17118	16.17118
25ms, 40Hz	0.00000	0.14908	0.14908
	4.00000	3.88117	3.87703
	8.00000	8.02929	8.02929
	12.00000	12.25730	12.25730
	16.00000	16.04753	16.04753
12.5ms, 80Hz	0.00000	0.23492	0.22505
	4.00000	3.87479	3.87119
	8.00000	7.64512	7.64076
	12.00000	12.07039	12.06823
	16.00000	16.07881	16.07555
6.25ms, 160Hz	0.00000	0.92455	0.91833
	4.00000	3.89781	3.89488
	8.00000	8.03902	8.03655
	12.00000	11.91046	11.90827
	16.00000	16.05588	16.05412

Table C.15: Estimates of signal–to–noise ratio (dB) calculated from the $\mathcal{WT}$=2 buffer power spectra.

Signal Set	Desired SNR	Maximum SNR	Average SNR
400ms, 5Hz	−4.00000	−3.91857	−4.35315
	0.00000	−0.08684	−0.27410
	4.00000	4.33541	4.21120
	8.00000	7.80223	7.77945
	12.00000	11.94647	11.93248
200ms, 10Hz	−4.00000	−3.90949	−4.12258
	0.00000	−0.15371	−0.25501
	4.00000	4.05235	3.94920
	8.00000	7.79132	7.63917
	12.00000	12.22139	12.07296
100ms, 20Hz	−4.00000	−3.31973	−3.73603
	0.00000	0.44085	0.20146
	4.00000	3.98791	3.78112
	8.00000	8.20215	8.09127
	12.00000	11.98688	11.84985
50ms, 40Hz	−4.00000	−3.15207	−3.35671
	0.00000	0.75757	0.61468
	4.00000	3.71265	3.68189
	8.00000	7.97039	7.94178
	12.00000	12.08796	12.04652
25ms, 80Hz	−4.00000	−3.96125	−4.41907
	0.00000	0.30519	0.01521
	4.00000	4.12651	4.04806
	8.00000	7.71912	7.57076
	12.00000	11.95396	11.88666
12.5ms, 160Hz	−4.00000	−3.43759	−3.63833
	0.00000	0.21261	−0.31269
	4.00000	4.53558	4.49110
	8.00000	7.83869	7.82192
	12.00000	11.97870	11.85426

Table C.16: Estimates of signal–to–noise ratio (dB) calculated from the $\mathcal{WT}$=4 buffer power spectra.

Signal Set	Desired SNR	Maximum SNR	Average SNR
400ms, 10Hz	−8.00000	−5.47017	−7.55798
	−4.00000	−2.66808	−3.53568
	0.00000	0.38567	0.07302
	4.00000	4.16302	3.88451
	8.00000	8.16745	8.03860
200ms, 20Hz	−8.00000	−6.00610	−7.60815
	−4.00000	−3.54685	−3.83918
	0.00000	0.34170	−0.15450
	4.00000	4.32710	4.15577
	8.00000	8.48315	8.07886
100ms, 40Hz	−8.00000	−5.28031	−7.57393
	−4.00000	−3.45046	−4.38874
	0.00000	0.34885	0.05788
	4.00000	4.17974	3.84169
	8.00000	8.12801	7.90032
50ms, 80Hz	−8.00000	−6.38848	−7.77954
	−4.00000	−3.96285	−4.52925
	0.00000	0.48639	−0.16830
	4.00000	4.03657	3.81618
	8.00000	8.33893	8.00105
25ms, 160Hz	−8.00000	−6.42417	−8.26007
	−4.00000	−3.58829	−4.27207
	0.00000	0.39972	−0.00037
	4.00000	4.29254	4.03224
	8.00000	8.13553	7.99451

Appendix D

ROC, mROC, and GOC Analyses

This appendix presents the results of the ROC, mROC, and GOC analyses described in Chapter 3. *Attenuated* theoretical ROC curves (dashed curves) in each mROC and GOC space are the $\mathcal{WT}$=1 χ^2 energy detector for the $\mathcal{WT}$=1 for experimental condition; the $\mathcal{WT}$=3 χ^2 energy detector, for experimental condition $\mathcal{WT}$=2; and the $\mathcal{WT}$=5 energy detector, for experimental condition $\mathcal{WT}$=4. The explanation for fitting these particular models, and the method for doing so, is given in Chapter 3.

Measures of detectability

Tables D.1–D.9 show the measures of detectability, $\mathcal{A}$ and $\mathcal{D}_2$: for the energy detector (for the theoretical and empirical signal–to–noise ratios), and the mROC, GOC, and FORCE analyses for each observer. Tables D.10–D.18 show the measures of detectability, $\mathcal{A}$ and $\mathcal{D}_2$, averaged over the six replications, and their standard deviations.

ROC curves

Figures D.1–D.7 show the six single–replication ROC curves at each of the five signal–to–noise ratios completed by each observer. Remember that the false–alarm rate is common within a level.

mROC curves

Figures D.8–D.14 show the mROC curves calculated from the six single–replication ROC curves at each of the five signal–to–noise ratios completed by each observer.

GOC curves

Figures D.15–D.21 show the GOC curves calculated from the six single–replication ROC curves at each of the five signal–to–noise ratios completed by each observer.

Table D.1: $\mathcal{A}$ and $\mathcal{D}_2$ for Observer 1 after mROC, GOC, and FORCE analyses in condition $\mathcal{WT}$=1 compared with the χ^2 energy detector for $\mathcal{WT}$=1.

Experimental Level	Theory SNR (dB)	Theory $\mathcal{A}$	Theory $\mathcal{D}$	Empirical SNR (dB)	Empirical $\mathcal{A}$	Empirical $\mathcal{D}$	mROC $\mathcal{A}$	GOC $\mathcal{A}$	FORCE $\mathcal{A}$	mROC $\mathcal{D}$	GOC $\mathcal{D}$	FORCE $\mathcal{D}$
400ms, 2.5Hz	0	0.6667	0.0817	−0.5563	0.6528	0.0684	0.5127	0.5223	0.5291	0.0005	0.0014	0.0171
	4	0.7784	0.2369	4.1181	0.7817	0.2430	0.5367	0.5593	0.5865	0.0039	0.0102	0.1418
	8	0.8797	0.4697	8.3007	0.8859	0.4877	0.6159	0.6820	0.7473	0.0391	0.0978	0.3154
	12	0.9440	0.6885	11.8220	0.9419	0.6802	0.7032	0.7920	0.8523	0.1227	0.2623	0.4929
	16	0.9761	0.8371	15.6759	0.9743	0.8278	0.7917	0.8860	0.9207	0.2618	0.4883	0.6447
200ms, 5Hz	0	0.6667	0.0817	−0.4546	0.6552	0.0707	0.5319	0.5494	0.5597	0.0029	0.0070	0.0112
	4	0.7784	0.2369	4.0882	0.7809	0.2414	0.5801	0.6179	0.6316	0.0186	0.0405	0.0543
	8	0.8797	0.4697	7.8711	0.8769	0.4619	0.6543	0.7078	0.7299	0.0699	0.1285	0.1683
	12	0.9440	0.6885	12.0383	0.9444	0.6903	0.7859	0.8592	0.8797	0.2508	0.4138	0.4866
	16	0.9761	0.8371	16.1484	0.9768	0.8412	0.8595	0.9233	0.9326	0.4143	0.6097	0.6536
100ms, 10Hz	0	0.6667	0.0817	−1.2808	0.6356	0.0538	0.5377	0.5533	0.5635	0.0041	0.0082	0.0135
	4	0.7784	0.2369	4.0158	0.7788	0.2377	0.6334	0.6785	0.6945	0.0520	0.0940	0.1176
	8	0.8797	0.4697	8.1727	0.8833	0.4801	0.7356	0.7982	0.8188	0.1667	0.2744	0.3294
	12	0.9440	0.6885	11.8322	0.9420	0.6806	0.8332	0.8947	0.9076	0.3496	0.5146	0.5659
	16	0.9761	0.8371	16.0756	0.9765	0.8392	0.9107	0.9510	0.9541	0.5658	0.7179	0.7340
50ms, 20Hz	0	0.6667	0.0817	0.2443	0.6730	0.0881	0.5655	0.6051	0.6325	0.0124	0.0321	0.0690
	4	0.7784	0.2369	3.2400	0.7566	0.1994	0.6134	0.6585	0.6852	0.0374	0.0738	0.1142
	8	0.8797	0.4697	8.0137	0.8799	0.4705	0.7434	0.7993	0.8120	0.1784	0.2767	0.3088
	12	0.9440	0.6885	11.7777	0.9414	0.6780	0.8456	0.8964	0.9061	0.3793	0.5196	0.5570
	16	0.9761	0.8371	16.1712	0.9770	0.8418	0.9330	0.9642	0.9671	0.6452	0.7774	0.7939
25ms, 40Hz	0	0.6667	0.0817	0.1491	0.6705	0.0856	0.5511	0.5841	0.6402	0.0075	0.0205	0.0000
	4	0.7784	0.2369	3.8770	0.7749	0.2306	0.6206	0.6767	0.7238	0.0424	0.0921	0.1992
	8	0.8797	0.4697	8.0293	0.8803	0.4714	0.7583	0.8326	0.8594	0.2022	0.3483	0.4360
	12	0.9440	0.6885	12.2573	0.9469	0.7004	0.8538	0.9104	0.9221	0.3998	0.5648	0.6143
	16	0.9761	0.8371	16.0475	0.9763	0.8384	0.9250	0.9550	0.9582	0.6155	0.7354	0.7519
12.5ms, 80Hz	0	0.6667	0.0817	0.2250	0.6725	0.0876	0.5532	0.5770	0.5949	0.0082	0.0172	0.0315
	4	0.7784	0.2369	3.8712	0.7747	0.2303	0.6127	0.6608	0.6924	0.0369	0.0759	0.1279
	8	0.8797	0.4697	7.6408	0.8719	0.4479	0.7188	0.7839	0.8095	0.1430	0.2469	0.3140
	12	0.9440	0.6885	12.0682	0.9448	0.6917	0.8392	0.8935	0.9068	0.3638	0.5108	0.5622
	16	0.9761	0.8371	16.0755	0.9765	0.8392	0.9252	0.9557	0.9603	0.6163	0.7383	0.7622
6.25ms, 160Hz	0	0.6667	0.0817	0.9183	0.6909	0.1079	0.5658	0.5993	0.6199	0.0125	0.0286	0.0494
	4	0.7784	0.2369	3.8949	0.7754	0.2315	0.6102	0.6573	0.6792	0.0353	0.0726	0.1041
	8	0.8797	0.4697	8.0366	0.8804	0.4719	0.7069	0.7663	0.7855	0.1273	0.2156	0.2596
	12	0.9440	0.6885	11.9083	0.9429	0.6842	0.8302	0.8861	0.8955	0.3429	0.4884	0.5226
	16	0.9761	0.8371	16.0541	0.9764	0.8386	0.9083	0.9432	0.9479	0.5577	0.6855	0.7071

Table D.2: $\mathcal{A}$ and $\mathcal{D}_2$ for Observer 1 after mROC, GOC, and FORCE analyses in condition $\mathcal{WT}$=2 compared with the χ^2 energy detector theory for $\mathcal{WT}$=3.

Experimental Level	Theory SNR (dB)	Theory $\mathcal{A}$	Theory $\mathcal{D}$	Empirical SNR (dB)	Empirical $\mathcal{A}$	Empirical $\mathcal{D}$	mROC $\mathcal{A}$	GOC $\mathcal{A}$	FORCE $\mathcal{A}$	mROC $\mathcal{D}$	GOC $\mathcal{D}$	FORCE $\mathcal{D}$
400ms, 5Hz	−4	0.6528	0.0685	−4.3532	0.6431	0.0599	0.5344	0.5596	0.6844	0.0034	0.0103	0.0000
	0	0.7901	0.2588	−0.2741	0.7797	0.2394	0.5541	0.5900	0.6488	0.0085	0.0235	0.0000
	4	0.9241	0.6125	4.2112	0.9294	0.6317	0.6337	0.7187	0.8399	0.0522	0.1428	7.4702
	8	0.9856	0.8911	7.7795	0.9839	0.8812	0.7338	0.8577	0.9657	0.1640	0.4097	1.4789
	12	0.9984	0.9827	11.9325	0.9983	0.9821	0.8340	0.9557	0.9999	0.3516	0.7385	1.1301
200ms, 10Hz	−4	0.6528	0.0685	−4.1226	0.6494	0.0654	0.5179	0.5231	0.5234	0.0009	0.0015	0.0000
	0	0.7901	0.2588	−0.2550	0.7805	0.2407	0.5823	0.6194	0.6469	0.0196	0.0415	0.0792
	4	0.9241	0.6125	3.9492	0.9228	0.6078	0.7035	0.7795	0.8159	0.1230	0.2389	0.3468
	8	0.9856	0.8911	7.6392	0.9828	0.8747	0.8182	0.9181	0.9533	0.3161	0.5913	0.7955
	12	0.9984	0.9827	12.0730	0.9985	0.9833	0.8950	0.9799	0.9936	0.5153	0.8579	0.9784
100ms, 20Hz	−4	0.6528	0.0685	−3.7360	0.6604	0.0756	0.5615	0.5862	0.6003	0.0110	0.0216	0.0324
	0	0.7901	0.2588	0.2015	0.7978	0.2736	0.6419	0.6958	0.7213	0.0589	0.1137	0.1585
	4	0.9241	0.6125	3.7811	0.9184	0.5923	0.7465	0.8199	0.8423	0.1832	0.3197	0.3873
	8	0.9856	0.8911	8.0913	0.9862	0.8950	0.8792	0.9454	0.9560	0.4683	0.6945	0.7524
	12	0.9984	0.9827	11.8498	0.9982	0.9813	0.9450	0.9909	0.9940	0.6927	0.9255	0.9546
50ms, 40Hz	−4	0.6528	0.0685	−3.3567	0.6717	0.0869	0.5794	0.6195	0.6479	0.0183	0.0416	0.0821
	0	0.7901	0.2588	0.6147	0.8134	0.3056	0.6390	0.7044	0.7481	0.0565	0.1242	0.2253
	4	0.9241	0.6125	3.6819	0.9158	0.5831	0.7650	0.8463	0.8729	0.2134	0.3811	0.4754
	8	0.9856	0.8911	7.9418	0.9851	0.8885	0.9081	0.9674	0.9740	0.5572	0.7929	0.8364
	12	0.9984	0.9827	12.0465	0.9984	0.9831	0.9675	0.9960	0.9968	0.7935	0.9627	0.9718
25ms, 80Hz	−4	0.6528	0.0685	−4.4191	0.6413	0.0584	0.5463	0.5738	0.6008	0.0062	0.0158	0.0500
	0	0.7901	0.2588	0.0152	0.7907	0.2599	0.6239	0.6800	0.7150	0.0448	0.0957	0.1639
	4	0.9241	0.6125	4.0481	0.9253	0.6169	0.7387	0.8202	0.8522	0.1713	0.3203	0.4269
	8	0.9856	0.8911	7.5708	0.9822	0.8713	0.8660	0.9393	0.9534	0.4318	0.6700	0.7468
	12	0.9984	0.9827	11.8867	0.9983	0.9817	0.9499	0.9911	0.9937	0.7132	0.9265	0.9521
12.5ms, 160Hz	−4	0.6528	0.0685	−3.6383	0.6633	0.0783	0.5407	0.5687	0.5922	0.0048	0.0137	0.0356
	0	0.7901	0.2588	−0.3127	0.7783	0.2367	0.6046	0.6582	0.6961	0.0318	0.0735	0.1444
	4	0.9241	0.6125	4.4911	0.9359	0.6566	0.7345	0.8180	0.8514	0.1650	0.3156	0.4257
	8	0.9856	0.8911	7.8219	0.9843	0.8832	0.8584	0.9311	0.9453	0.4117	0.6383	0.7110
	12	0.9984	0.9827	11.8543	0.9982	0.9814	0.9508	0.9895	0.9926	0.7171	0.9158	0.9442

Table D.3: $\mathcal{A}$ and $\mathcal{D}_2$ for Observer 1 after mROC, GOC, and FORCE analyses in condition $\mathcal{WT}$=4 compared with the χ^2 energy detector theory for $\mathcal{WT}$=5.

Experimental Level	Theory SNR (dB)	Theory $\mathcal{A}$	Theory $\mathcal{D}$	Empirical SNR (dB)	Empirical $\mathcal{A}$	Empirical $\mathcal{D}$	mROC $\mathcal{A}$	GOC $\mathcal{A}$	FORCE $\mathcal{A}$	mROC $\mathcal{D}$	GOC $\mathcal{D}$	FORCE $\mathcal{D}$
400ms, 10Hz	−8	0.5897	0.0233	−7.5580	0.5984	0.0281	0.5069	0.5118	0.0000	0.0001	0.0004	0.0000
	−4	0.6969	0.1150	−3.5357	0.7136	0.1360	0.5477	0.5840	0.6630	0.0066	0.0204	0.0000
	0	0.8552	0.4032	0.0730	0.8580	0.4106	0.5970	0.6728	0.8882	0.0273	0.0880	0.0000
	4	0.9700	0.8054	3.8845	0.9681	0.7961	0.6959	0.8179	0.9430	0.1137	0.3154	2.3773
	8	0.9979	0.9784	8.0386	0.9980	0.9790	0.8125	0.9506	1.0083	0.3037	0.7160	1.2949
200ms, 20Hz	−8	0.5897	0.0233	−7.6082	0.5974	0.0275	0.5435	0.5671	0.5789	0.0055	0.0130	0.0196
	−4	0.6969	0.1150	−3.8392	0.7026	0.1220	0.5721	0.6099	0.6341	0.0150	0.0352	0.0651
	0	0.8552	0.4032	−0.1545	0.8490	0.3878	0.6791	0.7602	0.7987	0.0946	0.2053	0.3136
	4	0.9700	0.8054	4.1558	0.9723	0.8175	0.8125	0.9097	0.9360	0.3038	0.5624	0.6955
	8	0.9979	0.9784	8.0789	0.9980	0.9795	0.9196	0.9894	0.9967	0.5965	0.9151	0.9875
100ms, 40Hz	−8	0.5897	0.0233	−7.5739	0.5981	0.0279	0.5525	0.5785	0.5982	0.0080	0.0178	0.0334
	−4	0.6969	0.1150	−4.3887	0.6836	0.0996	0.5897	0.6303	0.6493	0.0233	0.0495	0.0709
	0	0.8552	0.4032	0.0579	0.8574	0.4091	0.7170	0.7866	0.8098	0.1405	0.2520	0.3136
	4	0.9700	0.8054	3.8417	0.9674	0.7926	0.8594	0.9350	0.9492	0.4141	0.6529	0.7283
	8	0.9979	0.9784	7.9003	0.9977	0.9769	0.9501	0.9949	0.9980	0.7140	0.9536	0.9874
50ms, 80Hz	−8	0.5897	0.0233	−7.7795	0.5939	0.0256	0.5525	0.5751	0.5869	0.0080	0.0164	0.0249
	−4	0.6969	0.1150	−4.5292	0.6789	0.0945	0.6069	0.6554	0.6837	0.0332	0.0709	0.1170
	0	0.8552	0.4032	−0.1683	0.8485	0.3864	0.7318	0.8120	0.8400	0.1611	0.3027	0.3896
	4	0.9700	0.8054	3.8162	0.9669	0.7905	0.8786	0.9525	0.9651	0.4666	0.7244	0.8017
	8	0.9979	0.9784	8.0010	0.9979	0.9784	0.9646	0.9970	0.9984	0.7791	0.9709	0.9870
25ms, 160Hz	−8	0.5897	0.0233	−8.2601	0.5849	0.0209	0.5384	0.5568	0.5671	0.0043	0.0093	0.0144
	−4	0.6969	0.1150	−4.2721	0.6875	0.1040	0.5994	0.6451	0.6752	0.0287	0.0617	0.1080
	0	0.8552	0.4032	−0.0004	0.8551	0.4032	0.6999	0.7718	0.8015	0.1186	0.2251	0.3029
	4	0.9700	0.8054	4.0322	0.9705	0.8079	0.8578	0.9449	0.9656	0.4101	0.6921	0.8195
	8	0.9979	0.9784	7.9945	0.9979	0.9783	0.9541	0.9946	0.9968	0.7313	0.9517	0.9758

Table D.4: $\mathcal{A}$ and $\mathcal{D}_2$ for Observer 2 after mROC, GOC, and FORCE analyses in condition $\mathcal{WT}$=1 compared with the χ^2 energy detector for $\mathcal{WT}$=1.

Experimental Level	Theory SNR (dB)	Theory $\mathcal{A}$	Theory $\mathcal{D}$	Empirical SNR (dB)	Empirical $\mathcal{A}$	Empirical $\mathcal{D}$	mROC $\mathcal{A}$	GOC $\mathcal{A}$	FORCE $\mathcal{A}$	mROC $\mathcal{D}$	GOC $\mathcal{D}$	FORCE $\mathcal{D}$
400ms, 2.5Hz	0	0.6667	0.0817	−0.5563	0.6528	0.0684	0.5188	0.5326	0.5424	0.0010	0.0031	0.0049
	4	0.7784	0.2369	4.1181	0.7817	0.2430	0.5509	0.5866	0.6422	0.0075	0.0217	0.4405
	8	0.8797	0.4697	8.3007	0.8859	0.4877	0.6270	0.6864	0.7337	0.0471	0.1028	0.2117
	12	0.9440	0.6885	11.8220	0.9419	0.6802	0.7475	0.8179	0.8480	0.1848	0.3154	0.4084
	16	0.9761	0.8371	15.6759	0.9743	0.8278	0.8555	0.9053	0.9173	0.4042	0.5480	0.5954
200ms, 5Hz	0	0.6667	0.0817	−0.4546	0.6552	0.0707	0.5326	0.5526	0.5853	0.0031	0.0080	0.0000
	4	0.7784	0.2369	4.0882	0.7809	0.2414	0.5933	0.6383	0.6705	0.0252	0.0560	0.1040
	8	0.8797	0.4697	7.8711	0.8769	0.4619	0.6683	0.7275	0.7641	0.0833	0.1549	0.2361
	12	0.9440	0.6885	12.0383	0.9444	0.6903	0.8064	0.8706	0.8902	0.2911	0.4443	0.5142
	16	0.9761	0.8371	16.1484	0.9768	0.8412	0.8966	0.9390	0.9509	0.5203	0.6687	0.7260
100ms, 10Hz	0	0.6667	0.0817	−1.2808	0.6356	0.0538	0.5364	0.5602	0.5822	0.0038	0.0105	0.0330
	4	0.7784	0.2369	4.0158	0.7788	0.2377	0.6212	0.6723	0.6982	0.0428	0.0875	0.1307
	8	0.8797	0.4697	8.1727	0.8833	0.4801	0.7332	0.7976	0.8189	0.1632	0.2733	0.3292
	12	0.9440	0.6885	11.8322	0.9420	0.6806	0.8343	0.8859	0.8988	0.3522	0.4877	0.5349
	16	0.9761	0.8371	16.0756	0.9765	0.8392	0.9212	0.9538	0.9580	0.6019	0.7300	0.7515
50ms, 20Hz	0	0.6667	0.0817	0.2443	0.6730	0.0881	0.5588	0.5940	0.6209	0.0100	0.0256	0.0606
	4	0.7784	0.2369	3.2400	0.7566	0.1994	0.6115	0.6662	0.7049	0.0362	0.0812	0.1524
	8	0.8797	0.4697	8.0137	0.8799	0.4705	0.7389	0.7917	0.8097	0.1715	0.2617	0.3055
	12	0.9440	0.6885	11.7777	0.9414	0.6780	0.8481	0.8974	0.9075	0.3855	0.5228	0.5613
	16	0.9761	0.8371	16.1712	0.9770	0.8418	0.9336	0.9575	0.9603	0.6477	0.7466	0.7608
25ms, 40Hz	0	0.6667	0.0817	0.1491	0.6705	0.0856	0.5662	0.6052	0.6656	0.0127	0.0322	0.2597
	4	0.7784	0.2369	3.8770	0.7749	0.2306	0.6394	0.7000	0.7402	0.0568	0.1187	0.2120
	8	0.8797	0.4697	8.0293	0.8803	0.4714	0.7789	0.8468	0.8676	0.2378	0.3823	0.4488
	12	0.9440	0.6885	12.2573	0.9469	0.7004	0.8764	0.9216	0.9307	0.4603	0.6035	0.6424
	16	0.9761	0.8371	16.0475	0.9763	0.8384	0.9347	0.9592	0.9642	0.6518	0.7542	0.7801
12.5ms, 80Hz	0	0.6667	0.0817	0.2250	0.6725	0.0876	0.5693	0.6008	0.6190	0.0139	0.0295	0.0453
	4	0.7784	0.2369	3.8712	0.7747	0.2303	0.6446	0.7073	0.7479	0.0612	0.1279	0.2191
	8	0.8797	0.4697	7.6408	0.8719	0.4479	0.7447	0.8054	0.8222	0.1805	0.2890	0.3338
	12	0.9440	0.6885	12.0682	0.9448	0.6917	0.8655	0.9165	0.9267	0.4303	0.5854	0.6293
	16	0.9761	0.8371	16.0755	0.9765	0.8392	0.9391	0.9660	0.9691	0.6689	0.7858	0.8033
6.25ms, 160Hz	0	0.6667	0.0817	0.9183	0.6909	0.1079	0.5712	0.6105	0.6475	0.0147	0.0356	0.0830
	4	0.7784	0.2369	3.8949	0.7754	0.2315	0.6277	0.6897	0.7237	0.0476	0.1065	0.1747
	8	0.8797	0.4697	8.0366	0.8804	0.4719	0.7210	0.7966	0.8239	0.1460	0.2713	0.3470
	12	0.9440	0.6885	11.9083	0.9429	0.6842	0.8407	0.9134	0.9285	0.3673	0.5748	0.6438
	16	0.9761	0.8371	16.0541	0.9764	0.8386	0.9070	0.9549	0.9646	0.5535	0.7349	0.7892

Table D.5: $\mathcal{A}$ and $\mathcal{D}_2$ for Observer 2 after mROC, GOC, and FORCE analyses in condition $\mathcal{WT}$=2 compared with the χ^2 energy detector theory for $\mathcal{WT}$=3.

Experimental Level	Theory SNR (dB)	Theory $\mathcal{A}$	Theory $\mathcal{D}$	Empirical SNR (dB)	Empirical $\mathcal{A}$	Empirical $\mathcal{D}$	mROC $\mathcal{A}$	GOC $\mathcal{A}$	FORCE $\mathcal{A}$	mROC $\mathcal{D}$	GOC $\mathcal{D}$	FORCE $\mathcal{D}$
400ms, 5Hz	−4	0.6528	0.0685	−4.3532	0.6431	0.0599	0.5421	0.5599	0.5653	0.0051	0.0104	0.0122
	0	0.7901	0.2588	−0.2741	0.7797	0.2394	0.5693	0.5983	0.6145	0.0139	0.0280	0.0433
	4	0.9241	0.6125	4.2112	0.9294	0.6317	0.6866	0.7471	0.7740	0.1029	0.1841	0.2440
	8	0.9856	0.8911	7.7795	0.9839	0.8812	0.8217	0.8970	0.9189	0.3236	0.5217	0.6162
	12	0.9984	0.9827	11.9325	0.9983	0.9821	0.9439	0.9766	0.9797	0.6881	0.8401	0.8607
200ms, 10Hz	−4	0.6528	0.0685	−4.1226	0.6494	0.0654	0.5311	0.5415	0.5436	0.0028	0.0050	0.0053
	0	0.7901	0.2588	−0.2550	0.7805	0.2407	0.5935	0.6320	0.6504	0.0254	0.0509	0.0718
	4	0.9241	0.6125	3.9492	0.9228	0.6078	0.7489	0.8236	0.8527	0.1870	0.3279	0.4187
	8	0.9856	0.8911	7.6392	0.9828	0.8747	0.8840	0.9407	0.9502	0.4821	0.6752	0.7241
	12	0.9984	0.9827	12.0730	0.9985	0.9833	0.9733	0.9939	0.9948	0.8223	0.9462	0.9551
100ms, 20Hz	−4	0.6528	0.0685	−3.7360	0.6604	0.0756	0.5454	0.5648	0.5757	0.0060	0.0122	0.0179
	0	0.7901	0.2588	0.2015	0.7978	0.2736	0.6422	0.6890	0.7043	0.0592	0.1056	0.1288
	4	0.9241	0.6125	3.7811	0.9184	0.5923	0.7530	0.8169	0.8356	0.1935	0.3132	0.3668
	8	0.9856	0.8911	8.0913	0.9862	0.8950	0.9128	0.9576	0.9642	0.5730	0.7470	0.7842
	12	0.9984	0.9827	11.8498	0.9982	0.9813	0.9792	0.9940	0.9946	0.8541	0.9468	0.9525
50ms, 40Hz	−4	0.6528	0.0685	−3.3567	0.6717	0.0869	0.5953	0.6371	0.6652	0.0264	0.0549	0.0961
	0	0.7901	0.2588	0.6147	0.8134	0.3056	0.6669	0.7371	0.7763	0.0819	0.1689	0.2679
	4	0.9241	0.6125	3.6819	0.9158	0.5831	0.7896	0.8632	0.8863	0.2577	0.4243	0.5081
	8	0.9856	0.8911	7.9418	0.9851	0.8885	0.9294	0.9667	0.9710	0.6319	0.7895	0.8151
	12	0.9984	0.9827	12.0465	0.9984	0.9831	0.9870	0.9977	0.9980	0.8999	0.9766	0.9800
25ms, 80Hz	−4	0.6528	0.0685	−4.4191	0.6413	0.0584	0.5702	0.6037	0.6203	0.0143	0.0312	0.0478
	0	0.7901	0.2588	0.0152	0.7907	0.2599	0.6602	0.7261	0.7613	0.0754	0.1530	0.2337
	4	0.9241	0.6125	4.0481	0.9253	0.6169	0.7982	0.8691	0.8888	0.2744	0.4400	0.5117
	8	0.9856	0.8911	7.5708	0.9822	0.8713	0.9076	0.9544	0.9612	0.5557	0.7324	0.7707
	12	0.9984	0.9827	11.8867	0.9983	0.9817	0.9786	0.9950	0.9959	0.8508	0.9549	0.9634
12.5ms, 160Hz	−4	0.6528	0.0685	−3.6383	0.6633	0.0783	0.5463	0.5728	0.5857	0.0062	0.0153	0.0276
	0	0.7901	0.2588	−0.3127	0.7783	0.2367	0.6217	0.6751	0.7007	0.0431	0.0904	0.1331
	4	0.9241	0.6125	4.4911	0.9359	0.6566	0.7579	0.8338	0.8586	0.2014	0.3510	0.4322
	8	0.9856	0.8911	7.8219	0.9843	0.8832	0.8656	0.9348	0.9490	0.4306	0.6524	0.7260
	12	0.9984	0.9827	11.8543	0.9982	0.9814	0.9576	0.9888	0.9911	0.7469	0.9112	0.9311

Table D.6: $\mathcal{A}$ and $\mathcal{D}_2$ for Observer 2 after mROC, GOC, and FORCE analyses in condition $\mathcal{WT}$=4 compared with the χ^2 energy detector theory for $\mathcal{WT}$=5.

Experimental Level	Theory SNR (dB)	Theory $\mathcal{A}$	Theory $\mathcal{D}$	Empirical SNR (dB)	Empirical $\mathcal{A}$	Empirical $\mathcal{D}$	mROC $\mathcal{A}$	GOC $\mathcal{A}$	FORCE $\mathcal{A}$	mROC $\mathcal{D}$	GOC $\mathcal{D}$	FORCE $\mathcal{D}$
400ms, 10Hz	−8	0.5897	0.0233	−7.5580	0.5984	0.0281	0.5270	0.5416	0.5520	0.0021	0.0050	0.0070
	−4	0.6969	0.1150	−3.5357	0.7136	0.1360	0.5658	0.5951	0.6236	0.0125	0.0263	0.0571
	0	0.8552	0.4032	0.0730	0.8580	0.4106	0.6535	0.7107	0.7398	0.0691	0.1322	0.1893
	4	0.9700	0.8054	3.8845	0.9681	0.7961	0.8100	0.8864	0.9092	0.2985	0.4892	0.5836
	8	0.9979	0.9784	8.0386	0.9980	0.9790	0.9468	0.9879	0.9922	0.7003	0.9058	0.9430
200ms, 20Hz	−8	0.5897	0.0233	−7.6082	0.5974	0.0275	0.5418	0.5559	0.5630	0.0050	0.0090	0.0109
	−4	0.6969	0.1150	−3.8392	0.7026	0.1220	0.5887	0.6237	0.6447	0.0228	0.0446	0.0675
	0	0.8552	0.4032	−0.1545	0.8490	0.3878	0.7054	0.7662	0.7890	0.1254	0.2154	0.2688
	4	0.9700	0.8054	4.1558	0.9723	0.8175	0.8681	0.9293	0.9419	0.4374	0.6315	0.6927
	8	0.9979	0.9784	8.0789	0.9980	0.9795	0.9655	0.9945	0.9963	0.7837	0.9508	0.9695
100ms, 40Hz	−8	0.5897	0.0233	−7.5739	0.5981	0.0279	0.5729	0.6031	0.6171	0.0154	0.0309	0.0429
	−4	0.6969	0.1150	−4.3887	0.6836	0.0996	0.6225	0.6691	0.6940	0.0437	0.0842	0.1230
	0	0.8552	0.4032	0.0579	0.8574	0.4091	0.7535	0.8133	0.8314	0.1943	0.3055	0.3555
	4	0.9700	0.8054	3.8417	0.9674	0.7926	0.8992	0.9498	0.9577	0.5284	0.7129	0.7554
	8	0.9979	0.9784	7.9003	0.9977	0.9769	0.9812	0.9981	0.9987	0.8654	0.9799	0.9875
50ms, 80Hz	−8	0.5897	0.0233	−7.7795	0.5939	0.0256	0.5612	0.5877	0.5986	0.0108	0.0223	0.0302
	−4	0.6969	0.1150	−4.5292	0.6789	0.0945	0.6038	0.6435	0.6623	0.0313	0.0603	0.0836
	0	0.8552	0.4032	−0.1683	0.8485	0.3864	0.7493	0.8209	0.8447	0.1876	0.3218	0.3935
	4	0.9700	0.8054	3.8162	0.9669	0.7905	0.8910	0.9497	0.9596	0.5033	0.7123	0.7675
	8	0.9979	0.9784	8.0010	0.9979	0.9784	0.9772	0.9950	0.9956	0.8430	0.9548	0.9613
25ms, 160Hz	−8	0.5897	0.0233	−8.2601	0.5849	0.0209	0.5359	0.5558	0.5931	0.0037	0.0090	0.0000
	−4	0.6969	0.1150	−4.2721	0.6875	0.1040	0.5834	0.6184	0.6414	0.0202	0.0409	0.0657
	0	0.8552	0.4032	−0.0004	0.8551	0.4032	0.6966	0.7680	0.7989	0.1146	0.2185	0.2988
	4	0.9700	0.8054	4.0322	0.9705	0.8079	0.8395	0.9177	0.9367	0.3645	0.5898	0.6814
	8	0.9979	0.9784	7.9945	0.9979	0.9783	0.9501	0.9872	0.9903	0.7142	0.9011	0.9267

Table D.7: $\mathcal{A}$ and $\mathcal{D}_2$ for Observer 3 after mROC, GOC, and FORCE analyses in condition $\mathcal{WT}$=1 compared with the χ^2 energy detector for $\mathcal{WT}$=1.

Experimental Level	Theory SNR (dB)	Theory $\mathcal{A}$	Theory $\mathcal{D}$	Empirical SNR (dB)	Empirical $\mathcal{A}$	Empirical $\mathcal{D}$	mROC $\mathcal{A}$	GOC $\mathcal{A}$	FORCE $\mathcal{A}$	mROC $\mathcal{D}$	GOC $\mathcal{D}$	FORCE $\mathcal{D}$
25ms, 40Hz	0	0.6667	0.0817	0.1491	0.6705	0.0856	0.5523	0.5776	—	0.0079	0.0174	—
	4	0.7784	0.2369	3.8770	0.7749	0.2306	0.6322	0.6828	—	0.0510	0.0987	—
	8	0.8797	0.4697	8.0293	0.8803	0.4714	0.7542	0.8040	—	0.1955	0.2862	—
	12	0.9440	0.6885	12.2573	0.9469	0.7004	0.8685	0.9066	—	0.4385	0.5522	—
	16	0.9761	0.8371	16.0475	0.9763	0.8384	0.9312	0.9503	—	0.6385	0.7149	—
6.25ms, 160Hz	0	0.6667	0.0817	0.9183	0.6909	0.1079	0.5365	0.5496	—	0.0039	0.0071	—
	4	0.7784	0.2369	3.8949	0.7754	0.2315	0.5634	0.5869	—	0.0116	0.0219	—
	8	0.8797	0.4697	8.0366	0.8804	0.4719	0.6542	0.6924	—	0.0698	0.1096	—
	12	0.9440	0.6885	11.9083	0.9429	0.6842	0.7447	0.7933	—	0.1804	0.2649	—
	16	0.9761	0.8371	16.0541	0.9764	0.8386	0.8572	0.8953	—	0.4084	0.5163	—

Table D.8: $\mathcal{A}$ and $\mathcal{D}_2$ for Observer 3 after mROC, GOC, and FORCE analyses in condition $\mathcal{WT}$=2 compared with the χ^2 energy detector theory for $\mathcal{WT}$=3.

Experimental Level	Theory SNR (dB)	Theory $\mathcal{A}$	Theory $\mathcal{D}$	Empirical SNR (dB)	Empirical $\mathcal{A}$	Empirical $\mathcal{D}$	mROC $\mathcal{A}$	GOC $\mathcal{A}$	FORCE $\mathcal{A}$	mROC $\mathcal{D}$	GOC $\mathcal{D}$	FORCE $\mathcal{D}$
400ms, 5Hz	−4	0.6528	0.0685	−4.3532	0.6431	0.0599	0.5249	0.5371	—	0.0018	0.0040	—
	0	0.7901	0.2588	−0.2741	0.7797	0.2394	0.5488	0.5687	—	0.0069	0.0137	—
	4	0.9241	0.6125	4.2112	0.9294	0.6317	0.6416	0.6908	—	0.0586	0.1078	—
	8	0.9856	0.8911	7.7795	0.9839	0.8812	0.7789	0.8454	—	0.2379	0.3788	—
	12	0.9984	0.9827	11.9325	0.9983	0.9821	0.9239	0.9614	—	0.6118	0.7642	—
50ms, 40Hz	−4	0.6528	0.0685	−3.3567	0.6717	0.0869	0.5700	0.5935	—	0.0142	0.0254	—
	0	0.7901	0.2588	0.6147	0.8134	0.3056	0.6188	0.6512	—	0.0411	0.0670	—
	4	0.9241	0.6125	3.6819	0.9158	0.5831	0.7470	0.7996	—	0.1840	0.2773	—
	8	0.9856	0.8911	7.9418	0.9851	0.8885	0.9023	0.9393	—	0.5383	0.6696	—
	12	0.9984	0.9827	12.0465	0.9984	0.9831	0.9788	0.9940	—	0.8516	0.9467	—
12.5ms, 160Hz	−4	0.6528	0.0685	−3.6383	0.6633	0.0783	0.5342	0.5461	—	0.0034	0.0061	—
	0	0.7901	0.2588	−0.3127	0.7783	0.2367	0.5987	0.6384	—	0.0283	0.0560	—
	4	0.9241	0.6125	4.4911	0.9359	0.6566	0.7248	0.7940	—	0.1511	0.2662	—
	8	0.9856	0.8911	7.8219	0.9843	0.8832	0.8364	0.8992	—	0.3572	0.5285	—
	12	0.9984	0.9827	11.8543	0.9982	0.9814	0.9434	0.9804	—	0.6861	0.8606	—

Table D.9: $\mathcal{A}$ and $\mathcal{D}_2$ for Observer 3 after mROC, GOC, and FORCE analyses in condition $\mathcal{WT}$=4 compared with the χ^2 energy detector theory for $\mathcal{WT}$=5.

Experimental Level	Theory SNR (dB)	Theory $\mathcal{A}$	Theory $\mathcal{D}$	Empirical SNR (dB)	Empirical $\mathcal{A}$	Empirical $\mathcal{D}$	mROC $\mathcal{A}$	GOC $\mathcal{A}$	FORCE $\mathcal{A}$	mROC $\mathcal{D}$	GOC $\mathcal{D}$	FORCE $\mathcal{D}$
400ms, 10Hz	−8	0.5897	0.0233	−7.5580	0.5984	0.0281	0.5339	0.5512	—	0.0033	0.0076	—
	−4	0.6969	0.1150	−3.5357	0.7136	0.1360	0.5558	0.5790	—	0.0090	0.0181	—
	0	0.8552	0.4032	0.0730	0.8580	0.4106	0.6082	0.6515	—	0.0340	0.0673	—
	4	0.9700	0.8054	3.8845	0.9681	0.7961	0.7551	0.8217	—	0.1969	0.3236	—
	8	0.9979	0.9784	8.0386	0.9980	0.9790	0.9283	0.9726	—	0.6276	0.8186	—
100ms, 40Hz	−8	0.5897	0.0233	−7.5739	0.5981	0.0279	0.5326	0.5451	—	0.0031	0.0059	—
	−4	0.6969	0.1150	−4.3887	0.6836	0.0996	0.5783	0.6080	—	0.0177	0.0339	—
	0	0.8552	0.4032	0.0579	0.8574	0.4091	0.6821	0.7290	—	0.0979	0.1571	—
	4	0.9700	0.8054	3.8417	0.9674	0.7926	0.8474	0.9056	—	0.3838	0.5491	—
	8	0.9979	0.9784	7.9003	0.9977	0.9769	0.9662	0.9910	—	0.7867	0.9258	—
25ms, 160Hz	−8	0.5897	0.0233	−8.2601	0.5849	0.0209	0.5412	0.5619	—	0.0049	0.0111	—
	−4	0.6969	0.1150	−4.2721	0.6875	0.1040	0.5757	0.6114	—	0.0166	0.0361	—
	0	0.8552	0.4032	−0.0004	0.8551	0.4032	0.6736	0.7368	—	0.0887	0.1684	—
	4	0.9700	0.8054	4.0322	0.9705	0.8079	0.8140	0.8959	—	0.3070	0.5180	—
	8	0.9979	0.9784	7.9945	0.9979	0.9783	0.9309	0.9789	—	0.6375	0.8524	—

Table D.10: Empirical average $\mathcal{A}$ and $\mathcal{D}_2$, and standard deviations for Observer 1 in the condition $\mathcal{WT}$=1.

Experimental Level	Theory SNR (dB)	Empirical SNR (dB)	Average $\mathcal{A}$	St. Dev. $\mathcal{A}$	Average $\mathcal{D}$	St. Dev. $\mathcal{D}$
400ms, 2.5Hz	0	−0.5563	0.51209	0.00998	0.00071	0.00000
	4	4.1181	0.53614	0.02057	0.00500	0.00002
	8	8.3007	0.61618	0.00869	0.03954	0.00004
	12	11.8220	0.70404	0.00901	0.12399	0.00013
	16	15.6759	0.79328	0.01410	0.26571	0.00076
200ms, 5Hz	0	−0.4546	0.53264	0.01201	0.00349	0.00001
	4	4.0882	0.57994	0.01236	0.01897	0.00004
	8	7.8711	0.65464	0.01428	0.07079	0.00017
	12	12.0383	0.78525	0.01478	0.25048	0.00080
	16	16.1484	0.85872	0.01685	0.41412	0.00196
100ms, 10Hz	0	−1.2808	0.53703	0.01146	0.00434	0.00001
	4	4.0158	0.63304	0.01292	0.05221	0.00010
	8	8.1727	0.73538	0.01161	0.16687	0.00029
	12	11.8322	0.83324	0.01302	0.35062	0.00084
	16	16.0756	0.91038	0.00740	0.56529	0.00060
50ms, 20Hz	0	0.2443	0.56529	0.00780	0.01252	0.00001
	4	3.2400	0.61331	0.01312	0.03789	0.00008
	8	8.0137	0.74321	0.00720	0.17833	0.00012
	12	11.7777	0.84511	0.00882	0.37849	0.00046
	16	16.1712	0.93271	0.00313	0.64440	0.00014
25ms, 40Hz	0	0.1491	0.55205	0.01898	0.00888	0.00005
	4	3.8770	0.62061	0.01065	0.04274	0.00006
	8	8.0293	0.75813	0.01034	0.20229	0.00030
	12	12.2573	0.85319	0.01246	0.39909	0.00098
	16	16.0475	0.92438	0.00928	0.61436	0.00114
12.5ms, 80Hz	0	0.2250	0.55265	0.00946	0.00827	0.00001
	4	3.8712	0.61261	0.01296	0.03742	0.00007
	8	7.6408	0.71821	0.01503	0.14293	0.00041
	12	12.0682	0.83846	0.00613	0.36223	0.00022
	16	16.0755	0.92501	0.00478	0.61595	0.00031
6.25ms, 160Hz	0	0.9183	0.56568	0.01195	0.01290	0.00002
	4	3.8949	0.60970	0.01008	0.03532	0.00005
	8	8.0366	0.70644	0.00818	0.12696	0.00011
	12	11.9083	0.82976	0.00849	0.34211	0.00038
	16	16.0541	0.90740	0.00972	0.55573	0.00105

Table D.11: Empirical average $\mathcal{A}$ and $\mathcal{D}_2$, and standard deviations for Observer 1 in the condition $\mathcal{WT}$=2.

Experimental Level	Theory SNR (dB)	Empirical SNR (dB)	Average $\mathcal{A}$	St. Dev. $\mathcal{A}$	Average $\mathcal{D}$	St. Dev. $\mathcal{D}$
400ms, 5Hz	−4	−4.3532	0.53454	0.01295	0.00393	0.00001
	0	−0.2741	0.55369	0.01038	0.00865	0.00001
	4	4.2112	0.63509	0.00980	0.05362	0.00006
	8	7.7795	0.73487	0.01775	0.16676	0.00064
	12	11.9325	0.83512	0.01495	0.35532	0.00123
200ms, 10Hz	−4	−4.1226	0.51816	0.02128	0.00226	0.00000
	0	−0.2550	0.58247	0.01142	0.02010	0.00003
	4	3.9492	0.70378	0.01652	0.12433	0.00046
	8	7.6392	0.81828	0.02038	0.31819	0.00194
	12	12.0730	0.89458	0.01836	0.51655	0.00297
100ms, 20Hz	−4	−3.7360	0.56120	0.00920	0.01108	0.00001
	0	0.2015	0.64241	0.01462	0.06001	0.00015
	4	3.7811	0.74722	0.00653	0.18450	0.00010
	8	8.0913	0.87965	0.00636	0.46993	0.00034
	12	11.8498	0.94505	0.00618	0.69346	0.00064
50ms, 40Hz	−4	−3.3567	0.57875	0.01425	0.01857	0.00005
	0	0.6147	0.63848	0.01893	0.05717	0.00020
	4	3.6819	0.76550	0.01484	0.21512	0.00065
	8	7.9418	0.90732	0.01280	0.55603	0.00171
	12	12.0465	0.96708	0.00291	0.79136	0.00020
25ms, 80Hz	−4	−4.4191	0.54687	0.00876	0.00657	0.00001
	0	0.0152	0.62408	0.00892	0.04514	0.00005
	4	4.0481	0.73887	0.01238	0.17213	0.00034
	8	7.5708	0.86576	0.00821	0.43146	0.00048
	12	11.8867	0.94930	0.00647	0.71126	0.00075
12.5ms, 160Hz	−4	−3.6383	0.54039	0.01367	0.00526	0.00001
	0	−0.3127	0.60507	0.01375	0.03266	0.00008
	4	4.4911	0.73525	0.01180	0.16670	0.00032
	8	7.8219	0.85894	0.00799	0.41336	0.00042
	12	11.8543	0.95050	0.00828	0.71675	0.00117

Table D.12: Empirical average $\mathcal{A}$ and $\mathcal{D}_2$, and standard deviations for Observer 1 in the condition $\mathcal{WT}$=4.

Experimental Level	Theory SNR (dB)	Empirical SNR (dB)	Average $\mathcal{A}$	St. Dev. $\mathcal{A}$	Average $\mathcal{D}$	St. Dev. $\mathcal{D}$
400ms, 10Hz	-8	-7.5580	0.50674	0.00568	0.00022	0.00000
	-4	-3.5357	0.54737	0.01467	0.00711	0.00002
	0	0.0730	0.59625	0.00803	0.02709	0.00002
	4	3.8845	0.69546	0.01599	0.11409	0.00035
	8	8.0386	0.81293	0.02158	0.30683	0.00185
200ms, 20Hz	-8	-7.6082	0.54230	0.01584	0.00590	0.00002
	-4	-3.8392	0.57147	0.00661	0.01492	0.00001
	0	-0.1545	0.67841	0.01416	0.09456	0.00025
	4	4.1558	0.81360	0.00626	0.30631	0.00018
	8	8.0789	0.91958	0.00505	0.59657	0.00031
100ms, 40Hz	-8	-7.5739	0.55225	0.01891	0.00893	0.00003
	-4	-4.3887	0.59052	0.01179	0.02419	0.00004
	0	0.0579	0.71771	0.00414	0.14151	0.00003
	4	3.8417	0.86070	0.00809	0.41800	0.00046
	8	7.9003	0.95094	0.00462	0.71793	0.00038
50ms, 80Hz	-8	-7.7795	0.55205	0.01210	0.00826	0.00001
	-4	-4.5292	0.60635	0.01341	0.03343	0.00008
	0	-0.1683	0.73211	0.01311	0.16221	0.00034
	4	3.8162	0.87838	0.00977	0.46666	0.00078
	8	8.0010	0.96386	0.00863	0.77723	0.00165
25ms, 160Hz	-8	-8.2601	0.53969	0.00555	0.00464	0.00000
	-4	-4.2721	0.60055	0.00718	0.02953	0.00002
	0	-0.0004	0.69930	0.00622	0.11799	0.00006
	4	4.0322	0.85859	0.00938	0.41261	0.00060
	8	7.9945	0.95458	0.00618	0.73403	0.00075

Table D.13: Empirical average $\mathcal{A}$ and $\mathcal{D}_2$, and standard deviations for Observer 2 in the condition $\mathcal{WT}$=1.

Experimental Level	Theory SNR (dB)	Empirical SNR (dB)	Average $\mathcal{A}$	St. Dev. $\mathcal{A}$	Average $\mathcal{D}$	St. Dev. $\mathcal{D}$
400ms, 2.5Hz	0	−0.5563	0.52038	0.01214	0.00162	0.00000
	4	4.1181	0.55159	0.01392	0.00826	0.00002
	8	8.3007	0.63011	0.00830	0.04963	0.00004
	12	11.8220	0.75032	0.01180	0.18978	0.00037
	16	15.6759	0.85742	0.01114	0.40979	0.00085
200ms, 5Hz	0	−0.4546	0.53273	0.01056	0.00342	0.00000
	4	4.0882	0.59551	0.00864	0.02671	0.00002
	8	7.8711	0.67145	0.00995	0.08689	0.00010
	12	12.0383	0.80842	0.01267	0.29597	0.00067
	16	16.1484	0.89788	0.00482	0.52449	0.00023
100ms, 10Hz	0	−1.2808	0.53703	0.00912	0.00420	0.00000
	4	4.0158	0.62146	0.00474	0.04307	0.00001
	8	8.1727	0.73467	0.00784	0.16554	0.00014
	12	11.8322	0.83526	0.00943	0.35495	0.00050
	16	16.0756	0.92170	0.00399	0.60402	0.00020
50ms, 20Hz	0	0.2443	0.56070	0.01058	0.01099	0.00001
	4	3.2400	0.61492	0.01276	0.03895	0.00007
	8	8.0137	0.73955	0.01452	0.17335	0.00045
	12	11.7777	0.85000	0.00900	0.39062	0.00050
	16	16.1712	0.93404	0.00404	0.64954	0.00024
25ms, 40Hz	0	0.1491	0.56716	0.01568	0.01377	0.00003
	4	3.8770	0.64132	0.01282	0.05894	0.00011
	8	8.0293	0.78098	0.01187	0.24222	0.00046
	12	12.2573	0.87752	0.00555	0.46377	0.00025
	16	16.0475	0.93566	0.00312	0.65567	0.00014
12.5ms, 80Hz	0	0.2250	0.56656	0.01310	0.01332	0.00002
	4	3.8712	0.64346	0.01165	0.06065	0.00010
	8	7.6408	0.74370	0.00621	0.17904	0.00009
	12	12.0682	0.86590	0.00402	0.43152	0.00012
	16	16.0755	0.93889	0.00352	0.66833	0.00019
6.25ms, 160Hz	0	0.9183	0.57387	0.01500	0.01646	0.00004
	4	3.8949	0.63233	0.01124	0.05153	0.00008
	8	8.0366	0.72574	0.01171	0.15299	0.00027
	12	11.9083	0.84404	0.01006	0.37600	0.00060
	16	16.0541	0.90945	0.00942	0.56250	0.00101

Table D.14: Empirical average $\mathcal{A}$ and $\mathcal{D}_2$, and standard deviations for Observer 2 in the condition $\mathcal{WT}$=2.

Experimental Level	Theory SNR (dB)	Empirical SNR (dB)	Average $\mathcal{A}$	St. Dev. $\mathcal{A}$	Average $\mathcal{D}$	St. Dev. $\mathcal{D}$
400ms, 5Hz	−4	−4.3532	0.54239	0.02140	0.00652	0.00002
	0	−0.2741	0.56915	0.01594	0.01459	0.00004
	4	4.2112	0.68712	0.02122	0.10504	0.00055
	8	7.7795	0.82285	0.01271	0.32699	0.00078
	12	11.9325	0.94469	0.00536	0.69184	0.00047
200ms, 10Hz	−4	−4.1226	0.53114	0.01188	0.00321	0.00001
	0	−0.2550	0.59423	0.01396	0.02636	0.00006
	4	3.9492	0.74851	0.02557	0.18889	0.00164
	8	7.6392	0.88461	0.01090	0.48486	0.00101
	12	12.0730	0.97332	0.00517	0.82330	0.00075
100ms, 20Hz	−4	−3.7360	0.54441	0.01087	0.00604	0.00001
	0	0.2015	0.64233	0.01207	0.05972	0.00010
	4	3.7811	0.75292	0.01020	0.19379	0.00027
	8	8.0913	0.91308	0.00494	0.57412	0.00028
	12	11.8498	0.97878	0.00290	0.85204	0.00026
50ms, 40Hz	−4	−3.3567	0.59648	0.00669	0.02716	0.00001
	0	0.6147	0.66782	0.00983	0.08318	0.00010
	4	3.6819	0.79101	0.01410	0.26134	0.00079
	8	7.9418	0.93005	0.00848	0.63507	0.00103
	12	12.0465	0.98678	0.00228	0.89883	0.00020
25ms, 80Hz	−4	−4.4191	0.56958	0.01253	0.01448	0.00002
	0	0.0152	0.65936	0.01524	0.07532	0.00022
	4	4.0481	0.79846	0.01680	0.27627	0.00110
	8	7.5708	0.90709	0.01167	0.55508	0.00151
	12	11.8867	0.97825	0.00486	0.84960	0.00069
12.5ms, 160Hz	−4	−3.6383	0.54685	0.01487	0.00699	0.00002
	0	−0.3127	0.62189	0.01871	0.04437	0.00016
	4	4.4911	0.75842	0.01298	0.20300	0.00045
	8	7.8219	0.86597	0.01349	0.43273	0.00130
	12	11.8543	0.95713	0.00775	0.74571	0.00116

Table D.15: Empirical average $\mathcal{A}$ and $\mathcal{D}_2$, and standard deviations for Observer 2 in the condition $\mathcal{WT}$=4.

Experimental Level	Theory SNR (dB)	Empirical SNR (dB)	Average $\mathcal{A}$	St. Dev. $\mathcal{A}$	Average $\mathcal{D}$	St. Dev. $\mathcal{D}$
400ms, 10Hz	−8	−7.5580	0.52760	0.01797	0.00314	0.00001
	−4	−3.5357	0.56678	0.01601	0.01366	0.00003
	0	0.0730	0.65435	0.01657	0.07075	0.00024
	4	3.8845	0.81134	0.01620	0.30257	0.00113
	8	8.0386	0.94656	0.01002	0.70056	0.00167
200ms, 20Hz	−8	−7.6082	0.54107	0.01804	0.00582	0.00002
	−4	−3.8392	0.58948	0.01551	0.02394	0.00007
	0	−0.1545	0.70503	0.01406	0.12563	0.00032
	4	4.1558	0.86775	0.00902	0.43694	0.00063
	8	8.0789	0.96471	0.00526	0.78030	0.00060
100ms, 40Hz	−8	−7.5739	0.57285	0.00927	0.01562	0.00001
	−4	−4.3887	0.62294	0.01591	0.04484	0.00015
	0	0.0579	0.75355	0.00887	0.19471	0.00021
	4	3.8417	0.89900	0.00858	0.52840	0.00069
	8	7.9003	0.98035	0.00405	0.86116	0.00052
50ms, 80Hz	−8	−7.7795	0.56142	0.01593	0.01166	0.00003
	−4	−4.5292	0.60460	0.01407	0.03240	0.00007
	0	−0.1683	0.74936	0.01191	0.18826	0.00037
	4	3.8162	0.89068	0.00655	0.50245	0.00040
	8	8.0010	0.97630	0.00099	0.83831	0.00003
25ms, 160Hz	−8	−8.2601	0.53609	0.00597	0.00387	0.00000
	−4	−4.2721	0.58375	0.00857	0.02055	0.00002
	0	−0.0004	0.69744	0.00626	0.11574	0.00006
	4	4.0322	0.83970	0.01006	0.36551	0.00059
	8	7.9945	0.94978	0.00605	0.71322	0.00063

Table D.16: Empirical average $\mathcal{A}$ and $\mathcal{D}_2$, and standard deviations for Observer 3 in the condition $\mathcal{WT}$=1.

Experimental Level	Theory SNR (dB)	Empirical SNR (dB)	Average $\mathcal{A}$	St. Dev. $\mathcal{A}$	Average $\mathcal{D}$	St. Dev. $\mathcal{D}$
25ms, 40Hz	0	0.1491	0.55205	0.01860	0.00884	0.00004
	4	3.8770	0.63223	0.00357	0.05109	0.00001
	8	8.0293	0.75441	0.01482	0.19665	0.00060
	12	12.2573	0.86803	0.01120	0.43798	0.00096
	16	16.0475	0.93051	0.00489	0.63624	0.00033
6.25ms, 160Hz	0	0.9183	0.53619	0.01094	0.00413	0.00001
	4	3.8949	0.56332	0.01503	0.01226	0.00003
	8	8.0366	0.65304	0.00282	0.06870	0.00001
	12	11.9083	0.74443	0.01327	0.18069	0.00042
	16	16.0541	0.85581	0.00802	0.40526	0.00043

Table D.17: Empirical average $\mathcal{A}$ and $\mathcal{D}_2$, and standard deviations for Observer 3 in the condition $\mathcal{WT}$=2.

Experimental Level	Theory SNR (dB)	Empirical SNR (dB)	Average $\mathcal{A}$	St. Dev. $\mathcal{A}$	Average $\mathcal{D}$	St. Dev. $\mathcal{D}$
400ms, 5Hz	-4	-4.3532	0.52421	0.00389	0.00174	0.00000
	0	-0.2741	0.54821	0.01631	0.00749	0.00002
	4	4.2112	0.64097	0.00886	0.05837	0.00005
	8	7.7795	0.77760	0.01259	0.23615	0.00053
	12	11.9325	0.92382	0.00454	0.61165	0.00027
50ms, 40Hz	-4	-3.3567	0.56925	0.01480	0.01453	0.00004
	0	0.6147	0.61885	0.00774	0.04133	0.00003
	4	3.6819	0.74703	0.00925	0.18437	0.00021
	8	7.9418	0.90177	0.00546	0.53688	0.00031
	12	12.0465	0.97856	0.00151	0.85064	0.00007
12.5ms, 160Hz	-4	-3.6383	0.53369	0.02134	0.00460	0.00003
	0	-0.3127	0.59846	0.01561	0.02889	0.00009
	4	4.4911	0.72407	0.01327	0.15079	0.00033
	8	7.8219	0.83539	0.01668	0.35624	0.00151
	12	11.8543	0.94239	0.00854	0.68309	0.00122

Table D.18: Empirical average $\mathcal{A}$ and $\mathcal{D}_2$, and standard deviations for Observer 3 in the condition $\mathcal{WT}$=4.

Experimental Level	Theory SNR (dB)	Empirical SNR (dB)	Average $\mathcal{A}$	St. Dev. $\mathcal{A}$	Average $\mathcal{D}$	St. Dev. $\mathcal{D}$
400ms, 10Hz	-8	-7.5580	0.53375	0.01014	0.00359	0.00000
	-4	-3.5357	0.55580	0.00539	0.00909	0.00000
	0	0.0730	0.60841	0.00576	0.03429	0.00001
	4	3.8845	0.75495	0.00509	0.19676	0.00007
	8	8.0386	0.92767	0.00382	0.62560	0.00020
100ms, 40Hz	-8	-7.5739	0.53286	0.00688	0.00325	0.00000
	-4	-4.3887	0.57858	0.01188	0.01831	0.00003
	0	0.0579	0.68252	0.00260	0.09840	0.00001
	4	3.8417	0.84698	0.01550	0.38400	0.00150
	8	7.9003	0.96559	0.00716	0.78512	0.00124
25ms, 160Hz	-8	-8.2601	0.54152	0.01805	0.00593	0.00002
	-4	-4.2721	0.57572	0.02159	0.01798	0.00009
	0	-0.0004	0.67321	0.00763	0.08858	0.00006
	4	4.0322	0.81444	0.00807	0.30823	0.00030
	8	7.9945	0.93117	0.00374	0.63861	0.00020

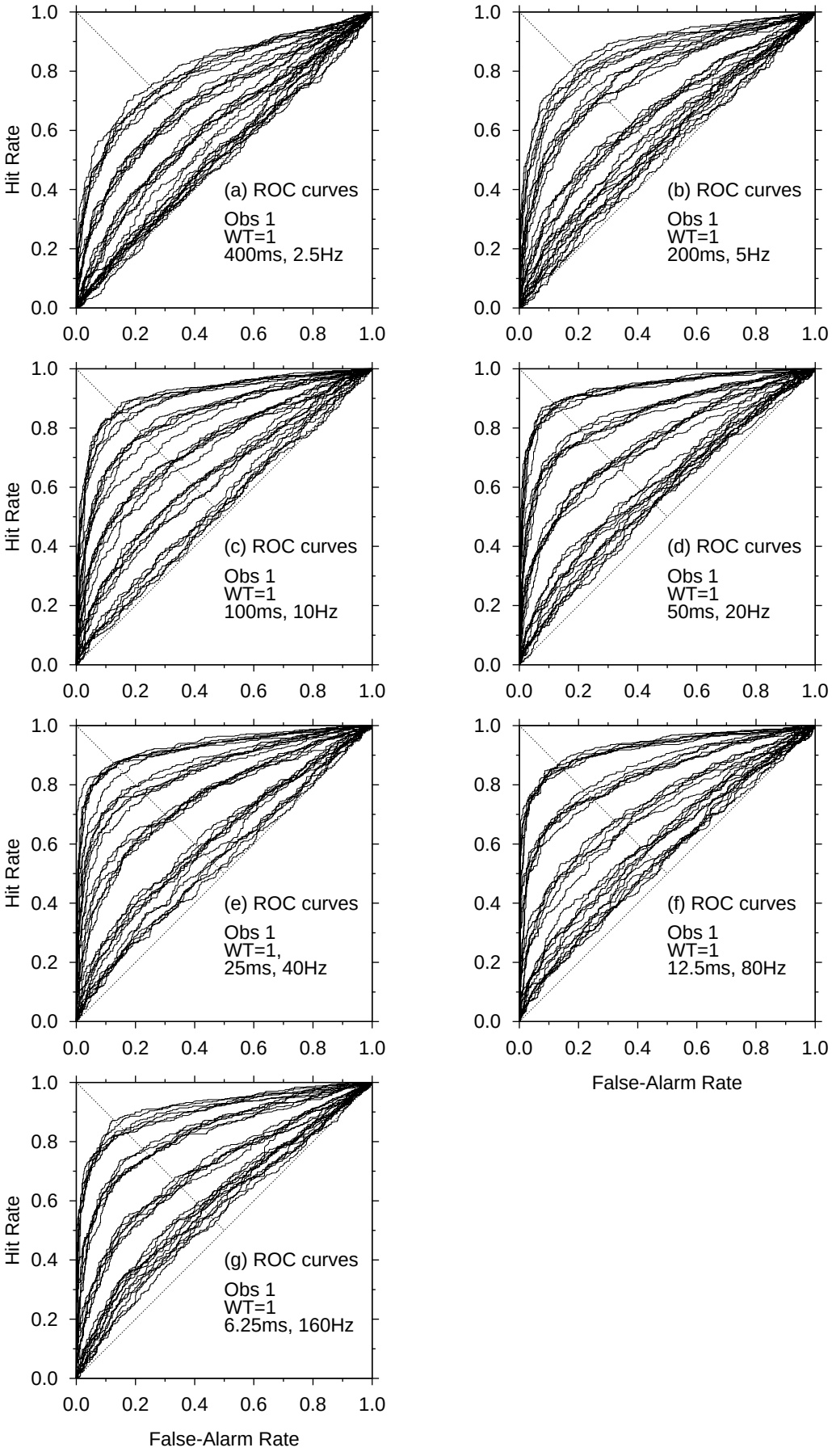

Figure D.1: ROC curves for Observer 1 in condition $\mathcal{WT}=1$.

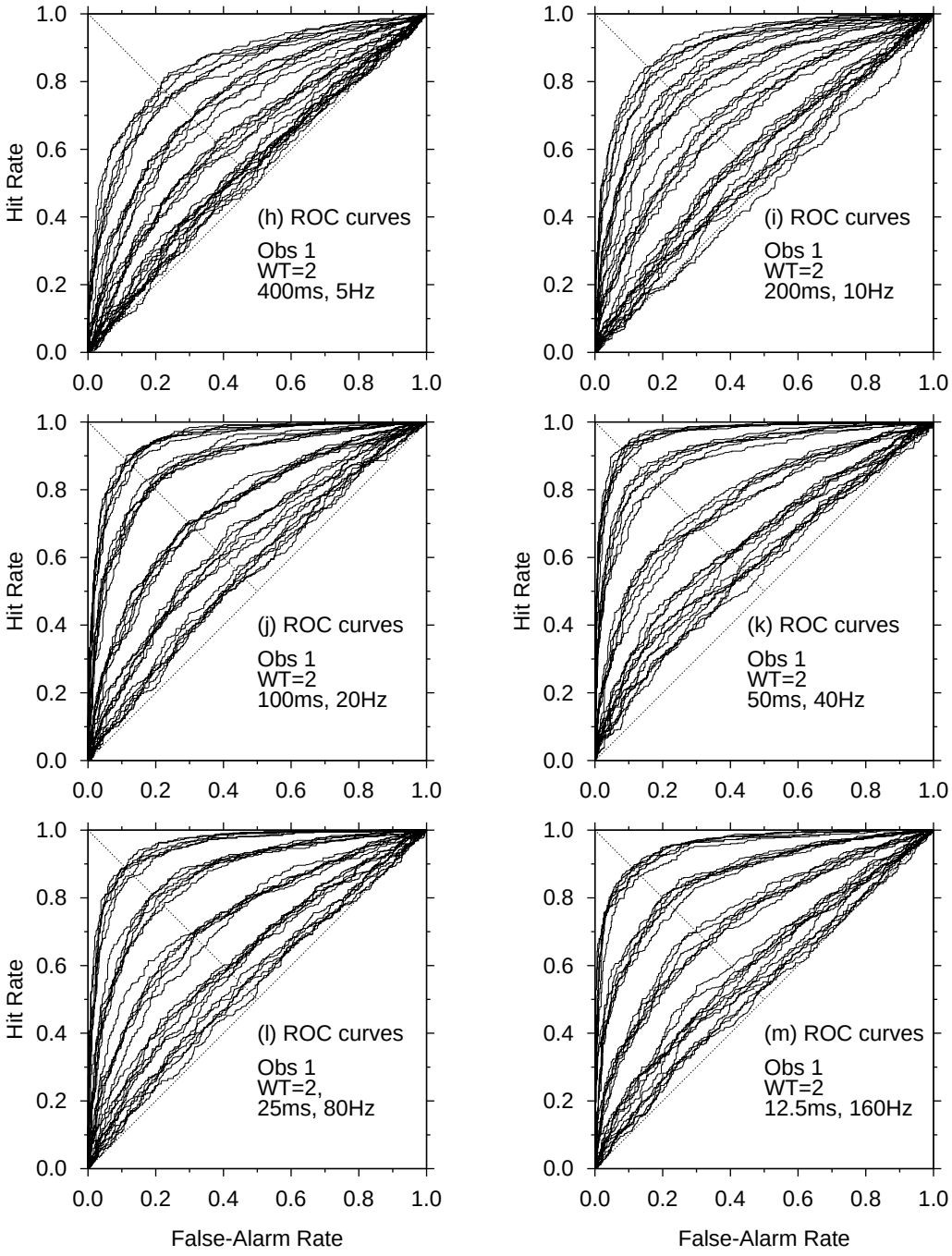

Figure D.2: ROC curves for Observer 1 in condition $\mathcal{WT}=2$.

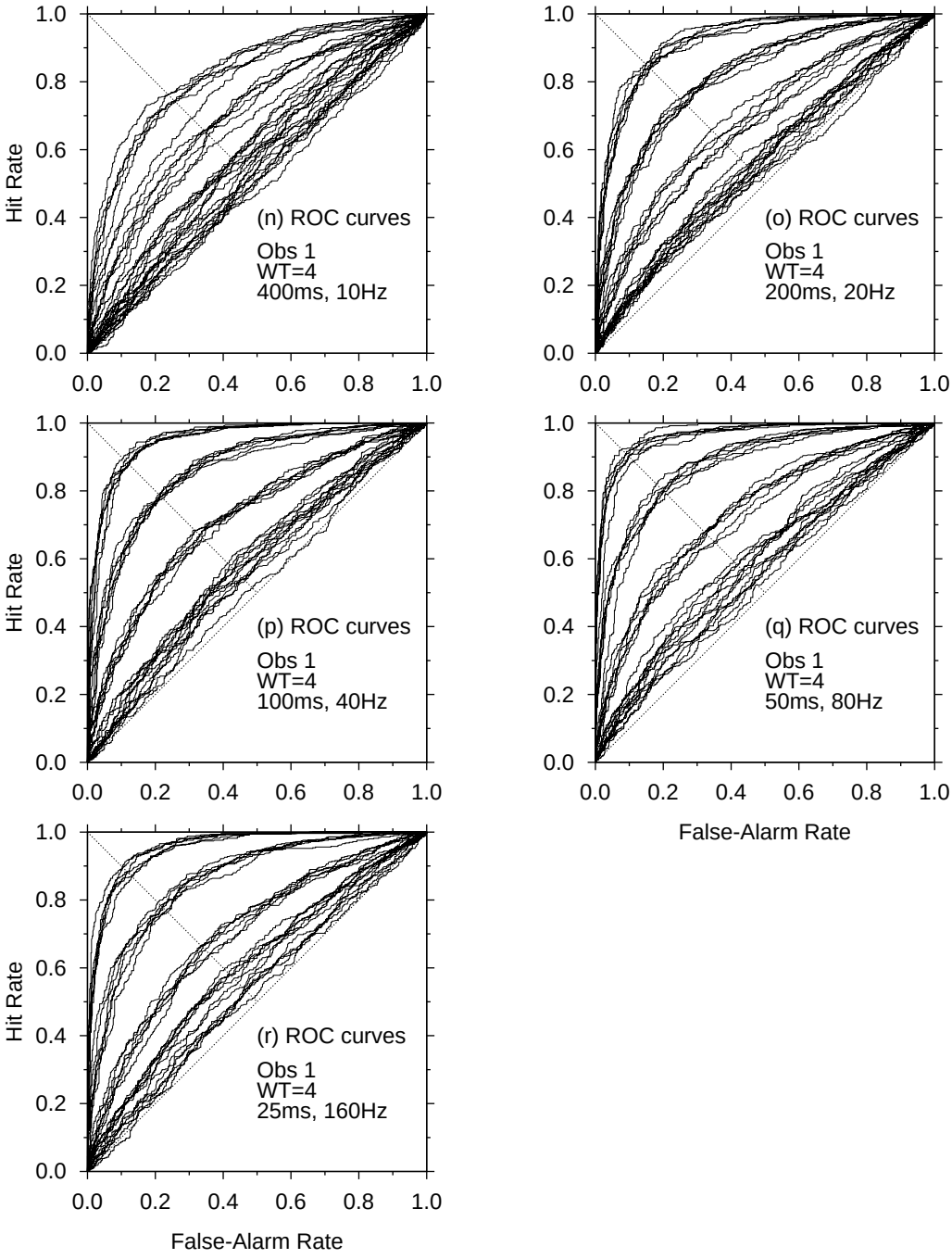

Figure D.3: ROC curves for Observer 1 in condition $\mathcal{WT}$=4.

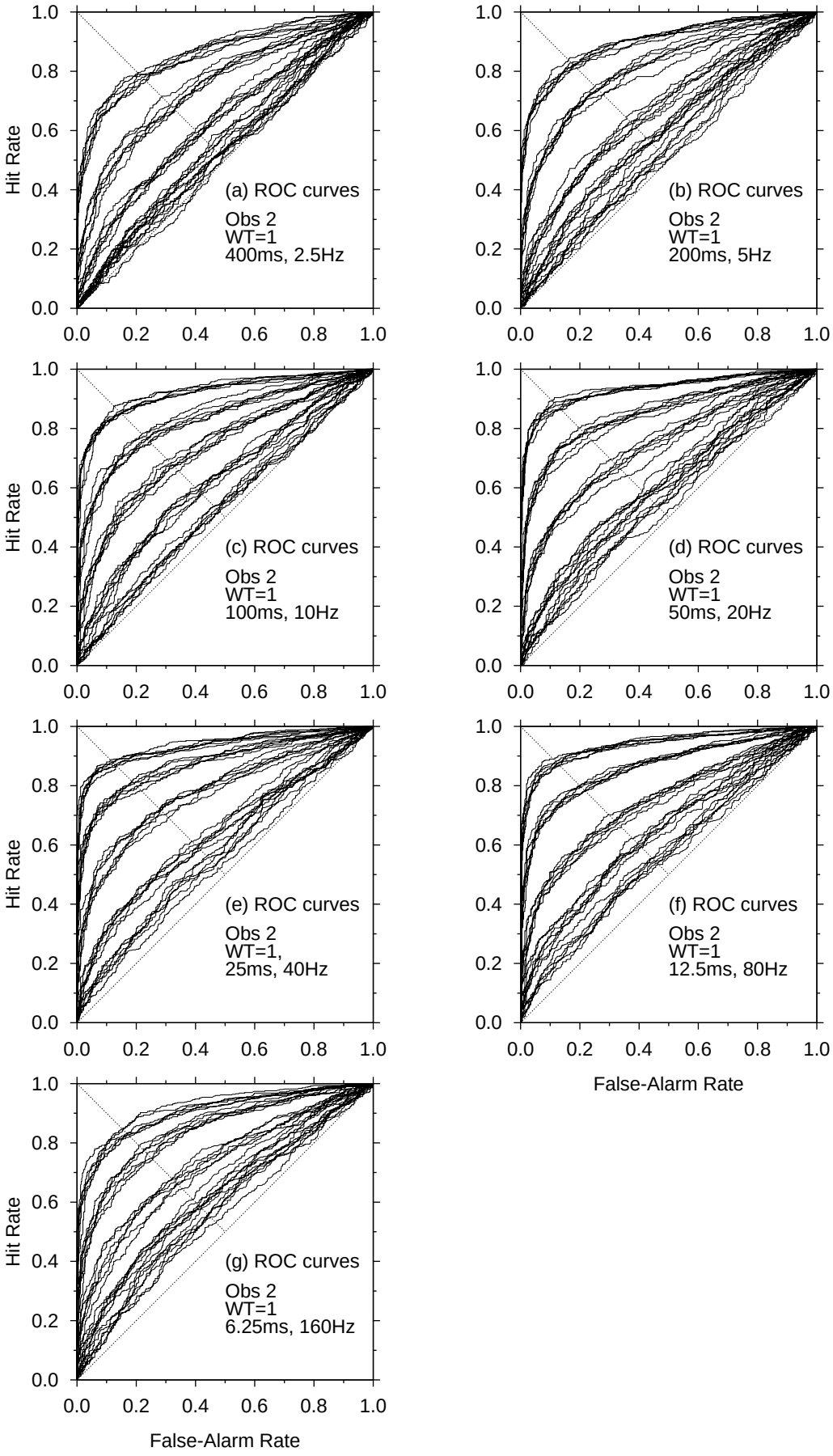

Figure D.4: ROC curves for Observer 2 in condition $\mathcal{WT}=1$.

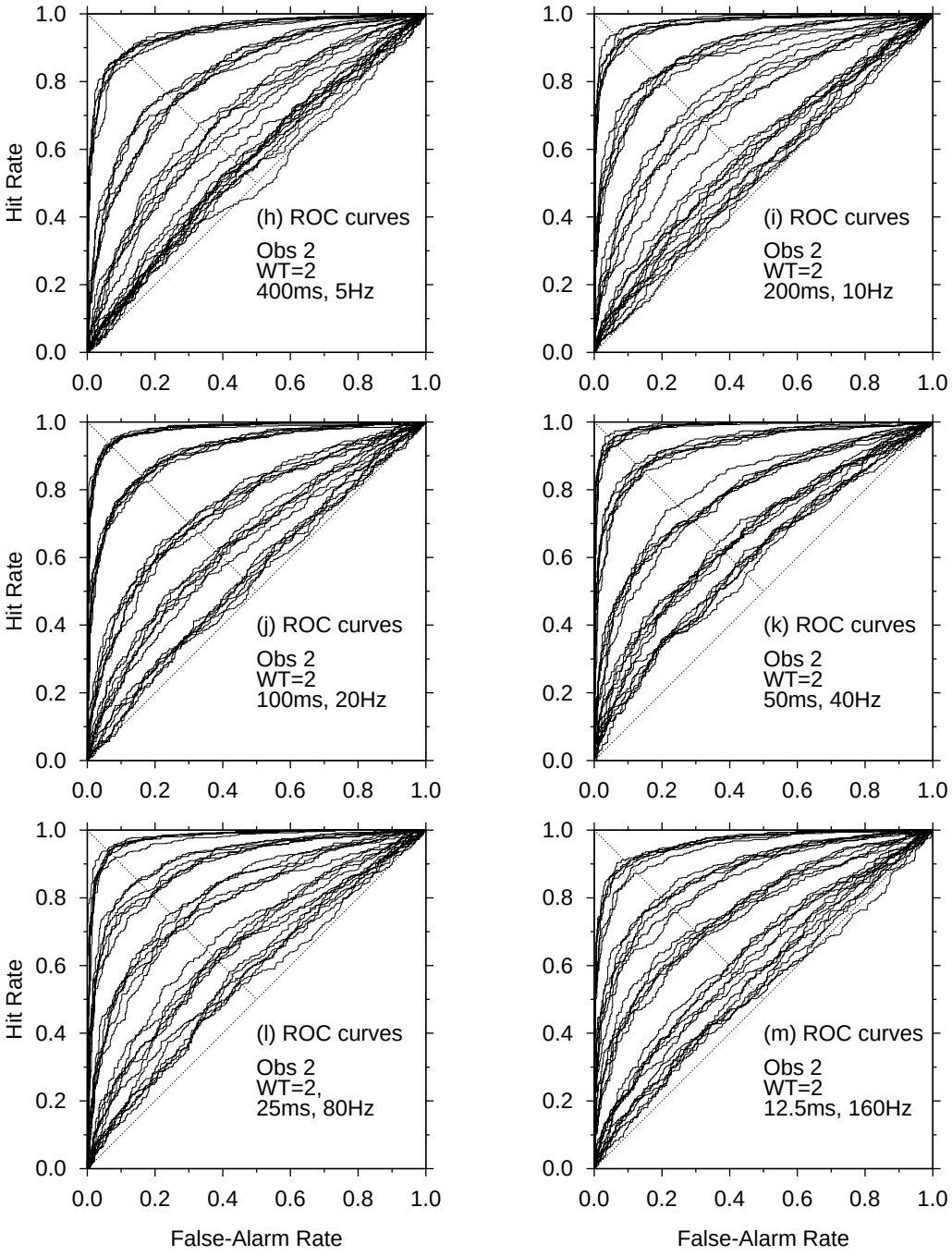

Figure D.5: ROC curves for Observer 2 in condition $\mathcal{WT}=2$.

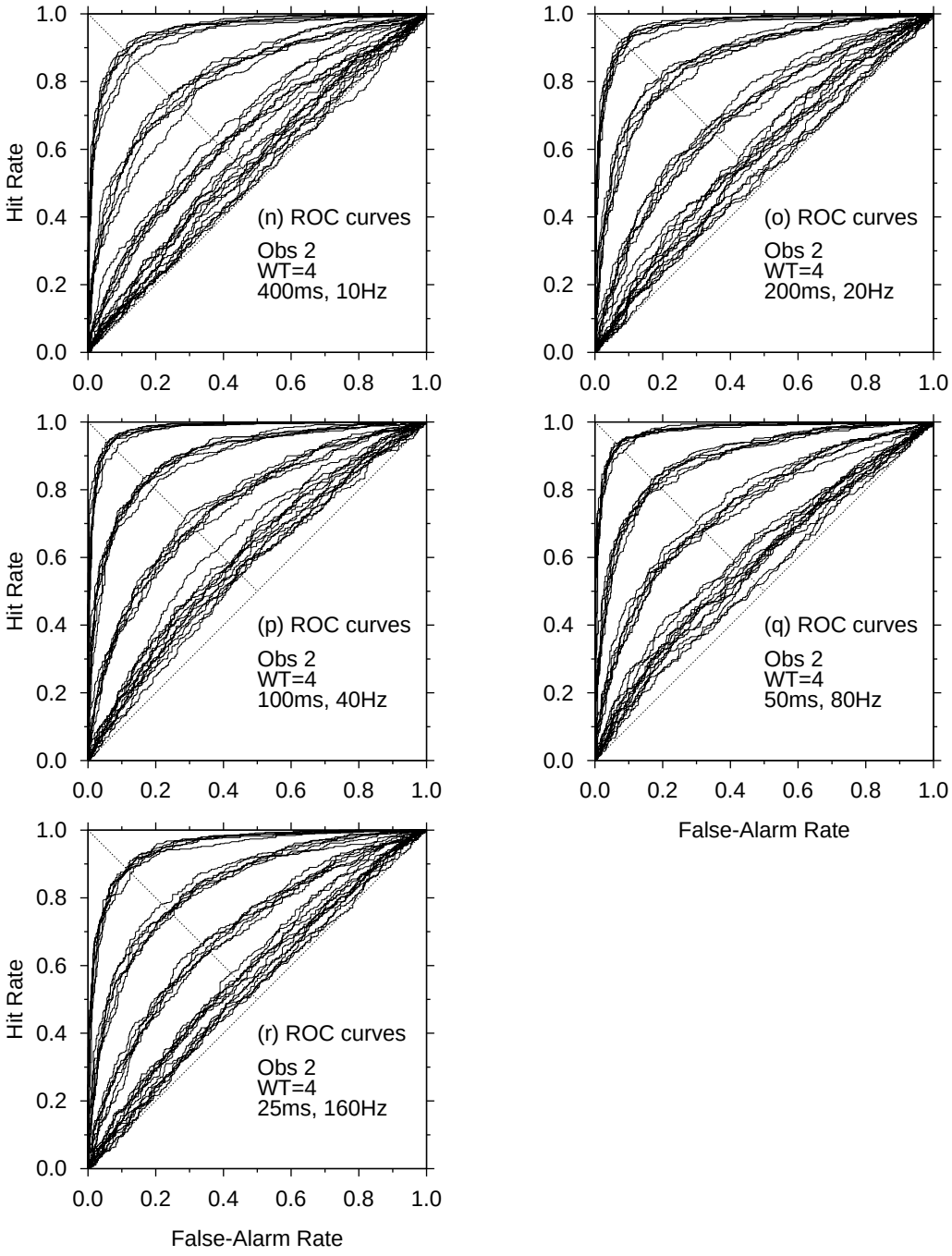

Figure D.6: ROC curves for Observer 2 in condition $\mathcal{WT}$=4.

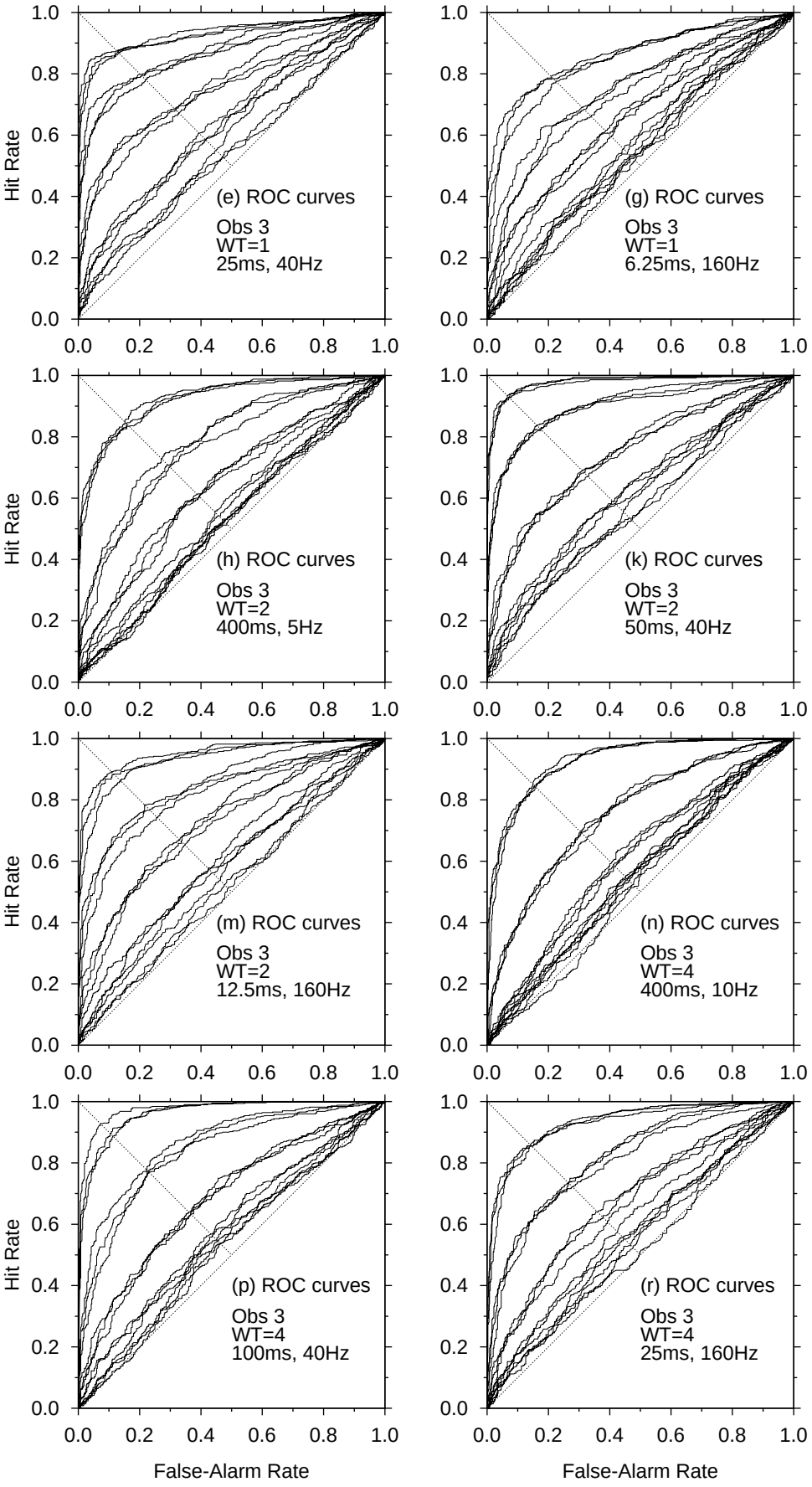

Figure D.7: ROC curves for Observer 3 in all conditions.

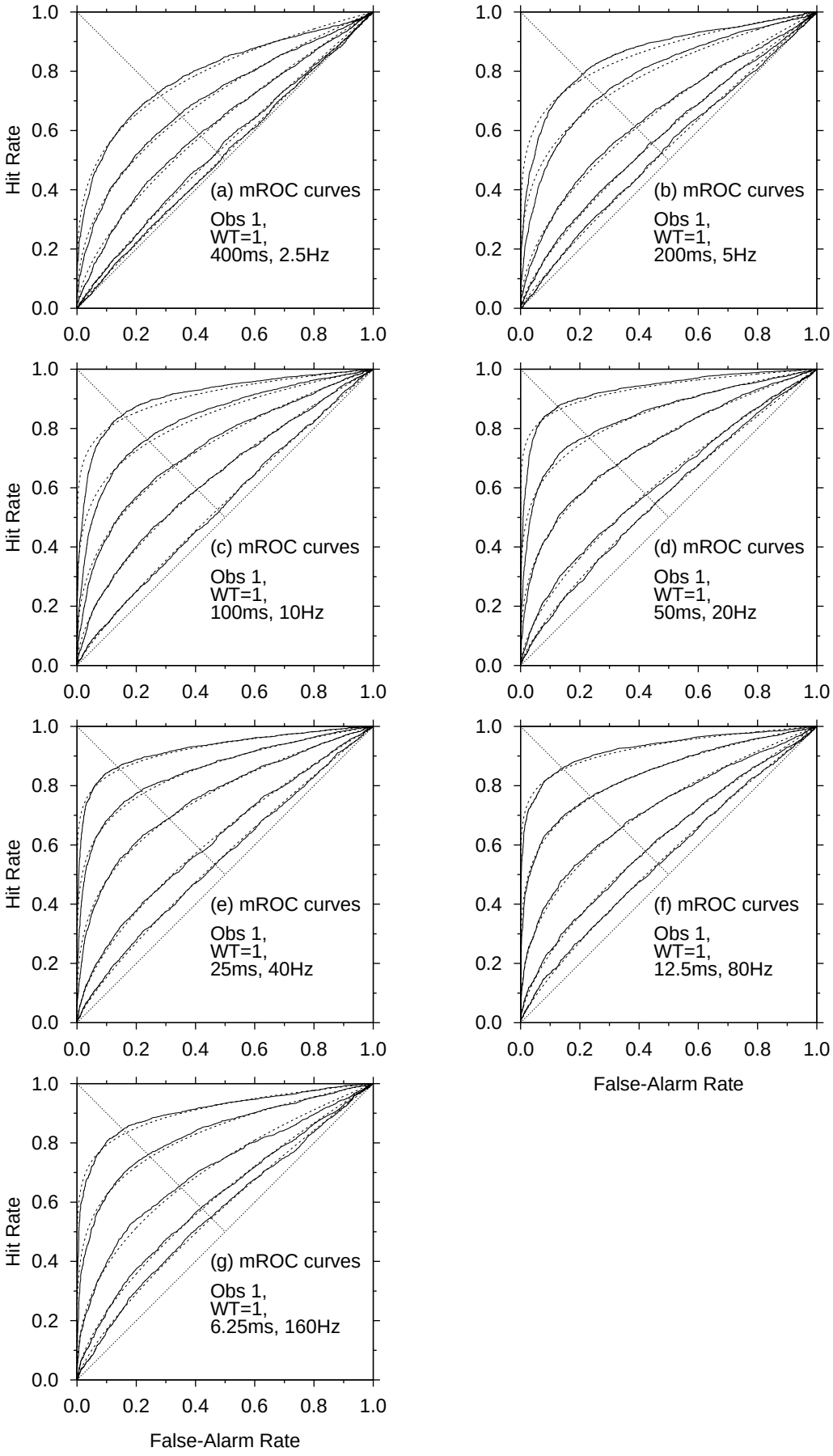

Figure D.8: The mean ROC curves for Observer 1 for condition $\mathcal{WT}=1$ compared with the χ^2 energy detector for $\mathcal{WT}=1$.

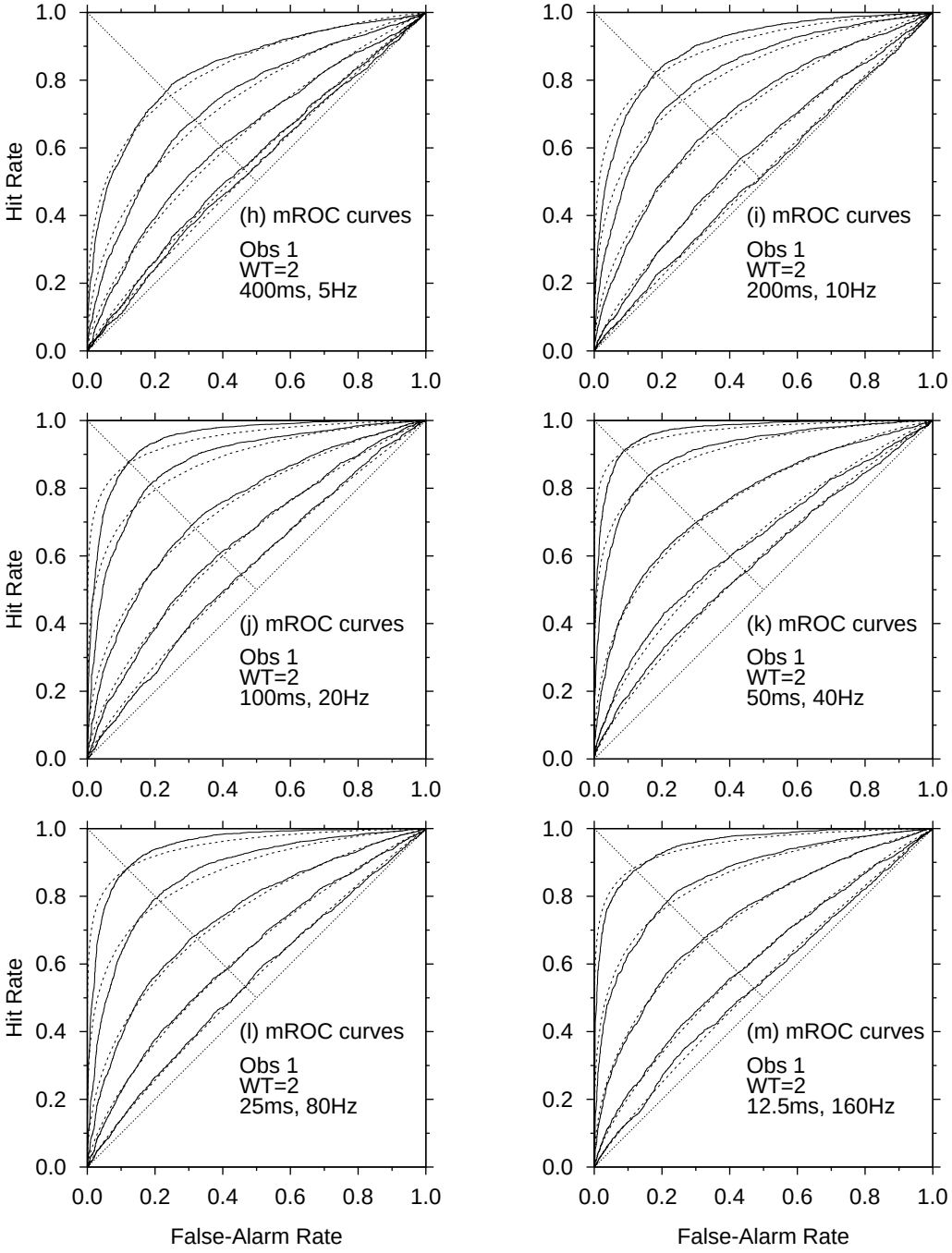

Figure D.9: The mean ROC curves for Observer 1 in condition $\mathcal{WT}=2$ compared with the χ^2 energy detector for $\mathcal{WT}=3$.

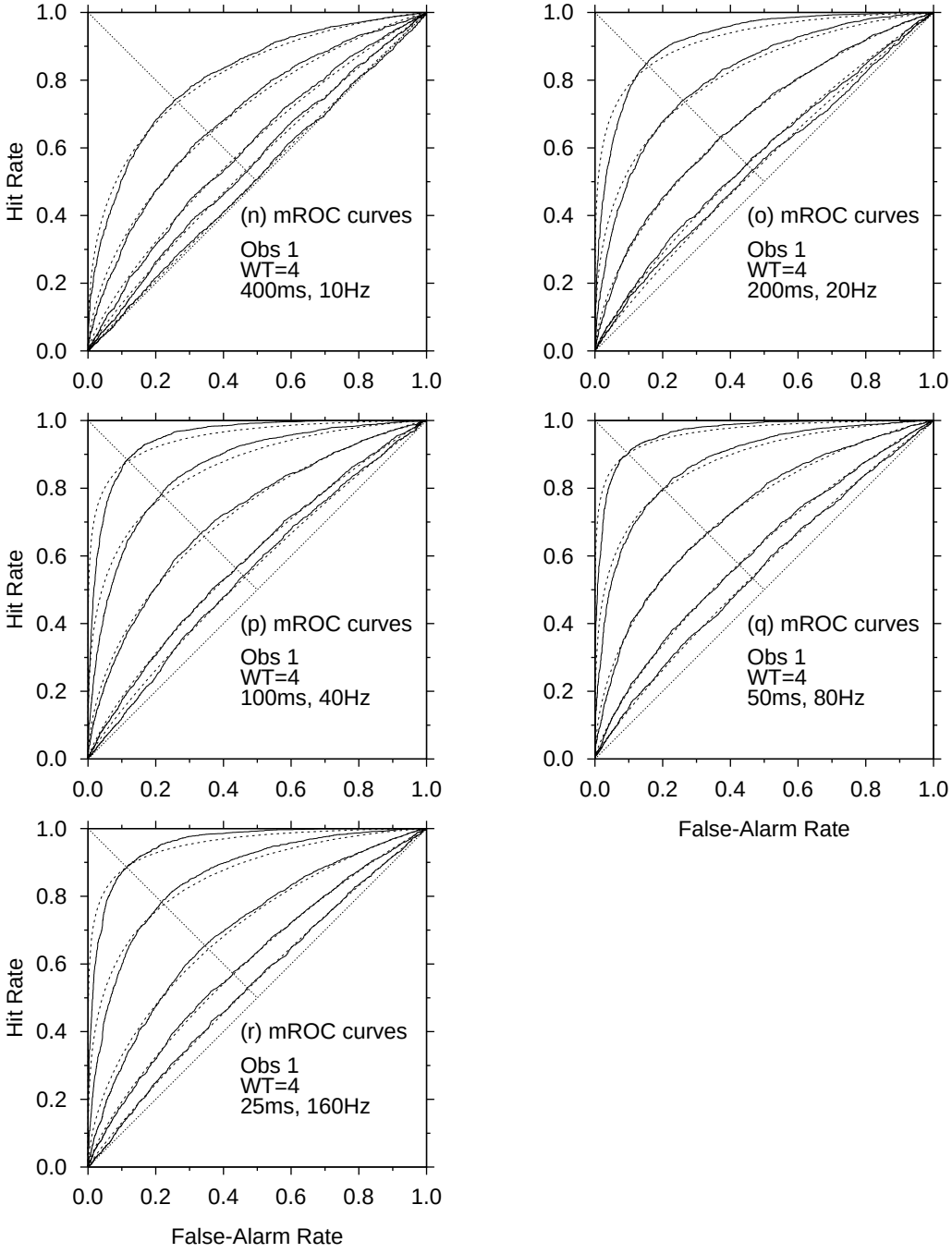

Figure D.10: The mean ROC curves for Observer 1 in condition $\mathcal{WT}$=4 compared with the χ^2 energy detector for $\mathcal{WT}$=5.

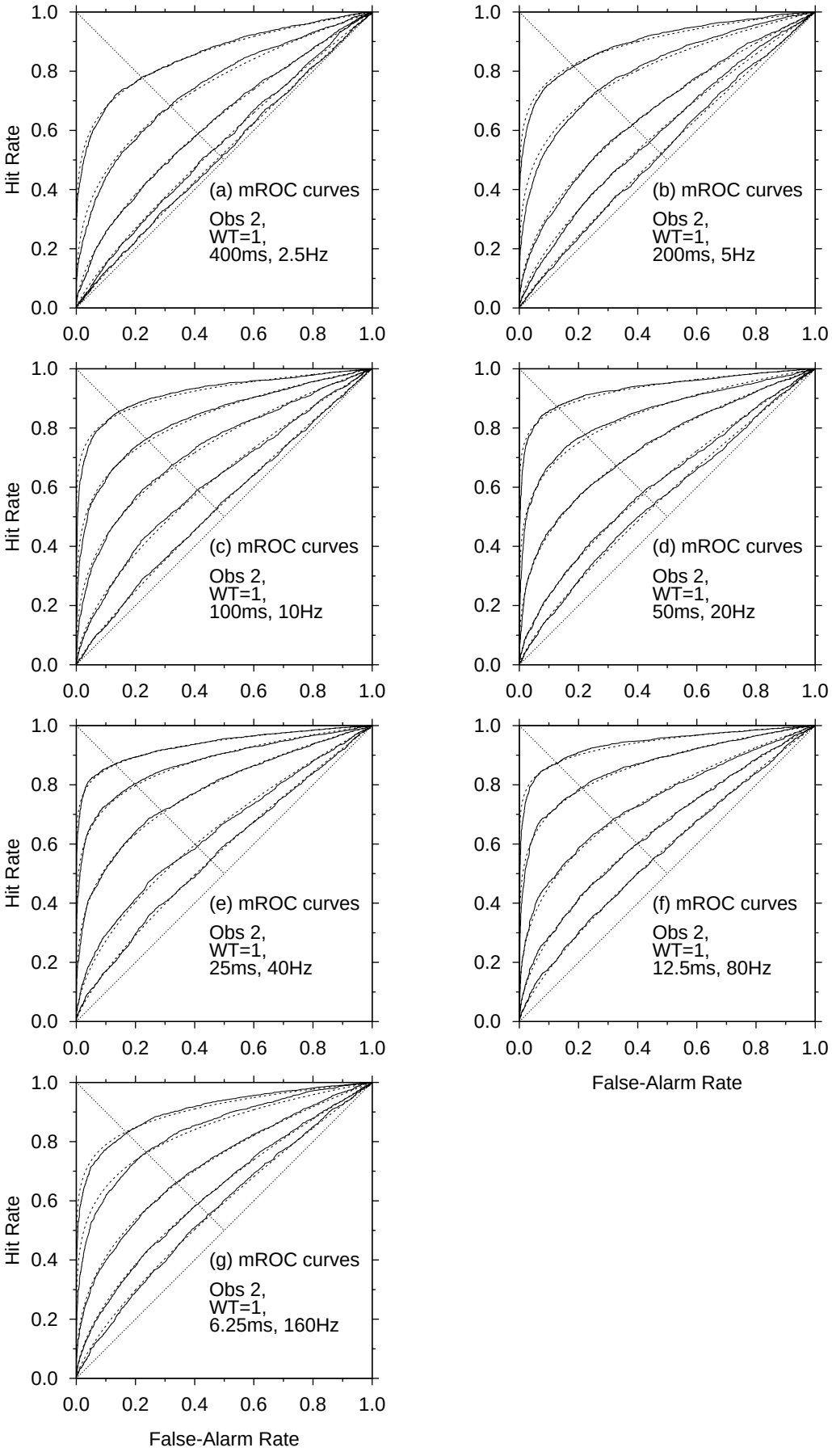

Figure D.11: The mean ROC curves for Observer 2 in condition $\mathcal{WT}=1$ compared with the χ^2 energy detector for $\mathcal{WT}=1$.

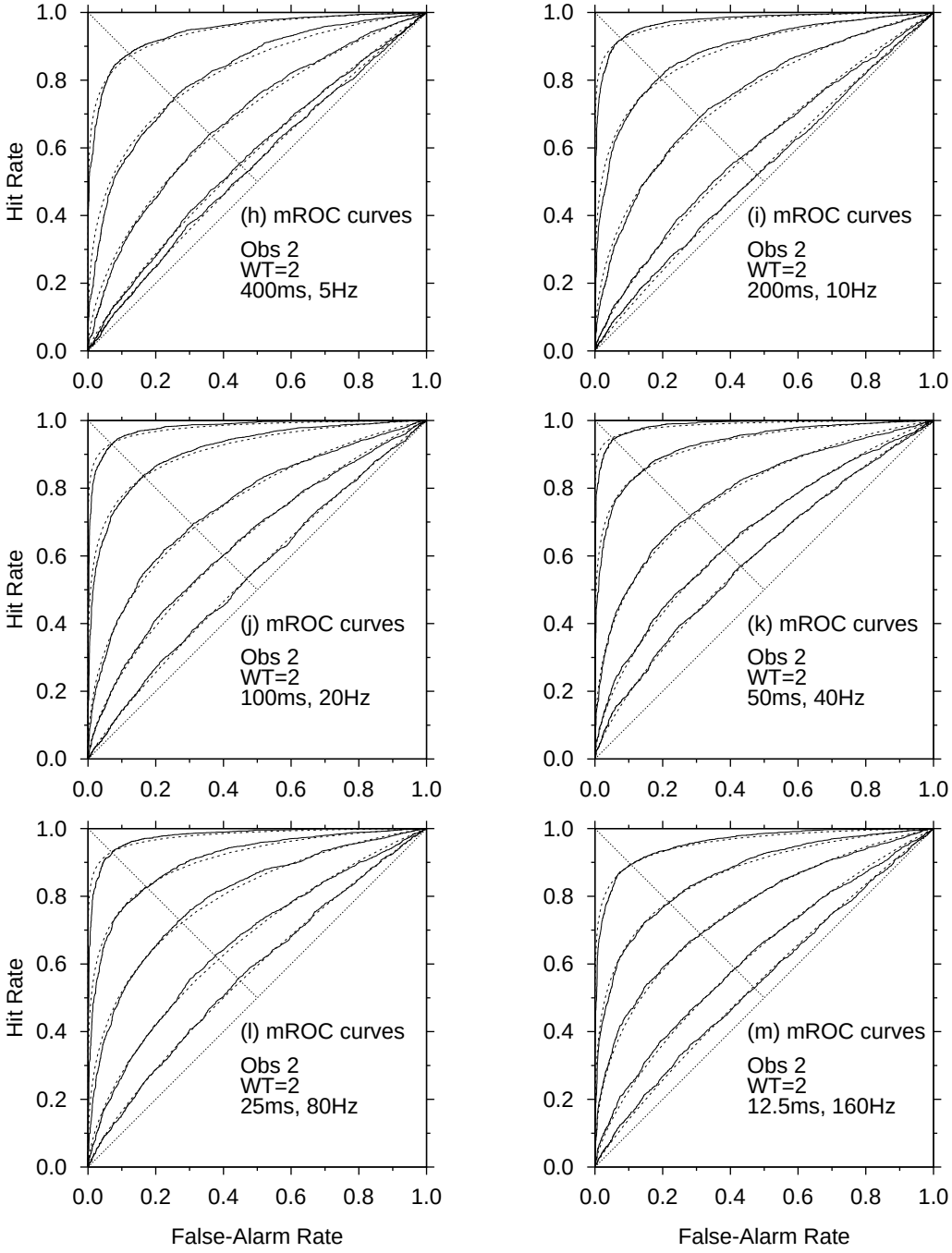

Figure D.12: The mean ROC curves for Observer 2 in condition $\mathcal{WT}=2$ compared with the χ^2 energy detector for $\mathcal{WT}=3$.

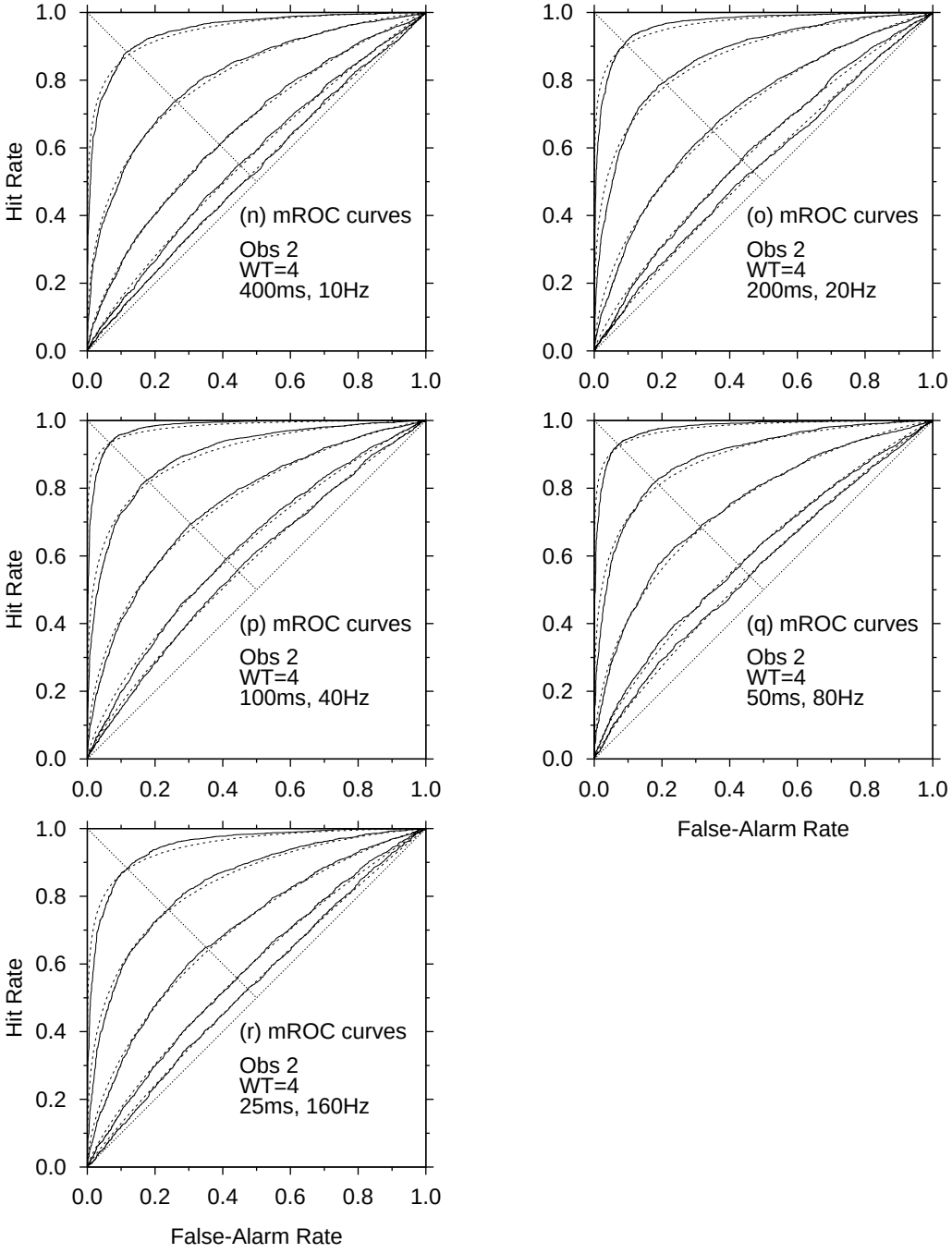

Figure D.13: The mean ROC curves for Observer 2 in condition $\mathcal{WT}=4$ compared with the χ^2 energy detector for $\mathcal{WT}=5$.

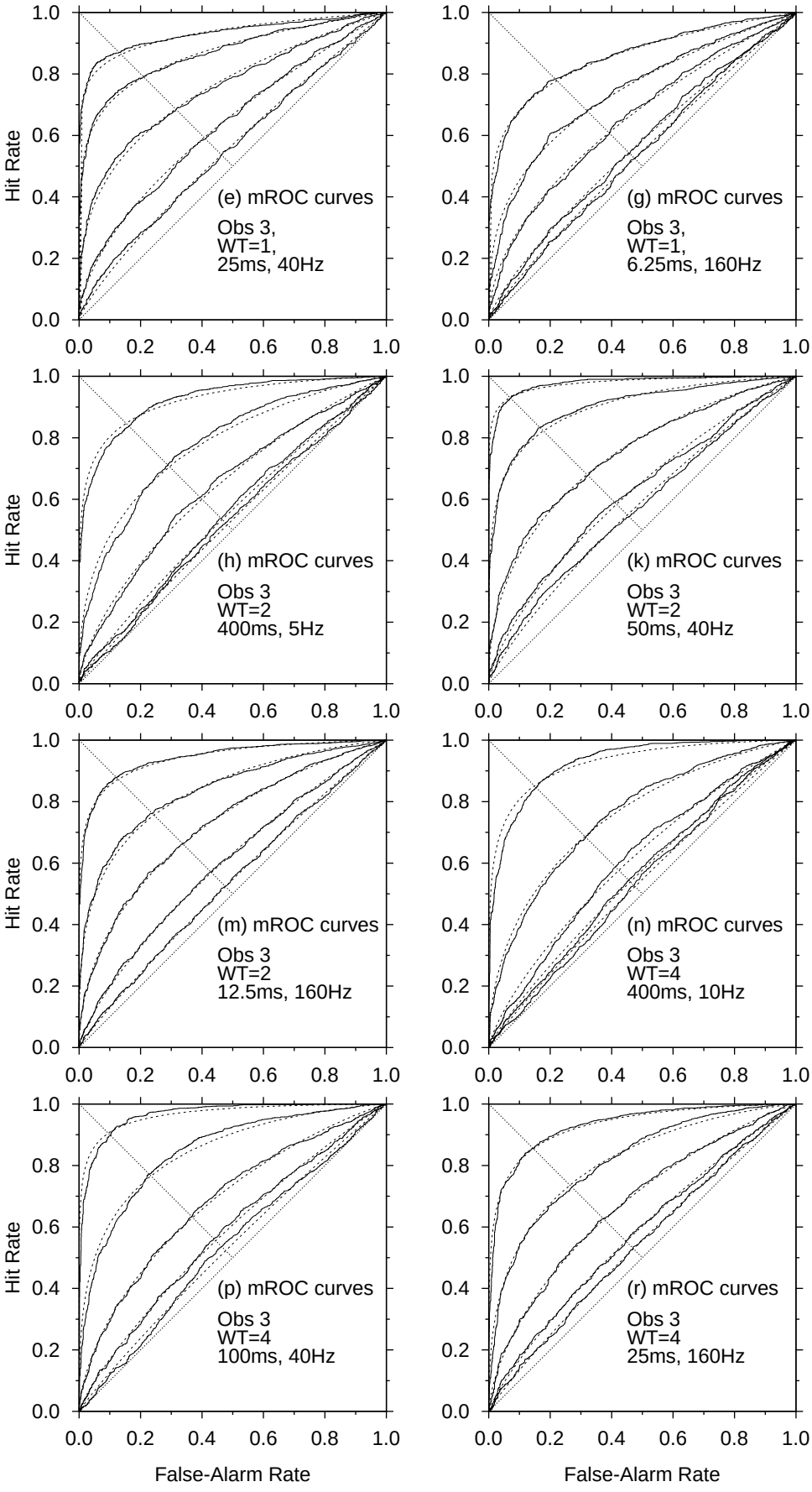

Figure D.14: The mean ROC curves for Observer 3 in all conditions compared with the χ^2 energy detector.

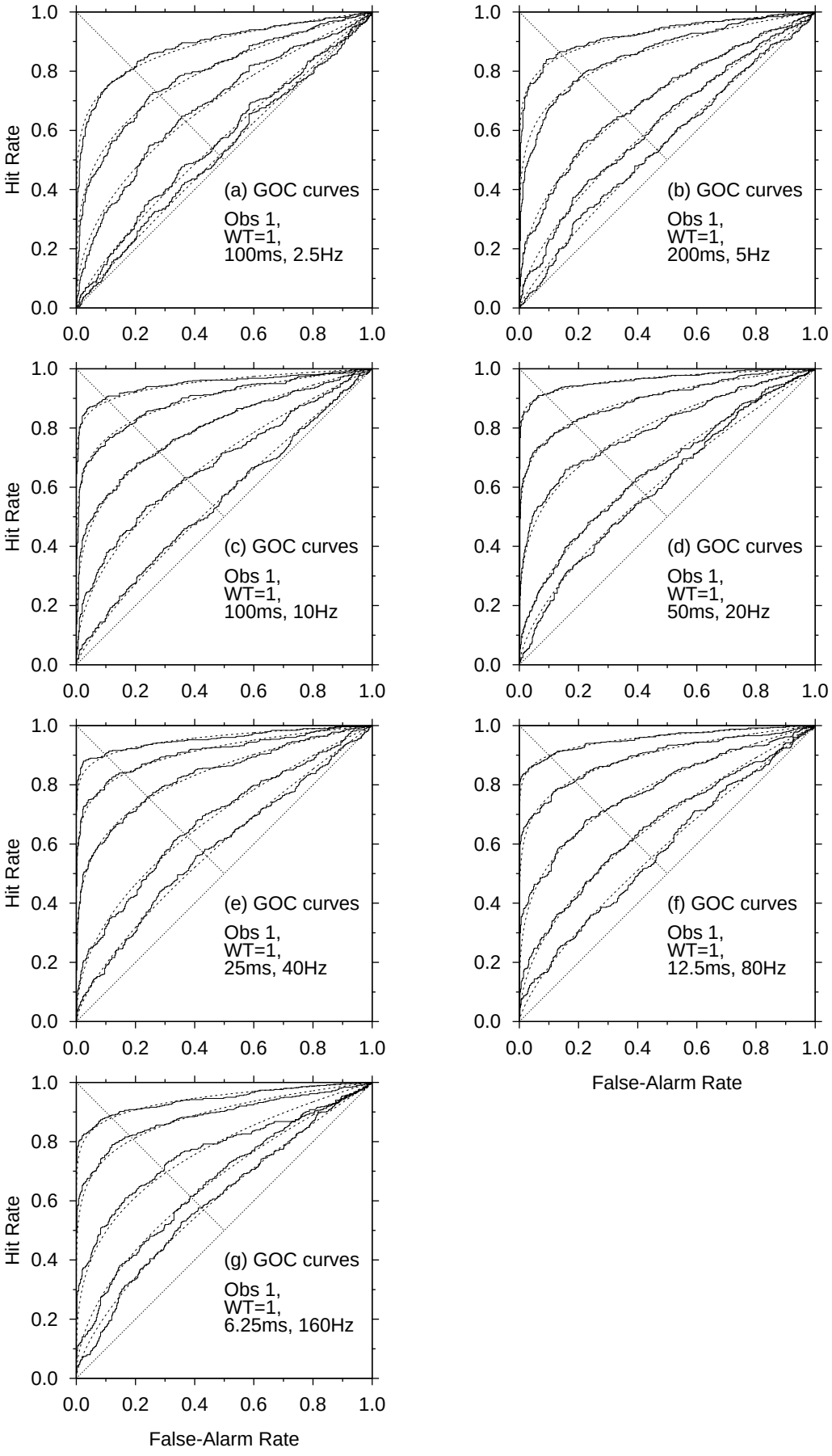

Figure D.15: The GOC curves for Observer 1 in condition $\mathcal{WT}=1$ compared with the χ^2 energy detector for $\mathcal{WT}=1$.

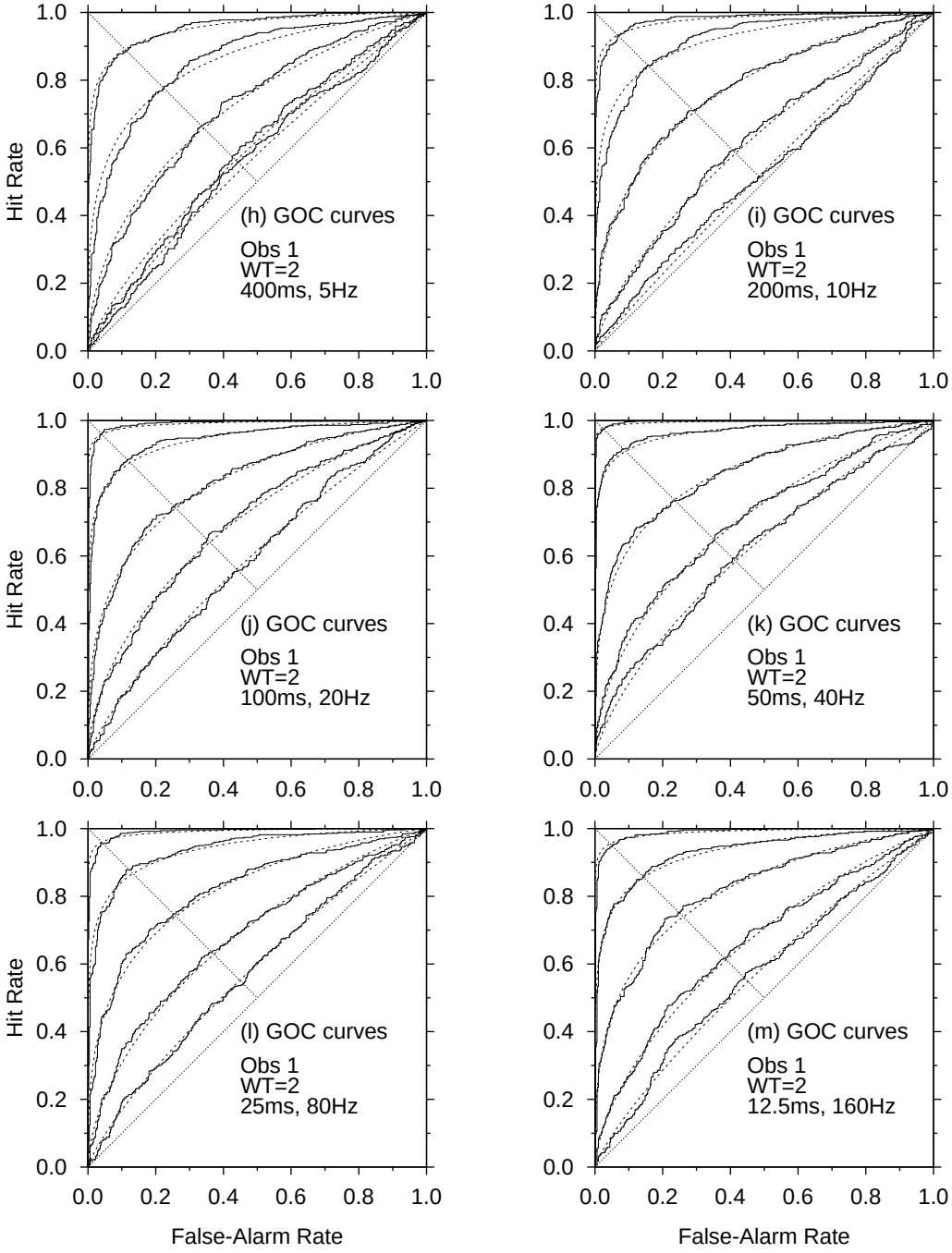

Figure D.16: The GOC curves for Observer 1 in condition $\mathcal{WT}$=2 compared with the χ^2 energy detector for $\mathcal{WT}$=3.

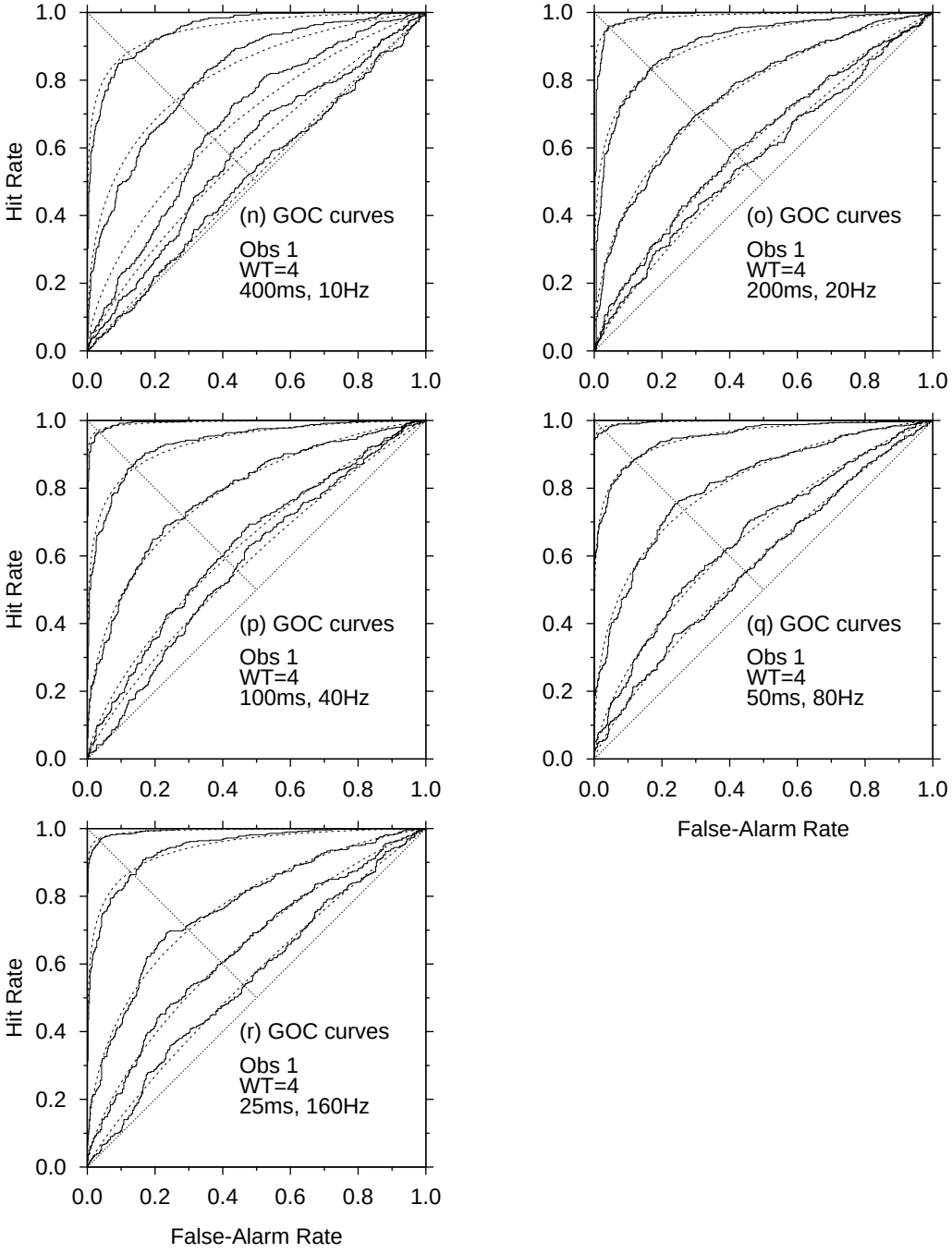

Figure D.17: The GOC curves for Observer 1 in condition $\mathcal{WT}$=4 compared with the χ^2 energy detector for $\mathcal{WT}$=5.

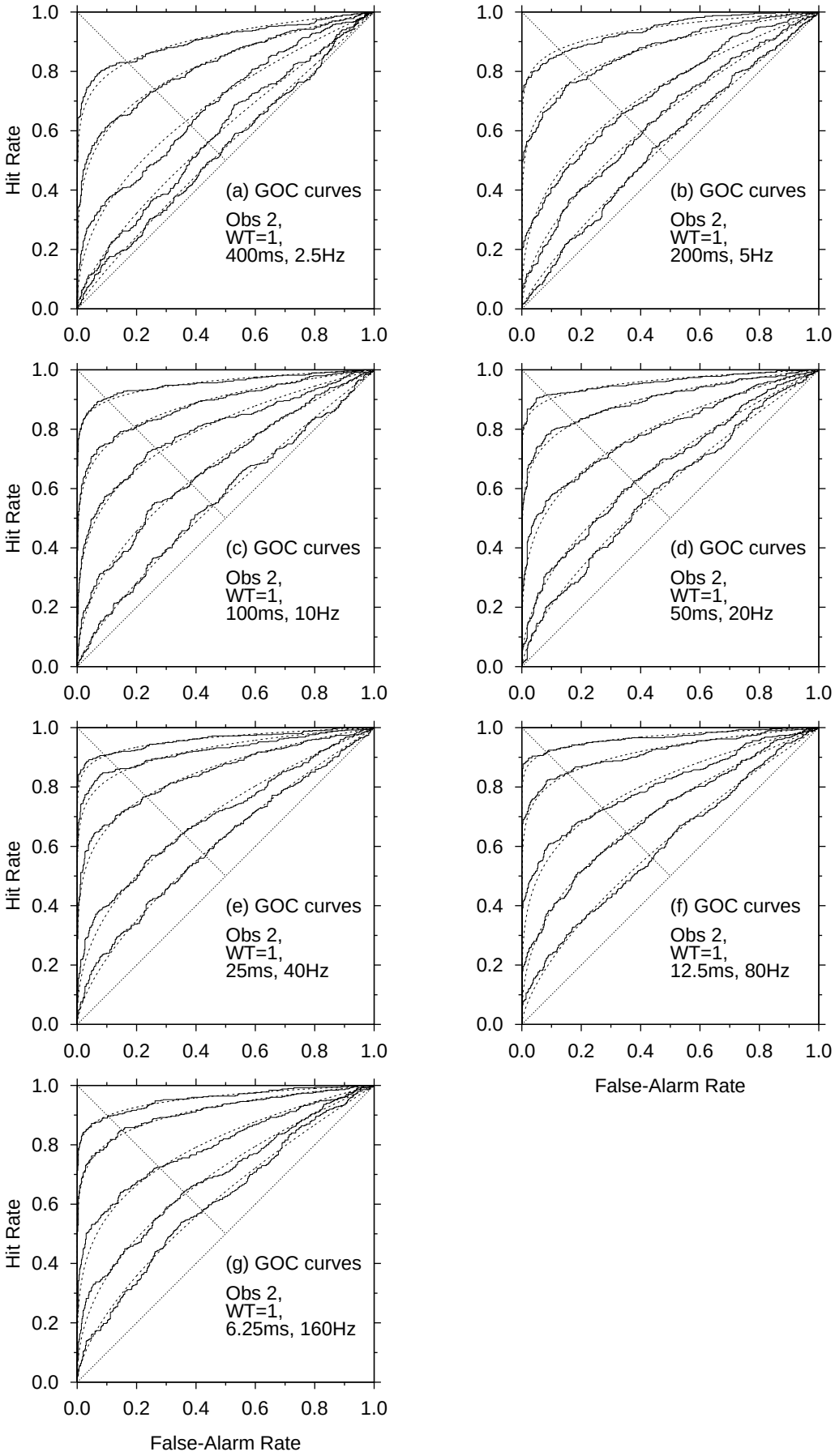

Figure D.18: The GOC curves for Observer 2 in condition $\mathcal{WT}=1$ compared with the χ^2 energy detector for $\mathcal{WT}=1$.

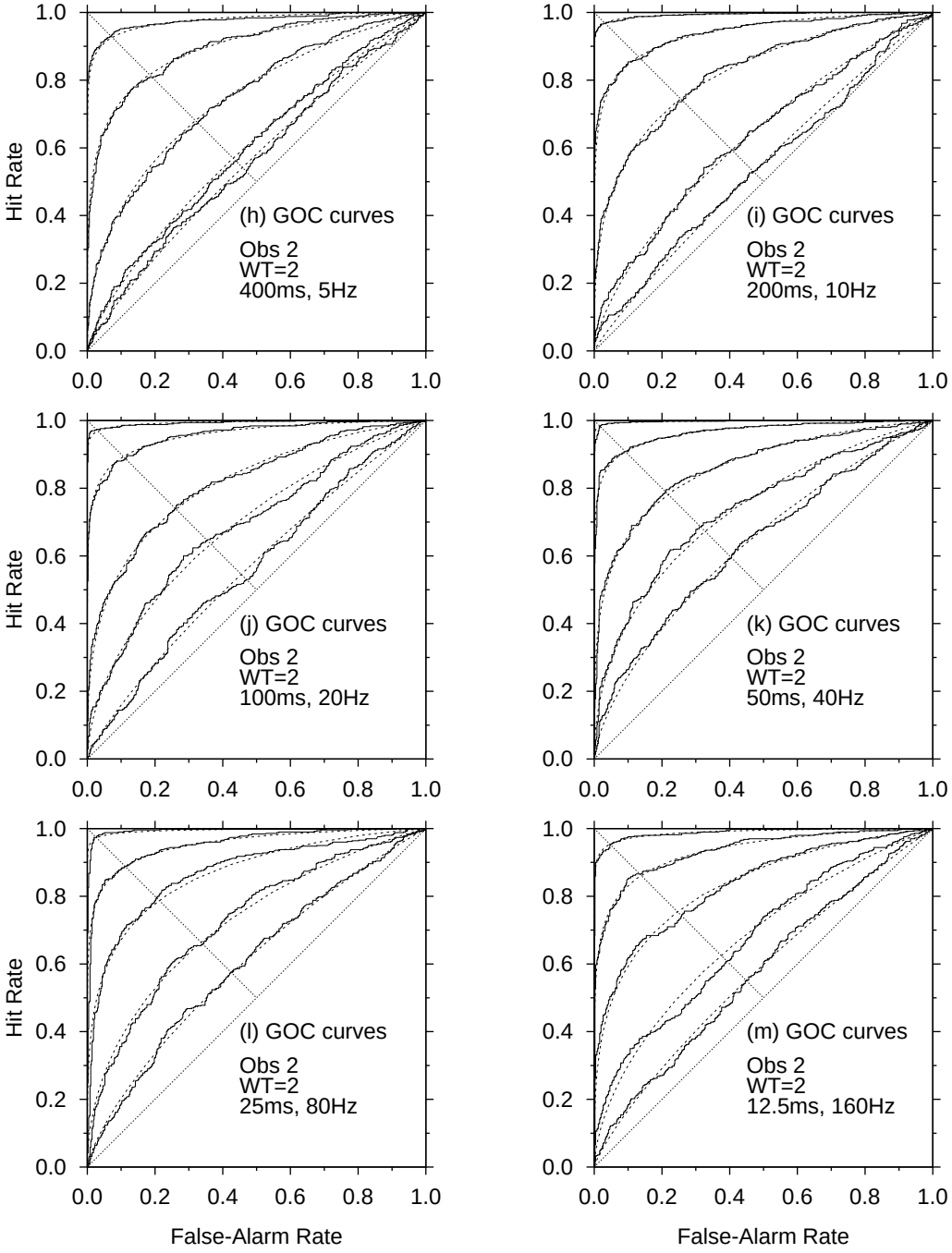

Figure D.19: The GOC curves for Observer 2 in condition $\mathcal{WT}$=2 compared with the χ^2 energy detector for $\mathcal{WT}$=3.

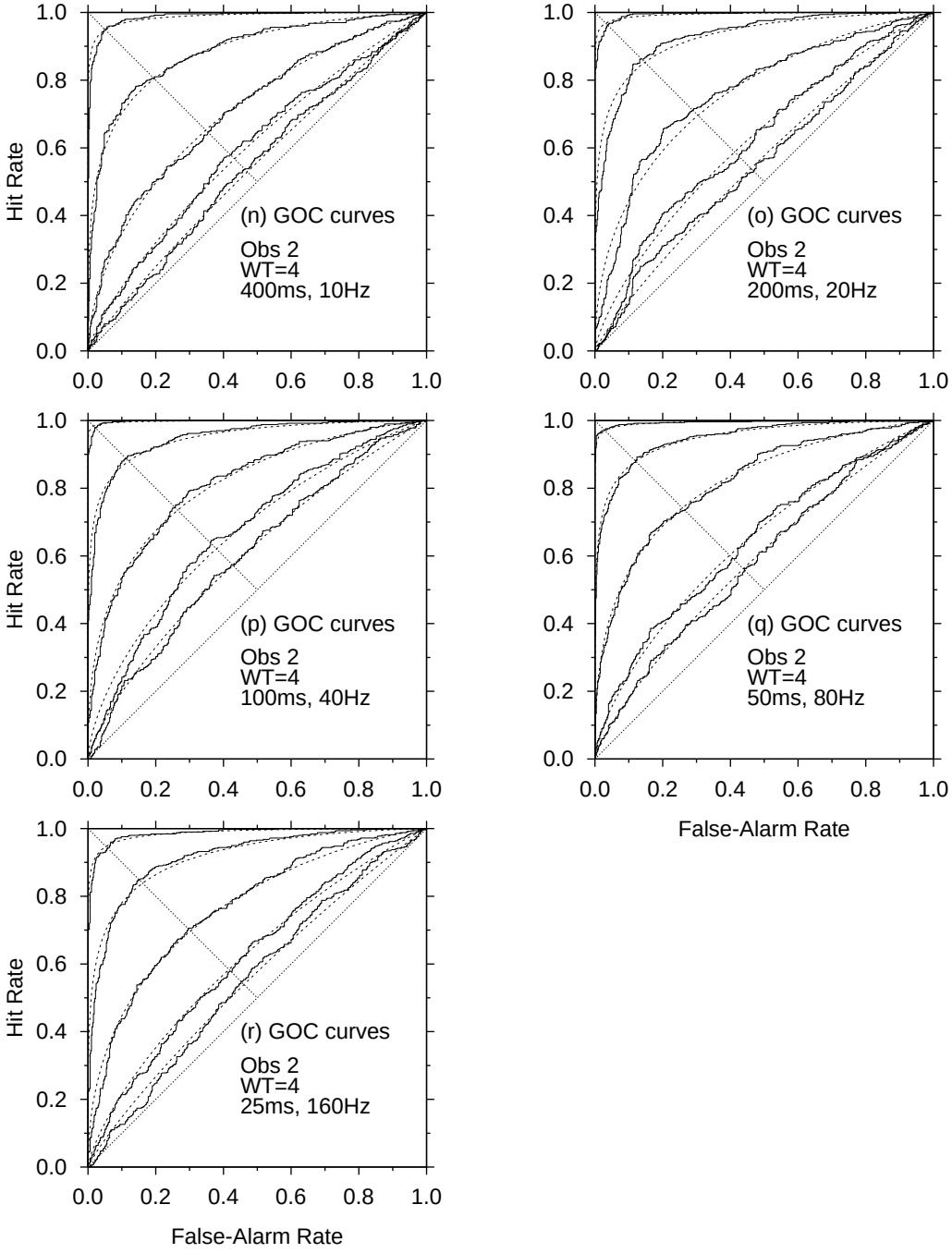

Figure D.20: The GOC curves for Observer 2 in condition $\mathcal{WT}$=4 compared with the χ^2 energy detector for $\mathcal{WT}$=5.

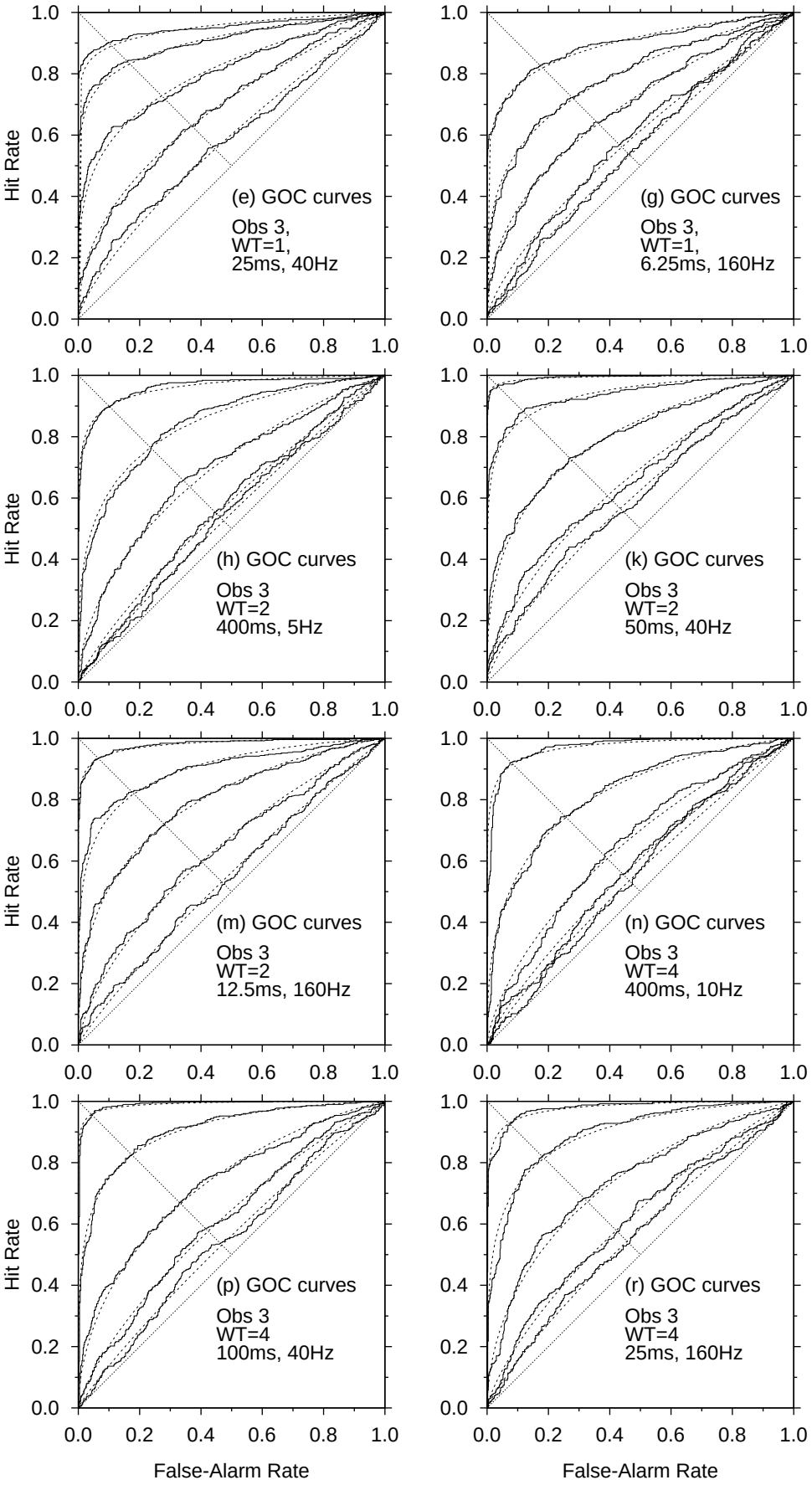

Figure D.21: The GOC curves for Observer 3 in all conditions compared with the χ^2 energy detector.

Appendix E

FORCE Analysis

The tables in this appendix show all the results from the FORCE analyses, including the parameters k and m used to fit the FORA, the asymptotic measure of detectability, and the correlation. The status of convergence is indicated with the codes: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications to do FORCE analysis.

Tables E.1–E.6 show the results for the measure $\mathcal{A}$, for Observer 1 and Observer 2, in each condition; Tables E.7–E.12 show the results for the measure $\mathcal{D}_2$, for Observer 1 and Observer 2, in each condition; and Tables E.13–E.18 show the results for the measure d', for Observer 1 and Observer 2, in each condition.

Figures E.1, E.3, E.5, E.7, E.9, and E.11 show the empirical FORA, the fitted FORA, and the asymptotic value for each signal–to–noise ratio for the measure $\mathcal{A}$. The error bars are ± 1 standard deviation. Figures E.2, E.4, E.6, E.8, E.10, and E.12 show the FORA in log–increment coordinates for the measure $\mathcal{A}$.

Tables E.19 and E.20 show the results for the measure $\mathcal{D}_6$ for all conditions for the $\mathcal{D}_6$ analysis from Chapter 6. Figures E.13–E.18 show the empirical and fitted FORA with asymptotic values. Note that the range of the $\mathcal{D}_6$ axis covers only 5 bits, out of a maximum value of approximately 9.5 bits, to accentuate changes in the FORA. Error bars are ± 1 standard deviation. Finally, Figure E.19 shows the results from the $\mathcal{D}_6$ analyses in log–increment coordinates for all conditions.

Table E.1: FORA $\mathcal{A}$ estimates for Observer 1 in condition $\mathcal{WT}$=1.

Expt. Level	SNR (dB)	m	k	$\mathcal{A}_1$	$\mathcal{A}_\infty$	Correlation	Status
400ms, 2.5Hz	0	−1.6191	0.0138	0.5121	0.5291	−0.9920	♡
	4	−1.4315	0.0262	0.5361	0.5865	−0.9973	♡
	8	−1.4869	0.0788	0.6162	0.7473	−0.9987	♡
	12	−1.6258	0.1217	0.7040	0.8523	−0.9995	♡
	16	−1.9167	0.1747	0.7933	0.9207	−0.9984	♡
200ms, 5Hz	0	−1.6635	0.0240	0.5326	0.5597	−0.9973	♡
	4	−1.9233	0.0714	0.5799	0.6316	−0.9997	♡
	8	−1.8590	0.0941	0.6546	0.7299	−0.9988	♡
	12	−2.0662	0.1609	0.7853	0.8797	−1.0000	♡
	16	−2.4302	0.1996	0.8587	0.9326	−0.9999	♡
100ms, 10Hz	0	−1.6643	0.0235	0.5370	0.5635	−0.9977	♡
	4	−1.9396	0.0871	0.6330	0.6945	−0.9998	♡
	8	−1.9711	0.1239	0.7354	0.8188	−0.9998	♡
	12	−2.2133	0.1540	0.8332	0.9076	−0.9995	♡
	16	−2.8160	0.1796	0.9104	0.9541	−0.9998	♡
50ms, 20Hz	0	−1.6242	0.0550	0.5653	0.6325	−0.9993	♡
	4	−1.6850	0.0662	0.6133	0.6852	−1.0000	♡
	8	−2.1740	0.1356	0.7432	0.8120	−0.9994	♡
	12	−2.2653	0.1349	0.8451	0.9061	−0.9996	♡
	16	−2.7056	0.1261	0.9327	0.9671	−0.9994	♡
25ms, 40Hz	0	−1.3169	0.0320	0.5520	0.6402	−0.9986	♡
	4	−1.5430	0.0708	0.6206	0.7238	−0.9999	♡
	8	−1.9254	0.1406	0.7581	0.8594	−0.9998	♡
	12	−2.2313	0.1462	0.8532	0.9221	−0.9994	♡
	16	−2.6448	0.1161	0.9244	0.9582	−0.9984	♡
12.5ms, 80Hz	0	−1.5935	0.0324	0.5527	0.5949	−1.0000	♡
	4	−1.6427	0.0677	0.6126	0.6924	−0.9999	♡
	8	−1.8748	0.1169	0.7182	0.8095	−0.9995	♡
	12	−2.1328	0.1275	0.8385	0.9068	−0.9998	♡
	16	−2.4011	0.0920	0.9250	0.9603	−0.9985	♡
6.25ms, 160Hz	0	−1.6706	0.0486	0.5657	0.6199	−0.9977	♡
	4	−1.8102	0.0802	0.6097	0.6792	−0.9998	♡
	8	−1.9811	0.1193	0.7064	0.7855	−0.9997	♡
	12	−2.3504	0.1616	0.8298	0.8955	−0.9998	♡
	16	−2.5021	0.1188	0.9074	0.9479	−0.9995	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.2: FORA $\mathcal{A}$ estimates for Observer 1 in condition $\mathcal{WT}$=2.

Expt. Level	SNR (dB)	m	k	$\mathcal{A}_1$	$\mathcal{A}_\infty$	Correlation	Status
400ms, 5Hz	-4	-1.1278	0.0202	0.5345	0.6844	-0.9954	♡
	0	-1.3356	0.0369	0.5537	0.6488	-0.9991	♡
	4	-1.3652	0.0875	0.6351	0.8399	-0.9991	♡
	8	-1.5293	0.1536	0.7349	0.9657	-0.9994	♡
	12	-1.9249	0.2286	0.8351	0.9999	-0.9994	♡
200ms, 10Hz	-4	-3.2242	0.0324	0.5182	0.5234	-0.9584	♡
	0	-1.5911	0.0492	0.5825	0.6469	-1.0000	♡
	4	-1.7821	0.1233	0.7038	0.8159	-1.0000	♡
	8	-1.9391	0.1914	0.8183	0.9533	-0.9998	♡
	12	-2.3901	0.2554	0.8946	0.9936	-1.0000	♡
100ms, 20Hz	-4	-1.7053	0.0375	0.5612	0.6003	-0.9995	♡
	0	-1.7927	0.0883	0.6424	0.7213	-0.9996	♡
	4	-2.0120	0.1500	0.7472	0.8423	-1.0000	♡
	8	-2.3862	0.1960	0.8797	0.9560	-1.0000	♡
	12	-2.9471	0.2297	0.9451	0.9940	-0.9999	♡
50ms, 40Hz	-4	-1.6229	0.0564	0.5787	0.6479	-0.9990	♡
	0	-1.6462	0.0938	0.6385	0.7481	-0.9998	♡
	4	-1.9725	0.1599	0.7655	0.8729	-1.0000	♡
	8	-2.6063	0.2195	0.9073	0.9740	-0.9999	♡
	12	-3.5020	0.2348	0.9671	0.9968	-0.9996	♡
25ms, 80Hz	-4	-1.4786	0.0317	0.5469	0.6008	-0.9990	♡
	0	-1.6641	0.0805	0.6241	0.7150	-0.9999	♡
	4	-1.8767	0.1457	0.7389	0.8522	-1.0000	♡
	8	-2.2744	0.1962	0.8658	0.9534	-0.9999	♡
	12	-2.9493	0.2091	0.9493	0.9937	-0.9997	♡
12.5ms, 160Hz	-4	-1.5551	0.0366	0.5404	0.5922	-0.9990	♡
	0	-1.6133	0.0728	0.6051	0.6961	-1.0000	♡
	4	-1.8701	0.1479	0.7353	0.8514	-0.9998	♡
	8	-2.2543	0.1885	0.8589	0.9453	-0.9998	♡
	12	-2.7840	0.1674	0.9505	0.9926	-0.9994	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.3: FORA $\mathcal{A}$ estimates for Observer 1 in condition $\mathcal{WT}$=4.

Expt. Level	SNR (dB)	m	k	$\mathcal{A}_1$	$\mathcal{A}_\infty$	Correlation	Status
400ms, 10Hz	-8	-0.3432	0.0016	0.5067	0.0000	-0.4416	♠
	-4	-1.2625	0.0340	0.5474	0.6630	-0.9974	♡
	0	-1.2112	0.0675	0.5963	0.8882	-0.9982	♡
	4	-1.4749	0.1443	0.6955	0.9430	-0.9995	♡
	8	-1.8533	0.2420	0.8129	1.0083	-0.9997	♣
200ms, 20Hz	-8	-1.7736	0.0396	0.5423	0.5789	-0.9958	♡
	-4	-1.6626	0.0553	0.5715	0.6341	-1.0000	♡
	0	-1.7963	0.1355	0.6784	0.7987	-0.9999	♡
	4	-2.0762	0.2115	0.8136	0.9360	-1.0000	♡
	8	-2.6508	0.2669	0.9196	0.9967	-1.0000	♡
100ms, 40Hz	-8	-1.5904	0.0350	0.5523	0.5982	-0.9963	♡
	-4	-1.7945	0.0661	0.5905	0.6493	-0.9998	♡
	0	-1.9589	0.1344	0.7177	0.8098	-0.9999	♡
	4	-2.2698	0.1971	0.8607	0.9492	-0.9998	♡
	8	-2.9013	0.2109	0.9509	0.9980	-0.9999	♡
50ms, 80Hz	-8	-1.7513	0.0363	0.5520	0.5869	-0.9999	♡
	-4	-1.7028	0.0738	0.6064	0.6837	-0.9999	♡
	0	-1.9416	0.1533	0.7321	0.8400	-0.9999	♡
	4	-2.3359	0.2095	0.8784	0.9651	-0.9999	♡
	8	-3.2619	0.2191	0.9639	0.9984	-0.9999	♡
25ms, 160Hz	-8	-1.6956	0.0258	0.5397	0.5671	-0.9933	♡
	-4	-1.6366	0.0626	0.6006	0.6752	-0.9988	♡
	0	-1.8656	0.1292	0.6993	0.8015	-0.9996	♡
	4	-2.1469	0.2034	0.8586	0.9656	-1.0000	♡
	8	-3.0398	0.2174	0.9546	0.9968	-0.9996	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.4: FORA $\mathcal{A}$ estimates for Observer 2 in condition $\mathcal{WT}$=1.

Expt. Level	SNR (dB)	m	k	$\mathcal{A}_1$	$\mathcal{A}_\infty$	Correlation	Status
400ms, 2.5Hz	0	−1.5606	0.0157	0.5204	0.5424	−0.9996	♡
	4	−1.3346	0.0350	0.5516	0.6422	−0.9992	♡
	8	−1.5408	0.0707	0.6301	0.7337	−0.9997	♡
	12	−1.8216	0.1148	0.7503	0.8480	−0.9999	♡
	16	−2.1226	0.1101	0.8574	0.9173	−0.9999	♡
200ms, 5Hz	0	−1.3322	0.0202	0.5327	0.5853	−0.9929	♡
	4	−1.5906	0.0572	0.5955	0.6705	−0.9994	♡
	8	−1.6438	0.0788	0.6715	0.7641	−0.9996	♡
	12	−1.9976	0.1263	0.8084	0.8902	−0.9999	♡
	16	−2.0503	0.0884	0.8979	0.9509	−1.0000	♡
100ms, 10Hz	0	−1.4984	0.0279	0.5370	0.5822	−0.9978	♡
	4	−1.7530	0.0801	0.6215	0.6982	−0.9999	♡
	8	−1.9696	0.1249	0.7347	0.8189	−0.9999	♡
	12	−2.1075	0.1146	0.8353	0.8988	−0.9998	♡
	16	−2.4850	0.1043	0.9217	0.9580	−0.9998	♡
50ms, 20Hz	0	−1.5576	0.0427	0.5607	0.6209	−0.9995	♡
	4	−1.5847	0.0677	0.6149	0.7049	−0.9993	♡
	8	−1.9485	0.1008	0.7396	0.8097	−0.9998	♡
	12	−2.1925	0.1160	0.8500	0.9075	−0.9993	♡
	16	−2.5605	0.0823	0.9340	0.9603	−0.9995	♡
25ms, 40Hz	0	−1.3391	0.0386	0.5672	0.6656	−0.9985	♡
	4	−1.6244	0.0809	0.6413	0.7402	−0.9989	♡
	8	−1.9970	0.1337	0.7810	0.8676	−1.0000	♡
	12	−2.2214	0.1112	0.8775	0.9307	−0.9996	♡
	16	−2.2038	0.0584	0.9357	0.9642	−0.9990	♡
12.5ms, 80Hz	0	−1.7346	0.0530	0.5666	0.6190	−0.9997	♡
	4	−1.6562	0.0911	0.6435	0.7479	−1.0000	♡
	8	−2.0793	0.1363	0.7437	0.8222	−0.9999	♡
	12	−2.2299	0.1286	0.8659	0.9267	−0.9998	♡
	16	−2.5678	0.0952	0.9389	0.9691	−0.9987	♡
6.25ms, 160Hz	0	−1.4796	0.0434	0.5739	0.6475	−0.9995	♡
	4	−1.6844	0.0842	0.6323	0.7237	−0.9995	♡
	8	−1.9011	0.1312	0.7257	0.8239	−0.9999	♡
	12	−2.1953	0.1712	0.8440	0.9285	−0.9999	♡
	16	−2.2056	0.1132	0.9095	0.9646	−0.9998	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.5: FORA $\mathcal{A}$ estimates for Observer 2 in condition $\mathcal{WT}$=2.

Expt. Level	SNR (dB)	m	k	$\mathcal{A}_1$	$\mathcal{A}_\infty$	Correlation	Status
400ms, 5Hz	−4	−2.0013	0.0355	0.5424	0.5653	−0.9989	♡
	0	−1.7148	0.0443	0.5691	0.6145	−0.9994	♡
	4	−1.8231	0.1024	0.6871	0.7740	−0.9999	♡
	8	−2.0316	0.1558	0.8229	0.9189	−0.9998	♡
	12	−2.6756	0.1243	0.9447	0.9797	−0.9993	♡
200ms, 10Hz	−4	−2.3073	0.0290	0.5311	0.5436	−0.9991	♡
	0	−1.7750	0.0610	0.5942	0.6504	−1.0000	♡
	4	−1.8875	0.1362	0.7485	0.8527	−1.0000	♡
	8	−2.3450	0.1601	0.8846	0.9502	−1.0000	♡
	12	−3.1539	0.1234	0.9733	0.9948	−0.9993	♡
100ms, 20Hz	−4	−1.7405	0.0320	0.5444	0.5757	−0.9966	♡
	0	−1.9710	0.0920	0.6423	0.7043	−0.9999	♡
	4	−2.0324	0.1343	0.7529	0.8356	−0.9999	♡
	8	−2.4193	0.1363	0.9131	0.9642	−0.9995	♡
	12	−3.2591	0.1001	0.9788	0.9946	−0.9993	♡
50ms, 40Hz	−4	−1.6155	0.0552	0.5965	0.6652	−0.9996	♡
	0	−1.7037	0.1036	0.6678	0.7763	−0.9999	♡
	4	−1.9851	0.1446	0.7910	0.8863	−0.9999	♡
	8	−2.5600	0.1282	0.9301	0.9710	−0.9990	♡
	12	−3.4059	0.0811	0.9868	0.9980	−0.9980	♡
25ms, 80Hz	−4	−1.7789	0.0555	0.5696	0.6203	−0.9979	♡
	0	−1.7404	0.1042	0.6594	0.7613	−1.0000	♡
	4	−2.0630	0.1532	0.7985	0.8888	−1.0000	♡
	8	−2.4281	0.1458	0.9071	0.9612	−0.9999	♡
	12	−3.0667	0.0930	0.9783	0.9959	−0.9995	♡
12.5ms, 160Hz	−4	−1.7627	0.0413	0.5469	0.5857	−0.9984	♡
	0	−1.7825	0.0867	0.6219	0.7007	−0.9997	♡
	4	−1.9682	0.1482	0.7584	0.8586	−1.0000	♡
	8	−2.2308	0.1760	0.8660	0.9490	−0.9998	♡
	12	−2.8426	0.1434	0.9571	0.9911	−0.9995	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.6: FORA $\mathcal{A}$ estimates for Observer 2 in condition $\mathcal{WT}$=4.

Expt. Level	SNR (dB)	m	k	$\mathcal{A}_1$	$\mathcal{A}_\infty$	Correlation	Status
400ms, 10Hz	-8	-1.6082	0.0193	0.5276	0.5520	-0.9988	♡
	-4	-1.4820	0.0337	0.5668	0.6236	-0.9997	♡
	0	-1.7522	0.0892	0.6544	0.7398	-0.9998	♡
	4	-2.0157	0.1553	0.8113	0.9092	-0.9998	♡
	8	-2.6528	0.1582	0.9466	0.9922	-0.9999	♡
200ms, 20Hz	-8	-1.7677	0.0235	0.5411	0.5630	-0.9986	♡
	-4	-1.6762	0.0502	0.5895	0.6447	-0.9982	♡
	0	-1.9093	0.1137	0.7050	0.7890	-0.9998	♡
	4	-2.2360	0.1583	0.8678	0.9419	-1.0000	♡
	8	-2.9810	0.1532	0.9647	0.9963	-0.9992	♡
100ms, 40Hz	-8	-1.8075	0.0508	0.5729	0.6171	-0.9996	♡
	-4	-1.7241	0.0705	0.6229	0.6940	-1.0000	♡
	0	-2.0180	0.1238	0.7536	0.8314	-0.9999	♡
	4	-2.4047	0.1538	0.8990	0.9577	-0.9997	♡
	8	-3.3350	0.1245	0.9804	0.9987	-0.9998	♡
50ms, 80Hz	-8	-1.8440	0.0454	0.5614	0.5986	-0.9993	♡
	-4	-1.7826	0.0635	0.6046	0.6623	-0.9999	♡
	0	-1.9687	0.1413	0.7494	0.8447	-0.9998	♡
	4	-2.3503	0.1693	0.8907	0.9596	-0.9999	♡
	8	-3.3202	0.1295	0.9763	0.9956	-0.9987	♡
25ms, 160Hz	-8	-1.2964	0.0192	0.5361	0.5931	-0.9877	♡
	-4	-1.6418	0.0489	0.5838	0.6414	-0.9994	♡
	0	-1.8299	0.1210	0.6974	0.7989	-1.0000	♡
	4	-2.1365	0.1819	0.8397	0.9367	-0.9999	♡
	8	-2.7695	0.1586	0.9498	0.9903	-0.9998	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.7: FORA $\mathcal{D}_2$ estimates for Observer 1 in condition $\mathcal{WT}$=1.

Expt. Level	SNR (dB)	m	k	$\mathcal{D}_1$	$\mathcal{D}_\infty$	Correlation	Status
400ms, 2.5Hz	0	−1.0305	0.0005	0.0007	0.0171	−0.9757	♡
	4	−1.0270	0.0037	0.0050	0.1418	−0.9813	♡
	8	−1.1656	0.0490	0.0395	0.3154	−0.9952	♡
	12	−1.3268	0.1388	0.1240	0.4929	−0.9981	♡
	16	−1.6230	0.3098	0.2657	0.6447	−0.9961	♡
200ms, 5Hz	0	−1.4219	0.0039	0.0035	0.0112	−0.9949	♡
	4	−1.6522	0.0306	0.0190	0.0543	−0.9997	♡
	8	−1.6271	0.0803	0.0708	0.1683	−0.9973	♡
	12	−1.8214	0.2775	0.2505	0.4866	−0.9998	♡
	16	−2.1741	0.4719	0.4141	0.6536	−1.0000	♡
100ms, 10Hz	0	−1.3840	0.0041	0.0043	0.0135	−0.9961	♡
	4	−1.7130	0.0636	0.0522	0.1176	−0.9999	♡
	8	−1.7542	0.1702	0.1669	0.3294	−1.0000	♡
	12	−1.9933	0.3303	0.3506	0.5659	−0.9998	♡
	16	−2.6141	0.5607	0.5653	0.7340	−0.9999	♡
50ms, 20Hz	0	−1.2961	0.0190	0.0125	0.0690	−0.9980	♡
	4	−1.4387	0.0404	0.0379	0.1142	−0.9998	♡
	8	−1.9780	0.1958	0.1783	0.3088	−0.9997	♡
	12	−2.0796	0.3096	0.3785	0.5570	−0.9998	♡
	16	−2.5273	0.4513	0.6444	0.7939	−0.9997	♡
25ms, 40Hz	0	−0.9283	0.0075	0.0089	0.0000	−0.9870	♠
	4	−1.2623	0.0459	0.0427	0.1992	−0.9994	♡
	8	−1.6834	0.2149	0.2023	0.4360	−0.9992	♡
	12	−2.0257	0.3462	0.3991	0.6143	−0.9997	♡
	16	−2.4765	0.3918	0.6144	0.7519	−0.9989	♡
12.5ms, 80Hz	0	−1.3343	0.0089	0.0083	0.0315	−0.9999	♡
	4	−1.3846	0.0411	0.0374	0.1279	−0.9998	♡
	8	−1.6464	0.1462	0.1429	0.3140	−0.9998	♡
	12	−1.9425	0.2846	0.3622	0.5622	−1.0000	♡
	16	−2.2494	0.3172	0.6160	0.7622	−0.9988	♡
6.25ms, 160Hz	0	−1.3913	0.0169	0.0129	0.0494	−0.9939	♡
	4	−1.5485	0.0479	0.0353	0.1041	−0.9998	♡
	8	−1.7679	0.1422	0.1270	0.2596	−0.9998	♡
	12	−2.1594	0.3490	0.3421	0.5226	−0.9999	♡
	16	−2.3580	0.3756	0.5558	0.7071	−0.9997	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.8: FORA $\mathcal{D}_2$ estimates for Observer 1 in condition $\mathcal{WT}$=2.

Expt. Level	SNR (dB)	m	k	$\mathcal{D}_1$	$\mathcal{D}_\infty$	Correlation	Status
400ms, 5Hz	-4	-0.7429	0.0033	0.0039	0.0000	-0.9860	♠
	0	-0.9946	0.0103	0.0086	0.0000	-0.9956	♠
	4	-1.0084	0.0626	0.0536	7.4702	-0.9956	♣
	8	-1.1425	0.1988	0.1668	1.4789	-0.9971	♣
	12	-1.4782	0.4558	0.3553	1.1301	-0.9977	♣
200ms, 10Hz	-4	-0.6917	0.0003	0.0023	0.0000	-0.8928	♠
	0	-1.3130	0.0211	0.0201	0.0792	-0.9996	♡
	4	-1.5037	0.1393	0.1243	0.3468	-0.9996	♡
	8	-1.5931	0.3661	0.3182	0.7955	-0.9989	♡
	12	-1.9451	0.6607	0.5166	0.9784	-0.9996	♡
100ms, 20Hz	-4	-1.4694	0.0122	0.0111	0.0324	-0.9982	♡
	0	-1.5517	0.0690	0.0600	0.1585	-0.9986	♡
	4	-1.7698	0.2183	0.1845	0.3873	-0.9999	♡
	8	-2.1156	0.5152	0.4699	0.7524	-0.9999	♡
	12	-2.5442	0.8037	0.6935	0.9546	-1.0000	♡
50ms, 40Hz	-4	-1.3148	0.0229	0.0186	0.0821	-0.9964	♡
	0	-1.3569	0.0700	0.0572	0.2253	-0.9988	♡
	4	-1.7071	0.2503	0.2151	0.4754	-0.9995	♡
	8	-2.2946	0.6441	0.5561	0.8364	-1.0000	♡
	12	-3.0768	0.9620	0.7914	0.9718	-0.9999	♡
25ms, 80Hz	-4	-1.1647	0.0077	0.0066	0.0500	-0.9965	♡
	0	-1.3848	0.0540	0.0451	0.1639	-0.9993	♡
	4	-1.6052	0.2004	0.1721	0.4269	-0.9998	♡
	8	-1.9848	0.4783	0.4315	0.7468	-1.0000	♡
	12	-2.5653	0.7586	0.7113	0.9521	-1.0000	♡
12.5ms, 160Hz	-4	-1.2289	0.0077	0.0053	0.0356	-0.9998	♡
	0	-1.3178	0.0407	0.0327	0.1444	-0.9994	♡
	4	-1.5980	0.2007	0.1667	0.4257	-0.9989	♡
	8	-1.9807	0.4486	0.4134	0.7110	-1.0000	♡
	12	-2.4363	0.6186	0.7168	0.9442	-0.9998	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.9: FORA $\mathcal{D}_2$ estimates for Observer 1 in condition $\mathcal{WT}$=4.

Expt. Level	SNR (dB)	m	k	$\mathcal{D}_1$	$\mathcal{D}_\infty$	Correlation	Status
400ms, 10Hz	-8	1.6665	0.0000	0.0002	0.0000	0.8794	♠
	-4	-0.8582	0.0078	0.0071	0.0000	-0.9918	♠
	0	-0.8038	0.0336	0.0271	0.0000	-0.9921	♠
	4	-1.0647	0.1506	0.1141	2.3773	-0.9967	♣
	8	-1.3732	0.4332	0.3068	1.2949	-0.9981	♣
200ms, 20Hz	-8	-1.5051	0.0086	0.0059	0.0196	-0.9906	♡
	-4	-1.3591	0.0210	0.0149	0.0651	-0.9996	♡
	0	-1.4923	0.1334	0.0946	0.3136	-0.9997	♡
	4	-1.7521	0.4063	0.3063	0.6955	-0.9994	♡
	8	-2.1865	0.7836	0.5966	0.9875	-0.9997	♡
100ms, 40Hz	-8	-1.3160	0.0089	0.0089	0.0334	-0.9905	♡
	-4	-1.5517	0.0327	0.0242	0.0709	-0.9992	♡
	0	-1.7163	0.1683	0.1415	0.3136	-1.0000	♡
	4	-1.9864	0.4716	0.4180	0.7283	-1.0000	♡
	8	-2.4561	0.7503	0.7180	0.9874	-1.0000	♡
50ms, 80Hz	-8	-1.4613	0.0094	0.0083	0.0249	-0.9999	♡
	-4	-1.4148	0.0415	0.0334	0.1170	-0.9998	♡
	0	-1.6717	0.2045	0.1622	0.3896	-0.9997	♡
	4	-2.0199	0.5346	0.4667	0.8017	-1.0000	♡
	8	-2.7971	0.8465	0.7773	0.9870	-1.0000	♡
25ms, 160Hz	-8	-1.4667	0.0056	0.0046	0.0144	-0.9871	♡
	-4	-1.3685	0.0339	0.0295	0.1080	-0.9969	♡
	0	-1.6071	0.1461	0.1180	0.3029	-0.9988	♡
	4	-1.8125	0.4712	0.4126	0.8195	-0.9999	♡
	8	-2.6110	0.8009	0.7341	0.9758	-0.9999	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.10: FORA $\mathcal{D}_2$ estimates for Observer 2 in condition $\mathcal{WT}$=1.

Expt. Level	SNR (dB)	m	k	$\mathcal{D}_1$	$\mathcal{D}_\infty$	Correlation	Status
400ms, 2.5Hz	0	−1.3919	0.0015	0.0016	0.0049	−0.9979	♡
	4	−1.0217	0.0095	0.0083	0.4405	−0.9995	♡
	8	−1.2739	0.0499	0.0496	0.2117	−0.9998	♡
	12	−1.5962	0.1687	0.1898	0.4084	−0.9999	♡
	16	−1.9532	0.2686	0.4098	0.5954	−1.0000	♡
200ms, 5Hz	0	−0.9359	0.0030	0.0034	0.0000	−0.9699	♠
	4	−1.3312	0.0295	0.0267	0.1040	−0.9981	♡
	8	−1.4225	0.0756	0.0869	0.2361	−0.9988	♡
	12	−1.7944	0.2451	0.2960	0.5142	−0.9996	♡
	16	−1.8841	0.2623	0.5245	0.7260	−1.0000	♡
100ms, 10Hz	0	−1.1703	0.0053	0.0042	0.0330	−0.9979	♡
	4	−1.4896	0.0530	0.0431	0.1307	−0.9999	♡
	8	−1.7577	0.1725	0.1656	0.3292	−0.9993	♡
	12	−1.9347	0.2531	0.3550	0.5349	−1.0000	♡
	16	−2.3240	0.3511	0.6040	0.7515	−0.9997	♡
50ms, 20Hz	0	−1.2425	0.0134	0.0110	0.0606	−0.9991	♡
	4	−1.3230	0.0421	0.0390	0.1524	−0.9994	♡
	8	−1.7715	0.1427	0.1734	0.3055	−0.9997	♡
	12	−2.0264	0.2746	0.3906	0.5613	−0.9996	♡
	16	−2.4319	0.3012	0.6496	0.7608	−0.9997	♡
25ms, 40Hz	0	−1.0540	0.0136	0.0138	0.2597	−0.9943	♡
	4	−1.3438	0.0610	0.0589	0.2120	−0.9994	♡
	8	−1.7923	0.2312	0.2422	0.4488	−0.9998	♡
	12	−2.0547	0.2992	0.4638	0.6424	−0.9998	♡
	16	−2.0828	0.2167	0.6557	0.7801	−0.9993	♡
12.5ms, 80Hz	0	−1.4921	0.0194	0.0133	0.0453	−0.9979	♡
	4	−1.3832	0.0716	0.0607	0.2191	−0.9998	♡
	8	−1.8697	0.1969	0.1790	0.3338	−0.9999	♡
	12	−2.0419	0.3255	0.4315	0.6293	−0.9999	♡
	16	−2.4091	0.3556	0.6684	0.8033	−0.9991	♡
6.25ms, 160Hz	0	−1.2349	0.0173	0.0165	0.0830	−0.9993	♡
	4	−1.4095	0.0602	0.0515	0.1747	−0.9979	♡
	8	−1.6634	0.1718	0.1530	0.3470	−0.9993	♡
	12	−1.9465	0.3838	0.3760	0.6438	−0.9999	♡
	16	−1.9926	0.3477	0.5625	0.7892	−0.9999	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.11: FORA $\mathcal{D}_2$ estimates for Observer 2 in condition $\mathcal{WT}$=2.

Expt. Level	SNR (dB)	m	k	$\mathcal{D}_1$	$\mathcal{D}_\infty$	Correlation	Status
400ms, 5Hz	−4	−1.7986	0.0064	0.0065	0.0122	−0.9995	♡
	0	−1.4409	0.0153	0.0146	0.0433	−0.9982	♡
	4	−1.5914	0.1062	0.1050	0.2440	−0.9993	♡
	8	−1.7803	0.3170	0.3270	0.6162	−0.9998	♡
	12	−2.4572	0.4704	0.6919	0.8607	−0.9997	♡
200ms, 10Hz	−4	−2.4673	0.0058	0.0032	0.0053	−0.9935	♡
	0	−1.5399	0.0310	0.0264	0.0718	−0.9993	♡
	4	−1.6465	0.1967	0.1889	0.4187	−0.9998	♡
	8	−2.1094	0.4325	0.4849	0.7241	−1.0000	♡
	12	−2.8770	0.5769	0.8233	0.9551	−0.9997	♡
100ms, 20Hz	−4	−1.5063	0.0075	0.0060	0.0179	−0.9931	♡
	0	−1.7608	0.0731	0.0597	0.1288	−1.0000	♡
	4	−1.8152	0.2012	0.1938	0.3668	−1.0000	♡
	8	−2.2007	0.4283	0.5741	0.7842	−0.9998	♡
	12	−3.0154	0.5046	0.8521	0.9525	−0.9997	♡
50ms, 40Hz	−4	−1.3552	0.0285	0.0272	0.0961	−0.9999	♡
	0	−1.4317	0.0960	0.0832	0.2679	−0.9993	♡
	4	−1.7513	0.2571	0.2614	0.5081	−1.0000	♡
	8	−2.3512	0.4428	0.6351	0.8151	−0.9994	♡
	12	−3.1030	0.4434	0.8989	0.9800	−0.9986	♡
25ms, 80Hz	−4	−1.4877	0.0201	0.0145	0.0478	−0.9953	♡
	0	−1.4699	0.0911	0.0753	0.2337	−0.9995	♡
	4	−1.8315	0.2815	0.2763	0.5117	−0.9998	♡
	8	−2.1998	0.4392	0.5551	0.7707	−1.0000	♡
	12	−2.7804	0.4512	0.8496	0.9634	−0.9998	♡
12.5ms, 160Hz	−4	−1.3606	0.0087	0.0070	0.0276	−0.9935	♡
	0	−1.5087	0.0562	0.0444	0.1331	−0.9996	♡
	4	−1.7212	0.2262	0.2030	0.4322	−0.9999	♡
	8	−1.9635	0.4309	0.4327	0.7260	−1.0000	♡
	12	−2.5403	0.5676	0.7457	0.9311	−0.9998	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.12: FORA $\mathcal{D}_2$ estimates for Observer 2 in condition $\mathcal{WT}$=4.

Expt. Level	SNR (dB)	m	k	$\mathcal{D}_1$	$\mathcal{D}_\infty$	Correlation	Status
400ms, 10Hz	-8	-1.4745	0.0023	0.0031	0.0070	-0.9981	♡
	-4	-1.2392	0.0115	0.0137	0.0571	-0.9987	♡
	0	-1.5104	0.0755	0.0708	0.1893	-0.9992	♡
	4	-1.7628	0.2989	0.3026	0.5836	-0.9995	♡
	8	-2.3129	0.5698	0.7006	0.9430	-0.9997	♡
200ms, 20Hz	-8	-1.6737	0.0046	0.0058	0.0109	-0.9983	♡
	-4	-1.4500	0.0238	0.0239	0.0675	-0.9964	♡
	0	-1.6895	0.1332	0.1256	0.2688	-0.9992	♡
	4	-2.0000	0.3966	0.4370	0.6927	-0.9999	♡
	8	-2.6080	0.6244	0.7803	0.9695	-0.9996	♡
100ms, 40Hz	-8	-1.5747	0.0201	0.0156	0.0429	-0.9988	♡
	-4	-1.4838	0.0465	0.0448	0.1230	-0.9997	♡
	0	-1.8155	0.1871	0.1947	0.3555	-0.9995	♡
	4	-2.1725	0.4463	0.5284	0.7554	-0.9999	♡
	8	-2.9427	0.5908	0.8612	0.9875	-1.0000	♡
50ms, 80Hz	-8	-1.5887	0.0141	0.0117	0.0302	-0.9986	♡
	-4	-1.5492	0.0356	0.0324	0.0836	-0.9996	♡
	0	-1.7364	0.2083	0.1883	0.3935	-0.9992	♡
	4	-2.0941	0.4693	0.5025	0.7675	-1.0000	♡
	8	-3.0138	0.6162	0.8383	0.9613	-0.9992	♡
25ms, 160Hz	-8	-0.9963	0.0036	0.0039	0.0000	-0.9756	♠
	-4	-1.4165	0.0225	0.0206	0.0657	-0.9984	♡
	0	-1.5757	0.1353	0.1157	0.2988	-0.9996	♡
	4	-1.8638	0.3980	0.3655	0.6814	-0.9995	♡
	8	-2.4649	0.6004	0.7132	0.9267	-1.0000	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.13: FORA d' estimates for Observer 1 in condition $\mathcal{WT}$=1.

Expt. Level	SNR (dB)	m	k	d'_1	d'_∞	Correlation	Status
400ms, 2.5Hz	0	−1.6186	0.0488	0.0429	0.1034	−0.9920	♡
	4	−1.4281	0.0930	0.1285	0.3089	−0.9973	♡
	8	−1.4394	0.2857	0.4180	0.9560	−0.9985	♡
	12	−1.4935	0.4706	0.7584	1.5290	−0.9994	♡
	16	−1.6412	0.7753	1.1581	2.0725	−0.9972	♡
200ms, 5Hz	0	−1.6617	0.0852	0.1159	0.2125	−0.9973	♡
	4	−1.9068	0.2566	0.2855	0.4758	−0.9997	♡
	8	−1.8086	0.3537	0.5631	0.8705	−0.9986	♡
	12	−1.8681	0.7199	1.1188	1.6864	−1.0000	♡
	16	−2.0825	1.0910	1.5241	2.1504	−1.0000	♡
100ms, 10Hz	0	−1.6616	0.0834	0.1315	0.2261	−0.9977	♡
	4	−1.9038	0.3226	0.4810	0.7214	−0.9999	♡
	8	−1.8595	0.5122	0.8904	1.3000	−0.9999	♡
	12	−1.9701	0.7899	1.3694	1.9020	−0.9997	♡
	16	−2.4374	1.3653	1.9015	2.4026	−0.9998	♡
50ms, 20Hz	0	−1.6089	0.1963	0.2326	0.4801	−0.9992	♡
	4	−1.6560	0.2416	0.4075	0.6849	−0.9999	♡
	8	−2.0684	0.5720	0.9241	1.2590	−0.9995	♡
	12	−2.0489	0.7363	1.4374	1.8805	−0.9998	♡
	16	−2.3114	1.1931	2.1168	2.6256	−0.9998	♡
25ms, 40Hz	0	−1.3070	0.1137	0.1853	0.5097	−0.9985	♡
	4	−1.5023	0.2586	0.4345	0.8491	−0.9998	♡
	8	−1.7661	0.5982	0.9910	1.5502	−0.9996	♡
	12	−1.9718	0.8152	1.4874	2.0354	−0.9996	♡
	16	−2.3089	1.0048	2.0342	2.4641	−0.9992	♡
12.5ms, 80Hz	0	−1.5871	0.1155	0.1872	0.3399	−1.0000	♡
	4	−1.6118	0.2468	0.4049	0.7145	−0.9999	♡
	8	−1.7735	0.4701	0.8176	1.2520	−0.9996	♡
	12	−1.9177	0.6787	1.3979	1.8932	−1.0000	♡
	16	−2.0925	0.8246	2.0371	2.5042	−0.9991	♡
6.25ms, 160Hz	0	−1.6592	0.1739	0.2340	0.4321	−0.9976	♡
	4	−1.7796	0.2920	0.3941	0.6608	−0.9998	♡
	8	−1.8969	0.4750	0.7682	1.1261	−0.9997	♡
	12	−2.1475	0.8408	1.3488	1.7907	−0.9999	♡
	16	−2.2496	0.9412	1.8771	2.3105	−0.9998	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.14: FORA d' estimates for Observer 1 in condition $\mathcal{WT}$=2.

Expt. Level	SNR (dB)	m	k	d'_1	d'_∞	Correlation	Status
400ms, 5Hz	-4	-1.1235	0.0718	0.1227	0.6744	-0.9954	♡
	0	-1.3249	0.1314	0.1910	0.5426	-0.9991	♡
	4	-1.2897	0.3179	0.4886	1.4575	-0.9990	♡
	8	-1.2691	0.5872	0.8888	2.8329	-0.9992	♡
	12	-1.2444	0.9420	1.3809	4.8519	-0.9991	♡
200ms, 10Hz	-4	-3.2399	0.1162	0.0644	0.0831	-0.9579	♡
	0	-1.5734	0.1766	0.2946	0.5350	-1.0000	♡
	4	-1.6676	0.4794	0.7579	1.2954	-1.0000	♡
	8	-1.5348	0.8555	1.2891	2.5585	-0.9995	♡
	12	-1.4378	1.2735	1.7779	4.1862	-0.9998	♡
100ms, 20Hz	-4	-1.6975	0.1340	0.2179	0.3598	-0.9995	♡
	0	-1.7476	0.3287	0.5165	0.8339	-0.9995	♡
	4	-1.8662	0.6274	0.9417	1.4379	-1.0000	♡
	8	-1.9402	1.1656	1.6601	2.4811	-0.9999	♡
	12	-1.7836	1.7661	2.2643	3.8656	-0.9996	♡
50ms, 40Hz	-4	-1.6042	0.2022	0.2812	0.5387	-0.9990	♡
	0	-1.5912	0.3460	0.5018	0.9547	-0.9998	♡
	4	-1.7829	0.6822	1.0253	1.6448	-0.9998	♡
	8	-1.9752	1.4474	1.8785	2.8464	-0.9998	♡
	12	-2.0185	2.3965	2.6036	4.1074	-0.9994	♡
25ms, 80Hz	-4	-1.4717	0.1129	0.1666	0.3620	-0.9990	♡
	0	-1.6212	0.2946	0.4473	0.8097	-0.9999	♡
	4	-1.7196	0.5941	0.9056	1.5095	-1.0000	♡
	8	-1.8325	1.0729	1.5661	2.4624	-0.9999	♡
	12	-1.8192	1.7451	2.3212	3.8114	-0.9999	♡
12.5ms, 160Hz	-4	-1.5499	0.1301	0.1435	0.3300	-0.9990	♡
	0	-1.5803	0.2637	0.3771	0.7305	-1.0000	♡
	4	-1.7160	0.6006	0.8899	1.5042	-0.9996	♡
	8	-1.8620	1.0193	1.5221	2.3338	-1.0000	♡
	12	-1.7802	1.4949	2.3401	3.7049	-0.9999	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.15: FORA d' estimates for Observer 1 in condition $\mathcal{WT}$=4.

Expt. Level	SNR (dB)	m	k	d'_1	d'_∞	Correlation	Status
400ms, 10Hz	-8	-0.3431	0.0058	0.0239	0.0000	-0.4417	♠
	-4	-1.2527	0.1207	0.1684	0.5971	-0.9974	♡
	0	-1.1627	0.2407	0.3447	1.7252	-0.9981	♡
	4	-1.2774	0.5304	0.7240	2.4217	-0.9994	♡
	8	-1.1920	0.9247	1.2606	5.6978	-0.9997	♡
200ms, 20Hz	-8	-1.7699	0.1410	0.1504	0.2815	-0.9958	♡
	-4	-1.6464	0.1979	0.2548	0.4861	-1.0000	♡
	0	-1.6942	0.5126	0.6557	1.2018	-0.9999	♡
	4	-1.7107	0.9596	1.2608	2.2519	-0.9997	♡
	8	-1.4421	1.4496	1.9842	4.6940	-0.9997	♡
100ms, 40Hz	-8	-1.5855	0.1250	0.1860	0.3516	-0.9963	♡
	-4	-1.7758	0.2388	0.3239	0.5433	-0.9998	♡
	0	-1.8472	0.5382	0.8148	1.2538	-1.0000	♡
	4	-1.8533	1.0651	1.5334	2.3934	-1.0000	♡
	8	-1.4789	1.5239	2.3416	4.9303	-0.9999	♡
50ms, 80Hz	-8	-1.7448	0.1295	0.1851	0.3109	-0.9999	♡
	-4	-1.6710	0.2674	0.3818	0.6796	-0.9999	♡
	0	-1.7950	0.6207	0.8764	1.4286	-0.9999	♡
	4	-1.7981	1.1727	1.6523	2.6903	-0.9999	♡
	8	-1.6313	1.8598	2.5571	4.7975	-0.9995	♡
25ms, 160Hz	-8	-1.6923	0.0918	0.1410	0.2389	-0.9932	♡
	-4	-1.6101	0.2267	0.3604	0.6456	-0.9987	♡
	0	-1.7632	0.5043	0.7389	1.2126	-0.9995	♡
	4	-1.6332	1.0306	1.5203	2.7573	-1.0000	♡
	8	-1.6662	1.7477	2.3967	4.3616	-0.9999	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.16: FORA d' estimates for Observer 2 in condition $\mathcal{WT}$=1.

Expt. Level	SNR (dB)	m	k	d'_1	d'_∞	Correlation	Status
400ms, 2.5Hz	0	−1.5609	0.0559	0.0723	0.1505	−0.9996	♡
	4	−1.3262	0.1246	0.1835	0.5161	−0.9992	♡
	8	−1.4964	0.2599	0.4699	0.8929	−0.9997	♡
	12	−1.6873	0.4836	0.9560	1.4781	−0.9999	♡
	16	−1.9107	0.6373	1.5135	1.9834	−0.9999	♡
200ms, 5Hz	0	−1.3284	0.0716	0.1162	0.3054	−0.9928	♡
	4	−1.5678	0.2068	0.3420	0.6269	−0.9993	♡
	8	−1.5842	0.3010	0.6281	1.0282	−0.9995	♡
	12	−1.8114	0.6095	1.2346	1.7619	−0.9997	♡
	16	−1.7688	0.6322	1.7961	2.3842	−1.0000	♡
100ms, 10Hz	0	−1.4939	0.0992	0.1315	0.2939	−0.9978	♡
	4	−1.7150	0.2932	0.4375	0.7383	−0.9999	♡
	8	−1.8602	0.5176	0.8870	1.3000	−0.9997	♡
	12	−1.9217	0.6089	1.3802	1.8216	−0.9999	♡
	16	−2.1637	0.8961	2.0042	2.4653	−0.9995	♡
50ms, 20Hz	0	−1.5460	0.1524	0.2161	0.4366	−0.9995	♡
	4	−1.5515	0.2470	0.4134	0.7666	−0.9993	♡
	8	−1.8602	0.4242	0.9088	1.2474	−0.9998	♡
	12	−1.9957	0.6542	1.4669	1.8917	−0.9996	♡
	16	−2.2860	0.8233	2.1318	2.4940	−0.9999	♡
25ms, 40Hz	0	−1.3272	0.1382	0.2394	0.6065	−0.9985	♡
	4	−1.5702	0.2984	0.5123	0.9215	−0.9989	♡
	8	−1.8426	0.6063	1.0976	1.5958	−0.9999	♡
	12	−1.9789	0.7095	1.6450	2.1172	−0.9999	♡
	16	−1.9412	0.5996	2.1494	2.5716	−0.9995	♡
12.5ms, 80Hz	0	−1.7247	0.1899	0.2372	0.4285	−0.9997	♡
	4	−1.6000	0.3376	0.5203	0.9544	−1.0000	♡
	8	−1.9628	0.5732	0.9262	1.3166	−0.9999	♡
	12	−1.9720	0.7642	1.5661	2.0799	−1.0000	♡
	16	−2.2048	0.9793	2.1867	2.6644	−0.9995	♡
6.25ms, 160Hz	0	−1.4678	0.1559	0.2636	0.5361	−0.9995	♡
	4	−1.6372	0.3095	0.4783	0.8469	−0.9994	♡
	8	−1.7839	0.5326	0.8490	1.3316	−0.9997	♡
	12	−1.8833	0.8944	1.4314	2.1196	−0.9999	♡
	16	−1.8010	0.8324	1.8947	2.6279	−0.9999	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.17: FORA d' estimates for Observer 2 in condition $\mathcal{WT}$=2.

Expt. Level	SNR (dB)	m	k	d'_1	d'_∞	Correlation	Status
400ms, 5Hz	-4	-2.0021	0.1268	0.1508	0.2323	-0.9989	♡
	0	-1.7040	0.1584	0.2466	0.4123	-0.9994	♡
	4	-1.7508	0.3963	0.6910	1.0714	-0.9998	♡
	8	-1.7595	0.7599	1.3115	2.0306	-0.9998	♡
	12	-2.1276	1.2781	2.2588	2.9488	-0.9999	♡
200ms, 10Hz	-4	-2.3089	0.1032	0.1106	0.1547	-0.9990	♡
	0	-1.7564	0.2207	0.3375	0.5475	-0.9999	♡
	4	-1.7472	0.5713	0.9504	1.5028	-1.0000	♡
	8	-1.9635	1.0003	1.6974	2.3779	-0.9999	♡
	12	-2.2083	2.0108	2.7430	3.7183	-0.9998	♡
100ms, 20Hz	-4	-1.7369	0.1138	0.1578	0.2699	-0.9965	♡
	0	-1.9332	0.3438	0.5160	0.7610	-0.9999	♡
	4	-1.9017	0.5712	0.9674	1.3947	-1.0000	♡
	8	-1.9882	1.0310	1.9243	2.6010	-0.9999	♡
	12	-2.3819	2.0133	2.8748	3.6628	-0.9999	♡
50ms, 40Hz	-4	-1.5924	0.1995	0.3455	0.6063	-0.9996	♡
	0	-1.6271	0.3910	0.6139	1.0891	-0.9999	♡
	4	-1.7891	0.6584	1.1469	1.7387	-1.0000	♡
	8	-2.1020	1.1312	2.0920	2.7243	-0.9997	♡
	12	-2.1276	1.9802	3.1463	4.2158	-0.9998	♡
25ms, 80Hz	-4	-1.7657	0.1985	0.2481	0.4337	-0.9978	♡
	0	-1.6729	0.3907	0.5814	1.0150	-0.9999	♡
	4	-1.8604	0.7116	1.1846	1.7525	-0.9999	♡
	8	-2.0022	1.0430	1.8759	2.5466	-0.9999	♡
	12	-2.0288	1.6311	2.8669	3.8761	-0.9999	♡
12.5ms, 160Hz	-4	-1.7549	0.1466	0.1666	0.3065	-0.9984	♡
	0	-1.7447	0.3172	0.4396	0.7477	-0.9997	♡
	4	-1.8053	0.6298	0.9926	1.5433	-1.0000	♡
	8	-1.8357	0.9814	1.5693	2.3851	-0.9999	♡
	12	-1.9543	1.5302	2.4384	3.4944	-1.0000	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.18: FORA d' estimates for Observer 2 in condition $\mathcal{WT}$=4.

Expt. Level	SNR (dB)	m	k	d'_1	d'_∞	Correlation	Status
400ms, 10Hz	-8	-1.6085	0.0686	0.0980	0.1846	-0.9988	♡
	-4	-1.4736	0.1207	0.2381	0.4458	-0.9997	♡
	0	-1.6994	0.3348	0.5621	0.9154	-0.9998	♡
	4	-1.7657	0.7331	1.2508	1.9366	-0.9997	♡
	8	-1.7099	1.3624	2.2896	3.6985	-0.9988	♡
200ms, 20Hz	-8	-1.7676	0.0837	0.1460	0.2241	-0.9986	♡
	-4	-1.6609	0.1813	0.3202	0.5258	-0.9982	♡
	0	-1.8244	0.4509	0.7628	1.1445	-0.9997	♡
	4	-1.8921	0.9142	1.5795	2.2732	-0.9999	♡
	8	-1.7322	1.5980	2.5626	4.1504	-0.9998	♡
100ms, 40Hz	-8	-1.7963	0.1825	0.2598	0.4217	-0.9996	♡
	-4	-1.6917	0.2586	0.4434	0.7204	-1.0000	♡
	0	-1.8977	0.5300	0.9702	1.3689	-0.9998	♡
	4	-1.9893	1.0438	1.8064	2.4901	-1.0000	♡
	8	-1.7491	1.8296	2.9255	4.6882	-0.9994	♡
50ms, 80Hz	-8	-1.8365	0.1622	0.2188	0.3534	-0.9993	♡
	-4	-1.7602	0.2310	0.3754	0.5937	-0.9999	♡
	0	-1.8285	0.5955	0.9517	1.4523	-0.9996	♡
	4	-1.9027	1.0722	1.7408	2.5412	-1.0000	♡
	8	-2.2050	2.0956	2.8047	3.8263	-1.0000	♡
25ms, 160Hz	-8	-1.2925	0.0682	0.1281	0.3337	-0.9877	♡
	-4	-1.6273	0.1764	0.2992	0.5134	-0.9994	♡
	0	-1.7322	0.4714	0.7314	1.1996	-0.9999	♡
	4	-1.7923	0.9221	1.4059	2.2298	-0.9996	♡
	8	-1.9128	1.5424	2.3270	3.4604	-0.9996	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

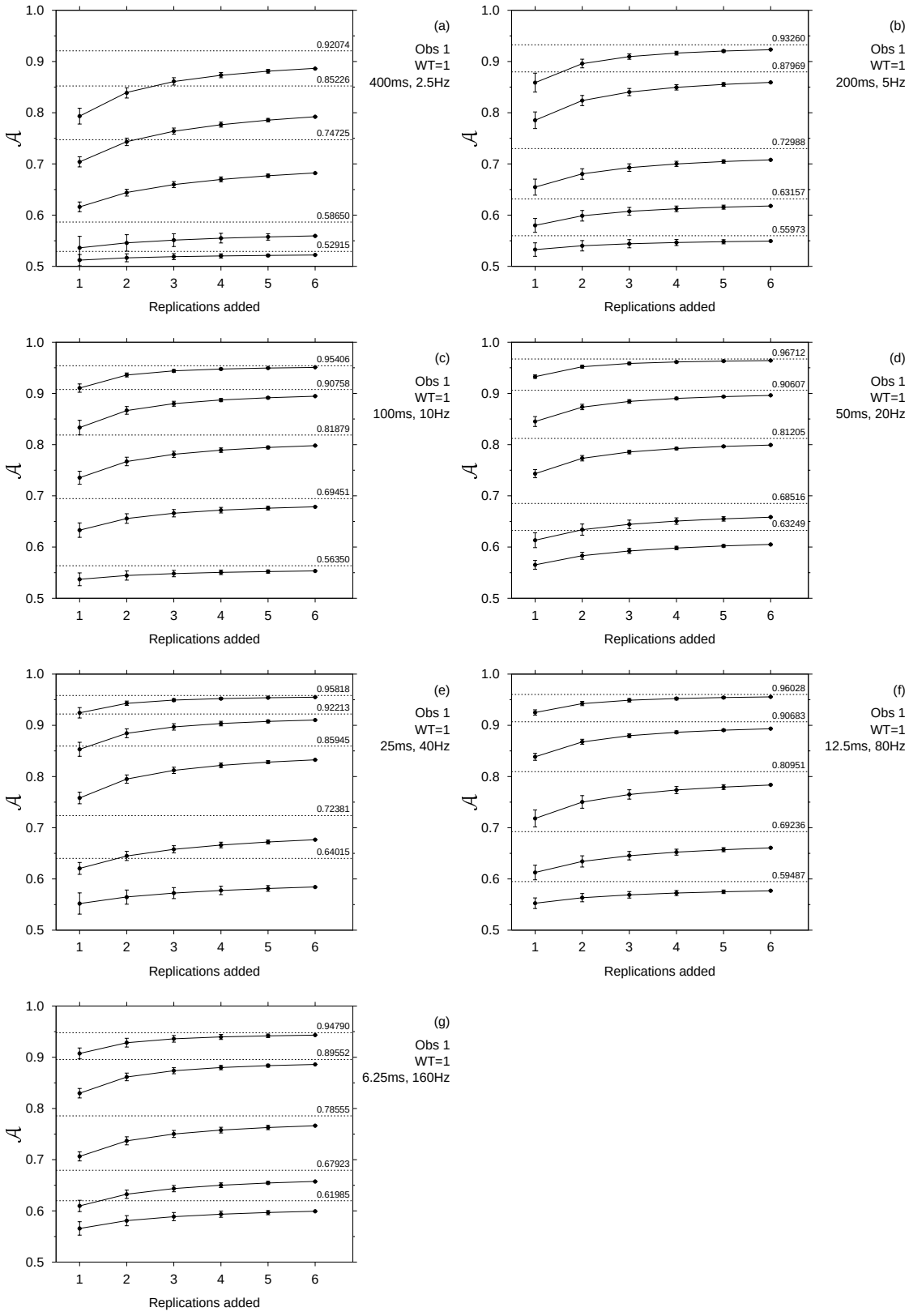

Figure E.1: FORA for each signal–to–noise ratio and level, analysed over Observer 1, in condition $\mathcal{WT}$=1. Dashed lines are the asymptotic $\mathcal{A}$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation.

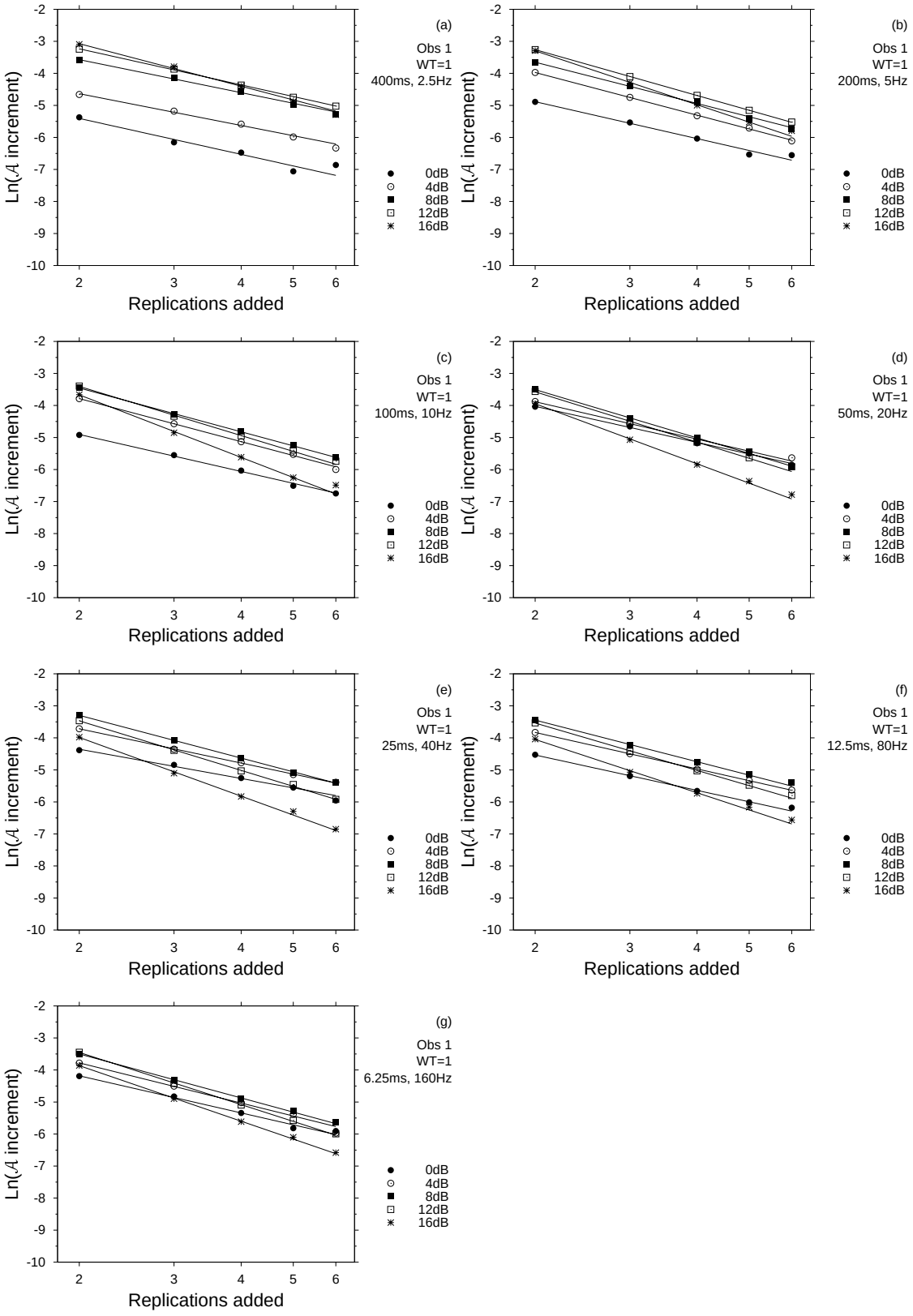

Figure E.2: FORA for each signal–to–noise ratio and level, in log–increment coordinates, analysed over Observer 1, in condition $\mathcal{WT}=1$. Solid lines are the fitted FORA.

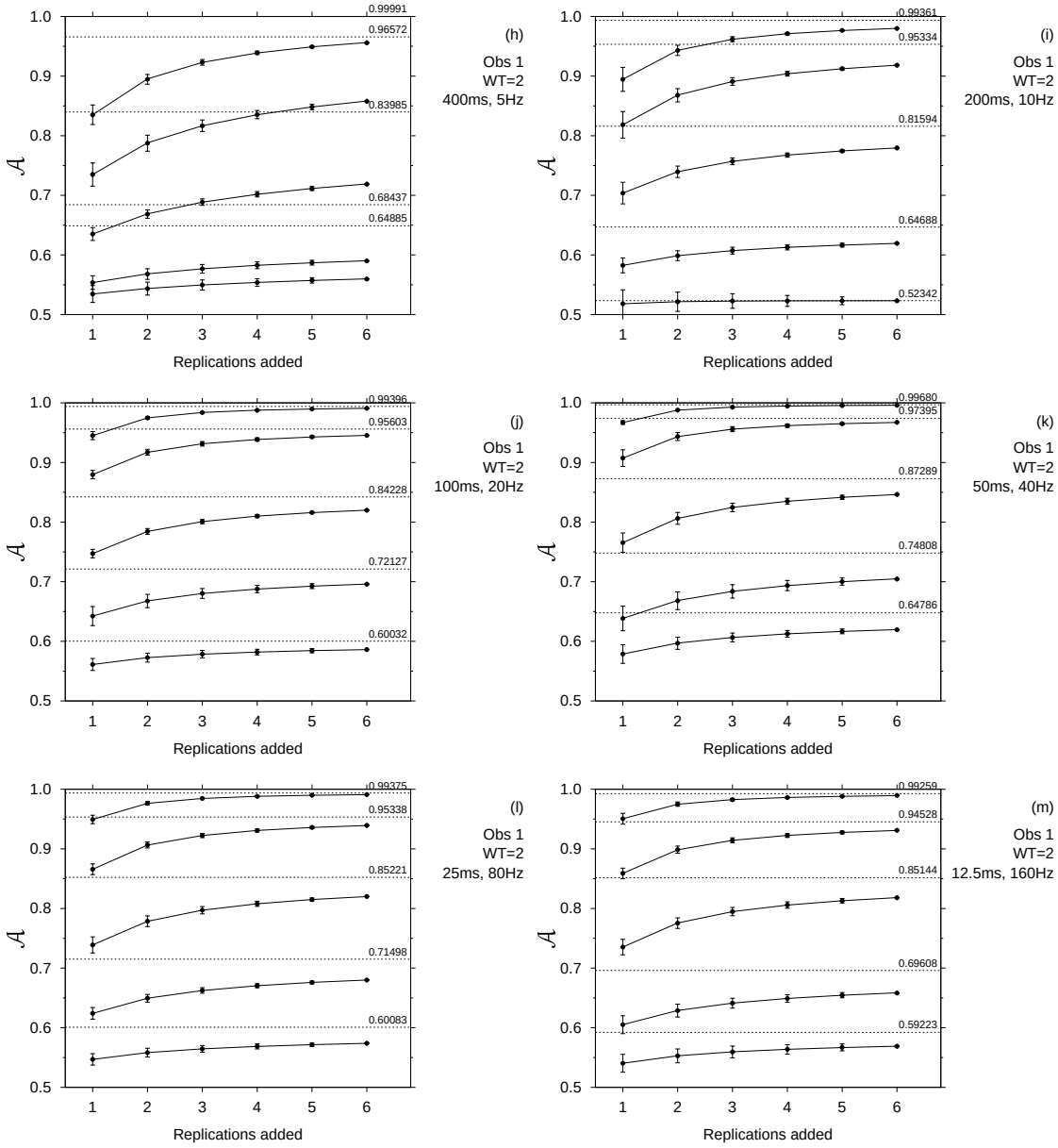

Figure E.3: FORA for each signal–to–noise ratio and level, analysed over Observer 1, in condition $\mathcal{WT}$=2. Dashed lines are the asymptotic $\mathcal{A}$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation.

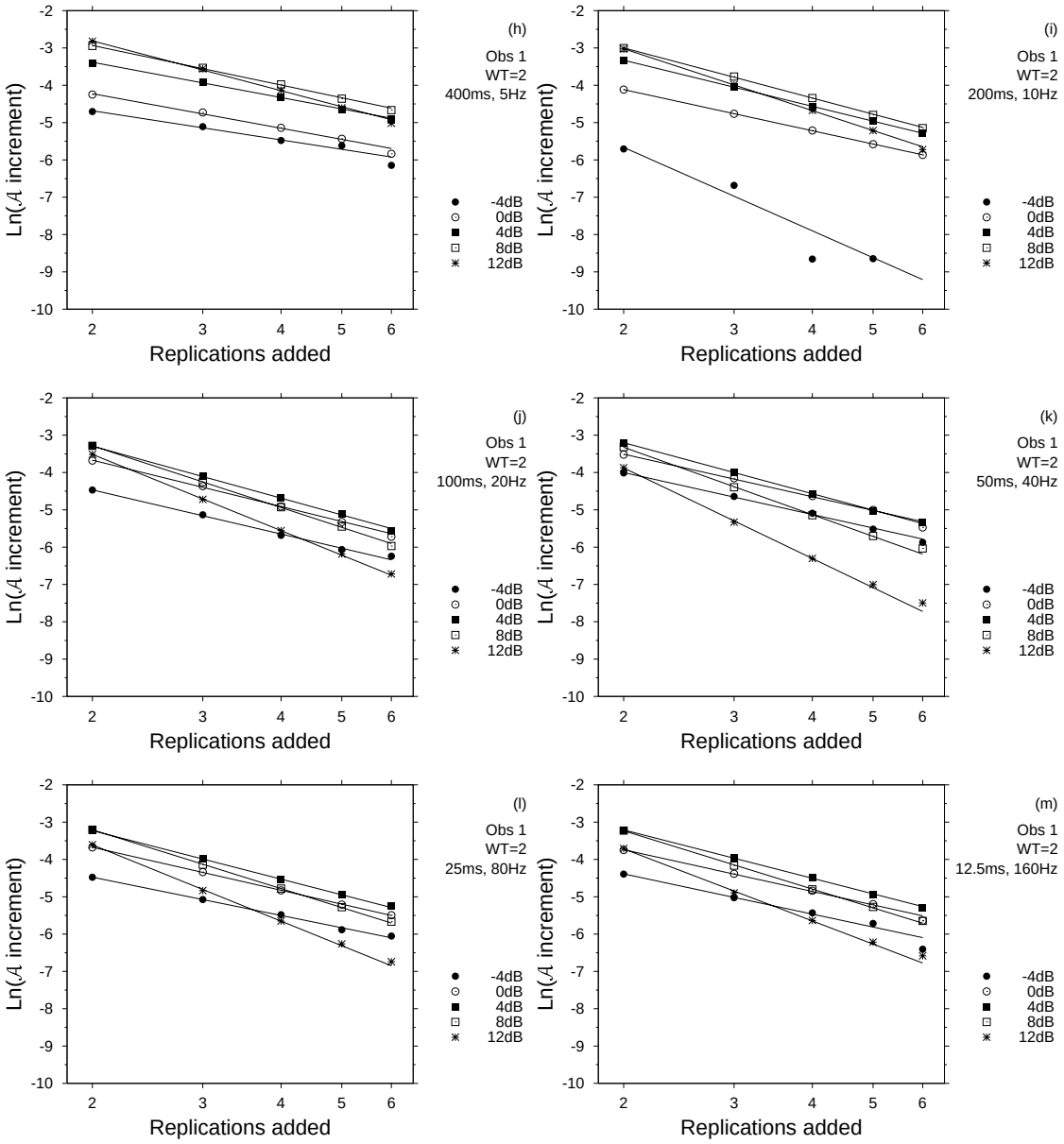

Figure E.4: FORA for each signal–to–noise ratio and level, in log–increment coordinates, analysed over Observer 1, in condition $\mathcal{WT}$=2. Solid lines are the fitted FORA.

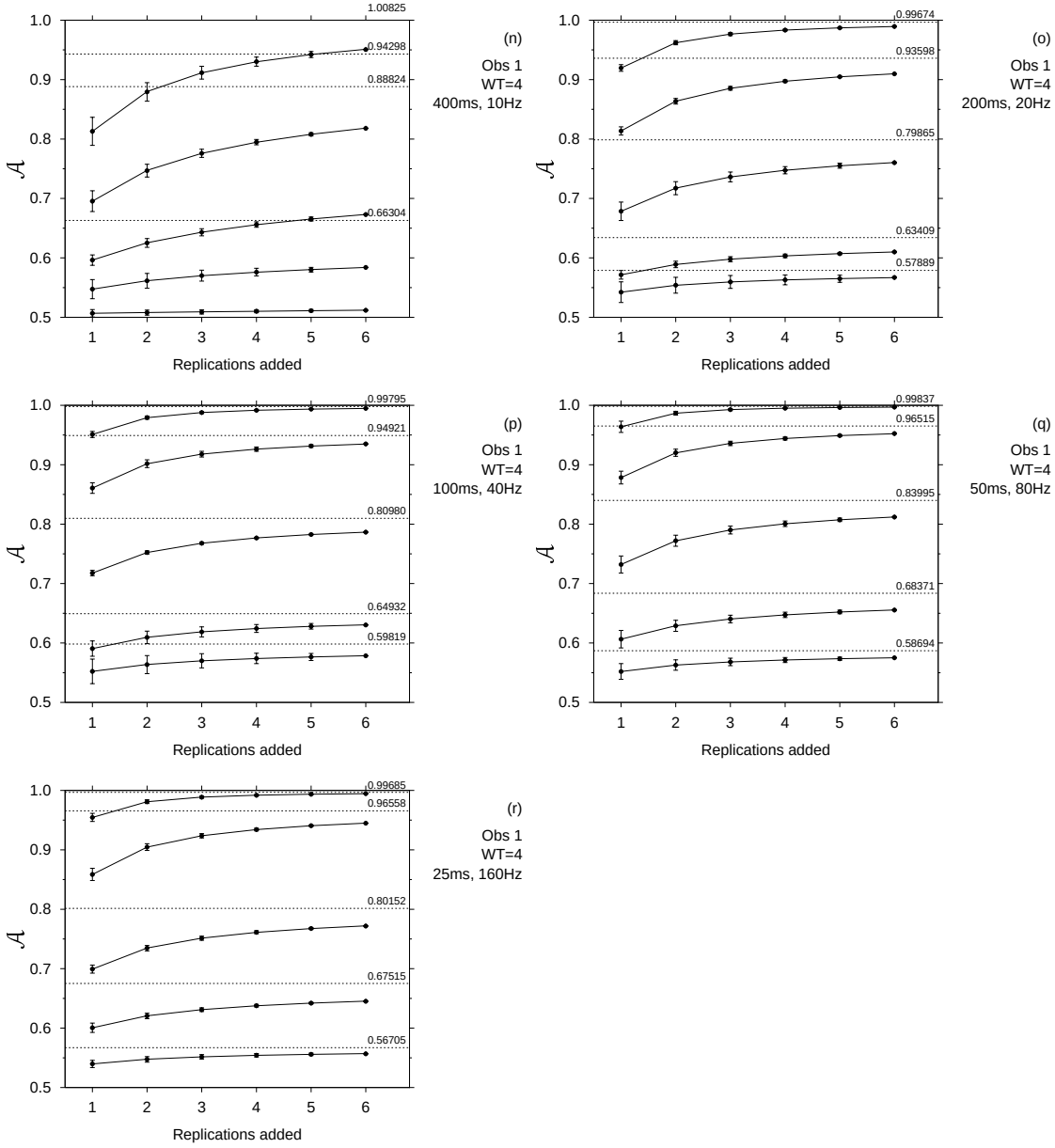

Figure E.5: FORA for each signal–to–noise ratio and level, analysed over Observer 1, in condition $\mathcal{WT}$=4. Dashed lines are the asymptotic $\mathcal{A}$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation.

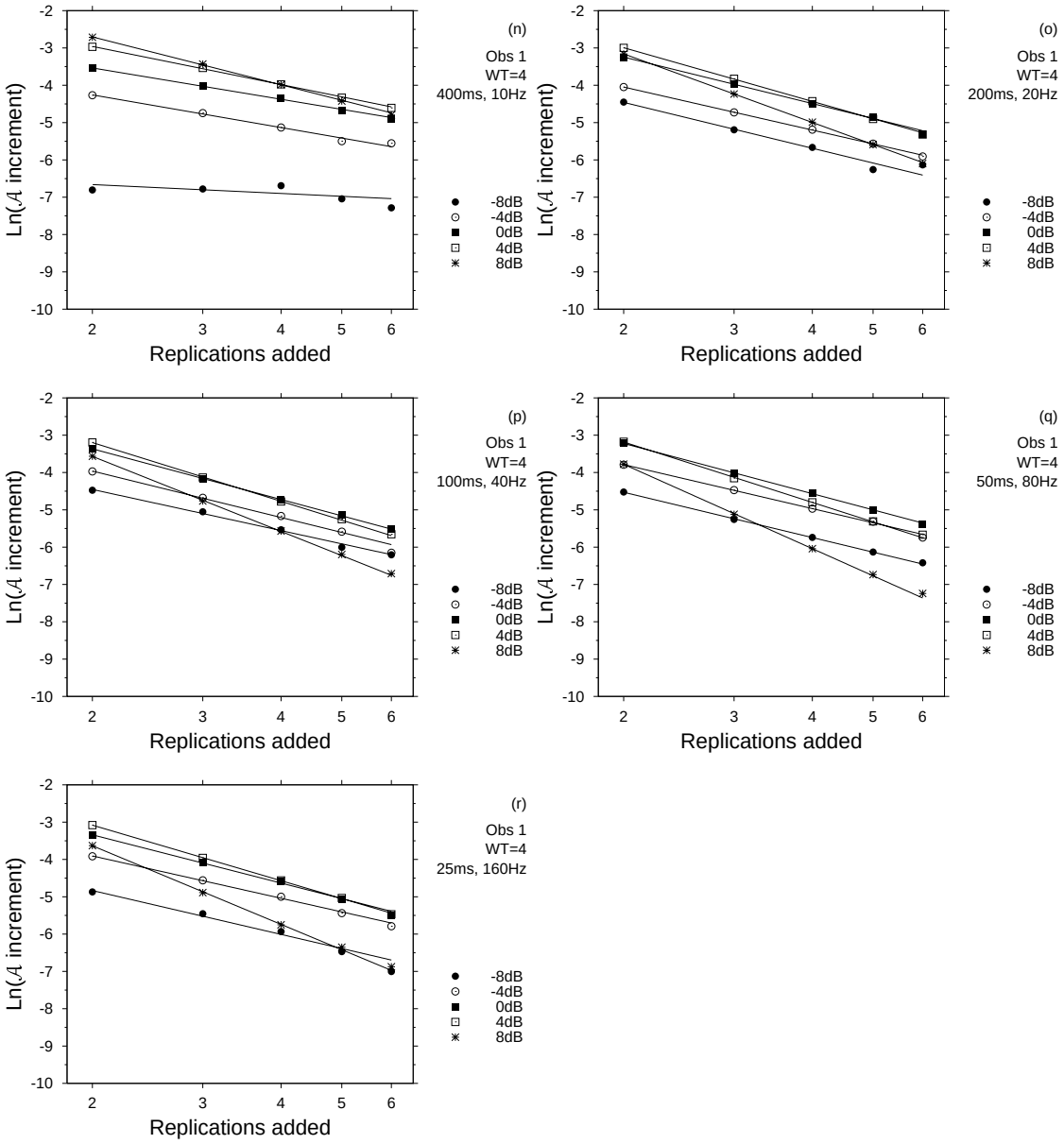

Figure E.6: FORA for each signal–to–noise ratio and level, in log–increment coordinates, analysed over Observer 1, in condition $\mathcal{WT}$=4. Solid lines are the fitted FORA.

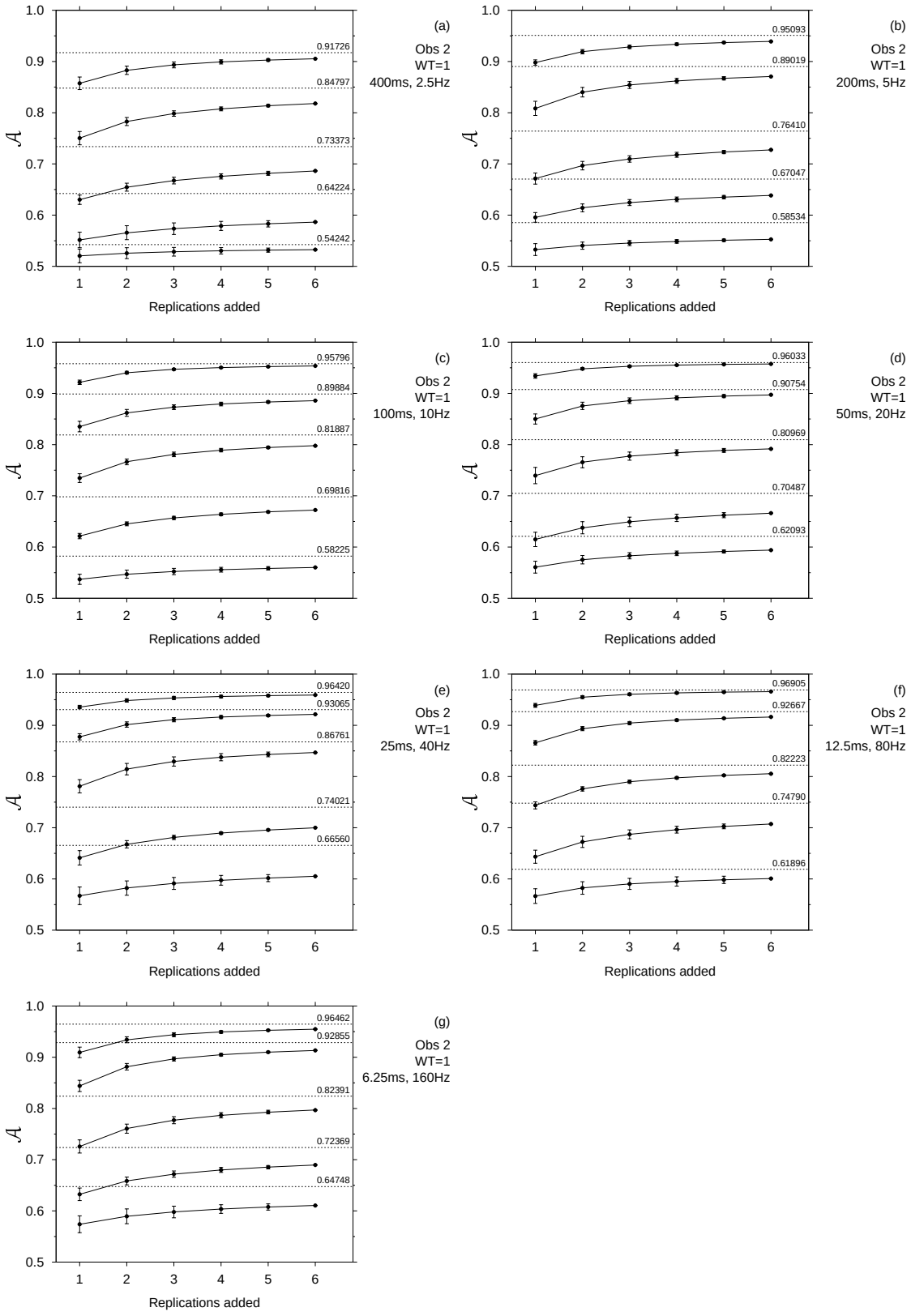

Figure E.7: FORA for each signal–to–noise ratio and level, analysed over Observer 2, in condition $\mathcal{WT}$=1. Dashed lines are the asymptotic $\mathcal{A}$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation.

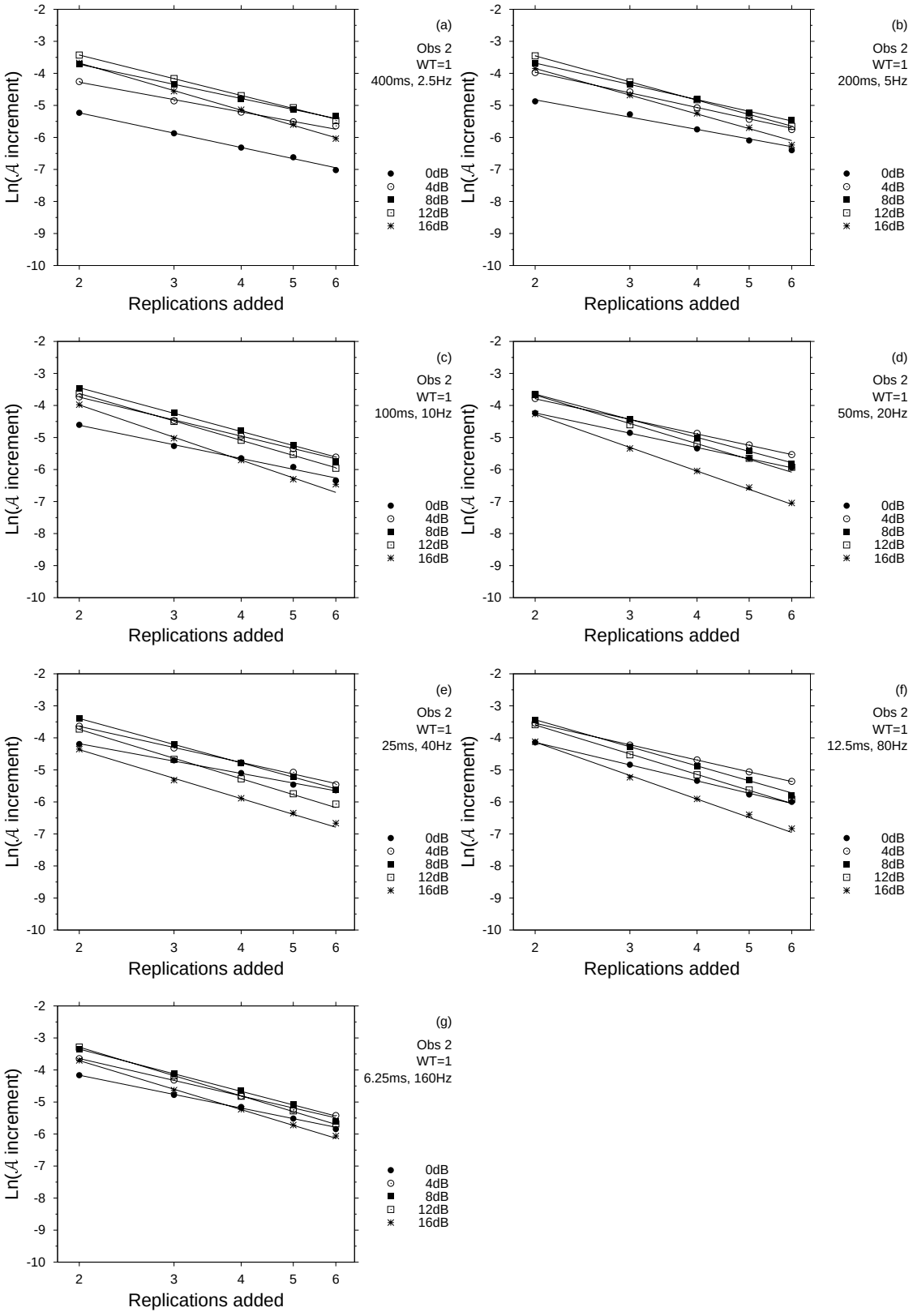

Figure E.8: FORA for each signal–to–noise ratio and level, in log–increment coordinates, analysed over Observer 2, in condition $\mathcal{WT}=1$. Solid lines are the fitted FORA.

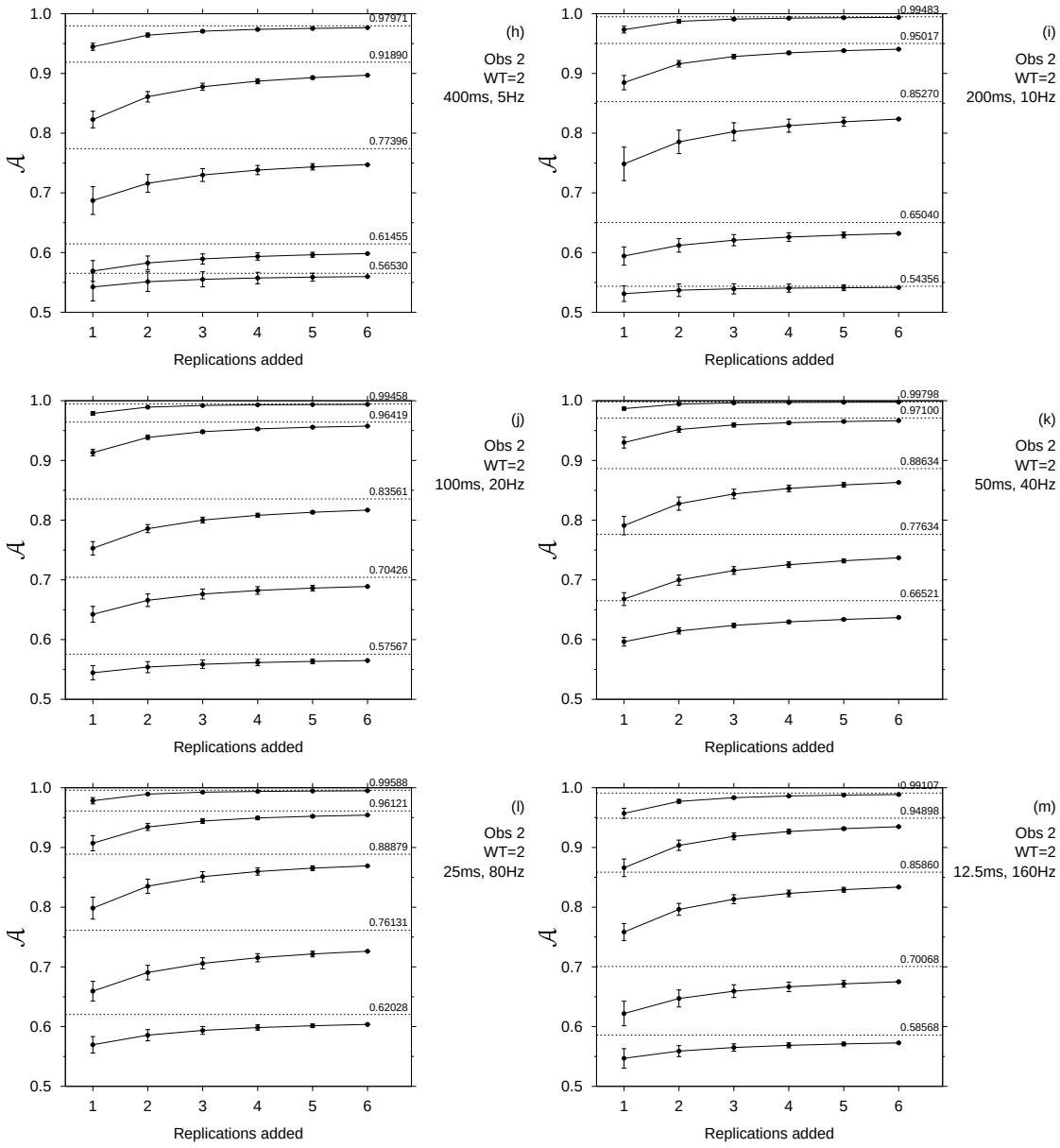

Figure E.9: FORA for each signal–to–noise ratio and level, analysed over Observer 2, in condition $\mathcal{WT}$=2. Dashed lines are the asymptotic $\mathcal{A}$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation. Error bars are ±1 standard deviation.

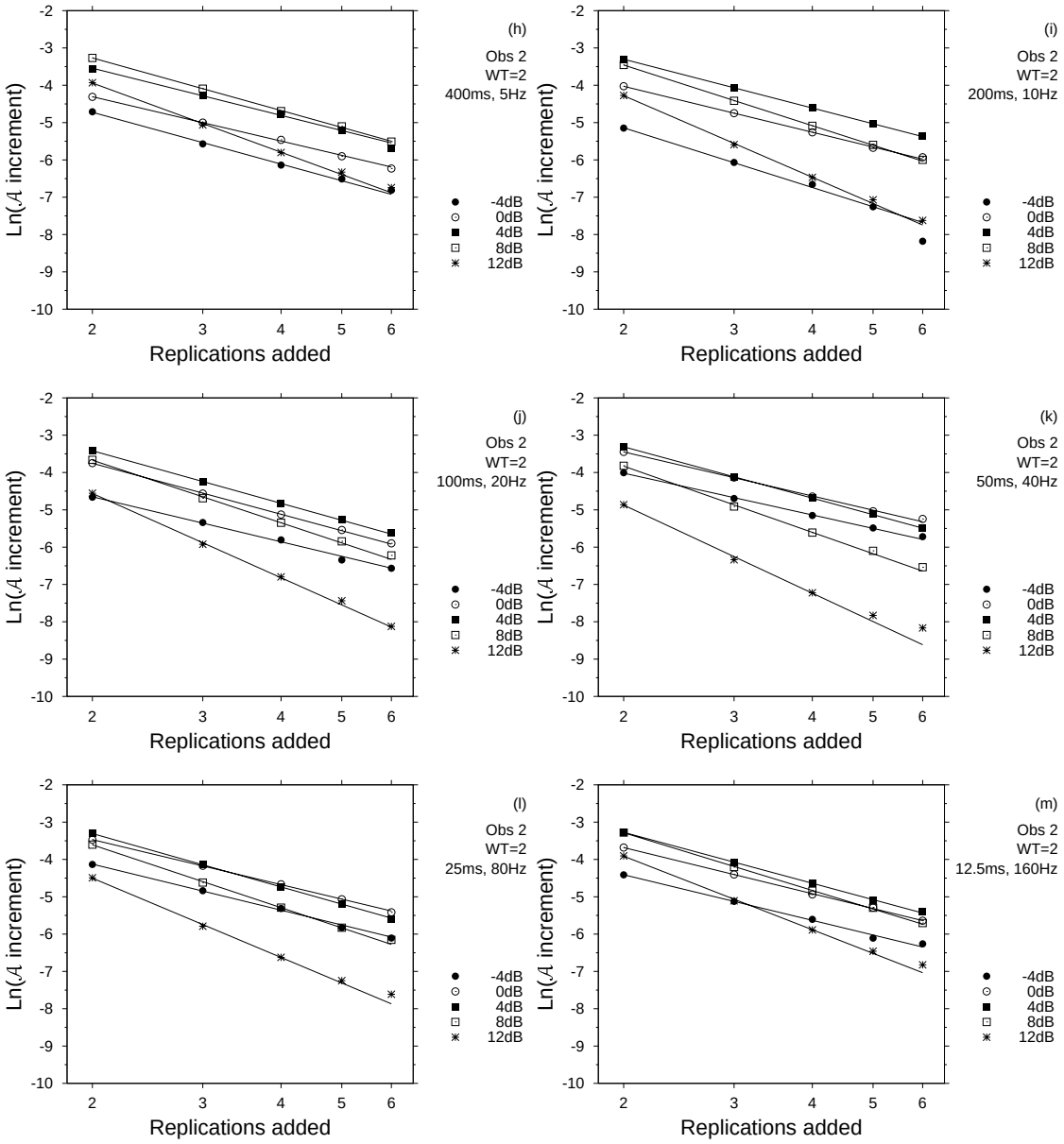

Figure E.10: FORA for each signal–to–noise ratio and level, in log–increment coordinates, analysed over Observer 2, in condition $\mathcal{WT}$=2. Solid lines are the fitted FORA.

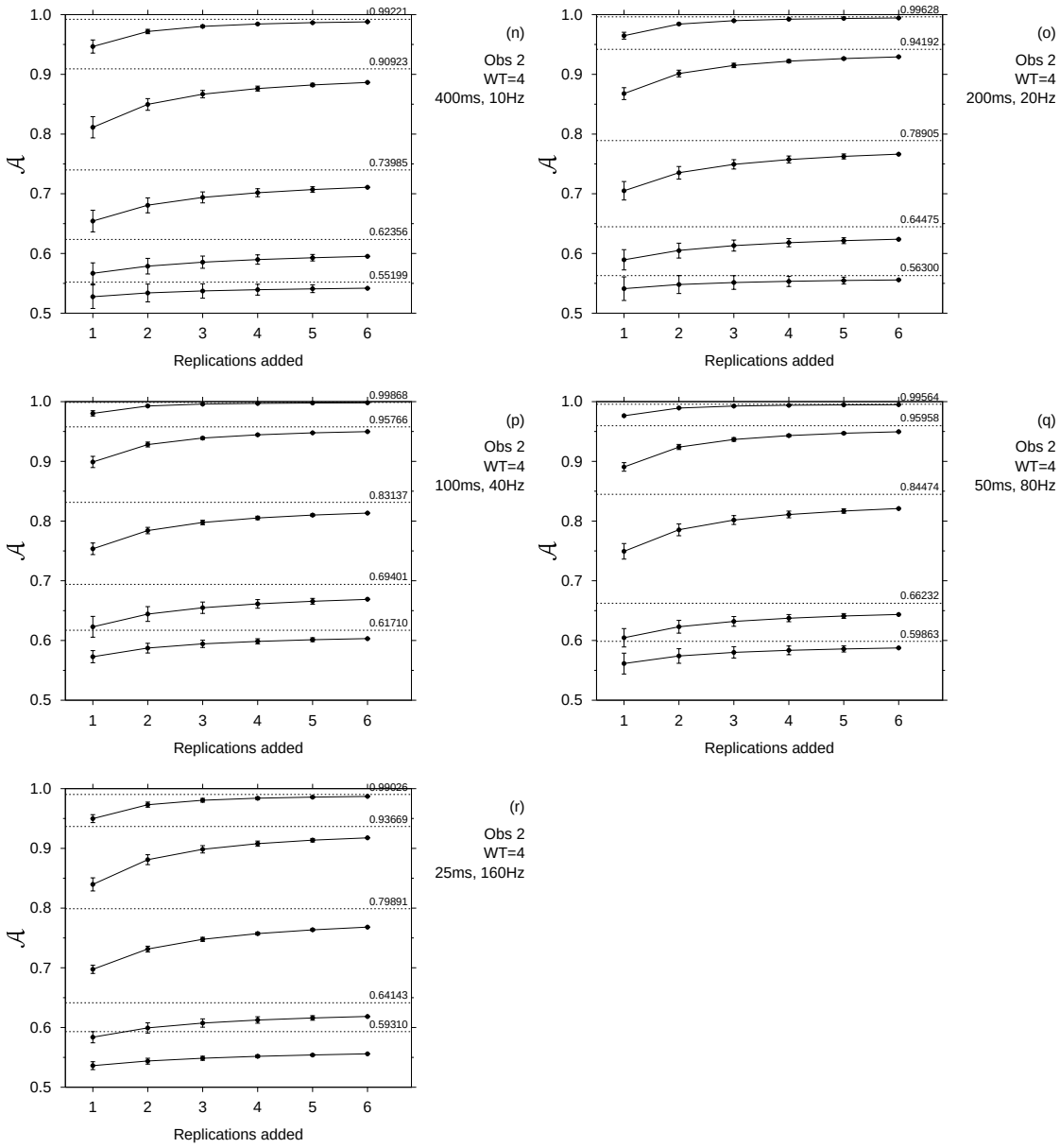

Figure E.11: FORA for each signal–to–noise ratio and level, analysed over Observer 2, in condition $\mathcal{WT}$=4. Dashed lines are the asymptotic $\mathcal{A}$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation.

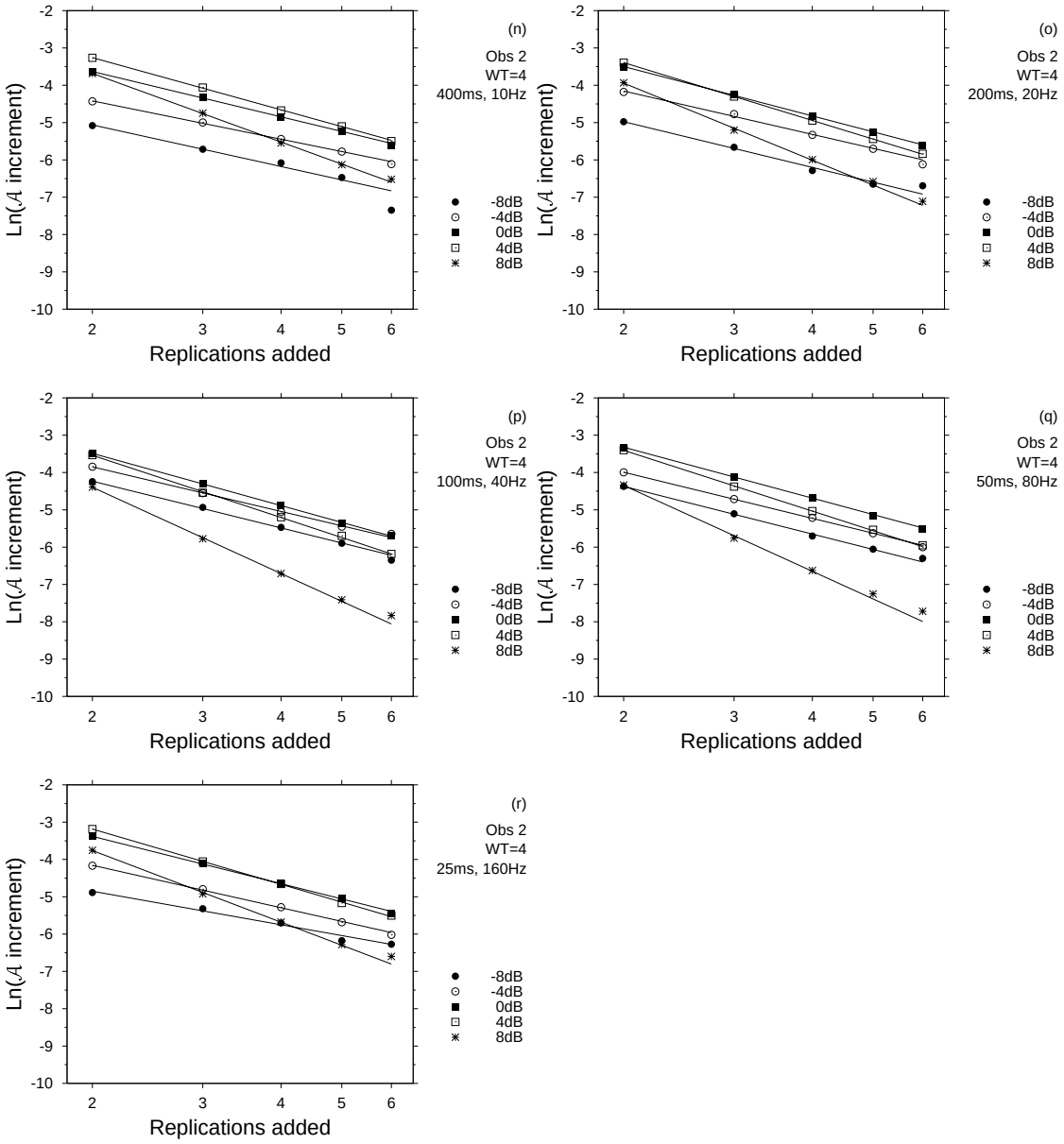

Figure E.12: FORA for each signal–to–noise ratio and level, in log–increment coordinates, analysed over Observer 2, in condition $\mathcal{WT}$=4. Solid lines are the fitted FORA.

Table E.19: FORA $\mathcal{D}_6$ estimates for Observer 1.

Expt. Level	m	k	$\mathcal{D}_{6,1}$	$\mathcal{D}_{6,\infty}$	Correlation	Status
400ms, 2.5Hz	−1.2940	0.5932	0.6488	2.4279	−0.9966	♡
200ms, 5Hz	−1.7191	0.9913	0.9742	1.9831	−0.9997	♡
100ms, 10Hz	−1.9125	1.3200	1.3300	2.3007	−1.0000	♡
50ms, 20Hz	−2.0188	1.3860	1.5519	2.4218	−0.9999	♡
25ms, 40Hz	−1.9386	1.3534	1.5772	2.5333	−0.9999	♡
12.5ms, 80Hz	−2.0253	1.2318	1.5230	2.2890	−0.9999	♡
6.25ms, 160Hz	−1.9933	1.2731	1.3775	2.2066	−1.0000	♡
400ms, 5Hz	−1.0810	0.8690	0.7924	11.1559	−0.9969	♣
200ms, 10Hz	−1.4370	1.3052	1.2509	3.7261	−0.9996	♡
100ms, 20Hz	−1.7211	1.7562	1.7193	3.4999	−1.0000	♡
50ms, 40Hz	−1.8254	2.0435	1.9756	3.7032	−1.0000	♡
25ms, 80Hz	−1.8006	1.9206	1.7480	3.4405	−0.9999	♡
12.5ms, 160Hz	−1.8837	1.8230	1.7861	3.1884	−1.0000	♡
400ms, 10Hz	−1.0276	0.8294	0.6684	30.3367	−0.9974	♣
200ms, 20Hz	−1.5159	1.5601	1.3331	3.7541	−0.9999	♡
100ms, 40Hz	−1.7157	1.8930	1.7770	3.7153	−1.0000	♡
50ms, 80Hz	−1.7931	2.1097	1.9836	3.8669	−1.0000	♡
25ms, 160Hz	−1.8056	2.0377	1.8537	3.6343	−0.9999	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

Table E.20: FORA $\mathcal{D}_6$ estimates for Observer 2.

Expt. Level	m	k	$\mathcal{D}_{6,1}$	$\mathcal{D}_{6,\infty}$	Correlation	Status
400ms, 2.5Hz	−1.7911	0.8548	1.0110	1.7766	−1.0000	♡
200ms, 5Hz	−1.9153	1.0105	1.3197	2.0595	−1.0000	♡
100ms, 10Hz	−1.9568	1.1482	1.5583	2.3476	−1.0000	♡
50ms, 20Hz	−2.0214	1.1805	1.6954	2.4336	−0.9999	♡
25ms, 40Hz	−1.9413	1.1873	1.7746	2.6100	−0.9999	♡
12.5ms, 80Hz	−2.0156	1.3042	1.6688	2.4912	−0.9999	♡
6.25ms, 160Hz	−1.8277	1.2835	1.3702	2.4512	−1.0000	♡
400ms, 5Hz	−1.9011	1.4392	1.7818	2.8589	−1.0000	♡
200ms, 10Hz	−1.9833	1.8051	2.2651	3.4580	−0.9999	♡
100ms, 20Hz	−2.0265	1.7909	2.3983	3.5101	−1.0000	♡
50ms, 40Hz	−1.9639	1.8271	2.4957	3.7384	−0.9999	♡
25ms, 80Hz	−1.9422	1.7603	2.3385	3.5754	−1.0000	♡
12.5ms, 160Hz	−1.8953	1.6451	1.8446	3.0871	−1.0000	♡
400ms, 10Hz	−1.7839	1.4277	1.7127	3.0075	−0.9998	♡
200ms, 20Hz	−1.8819	1.7216	2.0854	3.4135	−1.0000	♡
100ms, 40Hz	−1.9358	1.8657	2.3580	3.6818	−1.0000	♡
50ms, 80Hz	−1.9218	1.8709	2.2586	3.6148	−1.0000	♡
25ms, 160Hz	−1.8156	1.7020	1.7974	3.2600	−1.0000	♡

Code: ♡ converged, ♠ did not converge, ♣ converged to impossible number, and ♢ not enough replications.

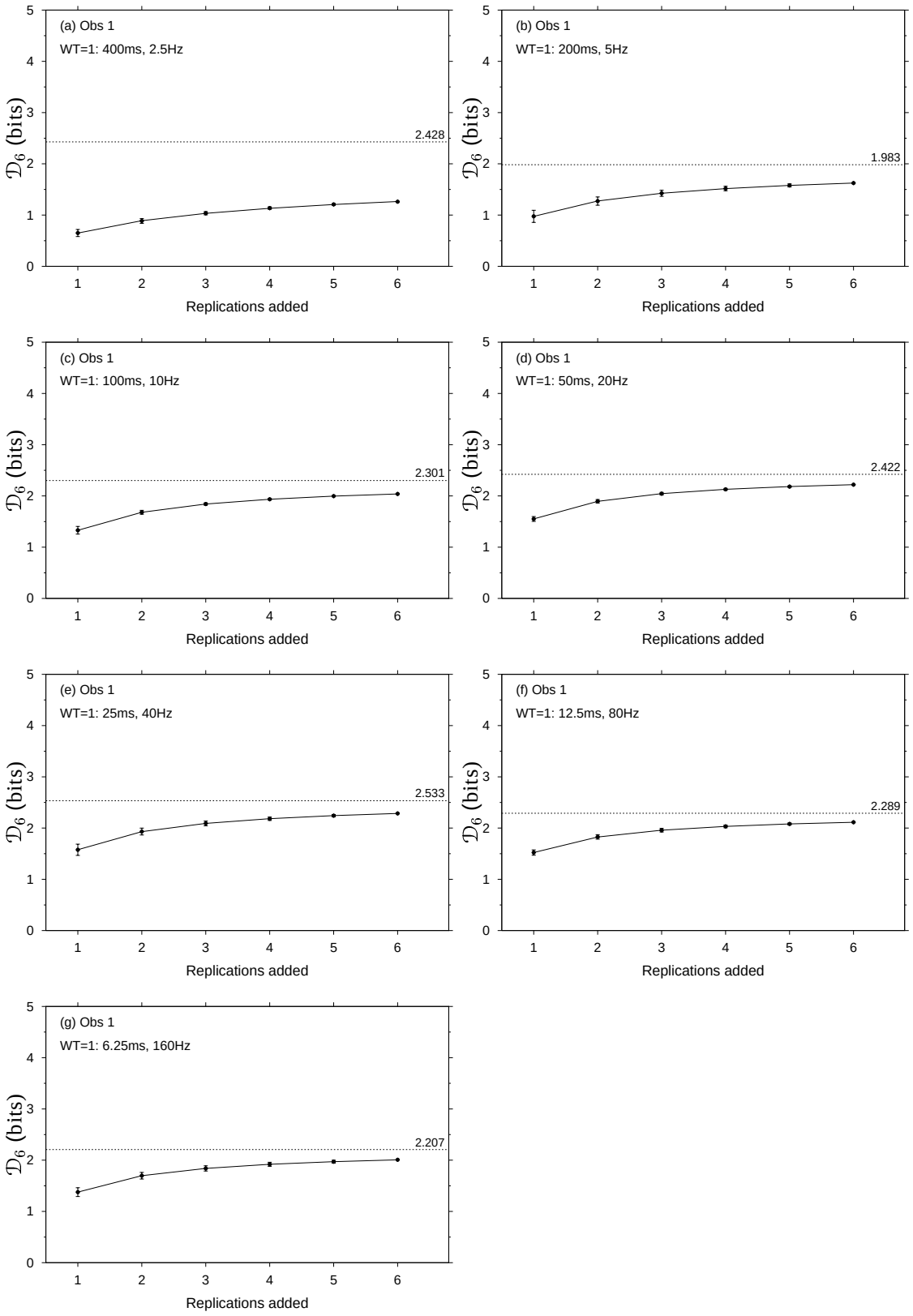

Figure E.13: $\mathcal{D}_6$ FORA for Observer 1, in condition $\mathcal{WT}$=1. Dashed lines are the asymptotic $\mathcal{D}_6$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation.

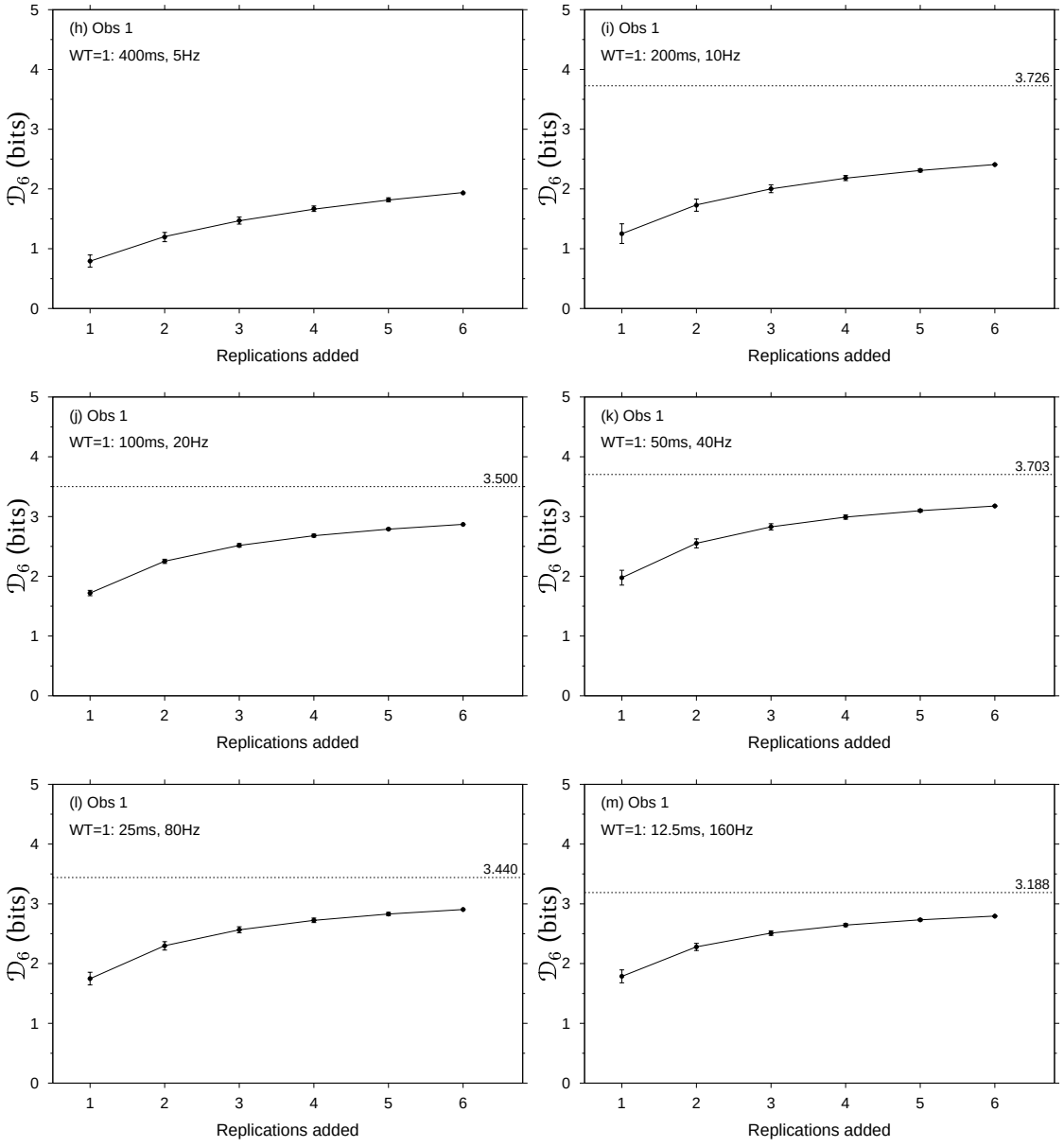

Figure E.14: $\mathcal{D}_6$ FORA for Observer 1, in condition $\mathcal{WT}$=2. Dashed lines are the asymptotic $\mathcal{D}_6$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation.

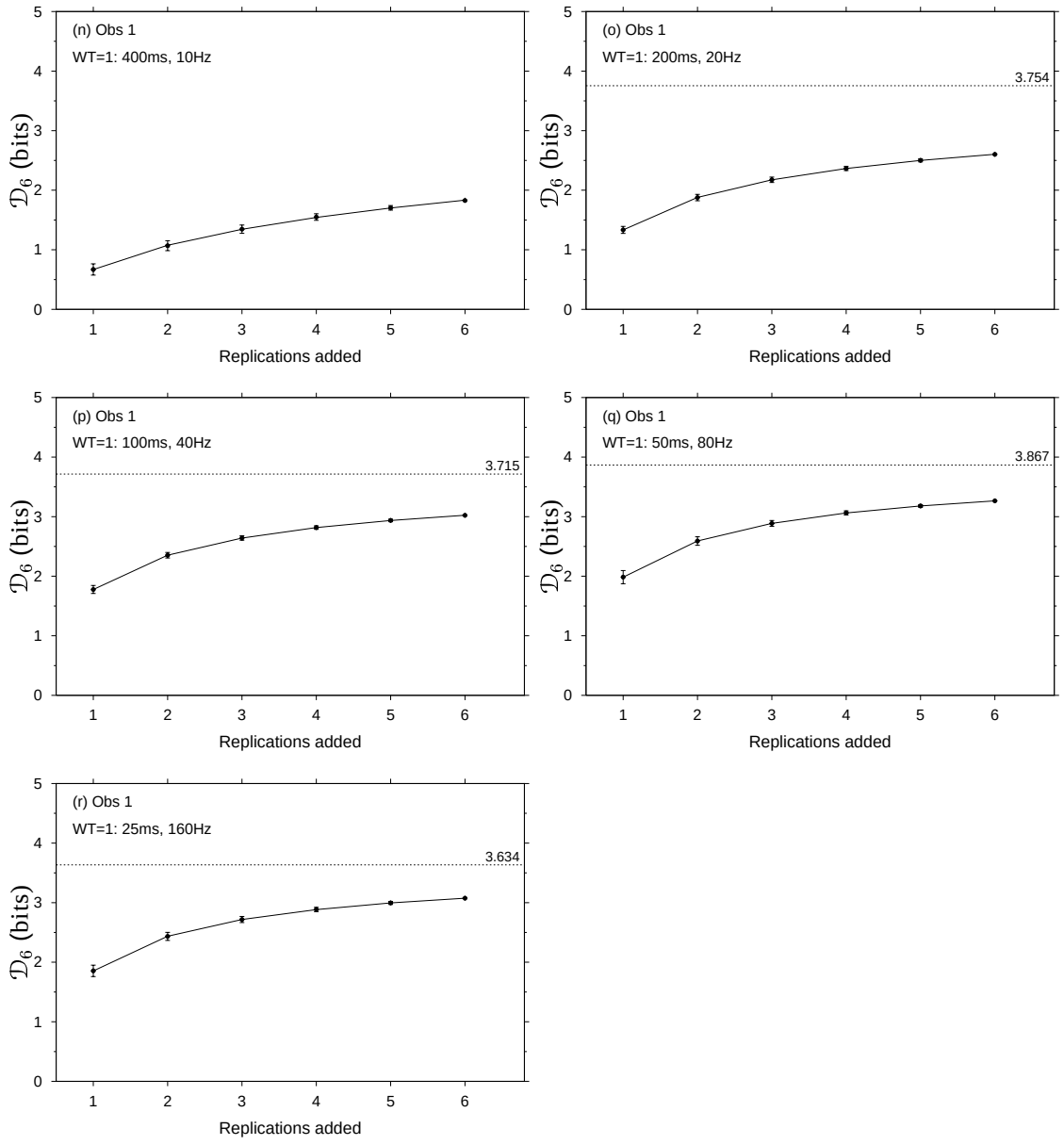

Figure E.15: $\mathcal{D}_6$ FORA for Observer 1, in condition $\mathcal{WT}$=4. Dashed lines are the asymptotic $\mathcal{D}_6$. Solid lines are the fitted FORA. Error bars are $\pm$1 standard deviation.

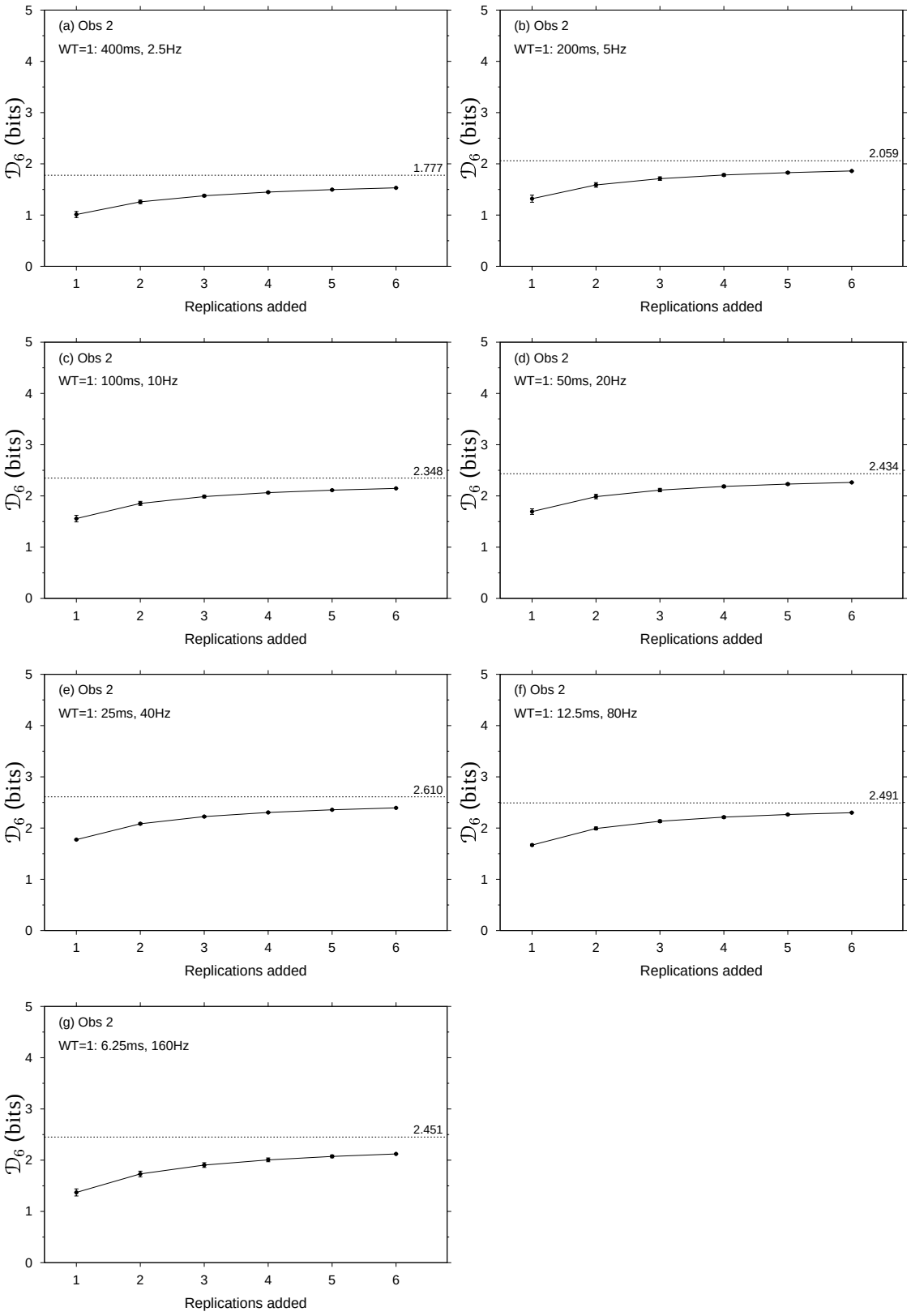

Figure E.16: $\mathcal{D}_6$ FORA for Observer 2, in condition $\mathcal{WT}$=1. Dashed lines are the asymptotic $\mathcal{D}_6$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation.

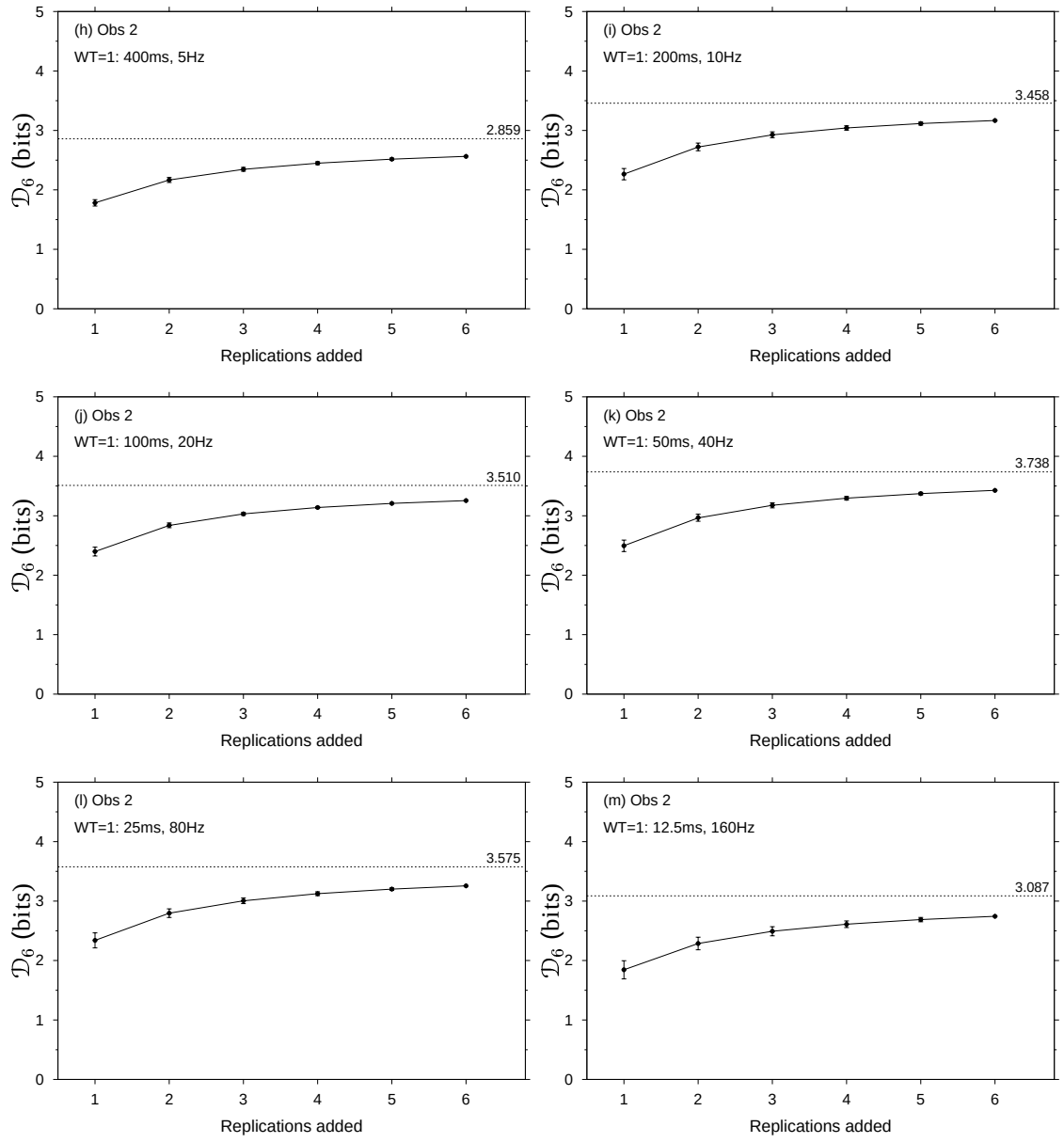

Figure E.17: $\mathcal{D}_6$ FORA for Observer 2, in condition $\mathcal{WT}$=2. Dashed lines are the asymptotic $\mathcal{D}_6$. Solid lines are the fitted FORA. Error bars are $\pm$1 standard deviation.

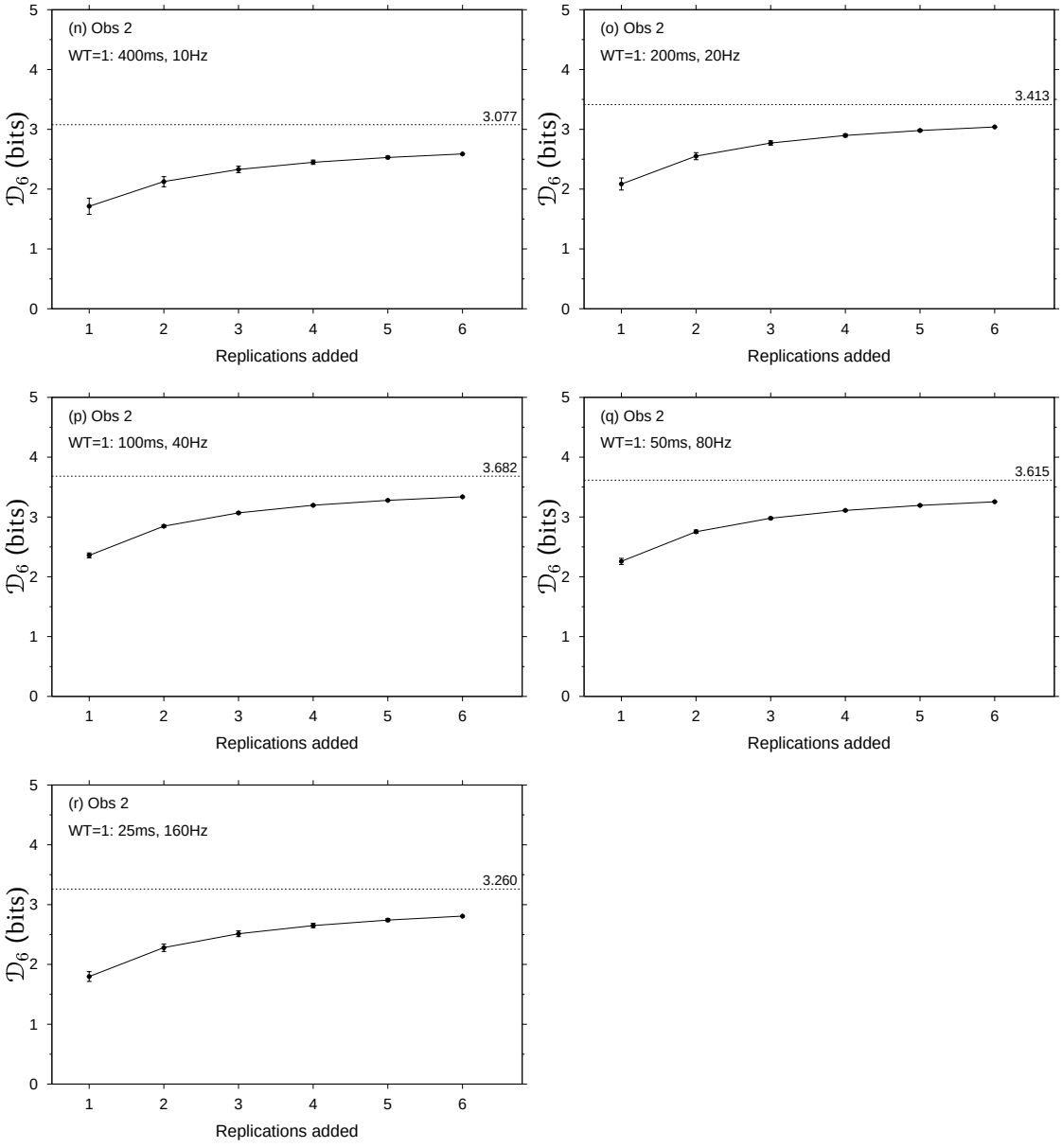

Figure E.18: $\mathcal{D}_6$ FORA for Observer 2, in condition $\mathcal{WT}$=4. Dashed lines are the asymptotic $\mathcal{D}_6$. Solid lines are the fitted FORA. Error bars are ±1 standard deviation.

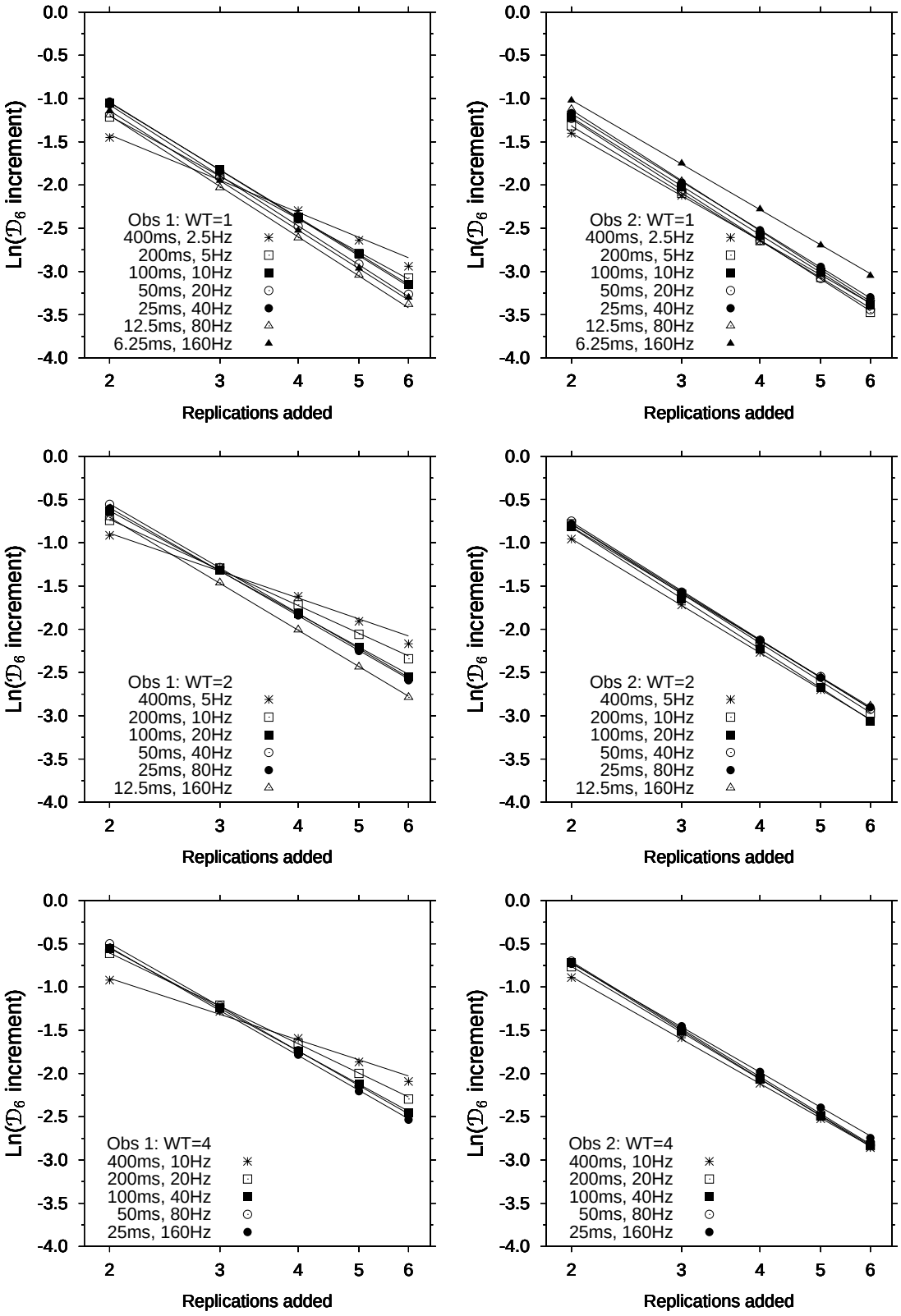

Figure E.19: $\mathcal{D}_6$ FORA, in log–increment coordinates, for Observer 1 and Observer 2 in each condition. Solid lines are the fitted model.

Appendix F

Psychometric and attenuation functions

Figures F.1–F.12 show the psychometric functions from the FORCE, GOC, and mROC analyses, calculated from the six replications completed by each observer. The plots on the left hand page show the human data compared to attenuated ideal energy detector psychometric functions, and the plots on the right hand side shows the same human data compared to the attenuated ideal (fitted polynomial) full–linear psychometric functions.

Each figure shows:

1. a theoretical psychometric function either the energy or the full–linear detector. Specifically, the parameters used for the energy detector were $\mathcal{WT}=1$ for experimental condition $\mathcal{WT}=1$, $\mathcal{WT}=3$ for experimental condition $\mathcal{WT}=2$, and $\mathcal{WT}=5$ for experimental condition $\mathcal{WT}=4$.

2. the empirical psychometric points for each signal–to–noise ratio from the FORCE, GOC, and mROC analyses for each observer, and

3. the theoretical psychometric function *attenuated* to fit the human data after FORCE, GOC, and mROC analysis. The attenuation is the linear average of attenuation in decibels for each five signal–to–noise ratios relative to the model being tested. For the FORCE results, some points are not included in the attenuation calculation if they are rogue or did not converge successfully.

Tables F.1–F.6 give the actual attenuation from the energy detector, for each signal–to–noise ratio, as well as the average attenuations on which the psychometric functions are based.

Chapters 3 and 4 explain the analyses in more detail.

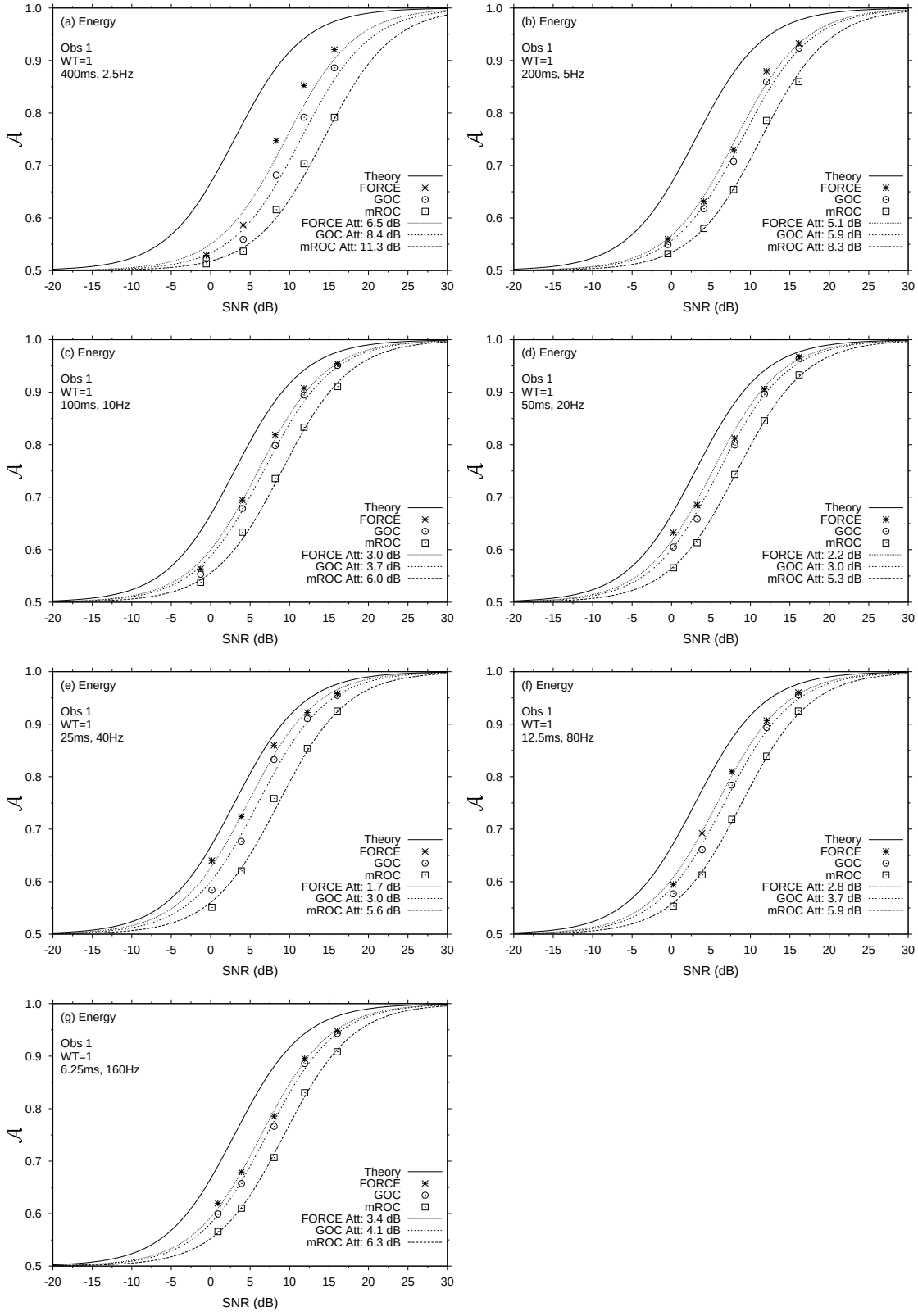

Figure F.1: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 1 in condition $\mathcal{WT}=1$. The fitted model is the energy detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

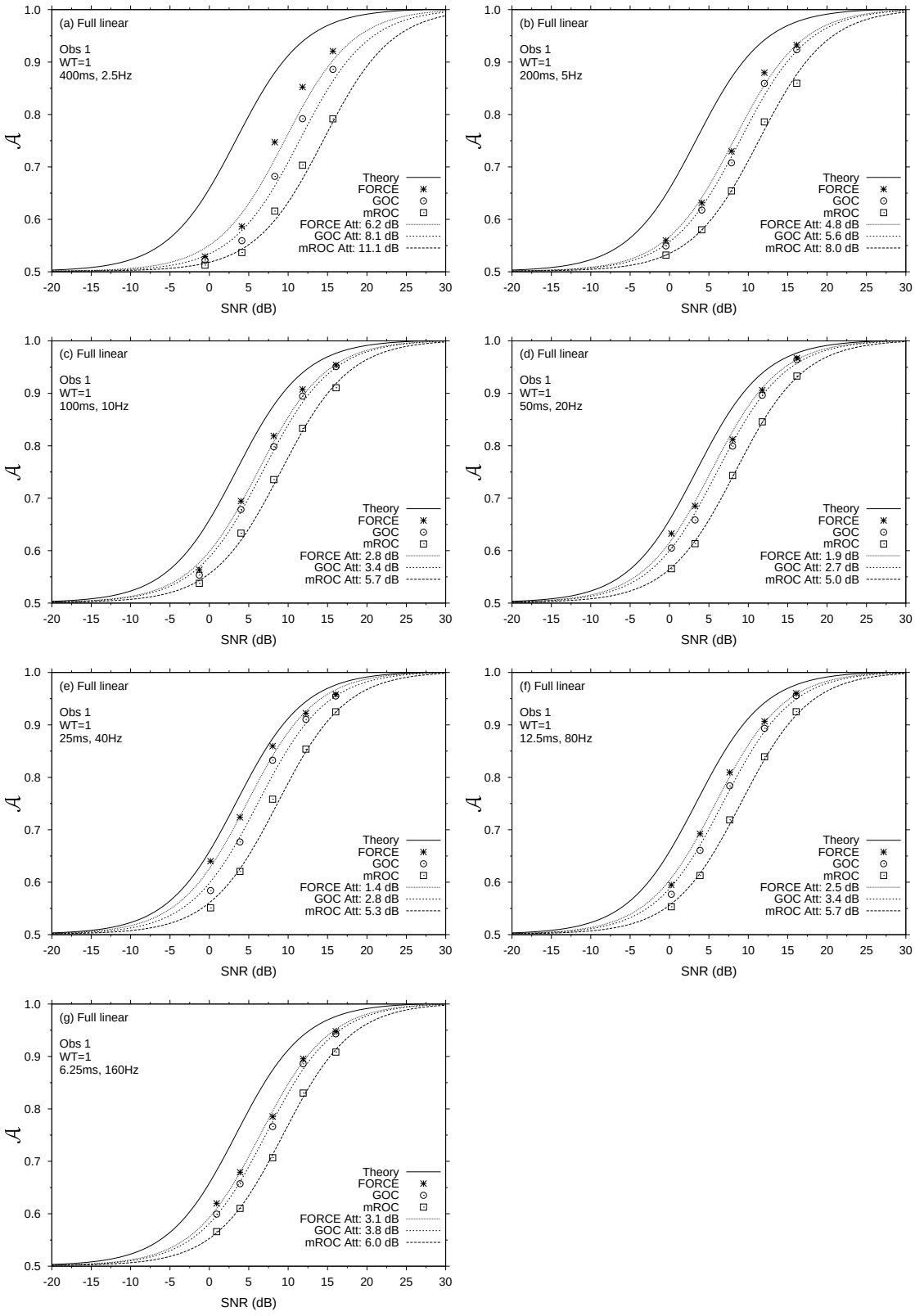

Figure F.2: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 1 in condition $\mathcal{WT}$=1. The fitted model is the full–linear detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

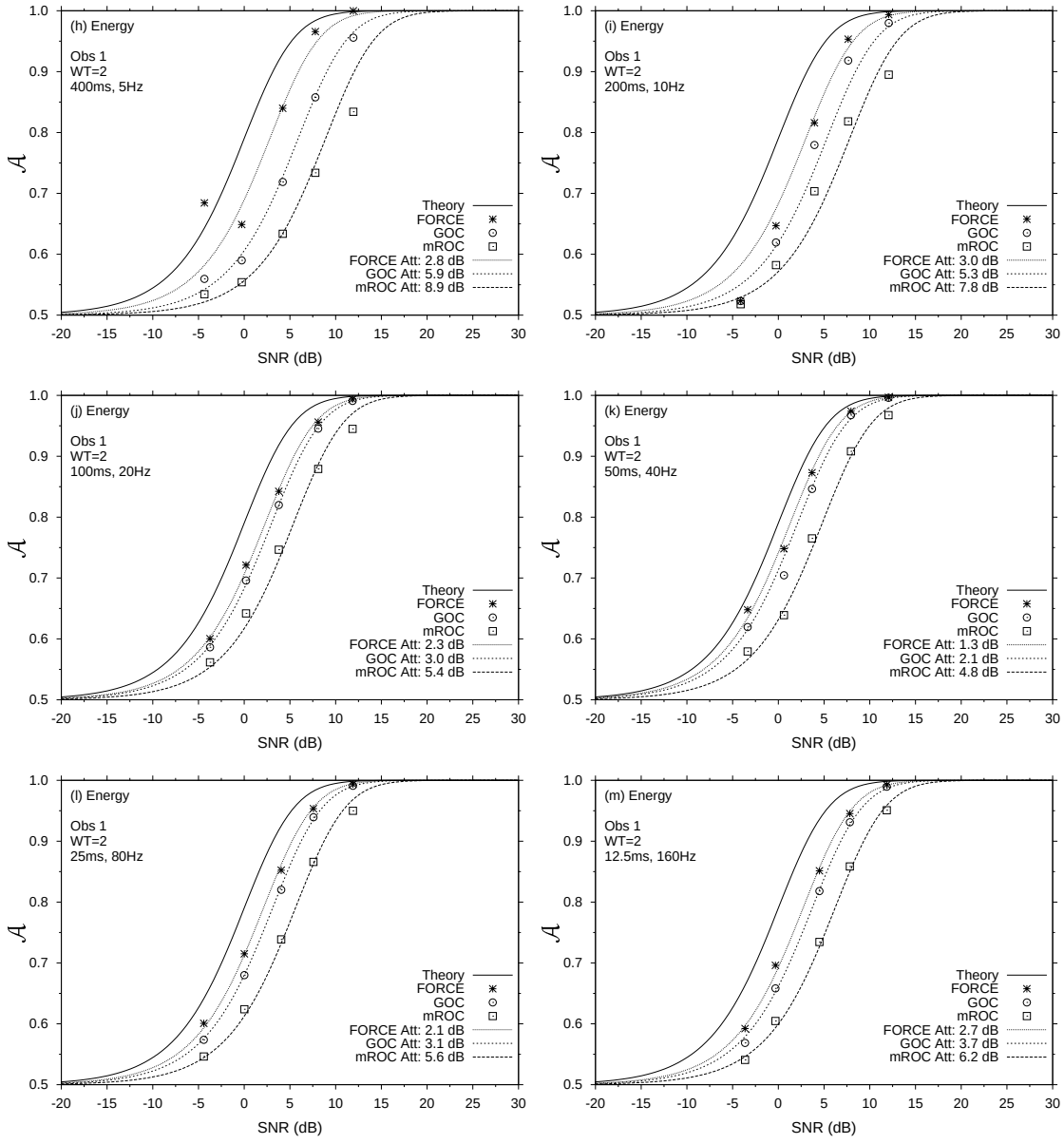

Figure F.3: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 1 in condition $\mathcal{WT}=2$. The fitted model is the energy detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

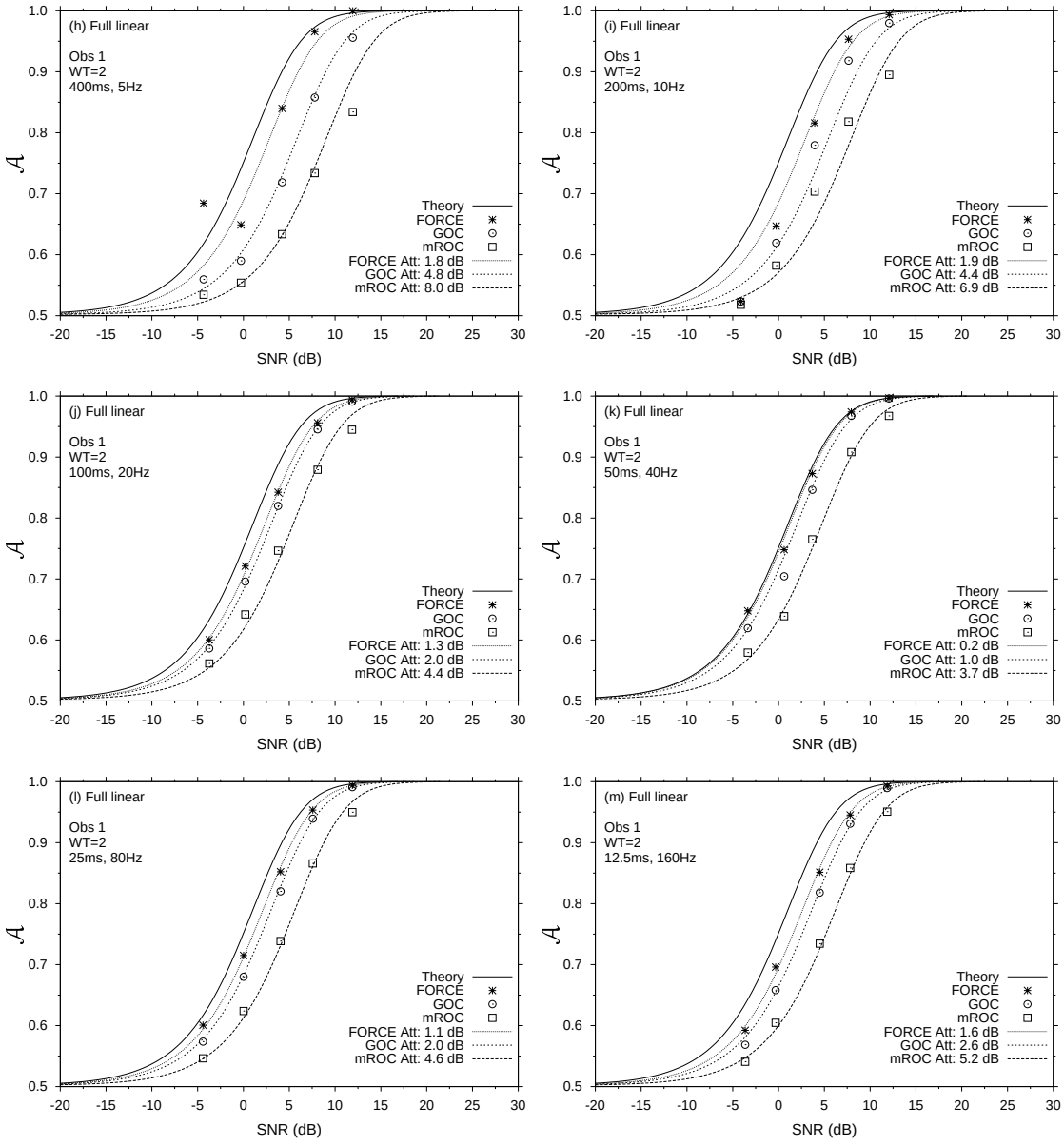

Figure F.4: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 1 in condition $\mathcal{WT}$=2. The fitted model is the full–linear detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

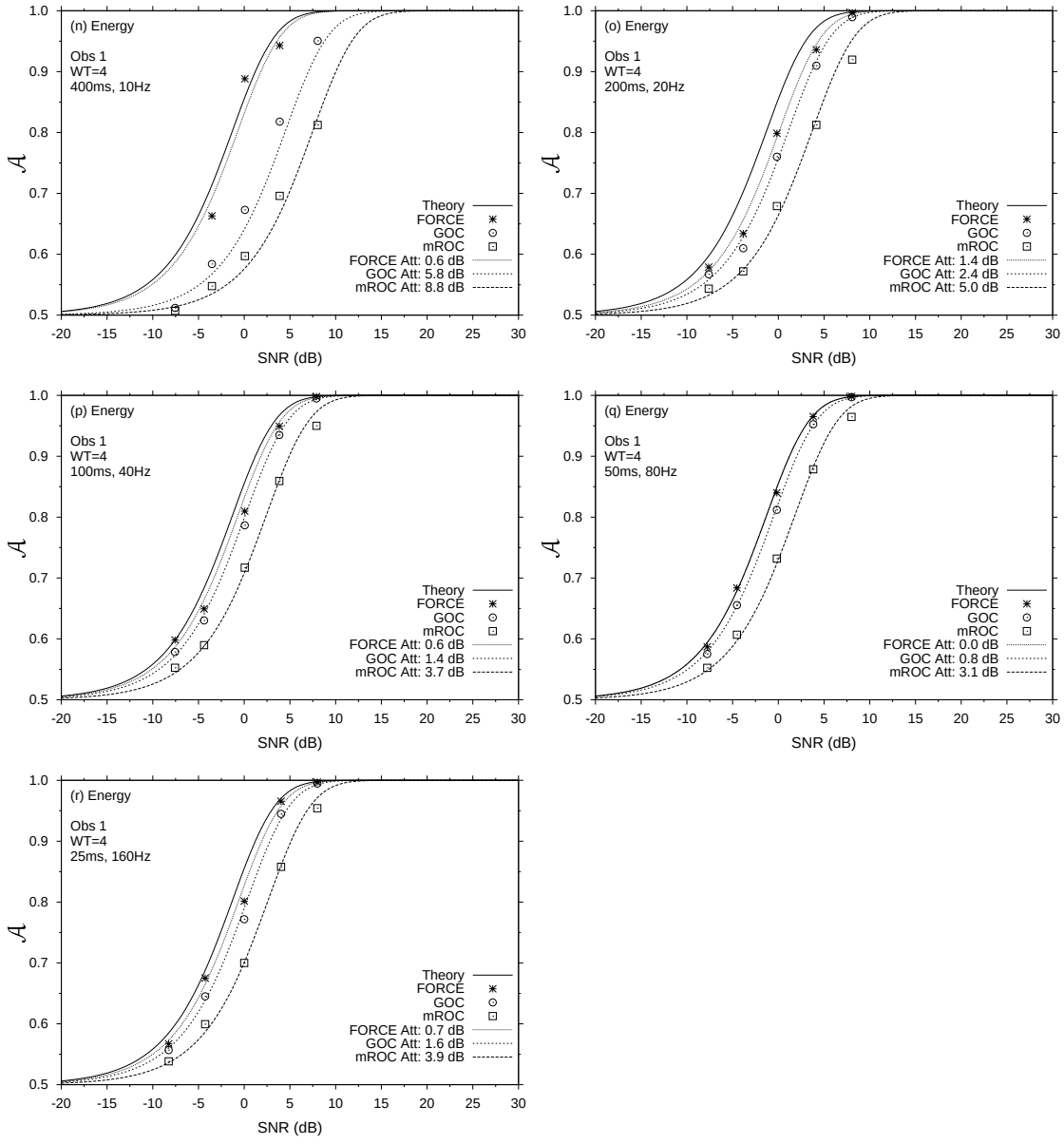

Figure F.5: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 1 in condition $\mathcal{WT}=4$. The fitted model is the energy detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

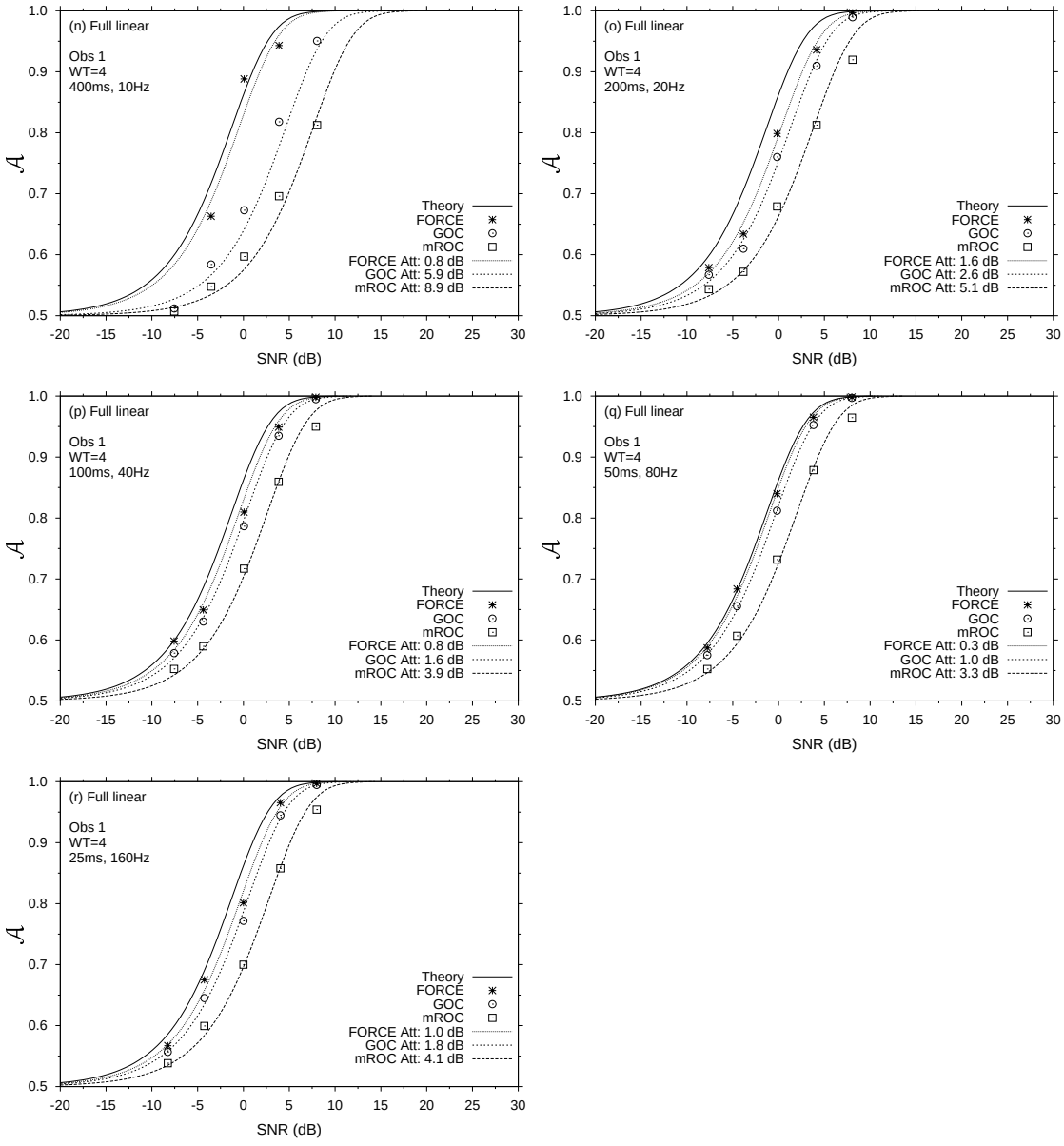

Figure F.6: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 1 in condition $\mathcal{WT}$=4. The fitted model is the full–linear detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

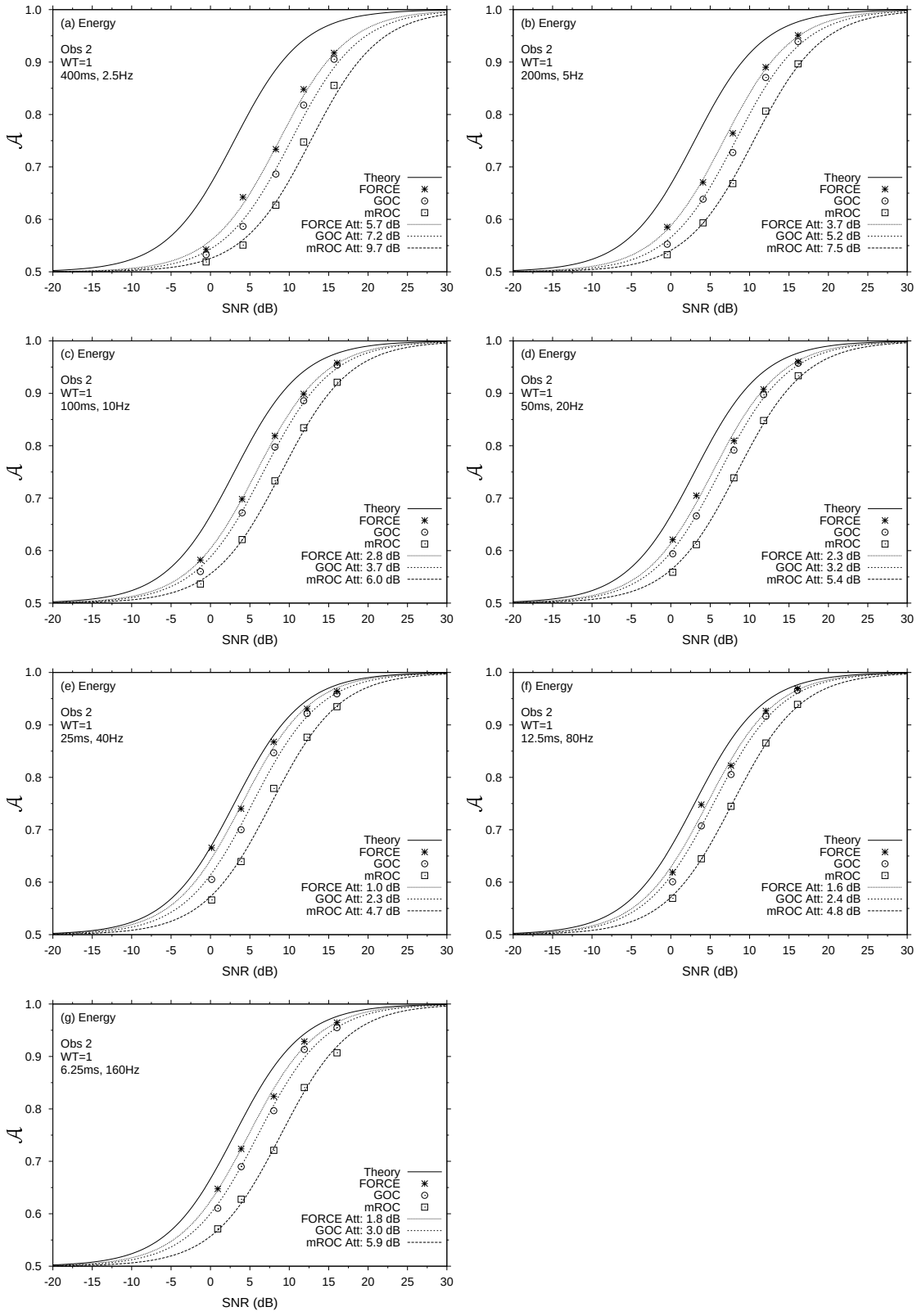

Figure F.7: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 2 in condition $\mathcal{WT}=1$. The fitted model is the energy detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

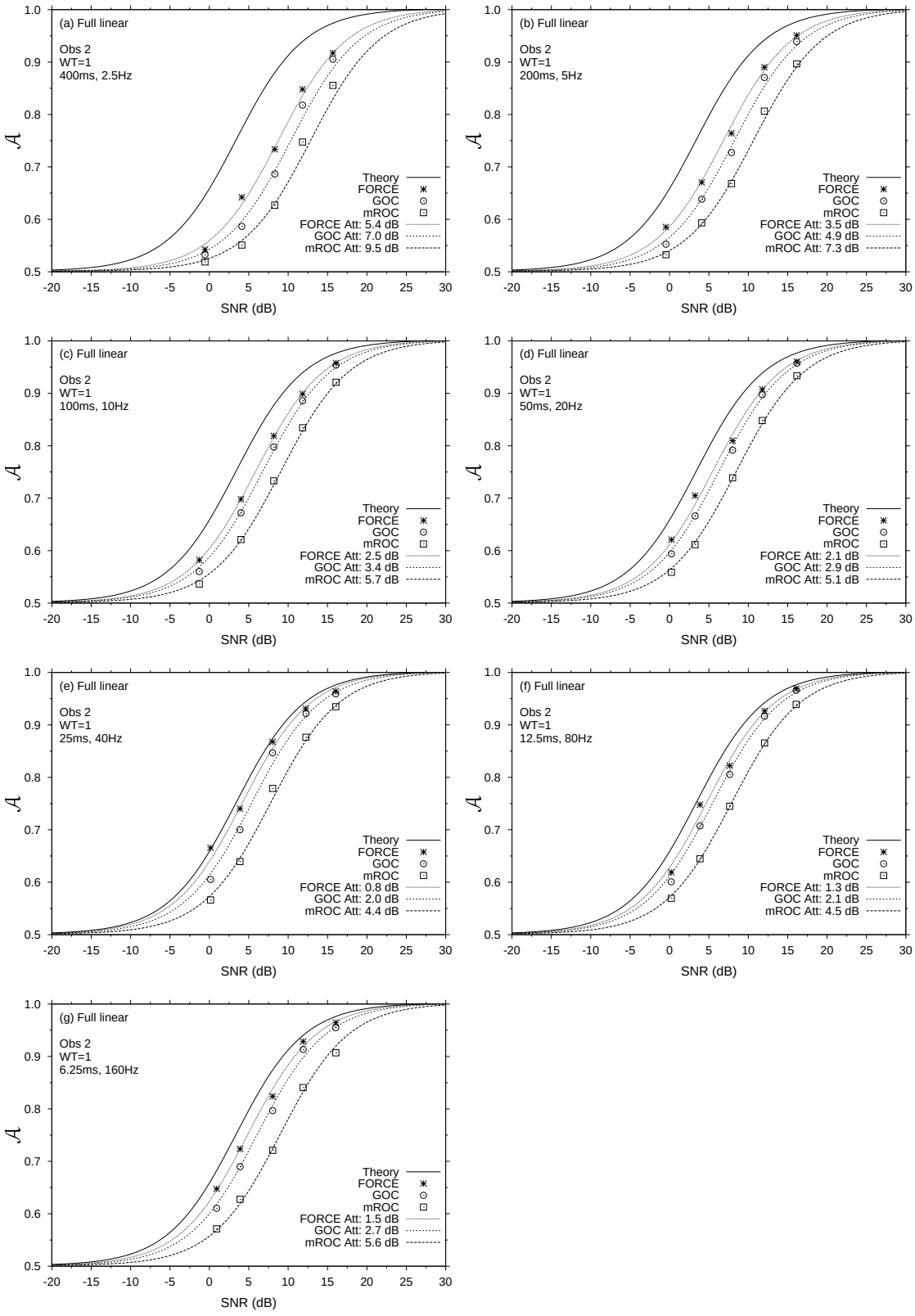

Figure F.8: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 2 in condition $\mathcal{WT}$=1. The fitted model is the full–linear detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

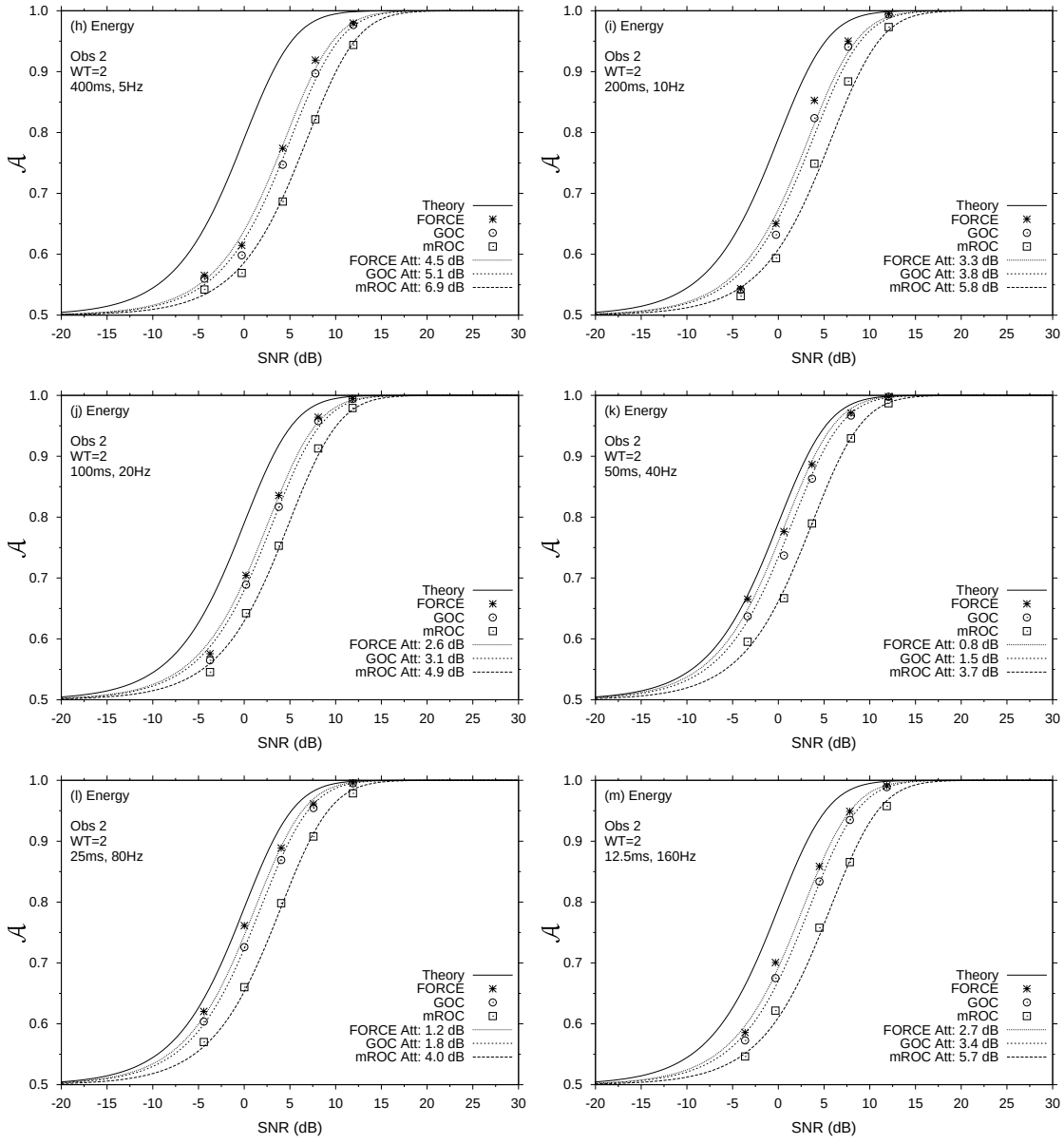

Figure F.9: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 2 in condition $\mathcal{WT}=2$. The fitted model is the energy detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

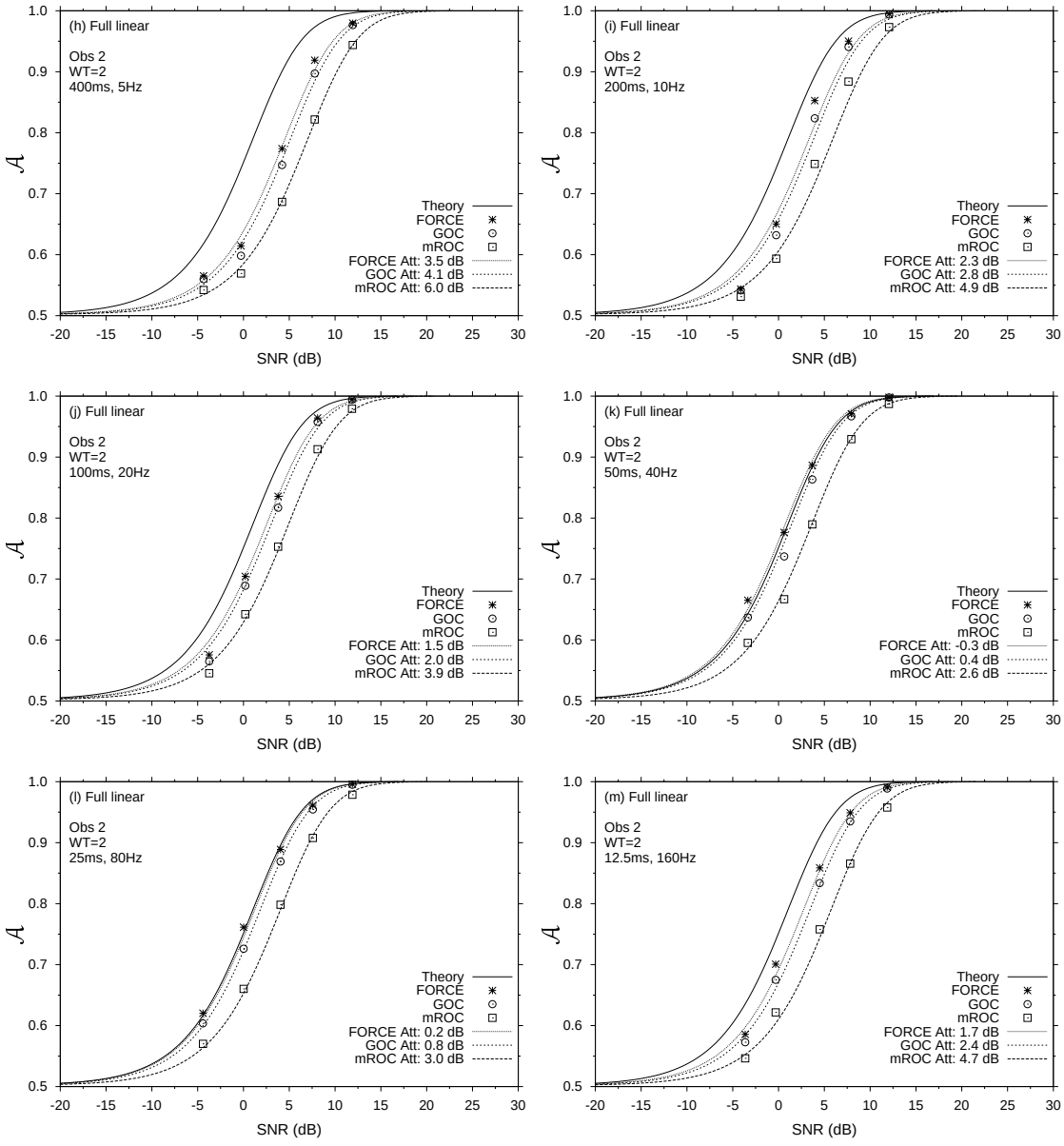

Figure F.10: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 2 in condition $\mathcal{WT}=2$. The fitted model is the full–linear detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

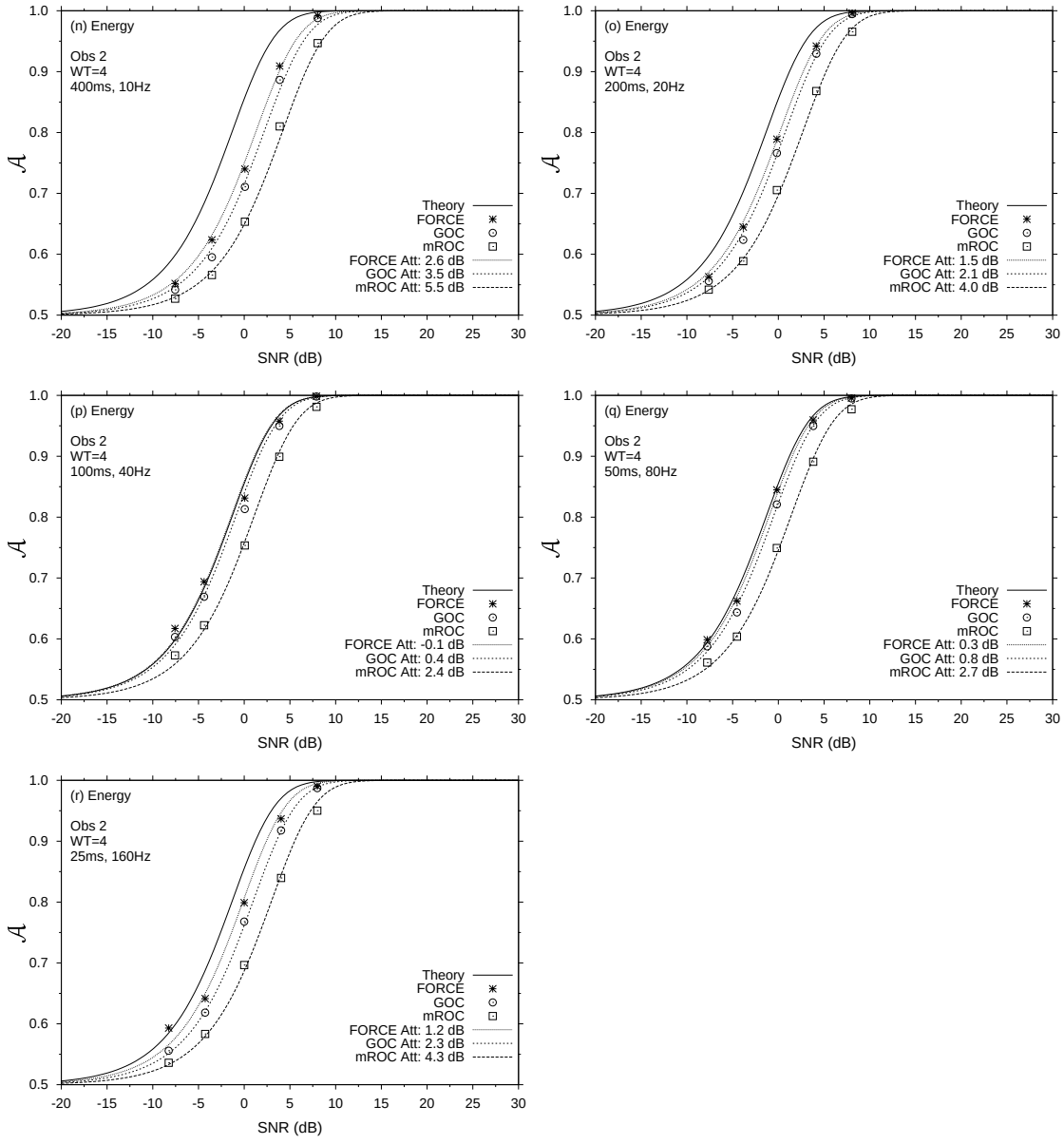

Figure F.11: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 2 in condition $\mathcal{WT}$=4. The fitted model is the energy detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

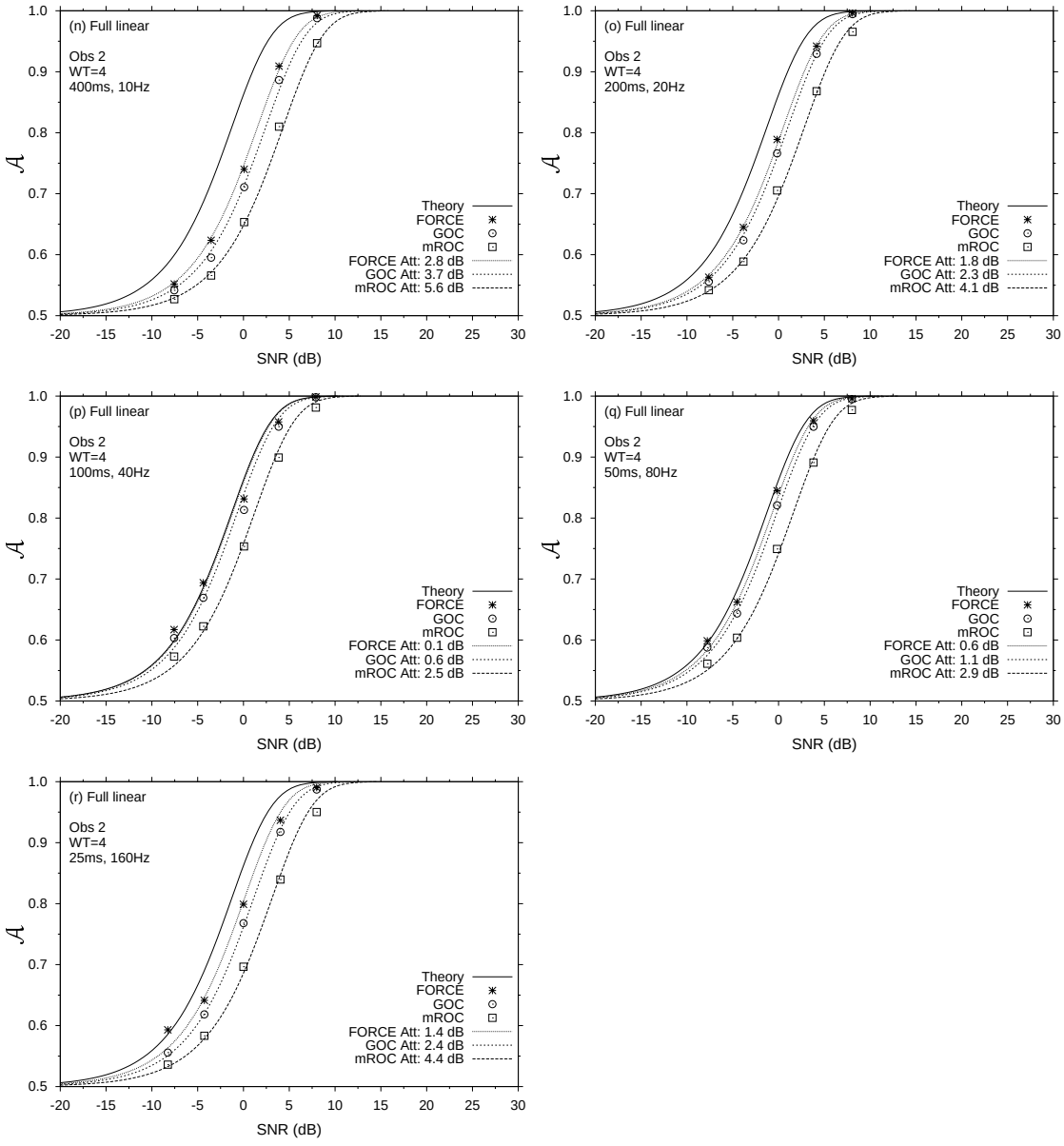

Figure F.12: Psychometric functions after FORCE, GOC and mROC analyses, for Observer 2 in condition $\mathcal{WT}=4$. The fitted model is the full–linear detector (a) with no attenuation (solid line) and (b) attenuated to fit the human data.

Table F.1: Attenuation, difference in attenuation, and average attenuation (dB) from the $\mathcal{WT}$=1 χ^2 energy detector, and the fitted polynomial full–linear detector, after each analysis for Observer 1 in condition $\mathcal{WT}$=1.

Experimental	Theory	Empirical	Energy Attenuation					Full–linear Attenuation				
Level	SNR (dB)	SNR (dB)	mROC	Diff	GOC	Diff	FORCE	mROC	Diff	GOC	Diff	FORCE
400ms, 2.5Hz	0	−0.5563	12.2735	2.5332	9.7403	1.2239	8.5164	12.2869	2.6969	9.5900	1.2733	8.3167
	4	4.1181	12.1206	2.3019	9.8187	1.9163	7.9023	11.8882	2.3491	9.5391	1.9393	7.5998
	8	8.3007	10.4959	2.7817	7.7142	2.3285	5.3858	10.1817	2.7911	7.3905	2.3278	5.0628
	12	11.8220	10.4561	3.1167	7.3394	2.3013	5.0381	10.1318	3.1106	7.0212	2.2864	4.7348
	16	15.6759	11.2033	3.8364	7.3668	1.9511	5.4157	10.8851	3.8048	7.0803	1.9180	5.1622
	Average		11.3099	2.9140	8.3959	1.9442	6.4517	11.0747	2.9505	8.1242	1.9490	6.1753
200ms, 5Hz	0	−0.4546	8.2031	2.0651	6.1381	0.9276	5.2104	7.9897	2.1160	5.8737	0.9433	4.9304
	4	4.0882	8.2739	2.0871	6.1868	0.6369	5.5499	7.9753	2.1033	5.8720	0.6401	5.2319
	8	7.8711	8.3630	2.0233	6.3396	0.7784	5.5612	8.0415	2.0262	6.0153	0.7780	5.2373
	12	12.0383	7.7715	2.8129	4.9586	0.9217	4.0369	7.4525	2.7944	4.6580	0.9116	3.7464
	16	16.1484	9.0598	3.3425	5.7173	0.6532	5.0641	8.7593	3.2918	5.4675	0.6377	4.8298
	Average		8.3343	2.4662	5.8681	0.7836	5.0845	8.0437	2.4663	5.5773	0.7821	4.7952
100ms, 10Hz	0	−1.2808	6.5918	1.6487	4.9431	0.8624	4.0807	6.3560	1.6839	4.6720	0.8758	3.7962
	4	4.0158	5.3947	1.8328	3.5619	0.5959	2.9660	5.0763	1.8380	3.2384	0.5965	2.6419
	8	8.1727	5.6635	2.1960	3.4674	0.7584	2.7091	5.3398	2.1896	3.1502	0.7543	2.3959
	12	11.8322	5.8172	2.7359	3.0813	0.7036	2.3777	5.5077	2.7068	2.8009	0.6923	2.1087
	16	16.0756	6.4401	3.0157	3.4243	0.3079	3.1164	6.1743	2.9359	3.2383	0.2963	2.9421
	Average		5.9814	2.2858	3.6956	0.6456	3.0500	5.6908	2.2708	3.4200	0.6430	2.7770
50ms, 20Hz	0	0.2443	5.4508	2.4661	2.9847	1.3198	1.6649	5.1642	2.4904	2.6739	1.3271	1.3468
	4	3.2400	5.5575	1.9959	3.5616	1.0263	2.5353	5.2439	2.0043	3.2397	1.0282	2.2114
	8	8.0137	5.2334	1.9663	3.2671	0.4654	2.8017	4.9101	1.9601	2.9501	0.4630	2.4871
	12	11.7777	5.2674	2.3265	2.9408	0.5314	2.4094	4.9617	2.3000	2.6617	0.5228	2.1389
	16	16.1712	5.0599	3.0296	2.0303	0.3941	1.6362	4.8264	2.9174	1.9089	0.3731	1.5358
	Average		5.3138	2.3569	2.9569	0.7474	2.2095	5.0213	2.3344	2.6869	0.7429	1.9440
25ms, 40Hz	0	0.1491	6.5777	2.4987	4.0790	2.8450	1.2340	6.3103	2.5325	3.7778	2.8634	0.9144
	4	3.8770	5.8458	2.3553	3.4906	1.7105	1.7800	5.5303	2.3632	3.1672	1.7112	1.4559
	8	8.0293	4.7289	2.6923	2.0365	1.0955	0.9411	4.4067	2.6798	1.7269	1.0863	0.6406
	12	12.2573	5.4089	2.7699	2.6390	0.7330	1.9059	5.1062	2.7333	2.3728	0.7185	1.6544
	16	16.0475	5.5066	2.5218	2.9848	0.3441	2.6407	5.2592	2.4447	2.8145	0.3297	2.4848
	Average		5.6136	2.5676	3.0460	1.3456	1.7003	5.3225	2.5507	2.7719	1.3418	1.4300
12.5ms, 80Hz	0	0.2250	6.4591	1.8485	4.6106	1.0914	3.5191	6.1882	1.8742	4.3141	1.1017	3.2124
	4	3.8712	6.2237	2.1201	4.1035	1.2032	2.9003	5.9103	2.1290	3.7814	1.2051	2.5762
	8	7.6408	5.7188	2.2718	3.4470	0.9244	2.5226	5.3946	2.2669	3.1276	0.9202	2.2075
	12	12.0682	5.8154	2.4345	3.3809	0.7243	2.6567	5.5077	2.4081	3.0996	0.7128	2.3869
	16	16.0755	5.5196	2.5758	2.9438	0.5193	2.4245	5.2726	2.4963	2.7763	0.4968	2.2795
	Average		5.9473	2.2501	3.6972	0.8925	2.8046	5.6547	2.2349	3.4198	0.8873	2.5325
6.25ms, 160Hz	0	0.9183	6.0990	2.1317	3.9673	1.0461	2.9211	5.8120	2.1534	3.6586	1.0528	2.6058
	4	3.8949	6.3705	2.1026	4.2679	0.8555	3.4124	6.0579	2.1118	3.9461	0.8572	3.0889
	8	8.0366	6.5403	2.0818	4.4585	0.6757	3.7828	6.2160	2.0789	4.1371	0.6734	3.4636
	12	11.9083	6.0077	2.4109	3.5969	0.4802	3.1167	5.6975	2.3871	3.3104	0.4734	2.8370
	16	16.0541	6.5601	2.4410	4.1191	0.4186	3.7005	6.2917	2.3829	3.9089	0.4049	3.5040
	Average		6.3155	2.2336	4.0819	0.6952	3.3867	6.0150	2.2228	3.7922	0.6924	3.0998

Table F.2: Attenuation, difference in attenuation, and average attenuation (dB) from the $\mathcal{WT}$=3 χ^2 energy detector, and the fitted polynomial full–linear detector, after each analysis for Observer 1 in condition $\mathcal{WT}$=2.

Experimental	Theory	Empirical	Energy Attenuation					Full–linear Attenuation				
Level	SNR (dB)	SNR (dB)	mROC	Diff	GOC	Diff	FORCE	mROC	Diff	GOC	Diff	FORCE
400ms, 5Hz	−4	−4.3532	6.8212	2.5145	4.3068	5.7062	−1.3995	5.9938	2.6127	3.3810	5.7964	−2.4154
	0	−0.2741	8.8359	2.4091	6.4268	2.5599	3.8670	7.9237	2.4674	5.4563	2.5948	2.8615
	4	4.2112	8.9210	2.7779	6.1431	3.2617	2.8814	7.9216	2.8017	5.1199	3.2786	1.8413
	8	7.7795	9.2905	3.3399	5.9506	4.2747	1.6758	8.2647	3.3567	4.9081	4.3116	0.5965
	12	11.9325	10.7621	4.3070	6.4550	11.1689	−4.7139	9.7227	4.3381	5.3846	9.8848	−4.5002
	Average		8.9261	3.0697	5.8564	5.3943	2.8081	7.9653	3.1153	4.8500	5.1732	−0.3233
200ms, 10Hz	−4	−4.1226	9.9710	1.1362	8.8348	0.0561	8.7787	9.3659	1.2422	8.1237	0.0606	8.0631
	0	−0.2550	6.8794	1.8417	5.0377	1.0805	3.9572	5.9171	1.8716	4.0455	1.0930	2.9525
	4	3.9492	6.3227	2.0928	4.2299	0.9642	3.2657	5.3025	2.1050	3.1975	0.9688	2.2287
	8	7.6392	6.8941	3.0250	3.8691	1.5730	2.2961	5.8568	3.0415	2.8152	1.5879	1.2274
	12	12.0730	9.1118	4.3444	4.7675	2.2769	2.4906	8.0632	4.3991	3.6642	2.3581	1.3060
	Average		7.8358	2.4880	5.3478	1.1901	3.0024	6.9011	2.5319	4.3692	1.2137	3.1555
100ms, 20Hz	−4	−3.7360	4.7729	1.6015	3.1714	0.7409	2.4305	3.8429	1.6383	2.2046	0.7537	1.4509
	0	0.2015	4.5985	1.7955	2.8031	0.7414	2.0617	3.5957	1.8113	1.7845	0.7464	1.0380
	4	3.7811	4.9426	1.9513	2.9913	0.6068	2.3845	3.9149	1.9611	1.9538	0.6097	1.3441
	8	8.0913	5.6284	2.4697	3.1587	0.5614	2.5974	4.5826	2.4880	2.0946	0.5679	1.5268
	12	11.8498	6.9384	4.0105	2.9278	0.7627	2.1652	5.8745	4.1021	1.7724	0.7967	0.9757
	Average		5.3762	2.3657	3.0105	0.6826	2.3278	4.3621	2.4001	1.9620	0.6949	1.2671
50ms, 40Hz	−4	−3.3567	3.9481	2.0161	1.9320	1.1119	0.8201	2.9893	2.0496	0.9397	1.1247	−0.1850
	0	0.6147	5.1200	2.1597	2.9603	1.2265	1.7338	4.1184	2.1785	1.9399	1.2341	0.7058
	4	3.6819	4.3461	2.1731	2.1730	0.7635	1.4095	3.3156	2.1836	1.1320	0.7673	0.3647
	8	7.9418	4.5371	2.8208	1.7163	0.5130	1.2033	3.4858	2.8508	0.6350	0.5224	0.1126
	12	12.0465	5.8132	4.2192	1.5940	0.3742	1.2198	4.7318	4.3661	0.3656	0.3936	−0.0279
	Average		4.7529	2.6778	2.0751	0.7978	1.2773	3.7282	2.7257	1.0024	0.8084	0.1940
25ms, 80Hz	−4	−4.4191	5.4070	2.1732	3.2338	1.5116	1.7222	4.5196	2.2371	2.2825	1.5403	0.7422
	0	0.0152	5.1166	2.0115	3.1051	1.0503	2.0548	4.1219	2.0317	2.0902	1.0579	1.0323
	4	4.0481	5.4237	2.1717	3.2520	0.8776	2.3744	4.3971	2.1827	2.2144	0.8818	1.3326
	8	7.5708	5.5009	2.5745	2.9264	0.7006	2.2258	4.4572	2.5920	1.8653	0.7082	1.1571
	12	11.8867	6.7279	3.7945	2.9334	0.6686	2.2649	5.6614	3.8847	1.7767	0.6983	1.0784
	Average		5.6352	2.5451	3.0901	0.9617	2.1284	4.6314	2.5856	2.0458	0.9773	1.0685
12.5ms, 160Hz	−4	−3.6383	6.7720	2.4159	4.3560	1.4148	2.9413	5.9082	2.4955	3.4127	1.4442	1.9685
	0	−0.3127	5.6465	2.1479	3.4986	1.2166	2.2821	4.6637	2.1737	2.4900	1.2266	1.2634
	4	4.4911	5.9826	2.2302	3.7524	0.9134	2.8391	4.9567	2.2416	2.7151	0.9177	1.7974
	8	7.8219	5.9713	2.4342	3.5371	0.6402	2.8969	4.9287	2.4495	2.4792	0.6463	1.8329
	12	11.8543	6.6475	3.4252	3.2223	0.6733	2.5490	5.5804	3.5022	2.0782	0.7010	1.3772
	Average		6.2040	2.5307	3.6733	0.9716	2.7017	5.2075	2.5725	2.6350	0.9871	1.6479

Table F.3: Attenuation, difference in attenuation, and average attenuation (dB) from the $\mathcal{WT}$=5 χ^2 energy detector, and the fitted polynomial full–linear detector, after each analysis for Observer 1 in condition $\mathcal{WT}$=4. Code: ♠ did not converge, ♣ converged to impossible number.

Experimental	Theory	Empirical	Energy Attenuation					Full–linear Attenuation				
Level	SNR (dB)	SNR (dB)	mROC	Diff	GOC	Diff	FORCE	mROC	Diff	GOC	Diff	FORCE
400ms, 10Hz	−8	−7.5580	11.9010	2.3283	9.5727	♠	♠	12.0136	2.3677	9.6459	♠	♠
	−4	−3.5357	7.3853	2.6089	4.7764	3.2848	1.4916	7.4250	2.6044	4.8206	3.2620	1.5586
	0	0.0730	7.6997	2.9103	4.7894	5.5942	−0.8049	7.7467	2.8869	4.8599	5.4519	−0.5921
	4	3.8845	7.9148	3.1080	4.8067	3.5895	1.2172	7.9942	3.0417	4.9524	3.4263	1.5262
	8	8.0386	9.0937	4.0424	5.0513	♣	♣	9.2355	3.8546	5.3808	♣	♣
	Average		8.7989	2.9996	5.7993	12.4686	0.6347	8.8830	2.9511	5.9319	12.1402	0.6309
200ms, 20Hz	−8	−7.6082	3.7354	1.9811	1.7543	0.7555	0.9987	3.7752	1.9799	1.7954	0.7535	1.0419
	−4	−3.8392	5.1910	2.0090	3.1820	0.9853	2.1967	5.2329	2.0006	3.2324	0.9784	2.2540
	0	−0.1545	4.3708	2.1940	2.1768	0.9413	1.2355	4.4436	2.1574	2.2862	0.9184	1.3678
	4	4.1558	5.2093	2.5589	2.6504	0.8898	1.7606	5.3511	2.4578	2.8932	0.8404	2.0529
	8	8.0789	6.2562	3.9389	2.3173	1.6692	0.6480	6.5158	3.6700	2.8458	1.6103	1.2355
	Average		4.9525	2.5364	2.4161	1.0482	1.3679	5.0637	2.4531	2.6106	1.0202	1.5904
100ms, 40Hz	−8	−7.5739	2.9125	1.8555	1.0570	1.0632	−0.0062	2.9522	1.8522	1.1000	1.0589	0.0412
	−4	−4.3887	3.6116	1.8194	1.7922	0.6945	1.0977	3.6569	1.8086	1.8484	0.6884	1.1600
	0	0.0579	3.5022	1.7613	1.7409	0.5634	1.1775	3.5905	1.7250	1.8655	0.5481	1.3174
	4	3.8417	3.7342	2.2496	1.4845	0.5712	0.9133	3.9150	2.1404	1.7745	0.5355	1.2390
	8	7.9003	4.9332	3.8536	1.0797	1.2056	−0.1259	5.2615	3.6012	1.6603	1.2120	0.4483
	Average		3.7387	2.3079	1.4309	0.8196	0.6113	3.8752	2.2255	1.6498	0.8086	0.8412
50ms, 80Hz	−8	−7.7795	2.7126	1.6576	1.0550	0.6863	0.3687	2.7524	1.6549	1.0975	0.6840	0.4135
	−4	−4.5292	2.6285	1.8800	0.7485	0.8919	−0.1434	2.6781	1.8653	0.8128	0.8816	−0.0689
	0	−0.1683	2.8852	1.9872	0.8980	0.6846	0.2134	2.9803	1.9408	1.0395	0.6630	0.3765
	4	3.8162	3.2037	2.4628	0.7409	0.6378	0.1031	3.4048	2.3286	1.0763	0.5937	0.4826
	8	8.0010	4.3199	3.8739	0.4460	0.7487	−0.3027	4.6972	3.6650	1.0322	0.7710	0.2612
	Average		3.1500	2.3723	0.7777	0.7299	0.0478	3.3026	2.2909	1.0116	0.7186	0.2930
25ms, 160Hz	−8	−8.2601	3.6426	1.7703	1.8723	0.7677	1.1046	3.6832	1.7709	1.9123	0.7665	1.1458
	−4	−4.2721	3.2361	1.8751	1.3609	0.9880	0.3729	3.2837	1.8619	1.4218	0.9775	0.4443
	0	−0.0004	3.9157	1.8709	2.0448	0.7248	1.3200	3.9967	1.8361	2.1607	0.7065	1.4542
	4	4.0322	3.9644	2.6761	1.2883	0.9936	0.2948	4.1437	2.5415	1.6021	0.9261	0.6760
	8	7.9945	4.8461	3.6052	1.2409	0.7198	0.5212	5.1864	3.3669	1.8195	0.7112	1.1083
	Average		3.9210	2.3595	1.5615	0.8388	0.7227	4.0587	2.2755	1.7833	0.8176	0.9657

Table F.4: Attenuation, difference in attenuation, and average attenuation (dB) from the $\mathcal{WT}$=1 χ^2 energy detector, and the fitted polynomial full–linear detector, after each analysis for Observer 2 in condition $\mathcal{WT}$=1.

Experimental	Theory	Empirical	Energy Attenuation					Full–linear Attenuation				
Level	SNR (dB)	SNR (dB)	mROC	Diff	GOC	Diff	FORCE	mROC	Diff	GOC	Diff	FORCE
400ms, 2.5Hz	0	−0.5563	10.5097	2.5075	8.0021	1.2395	6.7626	10.3988	2.6131	7.7857	1.2723	6.5134
	4	4.1181	10.5642	2.6667	7.8975	2.7841	5.1134	10.2972	2.7023	7.5949	2.8013	4.7936
	8	8.3007	9.9694	2.4212	7.5482	1.6918	5.8564	9.6524	2.4281	7.2243	1.6916	5.5327
	12	11.8220	8.8971	2.5061	6.3911	1.1755	5.2155	8.5742	2.4964	6.0777	1.1670	4.9107
	16	15.6759	8.7550	2.4029	6.3520	0.7134	5.6386	8.4529	2.3722	6.0807	0.7002	5.3805
	Average		9.7391	2.5009	7.2382	1.5209	5.7173	9.4751	2.5224	6.9527	1.5265	5.4262
200ms, 5Hz	0	−0.4546	8.1020	2.2693	5.8327	2.4322	3.4006	7.8855	2.3227	5.5628	2.4641	3.0987
	4	4.0882	7.4735	2.2224	5.2511	1.3110	3.9402	7.1675	2.2356	4.9319	1.3147	3.6172
	8	7.8711	7.8092	2.1634	5.6458	1.2752	4.3705	7.4863	2.1646	5.3218	1.2729	4.0489
	12	12.0383	7.0336	2.5768	4.4568	0.9351	3.5217	6.7179	2.5564	4.1615	0.9235	3.2380
	16	16.1484	7.3009	2.7368	4.5641	1.0592	3.5050	7.0220	2.6785	4.3434	1.0248	3.3187
	Average		7.5438	2.3937	5.1501	1.4025	3.7476	7.2558	2.3916	4.8643	1.4000	3.4643
100ms, 10Hz	0	−1.2808	6.7569	2.4111	4.3458	1.5788	2.7670	6.5254	2.4603	4.0652	1.5982	2.4670
	4	4.0158	5.9565	2.1603	3.7962	0.9631	2.8331	5.6408	2.1677	3.4731	0.9642	2.5089
	8	8.1727	5.7464	2.2591	3.4873	0.7811	2.7062	5.4226	2.2526	3.1700	0.7769	2.3931
	12	11.8322	5.7734	2.2415	3.5319	0.6677	2.8642	5.4642	2.2190	3.2452	0.6581	2.5871
	16	16.0756	5.7879	2.6463	3.1416	0.4476	2.6940	5.5350	2.5687	2.9663	0.4293	2.5370
	Average		6.0042	2.3436	3.6606	0.8877	2.7729	5.7176	2.3337	3.3840	0.8853	2.4986
50ms, 20Hz	0	0.2443	5.9846	2.3960	3.5885	1.3926	2.1960	5.7057	2.4235	3.2822	1.4018	1.8804
	4	3.2400	5.6504	2.3919	3.2585	1.4434	1.8152	5.3374	2.4015	2.9359	1.4450	1.4909
	8	8.0137	5.3906	1.8491	3.5415	0.6529	2.8886	5.0671	1.8438	3.2233	0.6498	2.5736
	12	11.7777	5.1645	2.2777	2.8868	0.5617	2.3251	4.8597	2.2511	2.6085	0.5525	2.0561
	16	16.1712	5.0108	2.1753	2.8355	0.3204	2.5151	4.7785	2.1019	2.6765	0.3063	2.3703
	Average		5.4402	2.2180	3.2222	0.8742	2.3480	5.1497	2.2044	2.9453	0.8711	2.0742
25ms, 40Hz	0	0.1491	5.3062	2.4261	2.8801	2.6892	0.1909	5.0190	2.4498	2.5692	2.7009	−0.1317
	4	3.8770	4.9930	2.3653	2.6276	1.4207	1.2069	4.6736	2.3702	2.3034	1.4199	0.8835
	8	8.0293	4.0101	2.5409	1.4691	0.8857	0.5835	3.6901	2.5262	1.1639	0.8773	0.2866
	12	12.2573	4.4113	2.4700	1.9413	0.6253	1.3160	4.1189	2.4299	1.6890	0.6108	1.0781
	16	16.0475	4.8056	2.2860	2.5196	0.6105	1.9091	4.5755	2.2064	2.3691	0.5814	1.7876
	Average		4.7052	2.4177	2.2876	1.2463	1.0413	4.4154	2.3965	2.0189	1.2381	0.7808
12.5ms, 80Hz	0	0.2250	5.1478	1.9549	3.1929	0.9224	2.2705	4.8575	1.9739	2.8837	0.9283	1.9554
	4	3.8712	4.7660	2.4080	2.3580	1.4242	0.9338	4.4458	2.4121	2.0337	1.4228	0.6108
	8	7.6408	4.8141	2.1412	2.6730	0.6256	2.0474	4.4910	2.1339	2.3571	0.6220	1.7350
	12	12.0682	4.7170	2.6353	2.0816	0.6720	1.4096	4.4191	2.5966	1.8225	0.6576	1.1650
	16	16.0755	4.4871	2.7866	1.7006	0.4412	1.2594	4.2665	2.6751	1.5914	0.4164	1.1751
	Average		4.7864	2.3852	2.4012	0.8171	1.5841	4.4960	2.3583	2.1377	0.8094	1.3283
6.25ms, 160Hz	0	0.9183	5.7031	2.3260	3.3771	1.6844	1.6926	5.4111	2.3468	3.0644	1.6923	1.3720
	4	3.8949	5.5306	2.5083	3.0223	1.2204	1.8019	5.2134	2.5151	2.6983	1.2205	1.4779
	8	8.0366	6.0371	2.6488	3.3883	1.0088	2.3795	5.7129	2.6421	3.0709	1.0033	2.0676
	12	11.9083	5.5968	3.4847	2.1121	0.9938	1.1183	5.2895	3.4402	1.8493	0.9726	0.8767
	16	16.0541	6.6342	3.6307	3.0035	1.1429	1.8605	6.3646	3.5318	2.8327	1.0909	1.7419
	Average		5.9004	2.9197	2.9807	1.2101	1.7706	5.5983	2.8952	2.7031	1.1959	1.5072

Table F.5: Attenuation, difference in attenuation, and average attenuation (dB) from the $\mathcal{WT}$=3 χ^2 energy detector, and the fitted polynomial full–linear detector, after each analysis for Observer 2 in condition $\mathcal{WT}$=2.

Experimental	Theory	Empirical	Energy Attenuation					Full–linear Attenuation				
Level	SNR (dB)	SNR (dB)	mROC	Diff	GOC	Diff	FORCE	mROC	Diff	GOC	Diff	FORCE
400ms, 5Hz	−4	−4.3532	5.9078	1.6289	4.2788	0.4002	3.8787	5.0376	1.6853	3.3523	0.4110	2.9413
	0	−0.2741	7.6760	1.6812	5.9948	0.7656	5.2292	6.7316	1.7147	5.0168	0.7771	4.2398
	4	4.2112	7.0958	1.7378	5.3580	0.7199	4.6381	6.0793	1.7492	4.3301	0.7237	3.6064
	8	7.7795	6.9423	2.1917	4.7506	0.7697	3.9808	5.9045	2.2029	3.7016	0.7748	2.9268
	12	11.9325	7.0763	2.1241	4.9522	0.3070	4.6452	6.0130	2.1566	3.8564	0.3141	3.5424
	Average		6.9396	1.8728	5.0669	0.5925	4.4744	5.9532	1.9018	4.0514	0.6001	3.4513
200ms, 10Hz	−4	−4.1226	7.5022	1.3028	6.1994	0.2165	5.9828	6.6996	1.3677	5.3318	0.2255	5.1063
	0	−0.2550	6.2552	1.7357	4.5195	0.6892	3.8303	5.2813	1.7604	3.5209	0.6966	2.8243
	4	3.9492	5.0459	1.9857	3.0602	0.7985	2.2617	4.0177	1.9956	2.0221	0.8023	1.2199
	8	7.6392	5.0295	2.0952	2.9343	0.4678	2.4665	3.9830	2.1104	1.8726	0.4728	1.3998
	12	12.0730	5.3924	2.9821	2.4103	0.3073	2.1030	4.3028	3.0809	1.2219	0.3225	0.8994
	Average		5.8451	2.0203	3.8247	0.4959	3.3289	4.8569	2.0630	2.7939	0.5039	2.2899
100ms, 20Hz	−4	−3.7360	6.1763	1.6479	4.5284	0.7302	3.7983	5.2922	1.7003	3.5919	0.7476	2.8443
	0	0.2015	4.5854	1.5724	3.0131	0.4605	2.5526	3.5825	1.5865	1.9960	0.4638	1.5322
	4	3.7811	4.7672	1.6956	3.0716	0.5040	2.5676	3.7384	1.7040	2.0344	0.5063	1.5281
	8	8.0913	4.5182	2.0147	2.5035	0.4113	2.0922	3.4657	2.0341	1.4316	0.4171	1.0145
	12	11.8498	4.6152	2.4536	2.1616	0.1962	1.9654	3.5136	2.5416	0.9720	0.2058	0.7662
	Average		4.9325	1.8768	3.0556	0.4604	2.5952	3.9185	1.9133	2.0052	0.4681	1.5370
50ms, 40Hz	−4	−3.3567	3.0620	1.8396	1.2225	1.0063	0.2161	2.0865	1.8649	0.2216	1.0163	−0.7947
	0	0.6147	4.1310	2.0974	2.0335	1.0548	0.9787	3.1196	2.1124	1.0072	1.0605	−0.0533
	4	3.6819	3.6969	2.0030	1.6939	0.6961	0.9978	2.6631	2.0126	0.6506	0.6998	−0.0492
	8	7.9418	3.7277	1.9608	1.7669	0.3196	1.4473	2.6703	1.9839	0.6864	0.3252	0.3612
	12	12.0465	3.8342	3.1900	0.6442	0.2194	0.4248	2.7049	3.3354	−0.6305	0.2278	−0.8583
	Average		3.6903	2.2181	1.4722	0.6593	0.8129	2.6489	2.2618	0.3871	0.6659	−0.2788
25ms, 80Hz	−4	−4.4191	3.4733	1.8880	1.5853	0.7493	0.8360	2.5275	1.9243	0.6032	0.7598	−0.1567
	0	0.0152	3.7582	2.0182	1.7401	0.9614	0.7787	2.7490	2.0333	0.7156	0.9668	−0.2512
	4	4.0481	3.8362	1.9475	1.8887	0.6022	1.2865	2.8014	1.9569	0.8445	0.6054	0.2391
	8	7.5708	4.1818	2.0099	2.1719	0.4070	1.7650	3.1306	2.0280	1.1025	0.4124	0.6901
	12	11.8867	4.7145	2.8740	1.8405	0.3358	1.5047	3.6144	2.9813	0.6331	0.3531	0.2800
	Average		3.9928	2.1475	1.8453	0.6111	1.2342	2.9646	2.1848	0.7798	0.6195	0.1603
12.5ms, 160Hz	−4	−3.6383	6.1841	2.1022	4.0820	0.7831	3.2989	5.2966	2.1644	3.1322	0.7995	2.3327
	0	−0.3127	4.8845	1.9507	2.9338	0.7894	2.1444	3.8910	1.9708	1.9202	0.7955	1.1247
	4	4.4911	5.3470	2.0190	3.3280	0.6918	2.6362	4.3176	2.0289	2.2887	0.6951	1.5936
	8	7.8219	5.7644	2.3864	3.3780	0.6669	2.7111	4.7208	2.4021	2.3187	0.6736	1.6451
	12	11.8543	6.2678	2.9179	3.3500	0.4456	2.9044	5.1960	2.9855	2.2106	0.4628	1.7478
	Average		5.6896	2.2752	3.4143	0.6754	2.7390	4.6844	2.3103	2.3741	0.6853	1.6888

Table F.6: Attenuation, difference in attenuation, and average attenuation (dB) from the $\mathcal{WT}$=5 χ^2 energy detector, and the fitted polynomial full–linear detector, after each analysis for Observer 2 in condition $\mathcal{WT}$=4.

Experimental	Theory	Empirical	Energy Attenuation					Full–linear Attenuation				
Level	SNR (dB)	SNR (dB)	mROC	Diff	GOC	Diff	FORCE	mROC	Diff	GOC	Diff	FORCE
400ms, 10Hz	−8	−7.5580	5.9161	1.9315	3.9846	1.0095	2.9751	5.9610	1.9363	4.0247	1.0099	3.0148
	−4	−3.5357	5.9138	1.7297	4.1841	1.2744	2.9097	5.9548	1.7241	4.2307	1.2669	2.9639
	0	0.0730	5.4178	1.7301	3.6877	0.7700	2.9177	5.4815	1.7082	3.7733	0.7566	3.0167
	4	3.8845	4.9994	1.9408	3.0587	0.6659	2.3927	5.1395	1.8703	3.2691	0.6342	2.6349
	8	8.0386	5.2121	2.7381	2.4740	0.6611	1.8129	5.5312	2.5418	2.9894	0.6204	2.3690
	Average		5.4919	2.0140	3.4778	0.8762	2.6016	5.6136	1.9561	3.6574	0.8576	2.7998
200ms, 20Hz	−8	−7.6082	3.9135	1.3158	2.5977	0.5535	2.0442	3.9535	1.3159	2.6377	0.5529	2.0848
	−4	−3.8392	4.2124	1.6099	2.6026	0.7952	1.8074	4.2576	1.6008	2.6568	0.7887	1.8681
	0	−0.1545	3.6088	1.5814	2.0274	0.5588	1.4686	3.6921	1.5520	2.1401	0.5454	1.5948
	4	4.1558	3.8213	1.8178	2.0035	0.4727	1.5308	4.0110	1.7294	2.2816	0.4444	1.8372
	8	8.0789	4.3438	2.9868	1.3570	0.5321	0.8249	4.7249	2.7903	1.9346	0.5220	1.4126
	Average		3.9800	1.8623	2.1176	0.5825	1.5352	4.1278	1.7977	2.3301	0.5706	1.7595
100ms, 40Hz	−8	−7.5739	1.4024	1.6427	−0.2403	0.6223	−0.8626	1.4445	1.6362	−0.1917	0.6185	−0.8103
	−4	−4.3887	2.1002	1.6557	0.4445	0.7490	−0.3045	2.1541	1.6405	0.5136	0.7395	−0.2259
	0	0.0579	2.5570	1.4646	1.0924	0.4408	0.6516	2.6629	1.4281	1.2348	0.4271	0.8076
	4	3.8417	2.6517	1.7661	0.8857	0.3630	0.5227	2.8790	1.6658	1.2132	0.3386	0.8746
	8	7.9003	3.0581	3.2603	−0.2022	0.4634	−0.6656	3.5214	3.1522	0.3692	0.4792	−0.1100
	Average		2.3539	1.9579	0.3960	0.5277	−0.1317	2.5324	1.9046	0.6278	0.5206	0.1072
50ms, 80Hz	−8	−7.7795	2.0100	1.6804	0.3296	0.5629	−0.2333	2.0504	1.6759	0.3745	0.5603	−0.1858
	−4	−4.5292	2.7707	1.6094	1.1612	0.6397	0.5216	2.8194	1.5979	1.2216	0.6333	0.5883
	0	−0.1683	2.4365	1.7542	0.6823	0.5881	0.0942	2.5402	1.7100	0.8302	0.5687	0.2614
	4	3.8162	2.8596	1.9925	0.8671	0.4656	0.4015	3.0759	1.8818	1.1941	0.4342	0.7600
	8	8.0010	3.4966	2.3577	1.1389	0.1794	0.9595	3.9345	2.2139	1.7206	0.1756	1.5450
	Average		2.7147	1.8788	0.8358	0.4871	0.3487	2.8841	1.8159	1.0682	0.4744	0.5938
25ms, 160Hz	−8	−8.2601	3.9407	1.9948	1.9460	2.3830	−0.4370	3.9818	1.9959	1.9859	2.3768	−0.3909
	−4	−4.2721	4.0735	1.6899	2.3836	0.8906	1.4930	4.1175	1.6812	2.4363	0.8836	1.5527
	0	−0.0004	4.0093	1.8711	2.1382	0.7548	1.3834	4.0890	1.8371	2.2519	0.7360	1.5159
	4	4.0322	4.4251	2.1523	2.2728	0.6623	1.6105	4.5878	2.0588	2.5290	0.6246	1.9044
	8	7.9945	5.0259	2.5003	2.5256	0.4239	2.1017	5.3542	2.3198	3.0344	0.3959	2.6386
	Average		4.2949	2.0417	2.2532	1.0229	1.2303	4.4261	1.9785	2.4475	1.0034	1.4441

Appendix G

Correlation analysis

This appendix presents the additional figures and tables from the Bester correlation method analyses of Chapter 5.

Correlation–bandwidth functions: Figures G.1–G.6 show the correlation–bandwidth functions for Observer 1 and Observer 2 in each experimental level, based on both the energy detector and the full–linear detector.

Correlations of six different detectors: Figures G.7 and G.8 show bar graphs of the correlations between six detectors and each observer, in each experimental level. The detectors are (a) the energy detector with the best correlated bandwidth, sampled at the end of the integrator [Energy–EOI], (b) the same energy detector, but with the best correlated sampling time—this is usually the same as the end of the integrator, but small statistical fluctuations may favour another time [Energy–ATT], (c) the full–linear detector with the best correlated bandwidth, sampled at the end of the integrator [Full–EOI], (d) the same full–linear detector, but with the best correlated sampling time [Full–ATT], (e) the envelope detector with the best correlated time constant (usually 1000 ms), sampled with a peak detector [Env.–Peak], and (f) the envelope detector with the best correlated time constant *and* sample time [Env.–ATT].

Signal analysis of best correlated full–linear simulation: Tables G.1–G.6 report the estimated $\mathcal{W}$, $\mathcal{T}$, $\mathcal{WT}$, and signal–to–noise ratio, from the signal analysis for the best correlated full–linear detector, for each observer and each condition.

Best correlated full–linear detector psychometric functions: Figures G.9 and G.10 show the best correlated full–linear detector psychometric functions compared to the human psychometric functions after GOC and FORCE analysis.

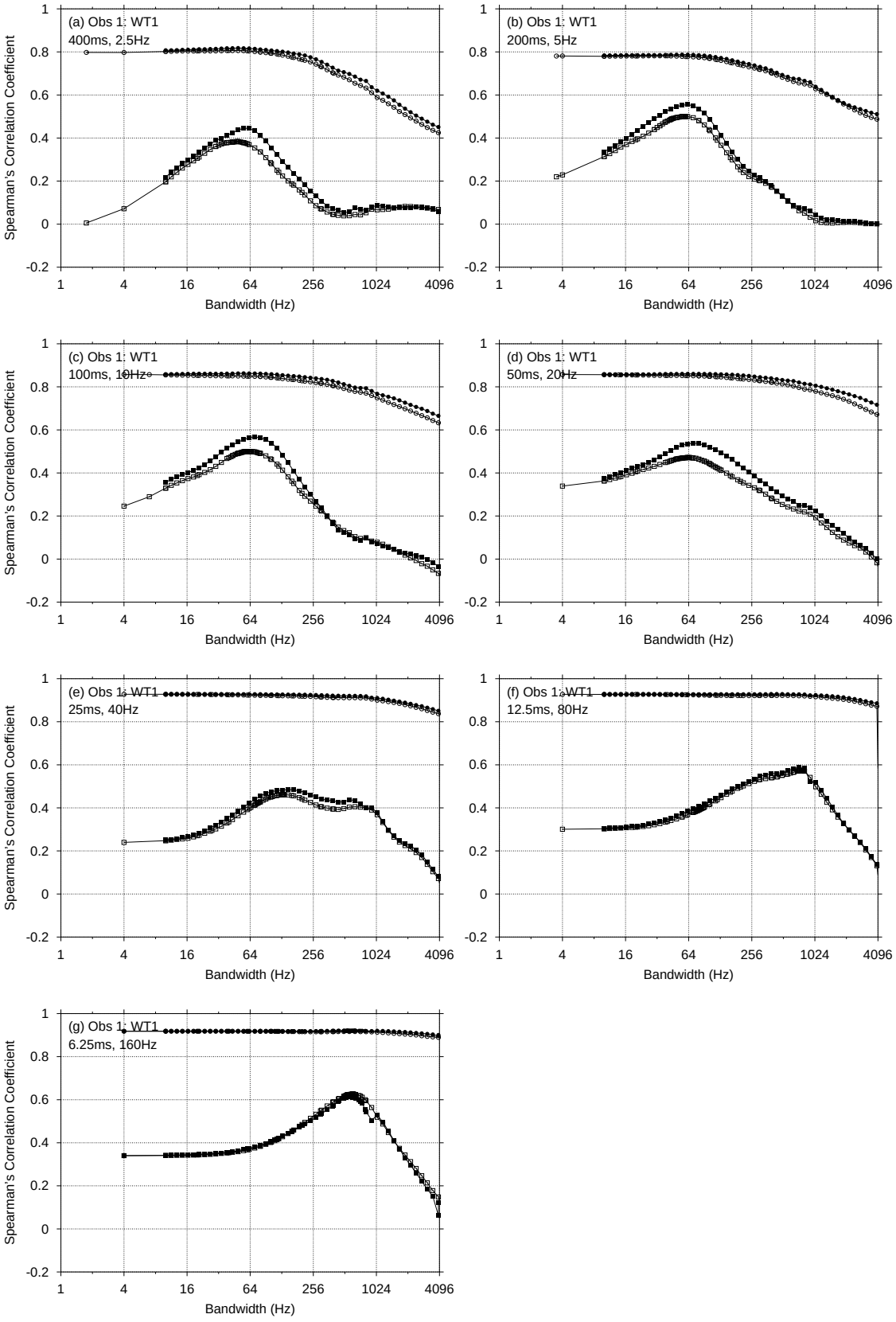

Figure G.1: Correlation–bandwidth functions of Observer 1 based on the energy detector simulation (open symbols) and the full–linear detector simulation (closed symbols), in condition $\mathcal{WT}$=1, for the $\mathcal{N}$ event (square symbols) and largest $\mathcal{SN}$ event (circle symbols).

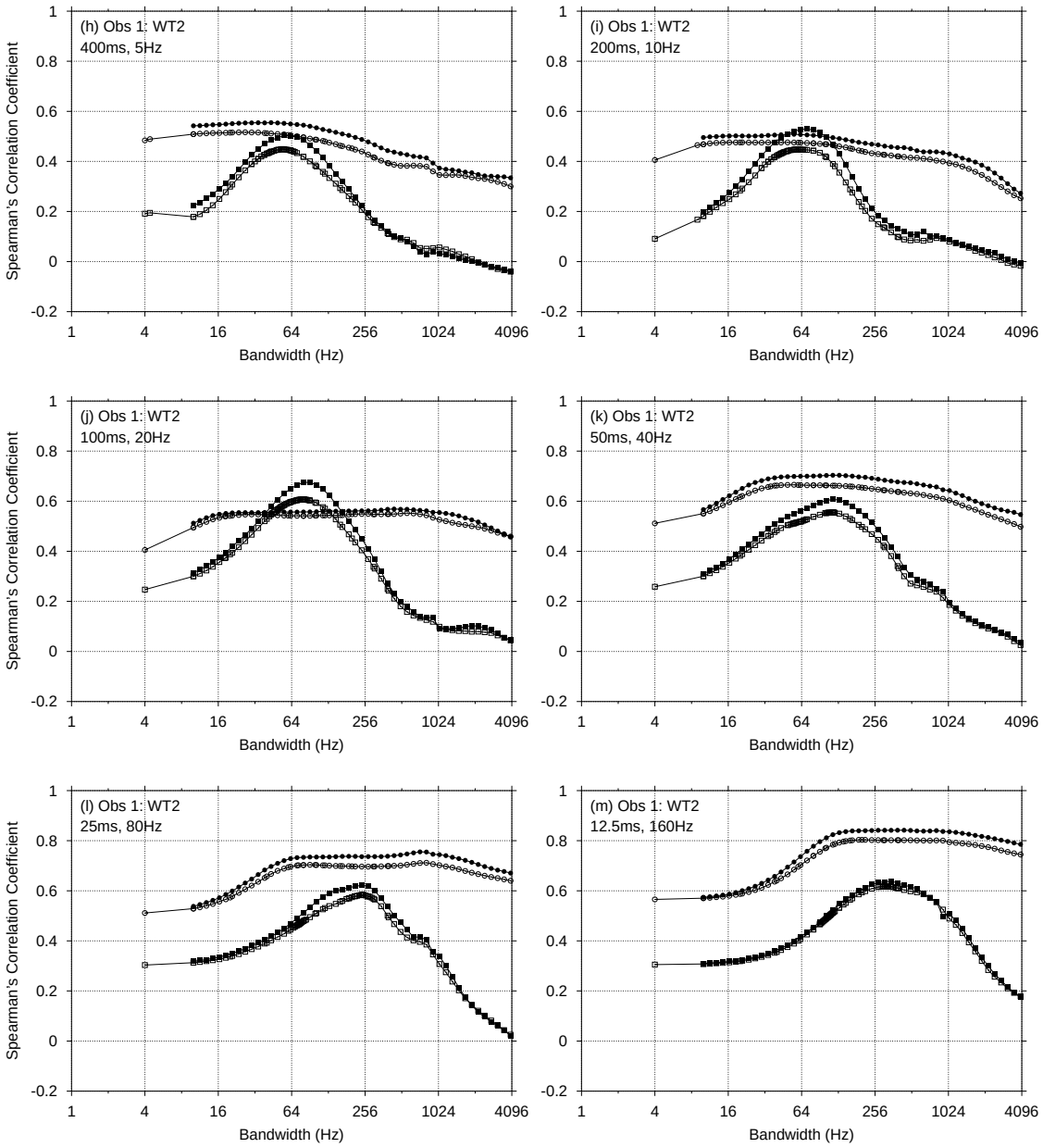

Figure G.2: Correlation–bandwidth functions of Observer 1 based on the energy detector simulation (open symbols) and full–linear detector simulation (closed symbols), in condition $\mathcal{WT}$=2, for the $\mathcal{N}$ event (square symbols) and largest $\mathcal{SN}$ event (circle symbols).

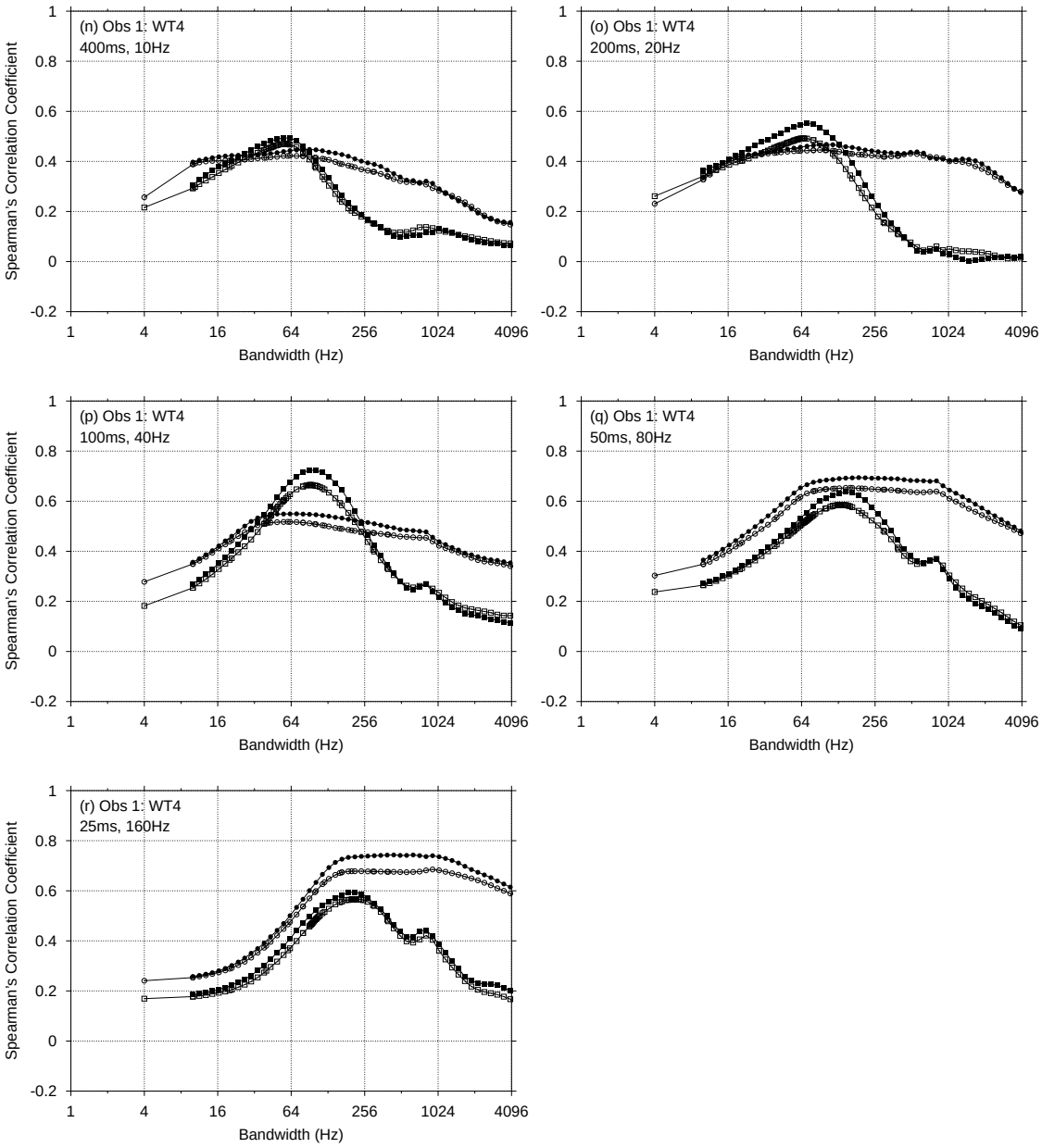

Figure G.3: Correlation–bandwidth functions of Observer 1 based on the energy detector simulation (open symbols) and full–linear detector simulation (closed symbols), in condition $\mathcal{WT}$=4, for the $\mathcal{N}$ event (square symbols) and largest $\mathcal{SN}$ event (circle symbols).

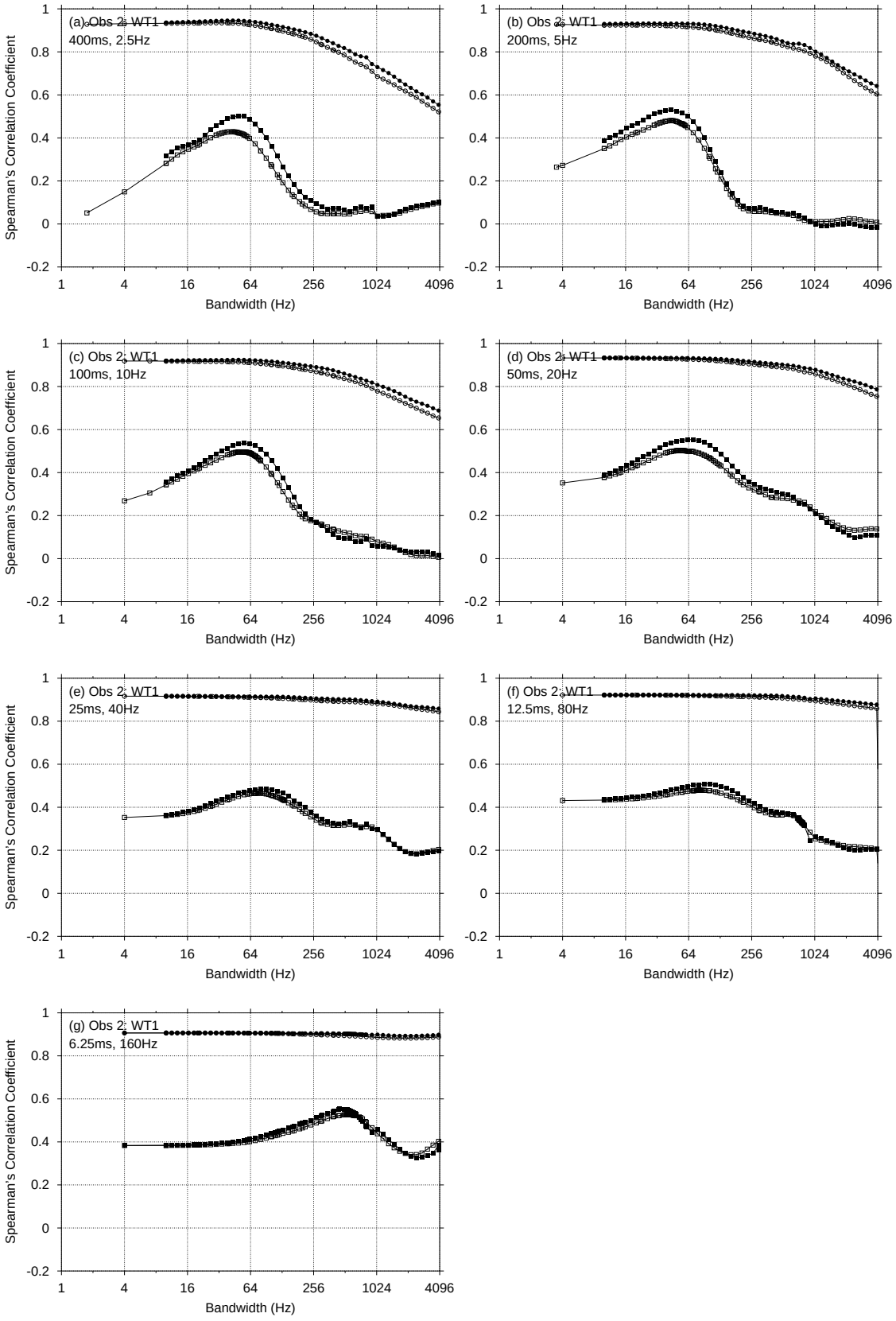

Figure G.4: Correlation–bandwidth functions of Observer 2 based on the energy detector simulation (open symbols) and full–linear detector simulation (closed symbols), in condition $\mathcal{WT}$=1, for the $\mathcal{N}$ event (square symbols) and largest $\mathcal{SN}$ event (circle symbols).

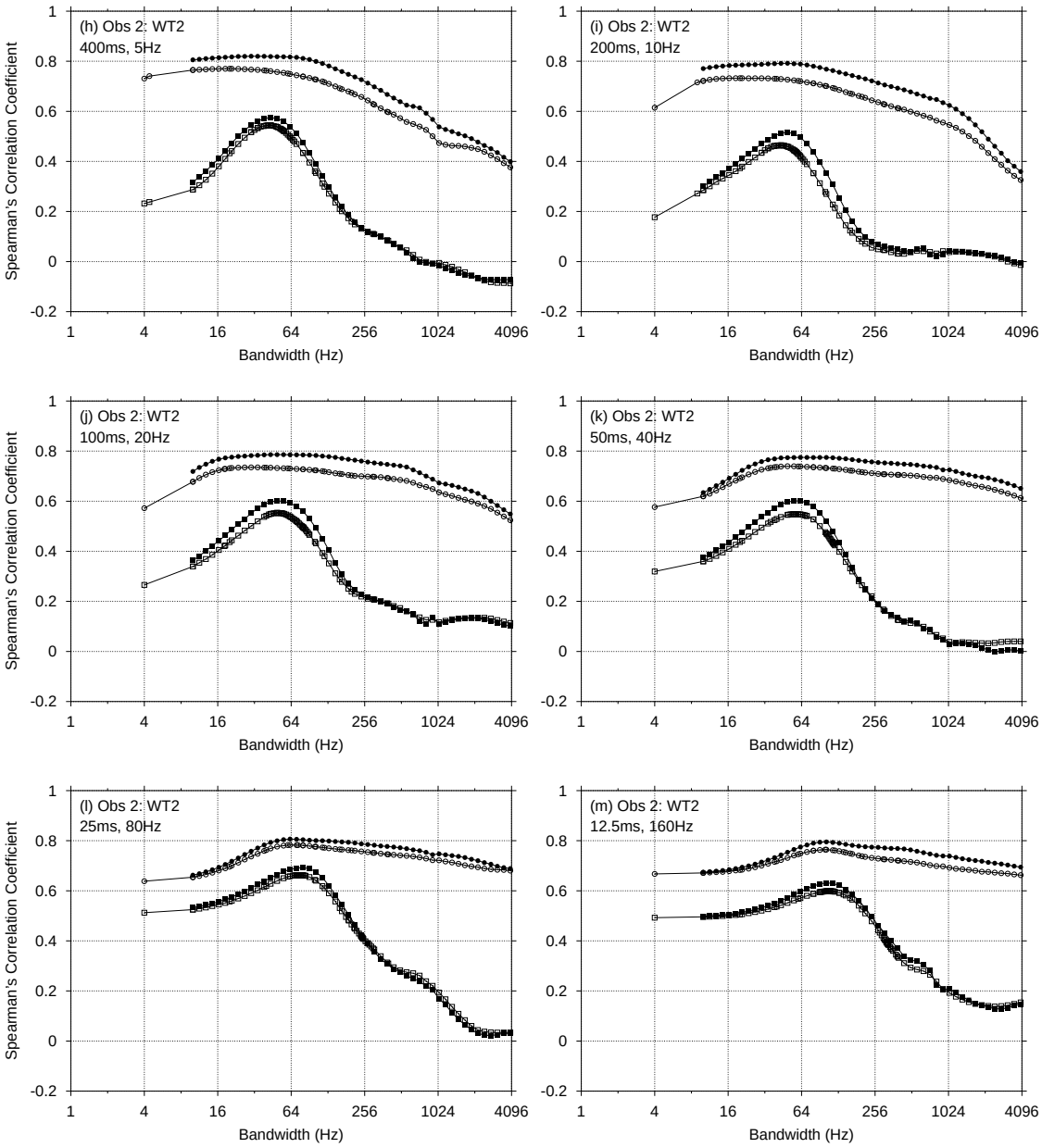

Figure G.5: Correlation–bandwidth functions of Observer 2 based on the energy detector simulation (open symbols) and full–linear detector simulation (closed symbols), in condition $\mathcal{WT}$=2, for the $\mathcal{N}$ event (square symbols) and largest $\mathcal{SN}$ event (circle symbols).

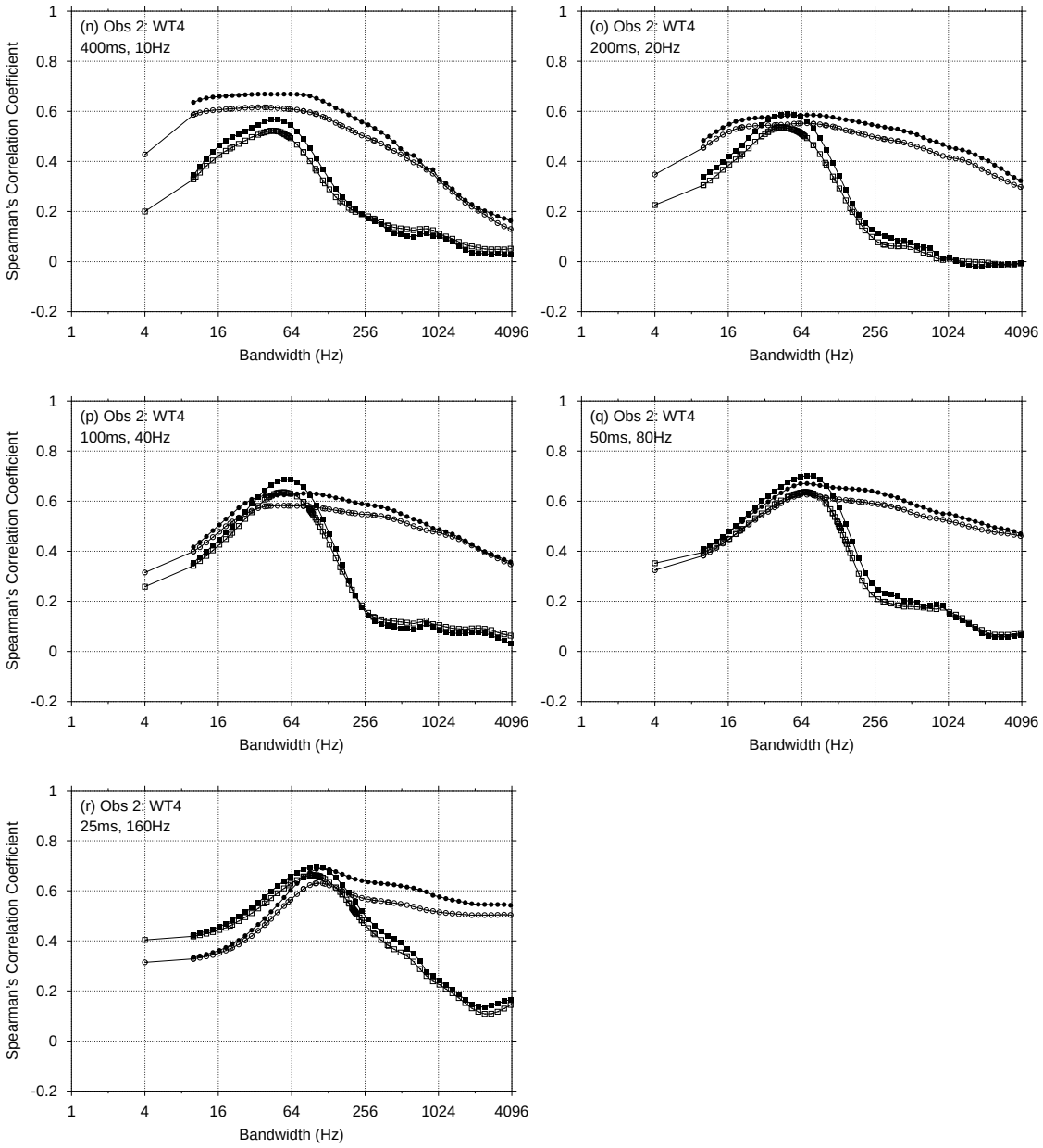

Figure G.6: Correlation–bandwidth functions of Observer 2 based on the energy detector simulation (open symbols) and full–linear detector simulation (closed symbols), in condition $\mathcal{WT}$=4, for the $\mathcal{N}$ event (square symbols) and largest $\mathcal{SN}$ event (circle symbols).

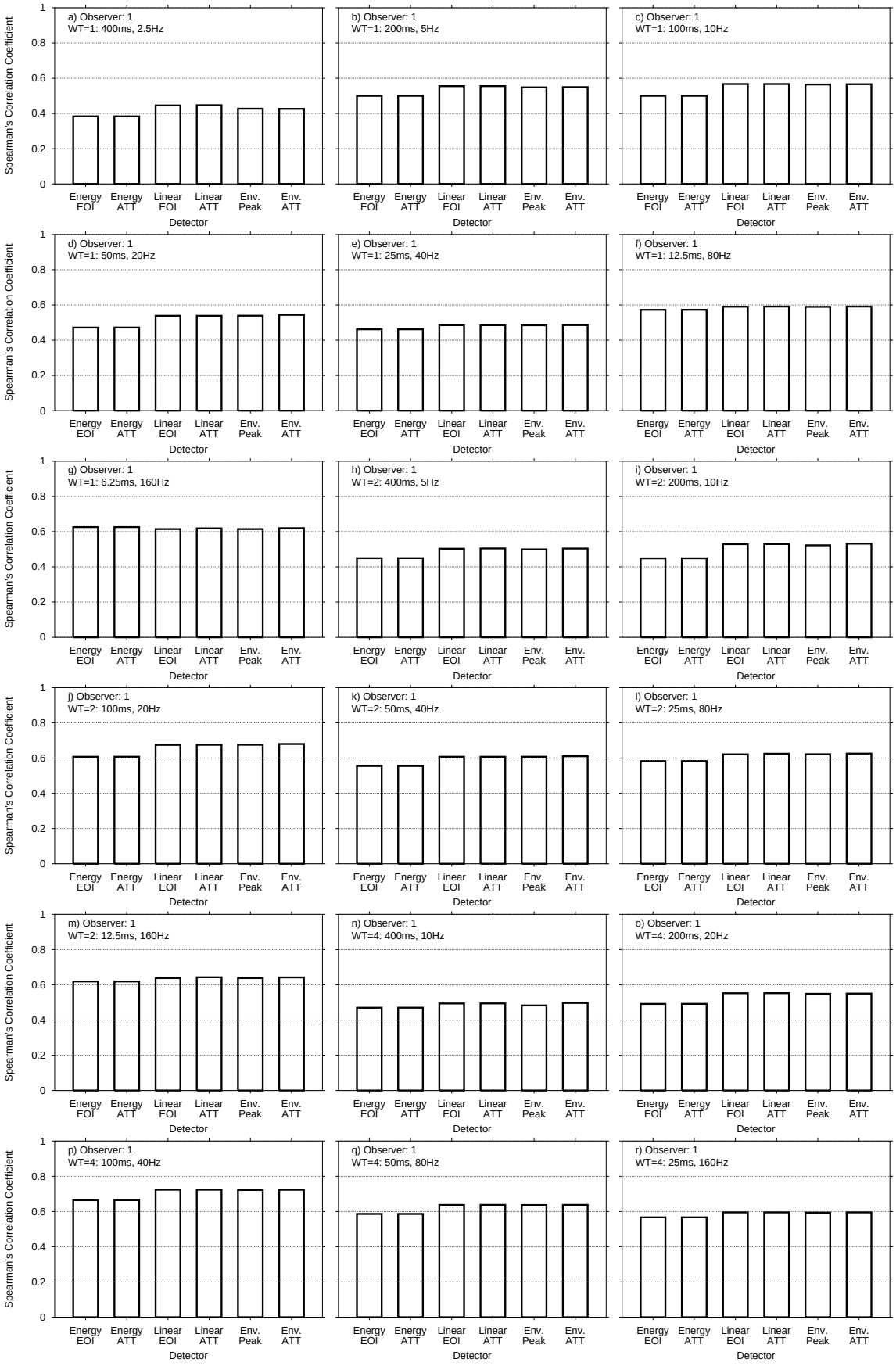

Figure G.7: The maximum Spearman's correlation coefficient for the relationship between the output of six different detectors and the sum–of–ratings from Observer 1 for all experimental levels.

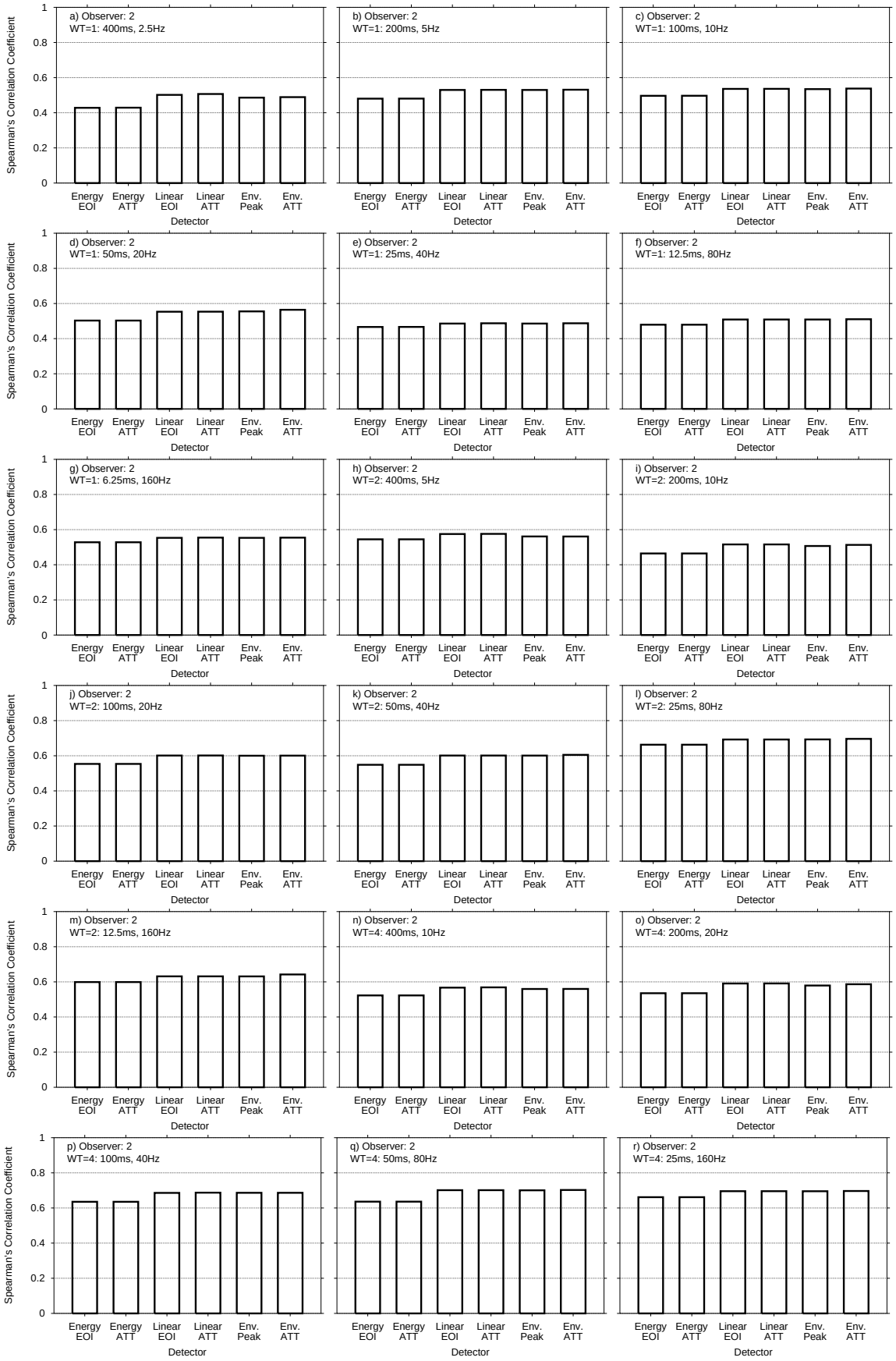

Figure G.8: The maximum Spearman's correlation coefficient for the relationship between the output of six different detectors and the sum–of–ratings from Observer 2 for all experimental levels.

Table G.1: Estimated parameters from the best–correlated full–linear simulation for Observer 1 in condition $\mathcal{WT}$=1.

Signal Set	SNR		$\mathrm{Ess}\mathcal{WT}_{92.4\%}$			$\mathcal{WT}_{\mathrm{AERB}\times\mathrm{ERD}}$			$\mathrm{Ess}\mathcal{WT}_{95\%}$			$\mathcal{WT}_{\mathrm{TdB}\times\mathrm{ERD}}$		
	Orig.	Est.	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 2.5Hz	$-\infty$	$-\infty$	64.87	402.45	26.11	63.55	368.12	23.40	71.29	442.65	31.56	53.88	368.12	19.84
	0	0.306	64.24	398.16	25.58	31.88	376.43	12.00	70.68	439.04	31.03	32.68	376.43	12.30
	4	4.684	63.54	399.41	25.38	17.48	359.46	6.29	70.10	440.40	30.87	2.23	359.46	0.80
	8	8.776	61.69	399.18	24.63	9.01	367.87	3.31	68.11	439.44	29.93	1.89	367.87	0.70
	12	12.270	58.31	399.60	23.30	5.32	374.99	1.99	64.81	440.46	28.55	1.74	374.99	0.65
	16	16.102	52.44	400.40	21.00	3.34	379.96	1.27	59.02	439.84	25.96	1.69	379.96	0.64
200ms, 5Hz	$-\infty$	$-\infty$	73.15	200.54	14.67	62.53	186.81	11.68	80.68	219.74	17.73	60.19	186.81	11.24
	0	−0.930	71.90	201.32	14.47	36.02	191.80	6.91	78.96	221.05	17.45	22.60	191.80	4.33
	4	3.773	70.32	200.78	14.12	20.99	187.95	3.94	77.53	221.25	17.15	4.59	187.95	0.86
	8	7.616	68.12	199.96	13.62	12.42	188.67	2.34	75.17	220.18	16.55	3.92	188.67	0.74
	12	11.810	61.78	199.70	12.34	7.34	186.16	1.37	68.88	219.20	15.10	3.60	186.16	0.67
	16	15.926	50.68	199.69	10.12	5.17	191.37	0.99	59.02	220.15	12.99	3.51	191.37	0.67
100ms, 10Hz	$-\infty$	$-\infty$	82.45	102.25	8.43	72.81	93.08	6.78	90.93	111.88	10.17	69.62	93.08	6.48
	0	−1.330	81.42	100.79	8.21	45.49	95.09	4.33	89.62	110.50	9.90	49.15	95.09	4.67
	4	4.000	77.96	100.25	7.82	26.76	94.47	2.53	86.38	110.22	9.52	9.06	94.47	0.86
	8	8.144	71.98	101.28	7.29	16.07	98.32	1.58	80.19	111.34	8.93	7.74	98.32	0.76
	12	11.816	62.49	100.28	6.27	11.46	95.98	1.10	71.69	110.25	7.90	7.31	95.98	0.70
	16	16.055	45.93	99.94	4.59	8.95	96.93	0.87	57.26	109.71	6.28	7.08	96.93	0.69
50ms, 20Hz	$-\infty$	$-\infty$	81.32	51.02	4.15	70.49	49.09	3.46	89.57	56.39	5.05	68.70	49.09	3.37
	0	−0.430	78.53	51.62	4.05	43.23	49.18	2.13	86.82	56.99	4.95	28.39	49.18	1.40
	4	2.769	76.06	51.84	3.94	33.25	50.03	1.66	84.65	56.95	4.82	18.47	50.03	0.92
	8	7.633	66.30	50.72	3.36	22.94	48.74	1.12	74.90	55.85	4.18	15.45	48.74	0.75
	12	11.429	52.35	50.93	2.67	18.13	50.79	0.92	62.76	56.01	3.52	14.44	50.79	0.73
	16	15.828	28.48	50.37	1.43	15.95	48.80	0.78	40.45	55.25	2.23	14.03	48.80	0.68
25ms, 40Hz	$-\infty$	$-\infty$	172.91	26.16	4.52	153.76	25.55	3.93	190.15	28.82	5.48	145.99	25.55	3.73
	0	0.280	164.67	25.49	4.20	91.75	24.89	2.28	181.57	28.16	5.11	73.90	24.89	1.84
	4	3.978	154.86	25.58	3.96	65.29	24.96	1.63	172.35	28.16	4.85	36.89	24.96	0.92
	8	8.114	137.37	25.48	3.50	46.34	24.66	1.14	155.19	28.09	4.36	30.67	24.66	0.76
	12	12.351	106.19	25.04	2.66	36.16	24.81	0.90	127.75	27.59	3.53	28.93	24.81	0.72
	16	16.131	60.45	25.13	1.52	32.20	24.55	0.79	88.52	27.64	2.45	28.15	24.55	0.69
12.5ms, 80Hz	$-\infty$	$-\infty$	841.92	12.80	10.78	728.40	12.63	9.20	923.49	14.17	13.09	686.56	12.63	8.67
	0	−0.103	823.83	12.76	10.51	400.37	12.10	4.84	903.89	14.12	12.76	125.24	12.10	1.52
	4	3.651	813.74	12.88	10.48	256.35	12.54	3.21	892.54	14.22	12.69	70.87	12.54	0.89
	8	7.458	765.34	12.77	9.77	159.32	12.58	2.00	850.51	14.05	11.95	61.69	12.58	0.78
	12	11.915	678.50	12.64	8.58	99.38	12.32	1.22	762.60	13.85	10.56	58.08	12.32	0.72
	16	15.923	535.45	12.58	6.74	75.50	12.38	0.93	636.16	13.85	8.81	56.11	12.38	0.69
6.25ms, 160Hz	$-\infty$	$-\infty$	657.04	6.87	4.51	570.05	6.93	3.95	721.52	7.59	5.48	569.11	6.93	3.94
	0	0.639	621.88	6.74	4.19	334.25	6.76	2.26	686.81	7.44	5.11	208.84	6.76	1.41
	4	3.708	593.61	6.66	3.96	255.06	6.81	1.74	657.56	7.35	4.83	143.91	6.81	0.98
	8	7.876	525.13	6.52	3.42	180.48	6.41	1.16	591.36	7.18	4.25	122.20	6.41	0.78
	12	11.796	418.07	6.41	2.68	143.65	6.42	0.92	505.21	7.07	3.57	114.80	6.42	0.74
	16	15.909	233.06	6.34	1.48	127.49	6.21	0.79	335.15	6.98	2.34	111.83	6.21	0.69

Table G.2: Estimated parameters from the best–correlated full–linear simulation for Observer 1 in condition $\mathcal{WT}$=2.

Signal Set	SNR		$\mathrm{Ess}\mathcal{WT}_{92.4\%}$			$\mathcal{WT}_{\mathrm{AERB}\times\mathrm{ERD}}$			$\mathrm{Ess}\mathcal{WT}_{95\%}$			$\mathcal{WT}_{\mathrm{TdB}\times\mathrm{ERD}}$		
	Orig.	Est.	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 5Hz	$-\infty$	$-\infty$	64.66	398.37	25.76	57.43	373.47	21.45	71.29	438.48	31.26	55.16	373.47	20.60
	-4	-4.382	63.51	398.82	25.33	44.25	365.54	16.17	70.08	437.80	30.68	47.70	365.54	17.43
	0	-0.369	63.74	400.30	25.52	31.99	369.06	11.81	70.34	439.89	30.94	7.80	369.06	2.88
	4	4.152	61.60	397.65	24.50	19.32	359.89	6.95	68.08	438.03	29.82	5.08	359.89	1.83
	8	7.709	58.85	399.18	23.49	12.21	372.08	4.54	65.19	439.30	28.64	4.63	372.08	1.72
	12	11.892	52.25	400.16	20.91	7.73	366.81	2.83	58.87	441.18	25.97	4.51	366.81	1.66
200ms, 10Hz	$-\infty$	$-\infty$	82.42	200.79	16.55	75.27	183.16	13.79	90.98	220.25	20.04	69.02	183.16	12.64
	-4	-4.021	81.82	201.65	16.50	55.58	180.91	10.06	90.29	220.17	19.88	60.70	180.91	10.98
	0	-0.166	80.56	199.57	16.08	42.84	188.77	8.09	88.76	220.06	19.53	25.66	188.77	4.84
	4	4.000	77.04	201.26	15.51	27.58	195.41	5.39	85.50	221.34	18.92	10.31	195.41	2.02
	8	7.689	71.71	200.64	14.39	18.83	182.78	3.44	80.12	220.87	17.70	9.64	182.78	1.76
	12	12.105	60.30	199.81	12.05	12.95	186.99	2.42	69.35	219.65	15.23	8.96	186.99	1.67
100ms, 20Hz	$-\infty$	$-\infty$	104.86	101.32	10.62	96.01	98.36	9.44	115.44	111.33	12.85	86.45	98.36	8.50
	-4	-3.516	103.24	100.81	10.41	71.25	94.90	6.76	113.75	110.75	12.60	77.11	94.90	7.32
	0	0.304	100.12	100.93	10.11	56.23	94.55	5.32	110.97	110.68	12.28	31.63	94.55	2.99
	4	3.860	95.96	100.82	9.67	40.81	97.35	3.97	106.72	110.63	11.81	20.94	97.35	2.04
	8	8.142	84.16	100.68	8.47	28.50	99.54	2.84	95.12	110.50	10.51	18.99	99.54	1.89
	12	11.921	66.20	100.16	6.63	23.37	95.79	2.24	78.91	110.17	8.69	18.70	95.79	1.79
50ms, 40Hz	$-\infty$	$-\infty$	135.11	51.14	6.91	126.68	49.69	6.30	148.47	56.08	8.33	117.25	49.69	5.83
	-4	-2.597	131.14	50.47	6.62	93.44	49.48	4.62	145.13	55.55	8.06	91.92	49.48	4.55
	0	1.040	125.96	50.48	6.36	74.91	48.22	3.61	139.16	55.62	7.74	51.15	48.22	2.47
	4	4.027	117.57	50.24	5.91	61.91	51.02	3.16	130.90	55.13	7.22	42.10	51.02	2.15
	8	8.235	98.46	50.00	4.92	48.68	48.30	2.35	113.05	54.88	6.20	39.24	48.30	1.90
	12	12.304	68.07	49.92	3.40	39.77	48.47	1.93	88.30	54.83	4.84	35.62	48.47	1.73
25ms, 80Hz	$-\infty$	$-\infty$	281.06	25.44	7.15	246.18	23.86	5.87	309.57	28.06	8.69	230.20	23.86	5.49
	-4	-4.742	272.71	25.55	6.97	205.82	25.08	5.16	301.16	28.12	8.47	209.12	25.08	5.24
	0	-0.183	263.57	25.27	6.66	166.12	24.04	3.99	290.99	27.90	8.12	104.30	24.04	2.51
	4	3.965	241.74	25.03	6.05	124.55	24.01	2.99	270.66	27.59	7.47	85.19	24.01	2.05
	8	7.476	210.52	25.50	5.37	102.70	24.55	2.52	239.31	27.95	6.69	79.14	24.55	1.94
	12	11.788	148.62	25.07	3.73	82.54	24.13	1.99	188.34	27.51	5.18	72.96	24.13	1.76
12.5ms, 160Hz	$-\infty$	$-\infty$	404.39	12.96	5.24	366.79	12.05	4.42	444.03	14.14	6.28	341.03	12.05	4.11
	-4	-3.561	391.12	12.99	5.08	289.89	12.69	3.68	431.90	14.23	6.15	298.16	12.69	3.79
	0	-0.233	371.84	13.03	4.84	255.20	12.50	3.19	411.82	14.31	5.89	205.66	12.50	2.57
	4	4.588	330.68	12.81	4.24	191.56	12.39	2.37	373.28	14.06	5.25	155.54	12.39	1.93
	8	7.851	275.53	12.74	3.51	171.69	12.30	2.11	322.35	14.06	4.53	149.61	12.30	1.84
	12	11.897	201.77	12.55	2.53	156.42	12.16	1.90	242.69	13.80	3.35	144.86	12.16	1.76

Table G.3: Estimated parameters from the best–correlated full–linear simulation for Observer 1 in condition $\mathcal{WT}$=4.

Signal Set	SNR		Ess$\mathcal{WT}_{92.4\%}$			$\mathcal{WT}_{\text{AERB}\times\text{ERD}}$			Ess$\mathcal{WT}_{95\%}$			$\mathcal{WT}_{\text{TdB}\times\text{ERD}}$		
	Orig.	Est.	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 10Hz	$-\infty$	$-\infty$	64.41	404.01	26.02	56.79	378.03	21.47	70.96	443.76	31.49	53.44	378.03	20.20
	−8	−8.383	63.88	400.48	25.58	51.27	370.73	19.01	70.48	439.85	31.00	50.24	370.73	18.62
	−4	−3.898	63.66	400.34	25.49	43.63	366.87	16.01	70.04	440.00	30.82	45.76	366.87	16.79
	0	−0.155	62.13	398.45	24.76	34.05	371.58	12.65	68.51	439.04	30.08	13.06	371.58	4.85
	4	3.716	58.81	399.66	23.51	24.46	377.13	9.23	65.24	439.25	28.66	11.26	377.13	4.25
	8	7.915	52.29	396.18	20.71	16.73	374.18	6.26	59.15	435.67	25.77	10.74	374.18	4.02
200ms, 20Hz	$-\infty$	$-\infty$	82.36	200.71	16.53	71.81	188.08	13.51	90.52	219.74	19.89	67.12	188.08	12.62
	−8	−8.245	81.58	200.89	16.39	65.63	188.94	12.40	89.62	220.93	19.80	63.40	188.94	11.98
	−4	−4.142	80.21	200.41	16.07	57.79	184.37	10.65	88.43	220.34	19.48	60.31	184.37	11.12
	0	−0.345	77.49	201.86	15.64	48.07	188.33	9.05	85.70	221.34	18.97	25.28	188.33	4.76
	4	4.030	72.06	199.62	14.39	35.33	184.54	6.52	79.99	219.05	17.52	22.00	184.54	4.06
	8	7.976	62.07	198.60	12.33	27.81	182.00	5.06	71.03	219.36	15.58	20.95	182.00	3.81
100ms, 40Hz	$-\infty$	$-\infty$	105.31	101.08	10.64	96.43	95.67	9.22	115.99	111.24	12.90	89.03	95.67	8.52
	−8	−7.029	102.60	100.40	10.30	87.30	94.92	8.29	113.34	110.37	12.51	81.22	94.92	7.71
	−4	−4.080	101.30	101.18	10.25	79.03	96.69	7.64	112.18	111.57	12.52	76.76	96.69	7.42
	0	0.204	96.14	100.21	9.63	67.31	96.11	6.47	107.10	110.41	11.82	49.11	96.11	4.72
	4	3.951	86.58	99.92	8.65	56.94	94.07	5.36	97.36	109.77	10.69	44.38	94.07	4.18
	8	8.001	69.85	99.92	6.98	49.40	94.39	4.66	81.95	110.06	9.02	42.74	94.39	4.03
50ms, 80Hz	$-\infty$	$-\infty$	172.45	50.59	8.72	153.44	47.61	7.30	190.08	55.80	10.61	142.74	47.61	6.80
	−8	−7.228	168.89	50.75	8.57	142.31	47.31	6.73	186.05	55.70	10.36	137.17	47.31	6.49
	−4	−4.266	165.53	50.55	8.37	136.05	48.33	6.58	182.60	55.43	10.12	128.71	48.33	6.22
	0	−0.026	153.38	50.29	7.71	119.37	46.73	5.58	170.88	55.34	9.46	97.55	46.73	4.56
	4	3.924	138.22	50.09	6.92	102.93	47.54	4.89	156.59	55.20	8.64	87.06	47.54	4.14
	8	8.075	105.40	50.43	5.32	92.50	47.64	4.41	126.32	55.15	6.97	83.63	47.64	3.98
25ms, 160Hz	$-\infty$	$-\infty$	247.04	25.97	6.42	225.26	24.66	5.56	272.91	28.51	7.78	211.42	24.66	5.21
	−8	−7.344	240.71	25.21	6.07	218.72	24.27	5.31	265.15	27.66	7.33	200.83	24.27	4.87
	−4	−3.814	234.19	25.51	5.97	206.53	23.91	4.94	259.08	27.97	7.25	188.53	23.91	4.51
	0	0.257	213.84	25.33	5.42	190.92	24.90	4.75	238.65	27.89	6.66	174.99	24.90	4.36
	4	4.208	190.58	25.57	4.87	178.02	24.11	4.29	213.11	28.06	5.98	166.21	24.11	4.01
	8	8.137	169.17	25.43	4.30	167.98	24.09	4.05	182.36	27.91	5.09	162.54	24.09	3.92

Table G.4: Estimated parameters from the best–correlated full–linear simulation for Observer 2 in condition $\mathcal{WT}$=1.

Signal Set	SNR		Ess$\mathcal{WT}_{92.4\%}$			$\mathcal{WT}_{\text{AERB}\times\text{ERD}}$			Ess$\mathcal{WT}_{95\%}$			$\mathcal{WT}_{\text{TdB}\times\text{ERD}}$		
	Orig.	Est.	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 2.5Hz	$-\infty$	$-\infty$	57.33	402.92	23.10	56.23	367.27	20.65	63.22	443.36	28.03	46.67	367.27	17.14
	0	0.306	56.73	398.43	22.60	28.30	376.45	10.65	62.63	439.23	27.51	30.87	376.45	11.62
	4	4.684	56.08	399.67	22.41	15.61	361.37	5.64	61.86	440.67	27.26	2.23	361.37	0.81
	8	8.776	54.23	399.32	21.65	8.15	368.58	3.01	60.01	439.58	26.38	1.89	368.58	0.70
	12	12.270	50.99	399.69	20.38	4.91	376.45	1.85	56.73	440.59	24.99	1.74	376.45	0.66
	16	16.102	45.56	400.58	18.25	3.16	382.02	1.21	51.38	439.91	22.60	1.69	382.02	0.65
200ms, 5Hz	$-\infty$	$-\infty$	50.42	201.52	10.16	43.42	183.78	7.98	55.47	220.56	12.23	42.55	183.78	7.82
	0	−0.930	50.05	202.33	10.13	25.42	192.49	4.89	55.20	222.53	12.28	21.45	192.49	4.13
	4	3.773	48.29	201.58	9.73	15.32	192.71	2.95	53.42	222.08	11.86	4.59	192.71	0.88
	8	7.616	45.44	200.15	9.09	9.59	191.56	1.84	50.53	220.64	11.15	3.92	191.56	0.75
	12	11.810	40.62	200.13	8.13	6.14	189.63	1.16	45.72	219.71	10.04	3.60	189.63	0.68
	16	15.926	30.96	199.76	6.18	4.69	192.95	0.90	37.26	220.17	8.20	3.51	192.95	0.68
100ms, 10Hz	$-\infty$	$-\infty$	64.66	102.68	6.64	56.92	93.54	5.32	71.24	112.55	8.02	54.89	93.54	5.13
	0	−1.330	63.09	101.13	6.38	36.35	95.41	3.47	69.72	110.99	7.74	39.74	95.41	3.79
	4	4.000	60.44	100.80	6.09	22.09	95.95	2.12	66.84	110.77	7.40	9.06	95.95	0.87
	8	8.144	55.37	101.59	5.62	13.96	99.54	1.39	61.73	111.71	6.90	7.74	99.54	0.77
	12	11.816	46.75	100.53	4.70	10.51	97.20	1.02	53.51	110.54	5.91	7.31	97.20	0.71
	16	16.055	30.62	99.97	3.06	8.57	97.49	0.84	40.82	109.78	4.48	7.08	97.49	0.69
50ms, 20Hz	$-\infty$	$-\infty$	81.32	51.02	4.15	70.49	49.09	3.46	89.57	56.39	5.05	68.70	49.09	3.37
	0	−0.430	78.53	51.62	4.05	43.23	49.18	2.13	86.82	56.99	4.95	28.39	49.18	1.40
	4	2.769	76.06	51.84	3.94	33.25	50.03	1.66	84.65	56.95	4.82	18.47	50.03	0.92
	8	7.633	66.30	50.72	3.36	22.94	48.74	1.12	74.90	55.85	4.18	15.45	48.74	0.75
	12	11.429	52.35	50.93	2.67	18.13	50.79	0.92	62.76	56.01	3.52	14.44	50.79	0.73
	16	15.828	28.48	50.37	1.43	15.95	48.80	0.78	40.45	55.25	2.23	14.03	48.80	0.68
25ms, 40Hz	$-\infty$	$-\infty$	105.89	27.65	2.93	93.55	27.45	2.57	116.46	30.49	3.55	90.13	27.45	2.47
	0	0.281	97.86	26.50	2.59	62.24	25.97	1.62	108.39	29.31	3.18	56.67	25.97	1.47
	4	3.979	90.02	26.41	2.38	48.42	25.87	1.25	100.68	29.11	2.93	36.72	25.87	0.95
	8	8.115	77.39	25.98	2.01	38.09	25.55	0.97	88.49	28.60	2.53	30.64	25.55	0.78
	12	12.352	55.52	25.38	1.41	32.80	24.86	0.82	66.95	27.91	1.87	28.90	24.86	0.72
	16	16.131	45.87	25.40	1.17	30.67	25.00	0.77	52.38	27.93	1.46	28.13	25.00	0.70
12.5ms, 80Hz	$-\infty$	$-\infty$	118.73	15.74	1.87	106.96	15.87	1.70	130.52	17.40	2.27	101.84	15.87	1.62
	0	−0.102	107.30	15.00	1.61	81.09	15.22	1.23	118.81	16.60	1.97	79.36	15.22	1.21
	4	3.651	97.29	14.84	1.44	69.54	14.59	1.01	108.19	16.36	1.77	66.20	14.59	0.97
	8	7.458	86.87	14.48	1.26	62.73	14.20	0.89	96.42	15.90	1.53	59.81	14.20	0.85
	12	11.915	79.32	14.02	1.11	58.81	13.87	0.82	87.16	15.42	1.34	56.81	13.87	0.79
	16	15.923	75.49	14.06	1.06	56.72	14.09	0.80	82.55	15.45	1.28	55.14	14.09	0.78
6.25ms, 160Hz	$-\infty$	$-\infty$	528.21	6.91	3.65	452.47	6.97	3.15	578.72	7.65	4.43	437.77	6.97	3.05
	0	0.637	490.98	6.81	3.35	280.12	6.78	1.90	542.72	7.51	4.08	206.21	6.78	1.40
	4	3.707	466.14	6.68	3.11	219.69	6.74	1.48	516.89	7.40	3.82	143.75	6.74	0.97
	8	7.875	402.68	6.49	2.62	164.06	6.27	1.03	457.52	7.18	3.29	122.15	6.27	0.77
	12	11.795	296.50	6.41	1.90	136.39	6.47	0.88	367.91	7.06	2.60	114.78	6.47	0.74
	16	15.908	199.35	6.34	1.26	124.44	6.24	0.78	242.31	6.97	1.69	111.80	6.24	0.70

Table G.5: Estimated parameters from the best–correlated full–linear simulation for Observer 2 in condition $\mathcal{WT}$=2.

Signal Set	SNR		Ess$\mathcal{WT}_{92.4\%}$			$\mathcal{WT}_{\text{AERB}\times\text{ERD}}$			Ess$\mathcal{WT}_{95\%}$			$\mathcal{WT}_{\text{TdB}\times\text{ERD}}$		
	Orig.	Est.	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 5Hz	$-\infty$	$-\infty$	50.54	398.87	20.16	44.88	373.97	16.78	55.61	438.73	24.40	43.07	373.97	16.11
	-4	-4.382	49.92	398.65	19.90	35.07	370.22	12.98	54.93	437.84	24.05	38.11	370.22	14.11
	0	-0.369	49.22	400.70	19.72	25.52	370.50	9.46	54.46	440.32	23.98	7.80	370.50	2.89
	4	4.152	47.77	396.92	18.96	15.85	360.48	5.71	52.80	437.48	23.10	5.08	360.48	1.83
	8	7.709	45.43	399.70	18.16	10.36	372.56	3.86	50.46	440.12	22.21	4.63	372.56	1.72
	12	11.892	38.89	400.45	15.57	6.97	369.93	2.58	44.25	441.62	19.54	4.51	369.93	1.67
200ms, 10Hz	$-\infty$	$-\infty$	57.31	201.76	11.56	51.99	186.55	9.70	63.07	221.27	13.96	48.34	186.55	9.02
	-4	-4.021	56.26	202.37	11.39	39.20	185.90	7.29	61.99	220.99	13.70	41.52	185.90	7.72
	0	-0.166	55.15	199.52	11.00	30.94	185.33	5.73	61.02	220.23	13.44	24.03	185.33	4.45
	4	4.000	52.11	201.88	10.52	21.13	197.56	4.17	57.73	221.70	12.80	10.31	197.56	2.04
	8	7.689	47.19	201.14	9.49	15.49	185.12	2.87	52.95	221.49	11.73	9.64	185.12	1.79
	12	12.105	37.97	199.75	7.58	11.58	191.62	2.22	44.56	219.60	9.79	8.96	191.62	1.72
100ms, 20Hz	$-\infty$	$-\infty$	56.96	103.25	5.88	52.24	100.25	5.24	62.52	113.57	7.10	49.83	100.25	5.00
	-4	-3.516	55.41	102.14	5.66	41.03	99.70	4.09	61.26	112.05	6.86	39.90	99.70	3.98
	0	0.304	52.84	102.10	5.40	35.18	98.79	3.47	58.48	111.91	6.54	28.76	98.79	2.84
	4	3.860	48.22	101.62	4.90	28.22	99.81	2.82	53.98	111.52	6.02	20.87	99.81	2.08
	8	8.142	39.74	100.74	4.00	22.75	99.83	2.27	46.22	110.85	5.12	18.96	99.83	1.89
	12	11.921	27.06	100.30	2.71	20.75	96.13	1.99	34.50	110.43	3.81	18.68	96.13	1.80
50ms, 40Hz	$-\infty$	$-\infty$	64.73	53.31	3.45	59.81	52.46	3.14	71.41	58.70	4.19	54.94	52.46	2.88
	-4	-2.597	60.98	52.42	3.20	50.34	52.16	2.63	67.14	57.69	3.87	48.83	52.16	2.55
	0	1.040	56.15	52.04	2.92	45.51	50.22	2.29	62.47	57.43	3.59	42.74	50.22	2.15
	4	4.027	51.66	52.03	2.69	42.01	52.54	2.21	57.86	57.24	3.31	39.29	52.54	2.06
	8	8.235	44.54	51.40	2.29	39.05	51.13	2.00	49.06	56.53	2.77	37.67	51.13	1.93
	12	12.304	40.78	51.36	2.09	35.20	49.75	1.75	44.34	56.46	2.50	34.64	49.75	1.72
25ms, 80Hz	$-\infty$	$-\infty$	92.30	27.74	2.56	83.43	27.51	2.30	101.72	30.63	3.12	78.48	27.51	2.16
	-4	-4.742	88.01	27.64	2.43	79.19	27.87	2.21	97.07	30.43	2.95	74.48	27.87	2.08
	0	-0.183	82.88	27.45	2.28	74.50	26.50	1.97	91.06	30.40	2.77	70.92	26.50	1.88
	4	3.965	77.74	27.12	2.11	70.73	26.88	1.90	84.65	29.98	2.54	68.59	26.88	1.84
	8	7.476	74.19	27.79	2.06	69.37	27.03	1.87	80.14	30.58	2.45	67.27	27.03	1.82
	12	11.788	70.84	27.14	1.92	64.88	26.49	1.72	76.23	29.91	2.28	64.30	26.49	1.70
12.5ms, 160Hz	$-\infty$	$-\infty$	119.43	15.62	1.87	107.83	15.68	1.69	131.41	17.35	2.28	102.68	15.68	1.61
	-4	-3.561	114.50	15.46	1.77	102.04	15.15	1.55	125.52	17.19	2.16	98.31	15.15	1.49
	0	-0.233	112.80	15.57	1.76	102.50	15.59	1.60	123.20	17.28	2.13	97.31	15.59	1.52
	4	4.588	108.14	15.34	1.66	97.98	15.39	1.51	117.49	17.07	2.01	95.68	15.39	1.47
	8	7.851	107.21	15.39	1.65	98.70	15.07	1.49	116.13	17.13	1.99	96.33	15.07	1.45
	12	11.897	106.53	15.38	1.64	98.31	15.37	1.51	115.19	17.04	1.96	95.59	15.37	1.47

Table G.6: Estimated parameters from the best–correlated full–linear simulation for Observer 2 in condition $\mathcal{WT}$=4.

Signal Set	SNR		Ess$\mathcal{WT}_{92.4\%}$			$\mathcal{WT}_{\text{AERB}\times\text{ERD}}$			Ess$\mathcal{WT}_{95\%}$			$\mathcal{WT}_{\text{TdB}\times\text{ERD}}$		
	Orig.	Est.	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$	$\mathcal{W}$	$\mathcal{T}$	$\mathcal{WT}$
400ms, 10Hz	$-\infty$	$-\infty$	57.07	404.38	23.08	50.29	380.77	19.15	62.75	444.15	27.87	47.29	380.77	18.01
	-8	-8.383	56.69	400.60	22.71	45.52	371.13	16.89	62.30	439.85	27.40	45.49	371.13	16.88
	-4	-3.898	56.20	400.80	22.52	38.93	365.31	14.22	62.01	440.43	27.31	38.85	365.31	14.19
	0	-0.155	54.62	398.44	21.76	30.72	371.24	11.40	60.45	439.06	26.54	13.04	371.24	4.84
	4	3.716	51.34	399.36	20.50	22.52	375.19	8.45	57.19	438.85	25.10	11.26	375.19	4.23
	8	7.915	45.24	395.88	17.91	15.84	373.55	5.92	51.22	435.41	22.30	10.74	373.55	4.01
200ms, 20Hz	$-\infty$	$-\infty$	56.60	200.86	11.37	50.15	187.04	9.38	62.49	219.76	13.73	47.31	187.04	8.85
	-8	-8.245	55.88	201.25	11.25	46.59	188.87	8.80	61.85	221.31	13.69	44.18	188.87	8.34
	-4	-4.142	55.25	200.90	11.10	41.76	187.90	7.85	61.04	220.74	13.47	41.32	187.90	7.76
	0	-0.345	52.62	202.03	10.63	36.48	194.58	7.10	58.38	221.46	12.93	24.96	194.58	4.86
	4	4.030	47.07	199.25	9.38	29.06	184.65	5.37	52.87	218.69	11.56	21.96	184.65	4.05
	8	7.976	38.00	198.45	7.54	24.73	186.31	4.61	44.87	219.28	9.84	20.92	186.31	3.90
100ms, 40Hz	$-\infty$	$-\infty$	73.04	101.71	7.43	66.97	97.49	6.53	80.46	111.82	9.00	61.17	97.49	5.96
	-8	-7.029	70.80	100.51	7.12	62.57	94.93	5.94	77.97	110.74	8.63	57.38	94.93	5.45
	-4	-4.079	69.52	101.55	7.06	58.02	97.74	5.67	76.77	111.94	8.59	50.86	97.74	4.97
	0	0.205	64.89	100.92	6.55	52.53	97.41	5.12	72.00	111.04	7.99	45.90	97.41	4.47
	4	3.951	56.38	100.22	5.65	48.11	95.88	4.61	63.90	110.10	7.04	43.31	95.88	4.15
	8	8.001	46.44	100.40	4.66	44.75	95.73	4.28	52.62	110.60	5.82	42.03	95.73	4.02
50ms, 80Hz	$-\infty$	$-\infty$	81.96	51.84	4.25	78.37	50.18	3.93	89.85	57.27	5.15	71.23	50.18	3.57
	-8	-7.222	79.73	52.13	4.16	76.54	50.05	3.83	87.62	57.38	5.03	68.88	50.05	3.45
	-4	-4.258	78.43	51.88	4.07	76.12	50.09	3.81	86.03	57.03	4.91	69.25	50.09	3.47
	0	-0.021	75.61	51.94	3.93	74.87	50.10	3.75	81.92	57.19	4.69	68.11	50.10	3.41
	4	3.931	72.01	51.52	3.71	72.92	50.85	3.71	77.30	56.72	4.38	67.88	50.85	3.45
	8	8.078	70.34	52.17	3.67	72.13	49.26	3.55	75.04	57.18	4.29	66.98	49.26	3.30
25ms, 160Hz	$-\infty$	$-\infty$	119.39	27.54	3.29	121.97	27.29	3.33	131.61	30.21	3.98	102.09	27.29	2.79
	-8	-7.396	119.16	26.70	3.18	122.04	25.99	3.17	130.79	29.46	3.85	102.57	25.99	2.67
	-4	-3.782	116.92	26.83	3.14	121.06	25.33	3.07	127.90	29.54	3.78	101.52	25.33	2.57
	0	0.303	115.14	26.77	3.08	120.25	26.75	3.22	125.52	29.48	3.70	99.22	26.75	2.65
	4	4.230	113.44	26.87	3.05	119.56	25.09	3.00	123.22	29.58	3.65	99.77	25.09	2.50
	8	8.154	112.59	26.71	3.01	119.20	26.16	3.12	121.90	29.40	3.58	100.43	26.16	2.63

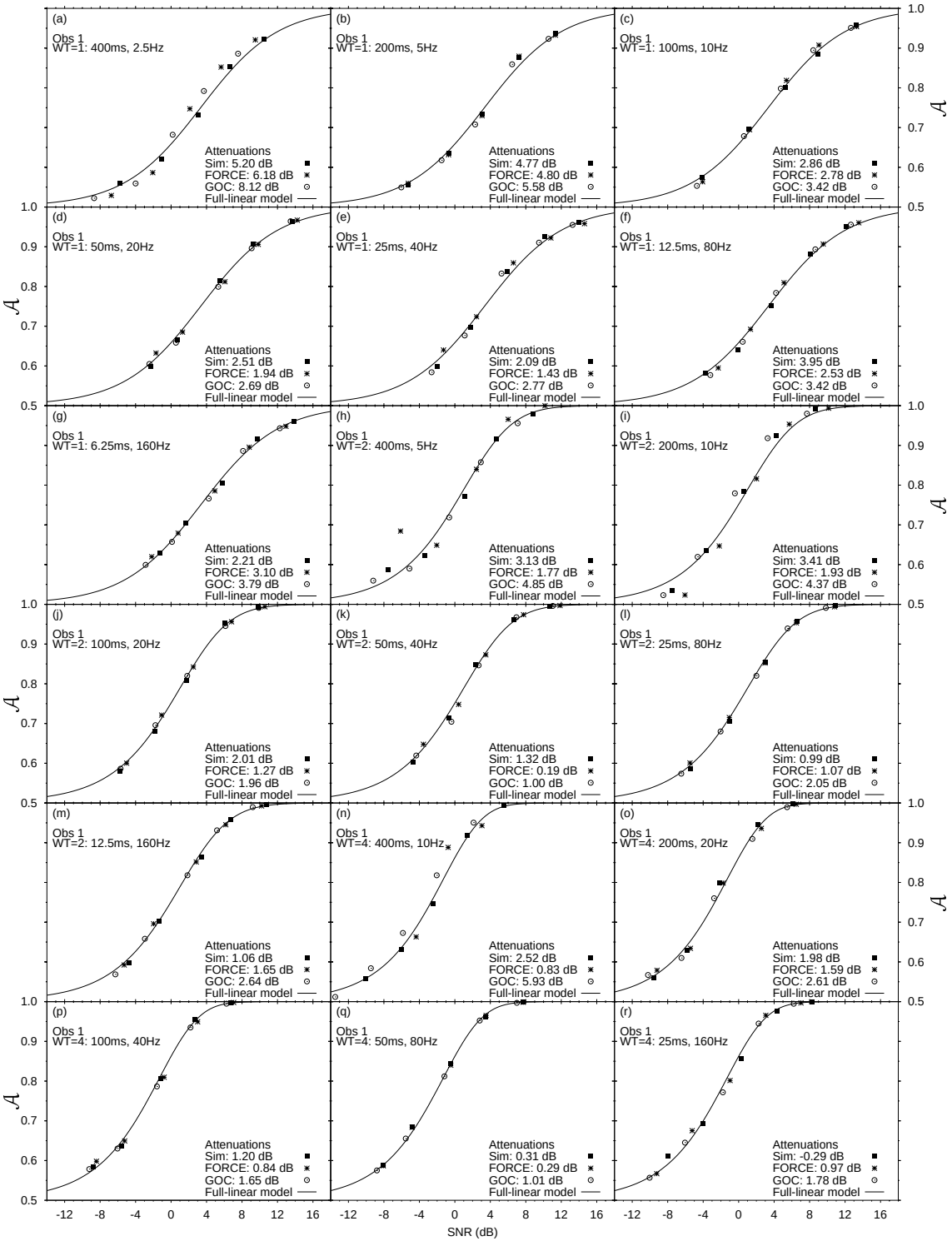

Figure G.9: Psychometric functions of the best correlated full–linear detector (filled squares), compared with FORCE (stars) and GOC (open circles) analyses for Observer 1. The data points for the simulated and human observers have been attenuated to fall onto the optimal full–linear detector psychometric function (line) so the shape, rather than the location, may be compared.

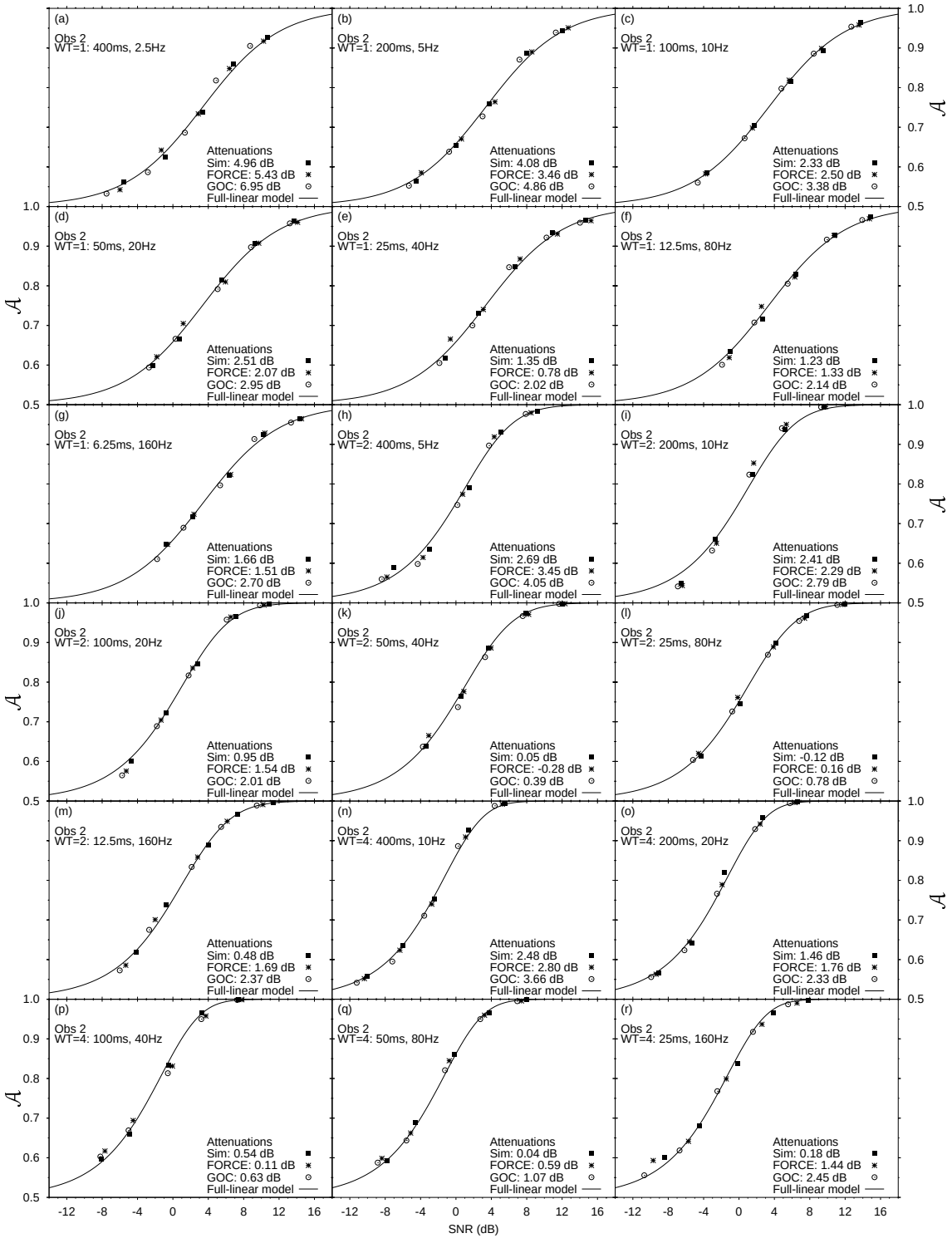

Figure G.10: Psychometric functions of the best correlated full–linear detector (filled squares), compared with FORCE (stars) and GOC (open circles) analyses for Observer 2. The data points for the simulated and human observers have been attenuated to fall onto the optimal full–linear detector psychometric function (line) so the shape, rather than the location, may be compared.