INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
NOTE TO USERS

This reproduction is the best copy available

UMI
ON AUDITORY DYNAMIC RANGE

Iftikhar Riaz Nizami

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Psychology
University of Toronto

© Copyright of Iftikhar Riaz Nizami, 1999
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

0-612-41258-X
Humans can detect intensity changes of a few decibels [dB], over a contiguous range of at least 100 dB. But an auditory neuron’s tone-provoked firing rate will change, on average, only over 40 dB. This smaller, neural dynamic range varies by afferent, as do threshold, saturation rate, and spontaneous rate (r_s). By r_s neurons fall into 3 groups whose average thresholds differ. For each group, within the frequency band presumably serving loudness, firing-rate equations were averaged over all varying quantities. These 3 results were then averaged by relative group size, allowing inference of overall dynamic range, 82 dB at 8 kHz. Many other groupings were investigated; isolating low-r_s neurons extends the upper limit by only 3 dB (to 92 dB SPL).

Dynamic range may be exposure-dependent. When an auditory stimulus (forward-masker) precedes a brief probe, the curve of neural response vs. probe level shifts up-level. Shifting may persist even at 100 msec post-masker for low-r_s neurons (Zeng, Turner, and Relkin, 1991), given that the average relative increase of the probe’s detection threshold recovers to zero much more slowly for low-r_s neurons than for high-r_s neurons (Relkin and Doucet, 1991). If subjective detection thresholds mimic lowest neural thresholds, a cusp should appear, the auditory equivalent of a rod-cone break. But none appeared for a 2 kHz Gaussian-shaped probe (σ=0.5 msec) centered 0-30 msec after a 200 msec, 97 dB SPL, 2 kHz tone.

Next, the probe/masker intensity ratio was fixed at 20 dB, and just-detectable intensity-increments were measured for seven probe levels at three post-masker gaps (3, 80 and ∞ msec). The results were independent of gap. A peak was ubiquitous at 70 dB SPL, as occurs for clicks. By analogy, change-detection may involve not alteration of firing-rate, but rate-of-growth of the pooled spike count (= the number of responding afferents). Distributing afferents into two disjoint threshold distributions broadens dynamic range; rate-of-growth suffers at intermediate probe levels. But this model fails for efficacious masking, and Zeng’s model also appears to fail for efficacious masking.

In sum, partitioning afferents into groups differing by average threshold does not explain auditory dynamic range for tones or brief forward-masked probes.
DEDICATION

These brief paragraphs have been re-worked many times, and still seem incomplete. My parents' appreciation of scholarship, and their constant financial help, has made the difference between success and failure. You have been very patient; I offer pride in return.

I thank my advisor, Bruce "Zen State" Schneider, who is well aware of his many contributions (see Preface).

I thank my hardworking experimental subjects, Brian Kobilke, Jane Carey, Jason Watchus, and Zoran Stojkov, without whose tolerant persistence I would have nothing to say. I thank Jane Carey and Dr. Peter Bascom for technical assistance. I thank Drs. M.C. Liberman, A.R. Palmer, William Rhode, and Fan-Gang Zeng for their gifts of original data. I thank Dr. Zeng once more for his earlier comments on the material presented in Chapters 5 and 6, and for his comments as the external appraiser of this dissertation. I also thank the members of my supervisory committee, Profs. C. Doug Creelman, Robert V. (Bob) Harrison, and Patrick Bennett; I also thank the Departmental reviewer, Prof. Mark Schmuckler. I thank Dept. Chairman and Professor A.M. "Marty" Wall for his diplomatic interventions. Julita Gonera (M.A., Slavic Studies, Toronto) receives my gratitude for her translation of Zaboeva (1966). A Grant-In-Aid Of Research from Sigma Xi, the Scientific Research Society, was used for partial remuneration of Jason Watchus and Zoran Stojkov.

Final mention goes to three people whose contributions may one day prove to be the most profound. In chronological order, I gratefully acknowledge Cheryl-Anne "Kit" Kowaluk, B. Arch., for reminding me of the value of risk; Barbara Lynn Zelsman, M.S.W., the "burning ring of fire" (see Carter and Kilgore, 1962), for reminding me of the value of ideals; and finally and most importantly my wife, Dr. Claire S. Barnes, Ph. D., "The Navigator", for reminding me of the value of myself; and for guiding me through not only the golden hills of California, but across the compass points of my doctorate as well. Despite the help of many others, my deepest feelings must go to the three of you. Words are inadequate.

Jane Withersteen: "Tell you! I? Never!"

Lassiter: "I reckon you will. An' I'll never ask you. I'm a man of strange beliefs an' ways of thinkin', an' I seem to see into the future an' feel things hard to explain. The trail I've been followin' for so many years was twisted an' tangled, but it's straightenin' out now."

- Zane Grey, Riders of the Purple Sage
# TABLE OF CONTENTS

List of Tables, Figures, and Appendices  viii  
Preface  1  

Chapter 1: Introduction  2  
Purpose  3  
The computations  5  
The experiments  11  
The first experiment  14  
The second experiment  19  
Notes  27  

Chapter 2: Computational methods  29  
The single-channel point model and the critical band  30  
Parameters and rate-level functions of single nerve fibers  30  
Obtaining a mean rate-intensity function  35  
The neural population: distributions of characteristics, and pool size  38  
Deriving a mean response-rate intensity function  44  
Computing just-detectable intensity-increments from the mean response-intensity function  47  
The ideal observer  52  
Notes  56  

Chapter 3: Actual dynamic range computations  61  
Auditory discriminability from the overall mean rate-intensity function  62  
One channel  62  
Two channels  64  
Discriminability for the ideal observer  66  
Discussion  68  
The validity of the assumptions  68  
Extending dynamic range by means of gain control  70  

iv
Other models that computed dynamic range 73
Rate-intensity functions, and possible correlations between their parameters 74
The role of sloping-saturating units 75
Would low-spontaneous-rate neurons have any advantage as an upper channel? 84
Possible dynamic ranges of auditory channels in man 87
Summary of results 88
Notes 90

Chapter 4: Experimental methods, and data analysis 93

Experiment methods 94
Subjects 94
Stimuli 94
Apparatus 96
Procedure 96
Performance 99
Data analysis 101
Notes 103

Chapter 5: Experiment 1: Forward-masked pip-detection thresholds 104

Additional note on methods 105
Results 105
Nonmasked pip-detection thresholds 105
Forward-masked pip-detection thresholds 106
The standard deviation $\sigma$ of the psychometric functions 106
Discussion 110
Regarding forward-masked thresholds 110
Regarding recovery times 117
A unique experimental problem at very early recovery times 123
Alternative explanations for the threshold oscillations 123
Model 125
Forward-masked thresholds as due to neural ringing 125
Uses of the ringing model: recovery times for forward-masked clicks 132
Limitations of the model: the fit to the data 133
Limitations of the model: the threshold shift in afferents vs. the threshold shift in psychophysics 134
Limitations of the model: regarding the rise in threshold over \( t=0-2 \) msec 135
Summary of the results 139
Notes 141

Chapter 6: Experiment 2: Forward-masked just-detectable intensity-increments 144
Recap of methods 145
Subjects 145
Stimuli and procedure 145
Results and Discussion 146
Nonmasked pip-detection thresholds 146
The detected intensity-increments 146
The standard deviation \( \sigma \) of the psychometric functions 149
Comparison of the just-detectable intensity-increments to Zeng's measure (1994) 151
Model 160
Background: just-detectable intensity-increments and \( N_1 \) variability 162
\( N_1 \) in terms of two neural populations 164
A counter-intuitive conclusion, and its resolution 167
Firing rate vs. rate of population growth: strategies for encoding intensity change 169
Mid-level humps: the role of spike pooling 172
Summary of the model 173
Summary of results 174
Notes 176

Chapter 7: Further discussion, and final summary of results 179
Can our computations (Chapters 2 and 3) be generalised to other species? The role of design

- Common features of fiber properties across species
- Differences in fiber properties across species: due to problems of definition?
- Summary of this Section

- Why and how might two auditory channels (Chapters 3 and 6) interact?
  - A possible locus of two-channel interaction
  - A potential model species for two-channel interaction: the moth

- Final summary of results
- Notes

Chapter 8: Suggestions for further work

- Ideas
- Notes

References
- Appendices
- Glossary of terms

vii
LIST OF TABLES, FIGURES, AND APPENDICES

TABLES

Table 1   Comparison of regression-fitted and visually-estimated parameters for the rate-intensity functions fit to individual neurons

Table 2   Comparison of regression-fitted versus visually estimated dynamic range for different values of the constant C

Table 3   Comparison of regression-fitted parameters for the logistic rate-intensity function fitted to sloping-saturating behaviour

Table 4   Parameter values used in initial simulations of computed auditory dynamic range

Table 5   Sources of evidence for independence between neural fiber parameters within a spontaneous-rate group

Table 6   Values of the Student-t statistic for comparison of the average dynamic ranges of sigmoidal vs. sloping-saturating fibers

Table 7   Just-noticeable-differences in level for detection thresholds over 0 ≤ t ≤ 2 msec

FIGURES

Figure 0   Relevant auditory structures

Figure 1   The human threshold of audibility and the threshold of feeling

Figure 2   Four parameters used to characterise an afferent fiber's rate-intensity function

Figure 3   Dynamic ranges for nerve fibers of one cochlear nerve of a single cat

Figure 4   The stimulus-response functions of two single neurons, and their sum

Figure 5   The two-alternative forced-choice procedure used to estimate a nerve fiber's threshold for detection of a tone-pip during post-masker recovery

Figure 6   Relative threshold shifts averaged over all fibers within each of two spontaneous rate classes

Figure 7   The Zeng et al. (1991) model of the effect of forward-masking on the operating ranges of the low- vs. high-spontaneous-rate neurons, and
just-noticeable-differences in intensity for three subjects

Figure 8  The Zeng et al. (1991) model of the effects of forward-masking on discrimination of increments in a short tone

Figure 9  Predictions of a model in which discriminability of an increment in pip-level is mediated by two neural channels

Figure 10  Conditions in which referential coding might occur in a two-interval forced-choice experiment

Figure 11  Predictions of a model in which the forward-masker provides a reference level against which changes in pip level can be more easily perceived than if the masker was absent

Figure 12  Variability in fiber properties of primary afferents of the cat's cochlear nerve

Figure 13  The fit of the logistic function to rate-level data

Figure 14  The fit of the weighted double-logistic function to neurons showing sloping-saturating behaviour

Figure 15  Averaging of actual firing-rate plots vs. averaging of the logistic function

Figure 16  The fit of the weighted double-logistic function and of the single-logistic function to seven sloping-saturating units

Figure 17  Distribution functions for fiber threshold

Figure 18  The distribution of auditory nerve fiber dynamic ranges

Figure 19  Fiber thresholds vs. their characteristic frequency

Figure 20  Auditory fiber dynamic ranges vs. their thresholds and their characteristic frequencies

Figure 21  Fiber dynamic range vs. spontaneous rate

Figure 22  Maximum discharge rates plotted vs. characteristic frequency and vs. threshold, and their histogram

Figure 23  Signal Detection Theory

Figure 24  ROC curves for the cases of equal, and unequal, variance

Figure 25  The overall mean response-intensity function and the intensity-increment function for the single channel model

Figure 26  Overall mean response-intensity functions and intensity-increment functions when the standard deviations of the dynamic ranges are halved or doubled

Figure 27  Overall mean response-intensity functions and intensity-increment functions when the mean dynamic ranges are increased or decreased by 20 dB
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>Overall mean response-intensity functions and intensity-increment functions when the standard deviations of the thresholds are either halved or doubled</td>
</tr>
<tr>
<td>29</td>
<td>Overall mean response-intensity functions and intensity-increment functions for two independent channels</td>
</tr>
<tr>
<td>30</td>
<td>Intensity-increment functions for multiple-channel ideal computers</td>
</tr>
<tr>
<td>31</td>
<td>Intensity-increment functions for a 938-channel ideal computer under different variance assumptions</td>
</tr>
<tr>
<td>32</td>
<td>Possible or confirmed pathways of the medial olivo-cochlear feedback system in the cat</td>
</tr>
<tr>
<td>33</td>
<td>The discriminability of individual sigmoidally-saturating auditory-nerve fibers</td>
</tr>
<tr>
<td>34</td>
<td>The discriminability of individual sloping-saturating auditory-nerve fibers</td>
</tr>
<tr>
<td>35</td>
<td>A comparison of neural discriminability calculated from sigmoidal units and sloping-saturating units</td>
</tr>
<tr>
<td>36</td>
<td>The stimuli, and the Gaussian probe pip's energy spectrum</td>
</tr>
<tr>
<td>37</td>
<td>Examples of relative energy spectra for tones without Gaussian ramping</td>
</tr>
<tr>
<td>38</td>
<td>Arrangement of masker and tone pip in Experiment 2</td>
</tr>
<tr>
<td>39</td>
<td>Forward-masked pip-detection thresholds (subject LN)</td>
</tr>
<tr>
<td>40</td>
<td>Forward-masked pip-detection thresholds (subject BAK)</td>
</tr>
<tr>
<td>41</td>
<td>Forward-masked pip-detection thresholds (subject JWC)</td>
</tr>
<tr>
<td>42</td>
<td>The psychometric functions for forward-masked thresholds in the neighbourhoods of sudden threshold rise</td>
</tr>
<tr>
<td>43</td>
<td>The standard deviations σ of the psychometric functions for threshold</td>
</tr>
<tr>
<td>44</td>
<td>Comparison of the forward-masked pip-detection thresholds to the literature</td>
</tr>
<tr>
<td>45</td>
<td>Least-squares fits of straight lines to pip-detection thresholds for t ≥ 13 msec</td>
</tr>
<tr>
<td>46</td>
<td>The fit of pip-detection thresholds to the function Threshold (dB SPL) = A/(B + ln(t))</td>
</tr>
<tr>
<td>47</td>
<td>Averaged forward-masked thresholds vs. those of a cochlear implantee, and average threshold shifts vs. those of a brief sound pulse preceded by an identical pulse</td>
</tr>
<tr>
<td>48</td>
<td>Model: pip-detection thresholds in terms of hypothetical impulse response functions of Type I and Type II afferents</td>
</tr>
<tr>
<td>49</td>
<td>Close-ups showing forward-masked thresholds (from Figures 39-41) and the standard deviations of their psychometric functions (from Figure 44)</td>
</tr>
</tbody>
</table>
for brief, common intervals of time where thresholds temporarily rise

Figure 50  Tone-pip intensity-increment thresholds for subject JW
Figure 51  Tone-pip intensity-increment thresholds for subject ZS
Figure 52  The standard deviations $\sigma$ of the psychometric functions for the
intensity-increment thresholds (JW & ZS)
Figure 53  Some of the psychometric functions for detection of the intensity
increment by subject JW
Figure 54  Some of the psychometric functions for detection of the intensity
increment by subject ZS
Figure 55  Both the intensity-increment thresholds and the standard deviations $\sigma$
of their psychometric functions (JW), for each respective stimulus
condition
Figure 56  Both the intensity-increment thresholds and the standard deviations $\sigma$
of their psychometric functions (ZS), for each respective stimulus
condition
Figure 57  The intensity-increment thresholds plotted vs. the standard deviations $\sigma$
of their psychometric functions (JW), with the correlation coefficients,
for each respective stimulus condition
Figure 58  The intensity-increment thresholds plotted vs. the standard deviations $\sigma$
of their psychometric functions (ZS), with the correlation coefficients,
for each respective stimulus condition
Figure 59  Intensity-increment thresholds for the nonmasked pip, vs. the variability
of loudness matches (data of Zeng, 1994)
Figure 60  The intensity-increment thresholds for nonmasked pips vs. those for
2 kHz, 500 msec tones
Figure 61  The intensity-increment thresholds for nonmasked pips vs. those for
nonmasked clicks
Figure 62  Model. The variability of neural counts vs. the growth in size of the
population of driven neurons, the presumed density of neural thresholds
vs. the growth in size of the population of driven neurons, and the
intensity-increment threshold vs. the presumed density of neural
thresholds
Figure 63  Rate-level curves for the $A_1$ and $A_2$ auditory receptors and the their
Figure 64 The neural discriminability computed from Equation 13 compared to the measured neural discriminability, for two sloping-saturating fibers of Delgutte (1987)  

Figure 65 The neural discriminability computed from Equation 13 compared to the measured neural discriminability, for two sigmoidal fibers of Delgutte (1987)  

APPENDICES  

Appendix A: The solution to the mean-rate equation using the wedge-shaped threshold distribution  

Appendix B: Siebert's (1965) "ideal computer"  

Appendix C: A comparison of our model of auditory neural discriminability to Delgutte's (1987) model  

Appendix D: Computing the dynamic ranges of single nerve fibers  

Appendix E: Mathematical expressions for sound intensity and discriminability  

Appendix F: Frishkopf's (1956) model of the N1 potential  

Appendix G: The mid-level hump: comparison of the Frishkopf-Radionova model to signal detection theory
Our senses are the interface between our minds and the outside world. Loss or absence of these senses reduces our ability to function. Although vision and taste allow us to interact with the world of objects, hearing also allows us to interact with the world of people. Impairment of this ability causes a loss of social activity. Thus hearing loss, if not prevented, must be compensated for in the best interests of all concerned, and the best compensation is the one that confers nothing more and nothing less than the missing ability. Thus the most desirable auditory prosthetic is that that most accurately mimics the damaged organ. To make such a device, we must understand how the organ works. The organ in question is the organ of Corti, which performs the first step in the encoding of sound intensity into the percept called loudness.

Despite decades of behavioural and neurophysiological study, this encoding process is not well understood. Our knowledge is fragmentary, consisting of experimental snippets that remain uncoordinated, in part because of a reluctance to embark on integrative theoretical studies, in part because work of a review nature is not considered “original” and resists publication. Here we present theoretical, integrative work in an attempt to circumvent these problems. We also present original experiments. I use the word “we” throughout this work, because most of the work reported here represents the creativity of more than one mind, although the effort is largely my own. The scholarly contributions of Prof. Bruce A. Schneider are so closely interwoven with my own that the individual contributions sometimes cannot be remembered. Hence the use of the pronoun “we” throughout the text.

Prof. Schneider also provided equipment and software that were necessary to certain stages of this project. These contributions were much appreciated.

This thesis is formatted in strict conformity with the “Guidelines for the preparation of theses for microfilming and binding”, School of Graduate Studies, University of Toronto.
CHAPTER 1

INTRODUCTION
PURPOSE

In this thesis we pursue one theoretical model and two experiments, all designed to ultimately help understand the encoding of sounds into neural impulses and hence into the percept that we call loudness. To do this we will focus on differences between the electrical performance of the peripheral auditory organs (Figure 0) and the performance of human subjects who make judgements about sounds heard in psychophysical experiments. This work was inspired by a problem that still begs solution, the “dynamic range problem” (Evans, 1981). Evans formulated the problem thus: “It is an everyday experience that our ears can operate over an amazingly large dynamic range approaching 100 dB, yet the dynamic range of the great majority of cochlear nerve fibers is remarkably limited...Cochlear nerve fibers therefore appear to be incapable of specifying the level of a stimulus in terms of their mean discharge rate over a sufficiently wide range to account for the psychophysical dynamic range”. We will be quantifying mean neural discharge rate in the first study in this thesis, using a novel method, and relating it to psychophysical dynamic range. But we first note that Evans appears to be using two definitions of dynamic range, one for the loudness of sounds, and one for the encoding range of the neuron. His first use of the term follows the definition used by Lawrence (1965, p.159): “The dynamic range expresses the extent of rising intensities over which hearing takes place, but the upper limit is difficult to define”. Lawrence describes this upper limit as that level at which the sensation of loudness mingles with the sensation of feeling. This limit appears to be constant despite the fact that the threshold for detection of sound, sometimes called the absolute threshold, varies with sound frequency (Figure 1; Licklider, 1966/1951). Lawrence (1965, p.161) notes that Licklider’s illustration implies psychophysical dynamic ranges of 90 dB at 100 and 10,000 Hz and 120 dB at 1 kHz.

There is another definition of dynamic range, the one that Evans applies to cochlear nerve fibers, henceforth referred to by the interchangeable terms “fiber”, “unit”, or “neuron”. The limited response ranges of single primary afferent neurons (the first, peripheral neurons to encode the sound signal) can be estimated by the range in dB over which the firing rate of the neuron shows an obvious increase as sound intensity increases. This usually corresponds to the steeply sloping portion of the neural unit’s rate-level curve, the plot of spikes fired within some common counting time, vs. the sound pressure level usually listed in dB SPL (see Glossary for definitions). Traditionally, single nerve fibers are characterized by four properties (see Figure 2): spontaneous rate, \( r_s \) (undriven activity in spikes/sec), saturation rate, \( r_{max} \) (maximum driven rate
Figure 0. Relevant auditory structures (from Gulick, Gescheider, and Frisina, 1989). a. The gross structure of the ear, with bone shown as stippled. Inset is a cross-section through the cochlea. b. Cross section through the scala media, showing the organ of Corti (below Reissner’s membrane and above the osseous spiral lamina). c. The afferent connections of the organ of Corti.
in spikes/sec), threshold intensity, $\epsilon$ (in dB SPL), and dynamic range, $\lambda$ (in dB). The spontaneous rate is the fiber's activity, in spikes/sec, in the absence of a stimulus. The unit reaches maximum (saturation) rate when raising the sound pressure produces no consistent change in the firing rate, which is expressed in spikes/sec and is measured by counting the number of spikes produced over some counting time, divided by that counting time. A unit's threshold is that sound pressure level (SPL) at which its firing rate can be reliably discriminated from its spontaneous rate. A unit cannot effectively discriminate intensity changes below its threshold or above the SPL that produces a rate just discriminably different from its saturation rate. Therefore one common assumption about the unit's dynamic range, the range over which the unit can code for changes in intensity, is that fiber dynamic range is the decibel difference between its threshold SPL and the SPL that produces a response just discriminable from saturation rate.

The dynamic range problem can now be presented graphically in terms of neural vs. perceptual dynamic range. The white space separating the tips of the lines in Figure 3 (Evans, 1981) from the upper axis (120 dB SPL) expresses the missing dynamic range, assuming that the upper limit of dynamic range is 120 dB. Conceivably, it could be higher.

The following computations and experiments explore how this missing range is filled in.

**THE COMPUTATIONS**

The ear's dynamic range can be measured more precisely than by Lawrence's definition. Towards this new definition, we first note that human subjects are capable of discriminating small increases in intensity on a given sound background of the same duration and frequency composition, sometimes called the pedestal, which is most often a brief pure tone (Scharf and Buus review, 1986). These small increases, as low as 0.5 decibels (dB), are sometimes called just-noticeable-differences (jnds) or difference limens. There are a variety of experimental techniques that measure such differences, such that these terms are not necessarily interchangeable. Thus to avoid ambiguity we will often refer to the *intensity-increment threshold* as the sound pressure level (SPL) at which one sound intensity can just be differentiated from another, according to the procedure that we detail in the Introduction and Methods. The continuous intensity range over which we can discriminate small intensity changes appears to lie within Lawrence's (1965) idea of dynamic range. Henceforth we will use the more restrictive definition of perceptual dynamic range.
Figure 1. (Licklider, 1966, Fig. 5) The human threshold of audibility (references available in Licklider, 1966) and the threshold of feeling (tickle, pain, etc.). MAP = minimum audible pressure at the eardrum, MAF = minimum audible pressure in a free sound field, measured at the position of the observer's head. The right-hand scale is in dB SPL. The difference between the lower plotted curves and the upper curves represents one measure of auditory dynamic range.

Figure 2. Four parameters used to characterise an afferent fiber's rate-intensity function. Shown are threshold $\varepsilon$, dynamic range $\lambda$, spontaneous rate $r_s$, and maximum rate $r_{max}$. 
range, in terms of discriminability, unless stated otherwise. Regarding single fibers, there are several ways in which a single fiber’s dynamic range has been assessed. Indeed, defining both thresholds and dynamic ranges for single nerve fibers is problematic, an issue which is reserved for Chapter 6. For this reason, we are compelled in Chapter 2 to develop our own definition of fiber dynamic range.

Often the mathematician does not have direct access to the neural recordings, only whatever published details are available, and may furthermore wish to manipulate the data in such a way as to make inferences or test models about neural behaviour. Thus theoretical constructs have been published that give equations for the intensity-increment threshold of a collection of fibers called a neural pool, computed thresholds that can be plotted vs. sound level in dB SPL in order to allow estimation of the pool’s dynamic range. There are more than one such model, and two (those of Siebert [1965] and Hellman and Hellman [1990]) are employed here, each as a necessary part of our own model of neural behaviour.

The theory section of this work examines the encoding of sound intensity by the afferent neurons of the cat, with particular emphasis on how neurons of limited encoding ability might nonetheless account for the wide responsiveness of the cat’s ear. The cat can respond to sounds that vary in intensity over a 100 decibel range, at least according to Lawrence’s definition of dynamic range (see Heffner and Heffner, 1985, Fig. 3). However, individual fibers of the cat, which leave the periphery by way of the 8th (auditory) nerve, typically have a range from threshold to saturation of only 15-40 dB (Kiang, Watanabe, Thomas, and Clark, 1965). Thus, theories of neural function that hope to capture the full behavioural range of hearing in the cat (and presumably in man) have to stipulate how fibers of such limited dynamic range can encode intensity over the full behavioural range. One approach has been to assume that sound intensity is proportional to activity in the whole nerve (Schiaffino, 1957; Barducci, 1961; McGill and Goldberg, 1968; Goldstein, 1974; Howes, 1974, 1979; Lachs, Al-Shaikh, Bi, Saia, and Teich, 1984). The general argument for this type of model is as follows. As the intensity of a pure tone is increased from threshold, neurons serving the region of the basilar membrane near the locus of maximum displacement increase their rate of firing until they reach saturation rate. (The part of the inner ear that encodes auditory events as neural spike trains is the organ of Corti, sometimes referred to by its substructure, the basilar membrane.) However, as the sound intensity of a pure tone increases further, there is a spread of excitation along the basilar membrane such that fibers adjacent to the point of maximal membrane displacement are recruited (e.g. Howes, 1974, based on Katsuki, Suga, and Kanno, 1962;
Thus increasing stimulus intensity means an increase in both fiber firing rate and the number of units responding (Whitfield, 1967). The total number of impulses elicited by the stimulus will thus continue to increase in such “whole nerve” models because neurons adjacent to the locus of maximal displacement will continue to be recruited as sound intensity increases.

However, this recruitment is found both within the region of already active units as well as on its edges (Dix, Hallpike, and Hood, 1948) and is replicated at regions closer to the brain where nerve-cell bodies are grouped together in nuclei sometimes referred to as “higher centers” (i.e. the lateral lemniscus; Kemp, Coppée, and Robinson, 1937). This suggests an approach in which we count accumulated neural voltage spikes not over the whole nerve, but instead over a narrow band of characteristic frequencies (CFs) corresponding to adjacent positions on the basilar membrane. Recent behavioural evidence (Viemeister, 1983) indicates that a wide range of intensity can be encoded even under conditions that prohibit the recruitment of unsaturated fibers. Viemeister used a high-intensity notched masking noise to eliminate potential contributions from remote fibers to the processing of a 6-14 kHz bandpass (“notched”) noise. Presumably, this high-intensity notched masker saturated the fibers whose center frequencies were less than 5 kHz and greater than 14 kHz. Thus, fibers outside of the 6-14 kHz bandwidth of the stimulus presumably could not contribute to the encoding of stimulus intensity. In the presence of this notched masker, Viemeister required subjects to detect an intensity increment to a 6-14 kHz noise. According to conventional whole-nerve theory, when this 6-14 kHz noise is presented at a high intensity, it should saturate the fibers in the 6-14 kHz region. If these fibers are fully saturated or nearly saturated, it should be difficult, if not impossible, to discriminate an intensity increment under these conditions, because the addition of an increment should not be able to produce any additional firing within this frequency region. Moreover, because fibers outside of this region have been saturated by the masker, any spread of excitation produced by the increment should not produce any increment in neural activity in the neighboring regions. Therefore, neural models that depend on recruitment of nearby neurons would predict that the notched masker would interfere substantially with intensity discrimination when the stimulus is presented at high levels. The fact that it did not suggests that the full dynamic range of hearing can be encoded by a limited portion of the organ of Corti surrounding the point of maximal displacement. In other words, a limited population of fibers in the region surrounding a point on the basilar membrane should be capable of coding intensity and accounting for intensity discrimination over a very wide dynamic range. We say “limited” in view of recent evidence suggesting that the physiological correlate of loudness is not the simple sum of the spike activity over the whole auditory nerve.
One way in which a limited section of the basilar membrane could allow a wide dynamic range is for the individual fibers to be selectively sensitive over different portions of the range. Let us provide a trivial example of what we mean. If, for example, different nerve fibers have different thresholds but similar dynamic ranges, then the dynamic range of the ensemble will be greater than the range of any individual unit. How this might occur for the simplest possible version of such a system, two fibers, is demonstrated graphically in Figure 4. As discussed in Chapter 6, this simple system may be realised in some species of moths, which have only two primary auditory afferents with overlapping dynamic ranges (i.e. Perez and Coro, 1985).

There is other psychoacoustic evidence suggesting that a small region surrounding the point of maximal displacement can encode a great dynamic range (Hellman, 1974). Intensity coding by recruitment of units with high thresholds at a given basilar position gains support from Hellman's discovery that using a masking noise to remove the neural firing contributions of fibers at higher CFs (positions further toward the base [near end] of the organ of Corti) does not prevent the growth of loudness of a 250 Hz tone, up to about 105 dB SPL. The growth of loudness vs. sound level could be fitted by a power function whose exponent was at best marginally different from that fitted to an unmasked tone.

We will explore what kind of models based on neural activity at a single locus on the basilar membrane can account for the full dynamic range observed in the cat. In order to do this we will propose a general equation to describe the rate-intensity function for an individual nerve fiber. This rate-intensity function for an individual neuron has four parameters that correspond to the neuron's threshold, spontaneous rate, maximum rate, and dynamic range. We will show that this rate-intensity function, with the appropriate selection of the four parameters, can account fairly well for any of the rate-intensity functions that have been found for individual neurons. We will then determine the distributions for each of these four parameters in a population of neurons using published data. Using these distributions to define a population of rate-intensity functions at a particular point along the basilar membrane, we then derive the function for mean neural activity, which we assume can be generalised to a limited population of fibers in a small basilar region.
Figure 3. (Evans, 1981, Fig. 2) Dynamic ranges (the lengths of the lines) for 121 nerve fibers of the cochlear (8th) nerve of a single cat, stimulated at their characteristic frequency (CF) by 100 msec tones. Arrows indicate fibers which had not reached saturation firing at the highest stimulus levels employed. The upper border of the graph is at 120 dB SPL, roughly the level at which physical feeling starts (Figure 1). The difference between the ends of the lines and the upper border expresses the dynamic range problem. As discussed later (Chapters 3 and 7), the lines with arrows, which represent a sub-class of nerve fibers, do not solve the problem.

Figure 4. (Lipetz, 1969, Fig. 16) The stimulus-response (rate-level) functions of two single neurons with differing thresholds but identical maximum rates (curves 1 and 2), and their sum (curve 3). If we take overall dynamic range to be the range over which firing rate does not saturate (i.e. does not plateau), then summing the fibers' responses extends overall dynamic range.
THE EXPERIMENTS

In addition to the theoretical work mentioned above, experimental work was conducted to investigate how the auditory system achieves such a large dynamic range. Starting from recent publications which offer clues to a solution, two forward-masking experiments were devised. As Zeng and Turner (1992) note, "forward masking is used to describe the effect of a preceding sound upon the detection and discrimination of a following sound". Alternatively, and more generally, "masking was the effect of one sound being smothered or concealed by another sound" (Harris and Dallos, 1979) and "physiological [forward] masking has acquired a similar definition as a shift in the threshold or a reduction in the magnitude of the response evoked by the probe stimulus caused by the introduction of the masking stimulus" (Harris and Dallos, 1979). Generally, after an intense preceding sound, probe-evoked firing rates of auditory neurons decrease relative to neurons stimulated in quiet (Smith, 1977, Mongolian gerbil; Harris and Dallos, 1979, chinchillas) and behavioural detection thresholds for weak sounds rise (Ehmer and Ehmer, 1969; Weber and Moore, 1981; Zwicker, 1984; Scharf and Buus, review, 1986; Carlyon, 1988; Champlin and Wright, 1993). In the following pages we will follow Harris and Dallos, such that the term "probe" applies to the sound to be detected (or to be discriminated from a similar sound), and the term "masker" refers to the sound introduced earlier to interfere with the subsequent detection (or discrimination) of the probe.

Our first experiment is a simple detection experiment to establish detection thresholds for forward-masked pure tones at very brief time gaps (1-40 milliseconds, abbreviated as "msec") between masker and tone. Our second experiment concerns establishing the thresholds for detection of increments in a probe tone under forward masking. The inspiration for the experiments comes from two recent papers by Relkin and colleagues. Earlier work on Mongolian gerbils (Smith, 1977) and cats (Young and Sachs, 1973) had examined the recovery of tone-evoked firing rates after forward masking. These authors had found that for tonal intensities close to neural threshold, the low-spontaneous-rate neurons recovered more slowly than the high-spontaneous-rate units. Relkin and Doucet (1991) employed anesthetised chinchillas. A two-interval forced-choice procedure was used (2IFC; sometimes called two-alternative forced-choice or 2AFC; Figure 5), in which neural spikes were counted in two time intervals of equal duration, one of which (randomly chosen) contained the probe tone, one of which did not. The forward masker was of 102 msec total duration and the probe tone 25 msec total duration, with these durations including 2 msec onset and offset ramps.
Both masker and probe were at the fiber's CF (characteristic frequency), the sound frequency to which the fiber is most sensitive. Masker levels were adjusted for each unit, and averaged 52 dB SPL for high-spontaneous-rate units and 56 dB SPL for the low-spontaneous-rate units. This assured saturation firing rates for the high-spontaneous-rate units, those units with rates greater than 40/sec, and saturation or near-saturation for the low-spontaneous-rate units, those units with rates less than 20/sec.

Relkin and Doucet (1991) presented their neural-recovery data by computing a relative threshold shift for each neuron, producing a maximum value of unity for each neuron by dividing the threshold at each masker-probe time gap (ΔT in Figure 5) by the threshold shift that had occurred at a time gap of zero. This data was then averaged over all neurons within each of the two spontaneous-rate groups. These two recovery curves showed a brief crossover at a time gap of about 6 msec, in which the curve of relative threshold shift for high-spontaneous-rate units temporarily rises above that for low-spontaneous-rate units. This was followed by a progressive difference in the degree of threshold shift between the two curves, such that the time required for this shift to decay to zero - the extrapolated times for average fiber recovery - were 215 msec for high-spontaneous-rate units and 2.064 sec for low-spontaneous-rate units. Hints of a spontaneous-rate difference in the recovery rates of auditory afferents can also be seen for Mongolian gerbils exposed to forward-masked tones (Smith, 1977).

Thus it appears that only fibers of high spontaneous rate should be mediating recovery beyond perhaps ΔT=6 msec. We must be careful, however, because Relkin and Doucet (1991) defined masker-pip time gap as from the offset of the masker to the onset of the 25 msec probe, so that a masker-probe gap of 6 msec means that the center of the probe’s energy actually appears at about 18 msec. Thus it is possible that the low-spontaneous-rate units may contribute to the overall perceptual recovery curve at masker-pip time gaps below 18 msec. Unfortunately, Relkin and Doucet presented only 4 data points on each of their neural recovery curves for time gaps between 1 and 20 msec post-masker. This data shortage, along with the relatively long probe, suggests that the recovery curves have been undersampled. It is thus difficult if not impossible to assess which of the two defined spontaneous-rate groups actually accounts for recovery over 0-20 msec post-masker.

In a separate but related project, Zeng, Turner, and Relkin (1991) reasoned that separate populations of afferents with different mean thresholds, and therefore different overall intensity ranges, could extend the
The two-alternative forced-choice procedure used by Relkin and Doucet (1991) to estimate threshold for detection of a tone-pip during post-masker recovery in nerve fibers stimulated at their CFs, as used earlier by Relkin and Pelli (1987). The 25 msec probe appears randomly in either the first or second interval, and the number of neural spikes produced by the probe is counted and compared to the counts within an identical time window in the other interval. When the interval containing the probe shows a higher count, a correct detection is registered. When some criterion % of correct detections occurs, probe level is adjusted upwards or downwards in an adaptive tracking procedure until a threshold criterion of % correct detections is implied. The corresponding probe level is considered to be the neural probe-detection threshold.

Figure 6. (Relkin and Doucet, 1991, Fig. 4a) Relative threshold shifts (each fiber's forward-masked fiber threshold, divided by its threshold shift [threshold minus nonmasked threshold] at T=0), averaged over all fibers within each of two spontaneous rate classes. In the right-hand frame a logarithmic time scale was used so that straight lines could be fitted, from which recovery to zero relative-threshold-shift could be extrapolated. These are the recovery times mentioned in the text.
range of good discriminability, the auditory dynamic range, by acting as separate channels for loud and soft sounds. Zeng et al. (1991) suggested that the upper or "louder" channel consists of low-spontaneous-rate units, whereas the lower or "softer" channel consists of high-spontaneous-rate units. As a test, Zeng et al. measured intensity discrimination for a 1 kHz pure tone preceded by a 90 dB SPL, narrow-band-noise adapter using the same arrangement of stimuli as Relkin and Doucet (1991) and as shown in Figure 5. Because fibers in the two spontaneous-rate groups appear to have notably differing recovery rates from adaptation (Relkin and Doucet, 1991; see Figure 6), Zeng et al. chose the 100 msec delay between masker offset and tone onset such that high-spontaneous-rate units, but not low-spontaneous-rate units, should have recovered from the forward masker (the adapter; Figure 7a). Thus, intensity-increment thresholds for soft tones should be relatively unaffected by the masker because the "soft" (high-spontaneous-rate) channel should have recovered from adaptation. For very high intensity tones, the "loud" (low-spontaneous-rate) channel has not completely recovered from adaptation, but the test pip is sufficiently above this channel's threshold to permit good discrimination. Poor discrimination, however, should occur for intermediate-level pips. These are out of the lower channel's range of good discriminability, and close to the "louder" channel's threshold, which is presumably still elevated due to the lingering effects of the masker. Zeng et al. found that intensity-increment thresholds were elevated above nonmasked values, particularly in the range of 40-60 dB SPL, in which intensity-increment thresholds might be elevated by as much as 5 dB. This is a very substantial change (Figure 7b). Figure 8 shows the Zeng model. The lower graphs illustrate the shifting of firing-rate curves as shown by Zeng et al. (Figure 7a) and the upper graphs illustrate the presumed intensity-increment thresholds.

The first experiment

The Zeng et al. experiment suggested that there may indeed be a dichotomy of neural recovery times in man. Ethical considerations forbid invasive neural studies in live humans; the characteristics of auditory afferent activity have not been recorded directly from the neuron in man, but only inferred from other mammals which have similar ultrastructure (the anatomical detail outside the cell), such as the cat, monkey, and Guinea pig. Hence we rely on psychophysics as did Zeng et al. We must start with Relkin and Doucet (1991), whose data showed a crossover between the mean recovery curves for low-spontaneous-rate units
Figure 7a. The Zeng et al. model of the effect of forward-masking on the operating ranges (shaded areas) of the low-vs. high-spontaneous-rate neurons (Zeng, Turner, and Relkin, 1991, Fig. 4). The thresholds of low-spontaneous-rate neurons are shifted upwards (arrow) relative to those of high-spontaneous-rate units, due to their slower recovery from the effects of the forward masker.

Figure 7b. (Zeng, Turner, and Relkin, 1991, Fig. 2) Just-noticeable-differences in intensity (in dB) for three human subjects, and their averages, either with a forward-masking narrowband noise (90 dB SPL, 900-1,100 Hz), or without. The jnds are expressed in units of intensity I rather than pressure P, but the resulting numbers in decibels are equivalent. The error bars express standard deviations across runs for each subject.
and high-spontaneous-rate units at very low time gaps (Figure 6). One might think that this crossover should be visible in psychophysical forward-masking data, in that it should produce a change of slope in the plot of forward-masked probe-tone detection threshold vs. the time gap between masker and probe. (This assumes that internal noise further up the neural chain is not strong enough to obscure such an effect.) However, there is a problem in attempting to infer recovery times from the Relkin and Doucet data. That is, their plotting procedure (noted above) renders the data relative and not absolute. Thus an apparent crossover may be illusory. The same plotting method was used, and the same caution applies, to later work by this group (Relkin, Doucet, and Sterns, 1995) showing recovery of the compound action potential (CAP) from forward-masking.

There is, nonetheless, some motivation for searching for a change of slope in auditory recovery. Auditory sensitivity to a brief tone (the probe stimulus) is decreased when the tone follows an intense conditioning stimulus (the forward-masker). Behaviourally, this loss of sensitivity results in an elevation of the probe's detection threshold (Ehmer and Ehmer, 1969; Weber and Moore, 1981; Zwicker, 1984; Scharf and Buus, 1986; Carlyon, 1988), an elevation that increases monotonically with rise in masker level (Zwicker, 1984; Fastl, 1979; Carlyon, 1988), and decreases monotonically with masker-probe time-gap as seen in all relevant literature. Physiologically, neural firing rates in response to a brief probe tone will be reduced following exposure to a forward masker whose frequency is the same as that of the probe (Smith, 1977, Mongolian gerbil; Harris and Dallos, 1979, chinchilla). We might therefore attempt to relate elevations in behavioural threshold to losses in neural sensitivity. Because the low-spontaneous-rate and high-spontaneous-rate populations of neurons have different post-masker recovery rates (Smith, 1977; Young and Sachs, 1973, cat; Relkin and Doucet, 1991, chinchilla), we might be able to identify, under the right conditions, the contributions of these two groups to the recovery of behavioural sensitivity.

The low-spontaneous-rate neurons typically have thresholds that are higher than those of the more numerous high-spontaneous-rate neurons, as we review in Chapter 2. It is sometimes speculated that low-spontaneous-rate neurons constitute a channel for high-intensity sounds. Such a channel would be reflected in the recovery curve for subjective detection thresholds, if it were possible to temporarily "disable" the high-spontaneous-rate neurons, for example by employing an intense forward masker. This conditioner should fully saturate the high-spontaneous-rate neurons because of their relatively low thresholds and limited
Figure 8. The Zeng et al. (1991) model of the effects of forward-masking on discrimination of increments in a short tone. The left-hand panels show the rate-level curves and corresponding discriminability limbs for the nonmasked tone, for pooled counts from populations of high-spontaneous-rate (lower threshold) and low-spontaneous-rate (higher threshold) units (based on actual computations of the kind presented in Chapter 3). The right-hand panels show the rate-level curves and corresponding discriminability limbs for the forward-masked tone. Note the loss of discriminability at moderate tone levels. Here, due to the fatigue induced by the forward-masker, a change in stimulus intensity causes less of a change in the firing rate of the neural population than for other tone levels.
dynamic ranges (as reviewed in Chapter 2). During and just following the forward-masker, the ability to detect a brief probe would thus be dependent on the high-intensity channel. Eventually the high-spontaneous-rate neurons will recover, ultimately determining the shape and asymptote of the plot of probe-detection threshold.

Thus, it might be possible to observe the auditory analog of the rod-cone break. Following an intense bleacher that saturates the rods, the curve describing recovering visual detection threshold for a small point of light shows a cusp midway through its course. The early part of the recovery curve reflects cone activity, whereas the later part reflects rod activity\textsuperscript{1,2}.

We assumed that the neural spontaneous-rate group that provides the lowest threshold for neural response to a probe is the one employed for psychophysical probe-detection. This should guarantee the lowest possible behavioural threshold. Under this assumption, any intersection between the recovery curves for two neural channels would appear as a cusp in the plot of recovering subjective probe-detection threshold. Because single-neuron recovery from a forward-masker can be substantial within a short duration post-masker, we sampled recovery with a high temporal resolution. An unambiguous masker-probe time-gap \( t \) demands rapid ramping of stimulus amplitude, introducing a spread in the stimulus' Fourier frequency components. We compensated by shaping the probe tone with a Gaussian envelope. For the tonal forward-masker the 200 msec duration, typical of forward-masking studies, ensures a fair deal of psychophysical forward-masking (Fastl, 1979, 2 msec probe with 1 msec Gaussian ramping; Kidd and Feth, 1982, 20 msec probe; Zwicker, 1984, 5 msec probe; Carlyon, 1988, 10 msec probe). We employed a high masker level, 97 dB SPL, to insure that all afferents probably suffered a good degree of forward-masking. Overall, psychophysical forward-masking should have been profound considering the exposure to this level, 10-30 dB higher than that of typical forward-masking studies, for a 200 msec duration.

Unlike the longer probes employed elsewhere (20-40 msec duration), the sort of very brief pips used here are probably not long enough to encode intensity as firing rate. Thus it might be thought that they do not obey the same rules as longer probes. The latter is unlikely, however. The very brief pips used here will undoubtedly provoke a CAP (compound action potential), usually reported as \( N_1 \) magnitude. The behaviour of \( N_1 \) can mimic that of firing rates and psychophysical thresholds. The \( N_1 \) for brief tone bursts can be forward-masked by tones (Abbas and Gorga, 1981, and Dolan et al., 1983; cats) with the same general kind
of dependence on masker frequency and intensity, and on probe frequency, as seen for tone-provoked neural firing rates (Smith, 1977, gerbils; Harris and Dallos, 1979, chinchillas) and for psychophysical detection thresholds (e.g. Fastl, 1979; Zwicker, 1984). For a stimulus similar to the present probe, human subjects produce a curve of \( N_1 \) vs. level that is similar in shape to that of other mammals (Gaussian pips of 3 msec total duration at 2,140 Hz; Hoke, 1974). Under forward-masking the \( N_1 \) curve shifts up-level (Cacace and Smith, 1986, Mongolian gerbil), with the degree of shift depending on masker level, just as found for curves that plot afferent firing rate vs. level for 40 msec tones (Smith, 1977).

In the first experiment, then, we charted forward-masked detection thresholds for a brief probe. The return of probe-detection thresholds to their nonmasked values (recovery) has not been examined unambiguously at small equal time increments even up to the first 10 msec post-masker. Thresholds are known at post-conditioner times spanning tens or hundreds of milliseconds but not in intimate detail just milliseconds after conditioner termination, when recovery rate is most crucial.

**The second experiment**

In the present study, the level of the forward-masker was varied while keeping the pip-to-masker intensity ratio constant at 20 dB. This ratio allowed the probe to stand out perceptually from the masker even at very brief time gaps. It also allowed the testing of a referential-coding model, as explained below. Note that certain conditions particular to the Zeng et al. (1991) experiment do not apply here. Therein, a narrow-band masker of 90 dB SPL preceded a 25 msec, 1 kHz tone by 100 msec. This masker presumably fatigued afferents of all spontaneous firing rates, which then recovered at different speeds (Figure 8). Here, it is assumed that the masker, even at its maximum intensity of 70 dB SPL, still does not shift the upper channel quite as far as shown in Figure 8. Also, it is possible that the relatively long probes used by Zeng et al. might have been long enough to affect the state of adaptation of the primary afferent neurons, causing more shifting of the mean-rate curves. To avoid this effect, a very brief pip was presently employed (see Chapter 4).

The forward-masker's actual effect on the intensity-increments can be anticipated as follows. The lowest employed pip level, 30 dB SPL, should be only a few dB above the absolute detection threshold for the nonmasked pip (confirmed in both Experiment 1 and Experiment 2). Thus the lowest pip accesses the lower channel, landing on the initial, steepening portion of the mean-rate curve (see Figure 8) and thus
Figure 9. Predictions of a model in which discriminability of an increment in pip-level is mediated by two neural channels. Each leftmost plot shows stimulus levels; moving from the bottom to the top, the pedestal level of the pip increases while maintaining a constant pip/masker intensity ratio. The open circles in the corresponding right-hand plots show the resulting discriminability limen (jnd) from 20 to 90 dB SPL, under three temporal conditions.
producing a reasonable intensity-increment. As pedestal level rises, the pip accesses higher and higher levels on the mean-rate curve of the lower channel, until this channel's firing rate starts to saturate, and discriminability starts to suffer. At some point the second (i.e. upper) channel will take over, again offering good discriminability until saturation sets in. This occurs as high as 80 dB SPL (Figure 8), an optimistic upper limit above which changes in mean-rate become negligible and the just-detectable intensity-increment should tend to infinity.

All this assumes that the pip itself is too short to provoke any adaptation. In the Zeng et al. model, adaptation caused by forward-masking shifts the mean-rate curves to higher sound pressure levels, so that the corresponding discriminability curve shifts accordingly (Figure 9). There is evidence that such shifting occurs for curves that plot afferent firing rate vs. level for 40 msec tones (Smith, 1977) as well as for curves expressing the growth of the N1 potential vs. level (Cacace and Smith, 1986). (The relevance of the latter reference will become clear in due course.) Thus, it is assumed from this point on that forward-maskers shift response-vs.-level curves for subsequent auditory stimuli, without change in shape, and with the degree of up-level shift being dependent on masker level. It is also assumed that a 10 dB increase in the masker level causes a corresponding 10-dB up-level shift in the response-vs.-level curves. This may be an exaggeration; examination of shifts in the plot of N1 potential vs. level (Cacace and Smith, 1986, Fig. 1) suggest a shift of perhaps 5 dB for a 10-dB change in masker level. Caution is advised, however, because the data are not sufficiently detailed to specify an equation expressing the degree of shift as a function of forward-masking level either for firing rates in response to pure tones (Smith, 1977), or for the N1 potential arising from initial rapid ramping of a pure tone (Cacace and Smith, 1986).

Now, in the Zeng et al. model, forward-masker-provoked shifting causes discriminability under forward-masking to differ from that in the absence of forward-maskers. To investigate this effect, intensity-increment thresholds were obtained at two widely separated times that allow vastly different degrees of recovery of detection threshold for this particular probe pip (as confirmed in Experiment 1) and for other very brief probe tones (Fastl, 1979; Zwicker, 1984). It is assumed, after Zeng et al. (1991), that fibers in the two spontaneous-rate groups have notably different recovery rates from adaptation. At a postmasker delay of t=3 msec, both channels should still be shifted up-level due to the forward masker. As the masker level is incremented in 10-dB steps from its lowest to its highest values, that is, from 10 to 70 dB SPL, the plots
of the discriminability afforded by the mean-rate curves will also shift up-level in 10-dB steps. Thus the pip, which starts out at 30 dB SPL, just a few dB above absolute threshold, will fall into and remain in the lower channel as masker and pip levels are made to simultaneously rise. Under the Zeng et al. model, the expected just-detectable intensity-increment at $t=3$ msec is thus a constant for the conditions of this experiment. It is however a higher constant than for the nonmasked just-detectable intensity-increment for the 30 dB SPL pip. The 10 dB SPL masker has presumably already started to shift the rate-level curves upward, so that the pip at first falls further onto the leftward tail of the curve, and then stays there, obtaining worse discriminability than for the nonmasked 30 dB SPL pip.

At $t=80$ msec, things change under the Zeng et al. model. The upper channel will still be up-shifted as in Figure 9 (right-hand side) but the recovery time will allow the lower channel to recover to its position for $t=\infty$ (no masker), regardless of the intensity of the forward-masker. Thus as the masker level increases, the lower channel affords the same discriminability as for the nonmasked pip. As pip level rises, this channel will saturate, and discriminability will presumably be taken over by the upper channel. For a mild masker, this channel will not be shifted too far up-level, and there can be a smooth transition between channels, assuring that the small just-detectable intensity-increment continues to be small. As the masker level increases, however, so does the overall degree of shift in the mean-rate curve. In the present experiment, as masker level rises, the pip's pedestal level rises by the same number of dB, so that the pip presumably provokes the same mean rate, corresponding to the same position on the mean-rate curve of the upper channel. Thus the just-detectable intensity-increment will stay low while the lower channel is being accessed, then rise as the lower channel saturates and the upper channel becomes responsible for discriminability. The just-detectable intensity-increment will then stay constant (but not very good) up to and perhaps beyond the highest level of the pip pedestal used in this experiment (90 dB SPL).

Thus the use of a fixed pip/masker intensity ratio produces different predictions about the just-detectable intensity-increment for each of the three different time gaps (Figure 9). There is a second motivation for using a constant pip/masker intensity ratio. Plack et al. (1995) investigated the concept of referential coding, which is very similar to "context coding" in which changes in pip level are encoded with reference to some fixed standard. Referential coding may occur in a two-interval forced-choice task, such as this one, when the subject detects an increment in a sound first by comparing the sound to a reference
Figure 10. The conditions in which referential coding might occur in a two-interval forced-choice experiment. The subject is assumed to perform a two-interval forced-choice experiment in which the target to be detected is presented in one of two comparison intervals (as in Figure 5). It is assumed that the forward-masker (dashed line) has a moderate-to-high intensity, as found in most forward-masking studies, and that the incremented pip occurs in the second interval. In a), the pip is near the threshold of hearing, and pip increments are encoded with respect to this level. In b), the threshold of feeling provides the reference point. In c), increments are encoded with respect to the level of the forward-masker.
level within each of the two intervals, and then comparing these impressions across the two intervals to make a choice (Plack et al., 1995). Within each interval, internal comparison levels may be available to low and high pip levels (Figure 10). Discrimination at low pip levels would be good because the pip would be referenced to the threshold of hearing, allowing small changes to be easily detected. Discrimination at high pip levels would be good because the pip could be referenced to the threshold of discomfort, with the same result. Alternatively, the pip could be encoded with reference to a forward-masker providing an external comparison level that may override the internal comparison level. The forward-masker could take the place of the internal reference when the latter is too far from the pip level to be an effective reference. Thus a low-level forward-masker might provide an adequate reference for a low-level pip, and a high-level forward-masker might provide an adequate reference for a high-level pip. However, a forward-masker whose level does not change as the pedestal level changes (pedestal=10 dB SPL, 20 dB SPL, 30 dB SPL, etc.) cannot provide a reference of constant character for judging when the pip’s intensity has been incremented on a given pedestal level. Thus, for instance, for a high-level forward-masker and moderate pip levels, a useful internal or external context is absent. Discriminability is weakened, as manifested by the mid-level hump. However, the mid-level hump should disappear when maintaining a constant context (not a constant level) for detection of the increment in the probe pip. Such a listening standard is provided by keeping the pedestal/masker intensity ratio constant for all pedestal levels.

Figure 11, done in the style of Figure 9, shows the putative effect of a constant pip/masker ratio. At t=0 (nonmasked pip), the reference is absent, and a mid-level hump appears that is assumed to have the same shape and magnitude as that for t=3 msec seen in Figure 9. As the masker and pip get closer together time-wise, the masker is presumed to provide an increasingly effective reference, perhaps due to better sensory integration of the neural responses to masker and pip. Discriminability thus improves relative to that for the nonmasked pip. The predictions of the two-channel Zeng et al. model and the referential-coding model are similar in character at t=3 msec (we do not know how the magnitudes compare), but differ for t=80 msec and at t=∞, so that the two models cannot both describe the just-detectable intensity-increments3.

It has been assumed, so far, that the forward-masker causes shifting of both the upper and lower auditory channels. But very low masker levels may fall below the upper channel’s threshold, so that the
Figure 11. Predictions of a model in which the forward-masker provides a reference level against which changes in pip level can be more easily perceived than if the masker was absent (= the column marked "t = ∞"). Each leftmost plot shows stimulus levels; moving from the bottom to the top, the pedestal level of the pip increases while maintaining a constant pip/masker intensity ratio. The open circles in the right-hand plots show the corresponding discriminability limen (jnd) from 20 to 90 dB SPL, under three temporal conditions. Predicted discriminability is flat at t=3 msec as for the Zeng et al. (1991) model (Figure 9); since it is not known how the magnitudes of the just-detectable intensity-increments compare across models, they are assumed, for simplicity's sake, to be equal.
upper channel has no impetus to shift. What happens under such circumstances? Let us look first at the Zeng et al. model under the assumption on which we have proceeded so far, that the auditory system attempts to maintain the best possible discriminability despite the forward-masking. For a time gap of 3 msec, the tone pip's pedestal level will fall first into the lower channel, which is assumed to shift up-level with increases in masker intensity. Because the upper channel is not shifting, the pedestal will start to move into the range of good discriminability for this channel. Thus discriminability will actually improve, beyond that shown in Figure 9, with increase in masker level. At some point, the upper channel will start to shift, and this improved discriminability will persist. Eventually, some forward-masker level will be reached beyond which there is no more shifting, and discriminability will progressively weaken with rise in pedestal level.

Now for the pip pedestal presented at \( t = 80 \) msec, a perceptual gap is still present, as before, due to different recovery rates of the two auditory channels. Thus the pattern of discriminability will not change, from that shown in Figure 9, for mid-to-high pedestal levels. At pedestal levels too low to cause significant shifting of the upper channel, there will be an actual improvement in discriminability as compared to Figure 9, for the same reasons as noted above for the \( t = 3 \) msec condition. For the nonmasked condition, there is no shifting of the rate-level curves, and hence no change in the just-detectable intensity-increment as compared to Figure 9.

Under the Plack et al. model, discriminability of the tone-pip's increment should not change at all, since the intensity ratio of the forward-masker to the tone-pip was not altered. The predictions of referential-coding are independent of the Zeng et al. model of channel shifting and recovery.
NOTES

1. We searched for different spontaneous-rate groups in man by analogy with human vision, in which receptors serving different intensity ranges produce a psychophysical recovery curve from intense-light bleaching which shows several stages for detection of a small, brief, dim target (Hood and Finkelstein, 1986). Coincidentally, Dewson (1967) had suggested that the system producing loudness was similar to the rod-cone system in vision. However, rods, the low-light cells used in night vision, have a fixed stimulus-encoding range, whereas the cones (day vision) shift their [group] mean operating point (roughly, the average of the centers of their individual dynamic ranges) based on ambient light level (Normann, Perlman, and Hallett, 1991). Mammalian auditory systems are analogous. The ear has a feedback control loop whose common mechanism (Brown, Nuttall, and Masta, 1983) is the modulation of the outer-hair-cell's amplification of basilar-membrane motion, a modulation done by the medial olivocochlear bundle (MOCB) (Dulon and Schacht, 1992). This control system's neural substrate exists in man (Rasmussen, 1946; Richter, Norris, Fullerton, Levine, and Kiang, 1983; Nadol, 1990) and other mammals (Rasmussen, 1946). In anesthetised cats, whose olivocochlear nerve bundles are electrically stimulated to offset the presumed effects of anesthesia, the auditory afferents shift their useful stimulus-encoding ranges with changes in ambient noise level (Dewson, 1967; Nieder and Nieder, 1970; Winslow and Sachs, 1988), which presumably is what allows behavioural tone detection at high noise levels (Costalupes, 1983). In response to tones presented alone, the electrically stimulated OCB shifts all the fibers' tone-evoked rate-response functions to higher sound pressure levels (plotted as firing rate vs. sound level in dB SPL), without (to a first approximation) changing their shapes (Wiederhold, 1970; Gifford and Guinan, 1983). This action is described further in Chapter 3. Similarly, shifting of cone potential with ambient light level preserves the shape of the plot of the cone's cellular potential vs. the logarithm of light intensity (Hood and Finkelstein, 1986; Normann, Perlman, and Hallett, 1991). Thus the auditory system's correct analog in vision is the cone system alone. In the retina's fovea, which has only cones, three cone mechanisms can appear in the recovery curve; careful inspection reveals separate branches of the recovery curve separated by two abrupt dislocations for white-light bleacher and test spot (Hood and Finkelstein, 1986; Cicerone, Hayhoe, and MacLeod, 1990). There appears an obvious single dislocation for pure, non-identical bleacher and test wavelengths (Eisner, 1986). Presumed cone-cone breaks can appear either as a sudden drop in the recovery curve (Auerbach and Wald, 1955; Hood and Finkelstein, 1986) or as a temporary threshold rise (Eisner, 1986; Cicerone, Hayhoe, and MacLeod, 1990). Even in the parafovea (the area surrounding the fovea), recovery can reveal different cone branches (Auerbach and Wald, 1955). Of course, a transition between two auditory channels must be neurally regulated at no lower than the cochlear nucleus, the first possible point of comparison of afferent firing (Osen, 1970, confirming earlier studies; Fekete, Rouiller, Liberman, and Ryugo, 1984), whereas MOCB-induced dynamic-range-shifting is ultimately a mechanical phenomenon of the basilar membrane itself (Wiederhold, 1986; Dulon and Schacht, 1992).

2. Visual and auditory recovery differ in the time required for complete recovery to nonmasked performance. Such auditory recovery is on the order of hundreds of milliseconds (Scharf and Buus, 1986) whereas foveal cone recovery (back to the dark-adapted state) takes minutes. The differences in orders-of-magnitude of time continue when considering the forward-maskers required; the rod-cone break in the psychophysical recovery plot appears only with higher bleaching intensities and long exposures (4-5 minutes) to the adapting light (Hood and Finkelstein, 1986). Even a series of plot discontinuities, that imply three cone mechanisms with different recovery times, requires a long, intense adaptor (Hood and Finkelstein, 1986). Regardless, the degree of forward masking depends systematically on adaptor intensity for both single cone mechanisms (bleach and test light of the same wavelength; Hollins and Alpern, 1973) and possible multiple mechanisms (differing wavelengths; Eisner, 1986).

3. Referential coding is hard to evaluate from the data provided by Plack et al. (1995), however, since their relatively broad error bars, combined with strong individual differences in their intensity-increment
thresholds, makes data interpretation difficult. However, some support for referential coding comes from Turner et al. (1994), who used a 102 msec narrowband noise to forward-mask a 2 kHz probe tone of about 25 msec duration, which had either very gradual ramps and a brief intensity plateau, or very short ramps and a long plateau. (The common factor for these tones was their very close threshold values in quiet.) With 2 msec separating masker offset and probe onset, the longer-ramped probe tones showed higher thresholds, suggesting that the shorter-ramped probe's greater Fourier energy splatter (due to shorter ramping) interacted with the spectrum of the narrowband noise, as an aid to detection.
CHAPTER 2

COMPUTATIONAL METHODS
We thank Dan Schnabel for the integral for the wedge-shaped threshold distribution, mentioned in the section below and shown in Appendix A. We thank Drs. M.C. Liberman and A.R. Palmer, for the use of their original data, and Suzanne P. McKee and Scott N.J. Watamaniuk for the probit analysis routine.

All subsequent neural references pertain to the cat unless stated otherwise. A recent model of discrimination in the cat (Delgutte, 1987) has the same aims of accounting for the statistical nature of certain fiber characteristics, but differs in the way that dynamic range is computed, and reaches somewhat different conclusions (see the Discussion at the end of this chapter).

THE SINGLE CHANNEL POINT MODEL AND THE CRITICAL BAND

Parameters and rate-level functions of single nerve fibers

We model the sigmoidal rate-level plot typical of most afferents (Kiang, Watanabe, Thomas, and Clark, 1965; Sachs and Abbas, 1974; Liberman, 1978) with a logistic function that includes threshold $\epsilon$, dynamic range $\lambda$, maximum (saturation) rate $r_{\text{max}}$, and spontaneous rate $r_s$ as parameters:

$$r(x) = \frac{r_{\text{max}} - r_s}{1 + \frac{100 - c}{c} \exp\left\{-2 \frac{\epsilon - x}{\lambda} \ln\left(\frac{100 - c}{c}\right)\right\}} + r_s$$

where $x$ is stimulus level in dB SPL.

Determining a unit's threshold requires a decision rule for deciding when the unit is firing above its spontaneous rate; here we will use the dB level at which the firing rate is equal to the spontaneous rate plus $c\%$ of the difference ($c = 2$) between the saturation rate and the spontaneous rate (point $\epsilon$ in Figure 2). Similarly, we define the dB level just short of saturation as the SPL where firing rate is equal to spontaneous rate plus $(100 - c)\%$ of the difference between the saturation rate and the spontaneous rate (indicated by $\epsilon + \lambda$ in Figure 2 for $(100 - c)\% = 98\%$). Firing rates at threshold and saturation are now respectively given by the decision rules

$$r_\epsilon = \frac{c}{100}(r_{\text{max}} - r_s) + r_s, \quad r_\epsilon + \lambda = \frac{100 - c}{100}(r_{\text{max}} - r_s) + r_s.$$  

Dynamic range $\lambda$, therefore, is presently defined as the dB difference between points $\epsilon$ and $\epsilon + \lambda$. It does not matter to any of our conclusions if we change the percentage values above to $5\%$ and $95\%$ respectively ($c=5$), or to $1\%$ and $99\%$.
respectively \((c=1)\). Therefore, defining threshold and dynamic range in this way will not substantially alter the conclusions in this paper. Unlike earlier equations for single-unit firing rate (Schiaffino, 1957; Barducci, 1961; McGill and Goldberg, 1968; Goldstein, 1974; Howes, 1974; Sachs and Abbas, 1974; Lachs, Al-Shaikh, Bi, Saia, and Teich, 1984; Sachs, Winslow, and Sokolowski, 1989; Yates, Winter, and Robertson, 1990), dynamic range \(\lambda\) (in decibels) now appears explicitly.

In spite of variety in fiber characteristics (Figure 12), Equation 1 fits typical rate-level plots quite well (Figure 13). It is important to determine the extent to which the parameter values estimated via least-squares fitting correspond to the parameter values obtained by visual estimation from the original rate-intensity plots. Table 1 shows that both ways of estimating parameters yield values that agree quite closely.

This curvefitting exercise also allows the opportunity to test the accuracy of the criterion that we assumed in defining dynamic range for a single fiber. Recall that the upper limit of \(\lambda\) is assumed as that SPL for which \(r - r_t = 0.98(r_{\max} - r_t)\), and that the lower limit of \(\lambda\) is at that SPL for which \(r - r_t = 0.02(r_{\min} - r_t)\). There is no guarantee that this definition is the most appropriate one; limits of \([.1,.9]\) or \([.05,.95]\) might produce better estimates of actual dynamic range. To find the most appropriate values for defining the dynamic range of a single unit, we repeated the curvefitting exercise illustrated in Figure 13. This time we successively varied the dynamic range criterion over four different pairs of values for each regression. The results (Table 2) show that the \([.02,.98]\) criterion comes closest to the values of dynamic range that were visually estimated from the digitised data. For \(c=1\) predicted dynamic ranges are closer to those of Liberman (Fig. 2 in Liberman, 1988); for \(c=2\) predicted ranges are closer to those typical of Evans and Palmer (1980). These differences likely represent different dynamic range criteria employed by these authors (as discussed in Chapter 6).

Afferents are classified by rate-level function as sigmoidal or sloping-saturating, and by spontaneous firing rate \(r_0\) (Liberman, 1978) as low \((r_0 < 0.5 \text{ spikes/sec})\), medium \((0.5 \leq r_0 \leq 18 \text{ spikes/sec})\), or high \((r_0 > 18 \text{ spikes/sec})\). Sloping-saturating fibers are a portion of low-\(r_0\) fibers whose rate-intensity curves show a change of slope at a particular sound pressure level, above which the slope becomes notably shallower but not flat, so that the apparent dynamic range of the fiber is increased relative to a sigmoidal fiber. The six units pictured in Sachs and Abbas (1974, Fig. 2) were digitised and fitted to Equation 1 using two different
Figure 12. Variability in fiber properties of primary afferents of the cat's cochlear nerve. Any or all units pictured here can in principle be at the same characteristic frequency (CF), corresponding to nearby places on the basilar membrane. Fibers have different spontaneous rates $r_s$ (compare a and b), different saturation rates $r_{sat}$, and hence different ranges of firing rate (c), and may not saturate at all in the dB range studied (such as the sloping-saturating fibers described in the text; d). Units with similar $r_s$ and $r_{sat}$ may have different dynamic ranges $\lambda$ (e) due in this case to differing slopes (f). Thresholds $\epsilon$ are not necessarily the same for fibers of the same CF (g, h).

Figure 13. The fit of Equation 1 to rate-level data (circles). Sources: CF=1.4 kHz: Units #225, 22, 27 (Liberman, 1978); CF=0.28, 1.12, 5.30, and 8.96 kHz: Sachs and Abbas, 1974, Fig. 2, right-hand side; CF=9.22 kHz: Sachs, Winslow, and Sokolowski, 1989, Fig. 1. Table 1 presents the least-squares estimates of the four parameters: fiber dynamic range, threshold, saturation rate, and spontaneous rate.
Table 1. Comparison of regression-fitted (F) and visually-estimated (VE) parameters for the rate-intensity functions fit to individual neurons (identified by characteristic frequency, CF) shown in Figure 13.

<table>
<thead>
<tr>
<th>CF</th>
<th>saturation rate (1/s)</th>
<th>spontaneous rate (1/s)</th>
<th>dynamic range in dB</th>
<th>threshold in dB SPL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>VE</td>
<td>F</td>
<td>VE</td>
</tr>
<tr>
<td>1.4</td>
<td>170.5</td>
<td>175</td>
<td>57.6</td>
<td>56.3</td>
</tr>
<tr>
<td>1.4b</td>
<td>178.9</td>
<td>182.5</td>
<td>3.9</td>
<td>3.83</td>
</tr>
<tr>
<td>1.4c</td>
<td>36.6</td>
<td>45</td>
<td>2.1</td>
<td>0.03</td>
</tr>
<tr>
<td>0.28d</td>
<td>135.3</td>
<td>135</td>
<td>1.42</td>
<td>2.85</td>
</tr>
<tr>
<td>1.12d</td>
<td>168.6</td>
<td>166.5</td>
<td>0</td>
<td>11.0</td>
</tr>
<tr>
<td>5.30d</td>
<td>242.8</td>
<td>238.5</td>
<td>80.1</td>
<td>87.3</td>
</tr>
<tr>
<td>9.22e</td>
<td>214.4</td>
<td>210</td>
<td>114.7</td>
<td>115.4</td>
</tr>
</tbody>
</table>

*Unit #225 from Liberman (1978). The visually estimated spontaneous rate was given by Liberman. ‡Unit #22 from Liberman (1978). The visually estimated spontaneous rate was given by Liberman. §Unit #227 from Liberman (1978). The visually estimated spontaneous rate was given by Liberman. †From Sachs and Abbas (1974), Figure 2, right-hand side. Units identified by CF. ‡From Sachs, Winslow, and Sokolowski (1989), Figure 1. Unit identified by CF.
Table 2. Comparison of regression-fitted versus visually estimated (VE) dynamic range for different values of the constant \(c\). The fitted dynamic ranges rise monotonically as the ratio (upper limit/lower limit) increases. When this ratio equals 9 (=.9/.1) the predicted dynamic ranges are closer to those for the units illustrated in Liberman (1988, fig. 2); predicted dynamic ranges for a ratio of 49 (=.98/.02) are closer to those typical of Evans and Palmer (1980). These differences presumably represent a more conservative definition of dynamic range by Liberman.

<table>
<thead>
<tr>
<th>CF</th>
<th>1-c, c</th>
<th>VE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4(^a)</td>
<td>90,10</td>
<td>95,5</td>
</tr>
<tr>
<td>1.4(^b)</td>
<td>18.6</td>
<td>24.9</td>
</tr>
<tr>
<td>1.4(^c)</td>
<td>6.10</td>
<td>8.20</td>
</tr>
<tr>
<td>0.28(^d)</td>
<td>24.2</td>
<td>32.4</td>
</tr>
<tr>
<td>1.12(^d)</td>
<td>27.7</td>
<td>37.1</td>
</tr>
<tr>
<td>5.30(^d)</td>
<td>15.4</td>
<td>20.6</td>
</tr>
<tr>
<td>8.96(^d)</td>
<td>19.9</td>
<td>26.6</td>
</tr>
<tr>
<td>9.22(^e)</td>
<td>24.3</td>
<td>32.6</td>
</tr>
</tbody>
</table>

\(^a\)Unit #225 from Liberman (1978). The visually estimated spontaneous rate was given by Liberman. \(^b\)Unit #22 from Liberman (1978). The visually estimated spontaneous rate was given by Liberman. \(^c\)Unit #227 from Liberman (1978). The visually estimated spontaneous rate was given by Liberman. \(^d\)From Sachs and Abbas (1974), Figure 2, right-hand side. Units identified by CF. \(^e\)From Sachs, Winslow, and Sokolowski (1989), Figure 1. Unit identified by CF.
strategies. First of all, the four parameters \( \{\epsilon, \lambda, r_{\text{max}}, r_s\} \) were allowed to float freely. In a subsequent exercise two parameters, \( \lambda \) and \( r_{\text{max}} \) were fixed. These were the two that were easiest to visually estimate from the digitised data. This second fitting was done to check for any unusual or unjustifiable closeness-of-fit that might be ascribed to having fully four parameters freely-floating. These two approaches produced similar results; neither one gave the kind of very good fit found above for ordinary sigmoidally-saturating fibers (Table 3). In either case, each fiber's fitted dynamic range corresponds more closely to that of the initial rapidly rising portion of the rate-level curve than to that of the entire curve. Thus the upper, more gently-sloping portion of the sloping-saturating firing-rate behaviour is poorly described.

Hence a weighted double-logistic function,

\[
r_{ss} = \gamma \cdot r_1(c, \epsilon, \lambda_1, r_{\text{max}}, r_s, x) + (1 - \gamma) \cdot r_2(c, \epsilon, \lambda_2, r_{\text{max}}, r_s, x),
\]

was used to fit the rate-level functions of these fibers (Figure 14). This novel function is not meant to imply the mechanism of sloping-saturation; however, other equations (i.e. Ohlmiller, Echteler, and Siegel, 1991), although they may provide a phenomenological account (i.e. Sachs, Winslow, and Sokolowski, 1989; Yates, 1990), do not fit the elbow as sharply. Also, Equation 3 extrapolates beyond the data points, which were typically absent above 90-95 dB. For example, the unit shown in the top left hand corner of Figure 14 was not subjected to intensities higher than 60 dB, yet Equation 3 extrapolates the firing rate to 115 dB SPL.

Earlier models of collected neural response (Schiaffino, 1957; Barducci, 1961; McGill and Goldberg, 1968; Goldstein, 1974; Howes, 1974; Lachs, Al-Shaikh, Bi, Saia, and Teich, 1984; Delgutte, 1987; Viemeister, 1988; Winslow and Sachs, 1988) did not account for the behaviour of sloping-saturating units.

**Obtaining a mean rate-intensity function**

Clearly, if we could measure the rate-intensity plots of all of the neurons in a given region along the basilar membrane, we could determine the average rate-intensity plot. Alternatively we could determine the average rate function by averaging the logistic functions fit to these rate-level functions, which we found works very well for a small population (Figure 15). We performed a similar exercise for 7 sloping-saturating units, by fitting both Equations 1 and 3 to the same data, then averaging the firing-rate values computed from the functions themselves (Figure 16). Although the double-logistic functions provide a better fit to
Figure 14. The fit of the weighted double-logistic rate function (Equation 3) to seven examples of sloping-saturating units (circles). Two dynamic ranges $\lambda_1$, $\lambda_2$ and a single weighting factor, $\gamma$, were fitted for each fiber. Threshold and saturation rate were visually estimated from the digitised fiber firing rates; $r = 0$, which improves the robustness of the parameters of the fitted function. The top two curves are from Sachs and Abbas (1974, Fig. 2); the bottom five, from Palmer and Evans (1979, Fig. 2).

Figure 15. Averaging of actual firing-rate plots vs. averaging of the logistic function (Equation 1). All fibers obey $r \geq 15$/sec and $2.5 \leq$ CF $\leq 25$ kHz, the CF range for which the computations are most appropriate (see Table 4). Actual rate-level curves are digitised from a), b) Sachs and Abbas, 1974, Fig. 2; c) Palmer and Evans, 1979, Fig. 2; d) Sachs, Winslow, and Sokolowski, 1989, Fig. 1; e) Palmer and Evans, 1979, Fig. 3a. They are fitted by least-squares regression to Equation 1. To facilitate averaging, 1) individual rate-intensity data-plots were blocked into 5 dB bins, as were the fitted functions; and 2) spontaneous rates and saturation rates were continued at their lowest and highest measured values respectively. The bottom-right panel shows the averaged firing rates (circles) and the averaged curve-fitted rates from Equation 1 (actual line).
Table 3. Comparison of regression-fitted parameters for the logistic rate-intensity function (Equation 1) fitted to sloping-saturating behaviour using four fitted parameters (F), vs. two fitted parameters (T) and two fixed parameters (in brackets, visually estimated from digitised data of Sachs and Abbas, 1974, Fig. 2, left-hand side).

<table>
<thead>
<tr>
<th>CF (kHz)</th>
<th>saturation rate (1/s)</th>
<th>spontaneous rate (1/s)</th>
<th>dynamic range in dB</th>
<th>threshold in dB SPL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>0.32</td>
<td>227.9</td>
<td>232.1</td>
<td>4·10⁻⁷ (0.8)</td>
<td>44.3</td>
</tr>
<tr>
<td>1.76</td>
<td>133.2</td>
<td>131.4</td>
<td>2·10⁻⁶ (0.0)</td>
<td>45.4</td>
</tr>
<tr>
<td>4.30</td>
<td>197.7</td>
<td>193.1</td>
<td>2·10⁻⁵ (9.8)</td>
<td>51.9</td>
</tr>
<tr>
<td>7.74</td>
<td>218.5</td>
<td>217.8</td>
<td>4.33 (15.6)</td>
<td>49.8</td>
</tr>
<tr>
<td>14.5</td>
<td>265.1</td>
<td>256.5</td>
<td>8·10⁻⁵ (50.2)</td>
<td>54.6</td>
</tr>
<tr>
<td>20.5</td>
<td>224.3</td>
<td>219.2</td>
<td>24.2 (32)</td>
<td>41.2</td>
</tr>
</tbody>
</table>
the data, single-logistic functions also provide a very good fit. Moreover, if we average the two sets of functions for the seven units in question, as shown in the lower right-hand panel, it can be seen that there is very little difference in the averages of these two functions except at the extreme edges of the range. Therefore, the single-logistic function, when averaged over a number of units, appears to provide an adequate description of the behaviour of an aggregate of sloping-saturating functions. However, as we shall see in Chapter 3, we do not wish to overestimate the useful dynamic ranges of sloping-saturating units. Hence we continue to describe these units using Equation 3.

To sidestep these kinds of computations, we can also measure fiber characteristics in the largest available sample of local neurons, and infer the probability distributions \( p(\varepsilon), p(\lambda), p(r_{\text{max}}), \) and \( p(r_s) \) for these characteristics. Given these probability distributions we can then use standard statistical techniques to determine the expected value of Equation 1 that defines the average rate-level function for the population in question (see below). In this way the behaviour of the fiber ensemble is described as if it were a single unit. The method is more elegant and inclusive of neural variation than others which average firing rates over a few “representative” units (e.g. Siebert, 1965; Goldstein, 1974; Howes, 1974; Viemeister, 1983; Winslow and Sachs, 1988).

The distribution of parameters may differ across the three recognised \( r_s \) groups, so that if \( \bar{r}_L(x) \), \( \bar{r}_M(x) \), \( \bar{r}_H(x) \) and \( \bar{r}_{SS}(x) \) are respectively the mean rate-intensity functions for the low-, medium-, and high-spontaneous-rate groups, and of the sloping-saturating fibers, the overall mean response-intensity function is

\[
\bar{r}(x) = w_L \bar{r}_L(x) + w_M \bar{r}_M(x) + w_H \bar{r}_H(x) + w_{SS} \bar{r}_{SS}(x)
\]  

(4)

where the \( \{w_i\} \) are the relative frequencies of the respective fiber groups. Within an \( r_s \) group \( p(\varepsilon), p(\lambda), p(r_{\text{max}}), \) and \( p(r_s) \) are assumed to be mutually independent. There may also be a high-threshold population of fibers with straight (rather than sigmoidal) rate-level functions, producing a linear plot vs. dB SPL (Winter, Robertson, and Yates, 1990). Unfortunately the properties of these units are not well-known.

**The neural population: distributions of characteristics, and pool size**

We need probability distributions for \( \varepsilon, \lambda, r_{\text{max}} \) and \( r_s \) within each \( r \) group, but only plots for spontaneous rates (Liberman, 1978; Kim and Molnar, 1979) and dynamic ranges (Evans and Palmer, 1980; Schalk and Sachs, 1980) have been published. We compiled new histograms and fitted probability densities
Figure 16. The fit of the weighted double-logistic rate function (Equation 3) and that of the single-logistic (Equation 1) to the seven examples of sloping-saturating units seen in Figure 14 (circles). The logistic function is the sigmoidal curve fitted to each data set. The lower right-hand plot shows the average of the rate-level data vs. the averages of the values computed from the seven fitted curves for the double-logistic and logistic functions.
Table 4. Parameter values used in initial simulations of computed auditory dynamic range.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>Symbol</th>
<th>Distribution mean and standard deviation by r, group</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low-r, Mid-r, High-r, Sloping</td>
<td></td>
</tr>
<tr>
<td>Spont. rate</td>
<td>spikes/s</td>
<td>r,</td>
<td>0.4</td>
<td>Liberman, 1978</td>
</tr>
<tr>
<td>Saturation rate</td>
<td>spikes/s</td>
<td>r_max</td>
<td>200</td>
<td>Liberman, 1978; Evans &amp; Palmer, 1980</td>
</tr>
<tr>
<td>Threshold</td>
<td>dB SPL</td>
<td>e</td>
<td>42.3</td>
<td>Evans &amp; Palmer, 1980; Geisler, Deng, &amp; Greenberg, 1985</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>32.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22.3</td>
<td></td>
</tr>
<tr>
<td>Dynamic range</td>
<td>dB</td>
<td>λ</td>
<td>40.2</td>
<td>Evans &amp; Palmer, 1980</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40.2</td>
<td></td>
</tr>
<tr>
<td>Weighting Factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*We assumed $r_\text{sloping} = 0.4/\text{sec}$ also for sloping-saturating fibers, which typically obey $r_s<0.5$ spikes/sec (Liberman, 1988).

Liberman (1978) observed 10 dB differences in $\bar{e}$ between low, medium, and high spontaneous rate groups; we maintain these differences, which may be upper estimates. There are differences in published thresholds which most likely stem from definitional differences of threshold. Liberman’s (1978) measurements were based on an observational count-difference criterion. Rhode and Smith (1985), using similar demarcations, report an intergroup difference of only 10 dB. The thresholds of Geisler, Deng, and Greenberg (1985) were based on spontaneous-rate count statistics, and showed an intergroup difference of at best 5 dB. Regarding sloping-saturating fibers, these typically have very low spontaneous rates (Liberman, 1988; Evans and Palmer, 1980; Schalk and Sachs, 1980) and thresholds within the range found for other low-r, units, we assumed a threshold distribution, $p(e)$, identical to that of the low-r, units. Pooling (Evans and Palmer, 1980; Geisler, Deng, and Greenberg, 1985) gave 365 high-r, units and 112 low-r, units (Figure 19; Geisler’s $r$, definitions [ibid.]). A smoothed curve (Cleveland, 1970) indicated that thresholds for high-r, units were relatively flat over 2.5-25 kHz. Consequently, high-r, units were pooled across CF. Threshold range was similar for each pooled $r$, group, so a common distribution (Figure 17) and common standard deviation were assumed for all $r_\lambda$. Sufficient data was available for high-r, units only. A common distribution pattern (Figure 18) and common standard deviation was assumed for the other two $r$, groups. Parameter values are the average obtained from the 7 plots of Figure 14 plus four others (Sachs and Abbas, 1974, Fig. 2, left-hand side; CFs = 4.30, 7.74, 14.5, and 20.5 kHz). The eleven $\lambda_1$ values were consistent with the assumption that they were sampled from $N(40.2,9.57)$. The mean and standard deviation of the $\lambda_1$ values were 88 and 22 dB, respectively. The mean $\gamma$ was 0.68; standard deviation $\sigma_\gamma = 0.08$. |
for ε (Figure 17) and λ (Figure 18) for these and other data (Table 4), and by examining the plotted data (Figures 19, 20, and 21) both here and in the published literature, we established mutual independence within each \( r_e \) group (Table 5) between \( ε, λ, r_{\text{max}} \) and \( r_e \), and between these parameters and CF (with the possible exception of \( r_{\text{max}} \); see below). Data was sometimes pooled to avoid errors of small sample size; when the data was insufficient, inferences were made based on behavior in larger samples\(^3\). Only mean values are needed for \( r_e, r_{\text{max}}, \) and \( γ \) (see below).

We referred above to the computations being a “point model”. This is because even the units driven by an individual inner hair cell, the basic unit of cochlear transduction, can have different thresholds. Therefore even a very localised point on the basilar membrane will have a distribution of threshold values, and it is not unreasonable to postulate a distribution of fiber thresholds at each CF (see the argument in Smith, 1985). This explains why the operating range of an IHC (sound pressure levels to which it responds) may be greater than that of any single contacting afferent (Mongolian gerbil; Goodman, Smith, and Chamberlain, 1982). Furthermore, primary afferents from the same IHC may belong to any of the spontaneous rate classifications (Liberman, 1982; Liberman and Oliver, 1984).

To obtain the jnd in dB, here called \( Δx \) where \( x \) is expressed in dB SPL, we need to know the size of the assumed neural pool. The basilar region corresponding to a critical band at 8 kHz (Nienhuys and Clark, 1979; Liberman, 1982) is approximately 0.7 mm in length in the cat\(^6\). With 105 cells/mm and 25 afferents/cell (Keithley and Schreiber, 1987) we obtain 1,875 neurons. Human studies (Green review, 1988) suggest a band about half as wide, giving 938 units, divided by \( r_e \) in the proportions 17%, 25%, and 58% (Liberman, 1978; Geisler, Deng, and Greenberg, 1985). About 1/2 of low-\( r_e \) units may be sloping-saturating (Liberman, 1988), corresponding to 80 units; the rest were 80 low-\( r_e, 234 \) medium-\( r_e, \) and 544 high-\( r_e^{7,8} \).

We assume that sound intensity is encoded by the total number of spikes summed over a pool of neurons (the auditory channel) within a specified time period. Thus, following Siebert (1965), we assume for a constant stimulus: (1) that spike timing does not contribute to intensity coding; (2) that initial neural transients, whatever their usefulness, can be ignored; (3) that individual units are mutually independent; (4) that the variance in each unit’s spike count (over a set counting time, the tone’s duration) at intensity \( x \) is approximately the same as at \( x + Δx \), if \( Δx \) is sufficiently small (Siebert, 1965); and (5) that no noise is added at or beyond the counting center (McGill and Goldberg, 1968; Lachs, Al-Shaikh, Bi, Saia, and Teich, 1985).
Figure 17. Probability density functions for threshold. The stepped function is a normalisation of the threshold histogram ($\bar{\epsilon} = 21.66$ dB SPL, $\sigma_\epsilon = 15.02$). The wedge-shaped approximation has been shifted 1 dB to the left for clarity, and is described by

$$p(\epsilon) = -a\epsilon + \frac{a}{2}(\epsilon_H + \epsilon_L) + \frac{1}{e_H - e_L}$$

where $\epsilon_L$ and $\epsilon_H$ are lower and upper boundaries of $p(\epsilon)$, and the fitted constant $a = 3.79 \times 10^4$ is invariant under shifts of $p(\epsilon)$ along the $\epsilon$ dimension. The wedge has mean value

$$\bar{\epsilon} = -\frac{a}{3}(\epsilon_H^3 - \epsilon_L^3) + \frac{a}{4}(\epsilon_H^2 - \epsilon_L^2)(\epsilon_H + \epsilon_L) + \frac{1}{2}(\epsilon_H + \epsilon_L)$$

and variance

$$\sigma_\epsilon^2 = -\frac{a}{4}(\epsilon_H - \epsilon_L)(\epsilon_H^3 + \epsilon_H^2 \epsilon_L + \epsilon_H \epsilon_L^2 + \epsilon_L^3) + \frac{a}{6}(\epsilon_H - \epsilon_L)(\epsilon_H^2 + \epsilon_L^2)(\epsilon_H + \epsilon_L) + \frac{1}{3}(\epsilon_H^2 + \epsilon_H \epsilon_L + \epsilon_L^2)$$

which for $\{\epsilon_L, \epsilon_H\} = \{0.55\}$ gives $\bar{\epsilon} = 22.24$ dB SPL and $\sigma_\epsilon = 14.99$ dB. A small proportion of the units supplied had thresholds in the range $-5 - 0$ dB SPL. This small probability step represents 1.83% of all units, which when included did not substantially alter the fitted value of $a$ or the mean threshold $\bar{\epsilon}$. The mean $\bar{\epsilon}$ from the wedge-shaped distribution is a good indicator of the mean trend of the threshold scatterplot (Figure 19) as it agrees well with the mean trend found by fitting a fifth-order polynomial (as a general utility function), in logarithmic scales, to the high-$\tau$ thresholds of Figure 19. The smooth curve is a Gaussian denoted $N(22.3$ dB, $15$ dB). Subsequent computations showed that these alternative distributions produced nearly equivalent overall dynamic range, so that the particular shape of the distribution does not appear to be crucial as long as the mean and standard deviation are the same.
Figure 18. The distribution of auditory nerve fiber dynamic ranges, a histogram from data supplied by Palmer (Evans and Palmer, 1980). All CFs are included. The fitted curve is a Gaussian probability density function (solid line) with a mean $\mu = 40.2$ dB and standard deviation $\sigma = 9.57$ dB, denoted $N(40.2, 9.57)$. We also tested the fit of the extreme-value distribution $\mathcal{E}(\mu, \sigma)$,

$$p(\lambda) = \frac{1}{\sigma} e^{-\frac{\lambda - \mu}{\sigma}} e^{-e^{-\frac{\lambda - \mu}{\sigma}}}$$

which is skewed (standard deviation $= 8.81$ dB) and hence provides a marginally better fit to the histogram, as judged by appearance alone. The means for these two probability densities were similar at 37.7 and 40.2 dB; as the data is not sufficiently plentiful to discriminate unequivocally between the two distributions, we adopt the more commonly used Gaussian.

Figure 19. Firing thresholds for single units pooled from Evans and Palmer (1980) and Geisler, Deng, and Greenberg (1985). Circles are 365 high-$\tau$ units and triangles are 112 low-$\tau$ units. Units with CFs above 50 kHz are not shown. Note that the range of thresholds covered by each group is similar, suggesting a similar variance in their probability distributions. The smooth curve is a fitted fifth-order polynomial which agrees closely with the data when the data is smoothed to show the mean trend (Cleveland routine, 1979, plot omitted for clarity; program written by Nizami). This mean trend is relatively flat over 2.5-25 kHz.
DERIVING A MEAN RESPONSE-RATE INTENSITY FUNCTION

Solving for the mean rate-intensity function will provide an equation in the independent variable "x" that is completely deterministic in that it describes neural response without the regression-fitting of rate-level curves to an equation. Assuming mutual independence of threshold, dynamic range, saturation rate, and spontaneous rate for the different $r_s$ groups (Table 5), the mean rate-intensity function for a fiber group is

$$
\overline{r(x)} = \int_{\varepsilon=\varepsilon_c}^{\varepsilon=\varepsilon_e} \int_{\lambda=\lambda^{min}}^{\lambda=\lambda^{max}} \int_{r_s=0}^{r_s=r_{max}} \int_{r_{max}=0}^{r_{max}=r_{max}} r(c,\varepsilon,\lambda, r_{max}, r_s, x) \ p(\varepsilon) p(\lambda) p(r_{max}) p(\varepsilon) \, d\varepsilon \, d\lambda \, dr_{max} \, dr_s
$$

(5)

where $p(\varepsilon), p(\lambda), p(r_{max})$, and $p(r_s)$ are the probability density functions for threshold, dynamic range, saturation rate, and spontaneous rate respectively. This stems from well-established rules of multivariate statistics (Bremaud, 1988). Equation 5 implicitly accounts for the rise in number of active units at a given CF, as it must. Integrating over spontaneous rate and saturation rate yields

$$
\overline{r(x)} = \int_{\varepsilon=\varepsilon_c}^{\varepsilon=\varepsilon_e} \int_{\lambda=\lambda^{min}}^{\lambda=\lambda^{max}} \left[ \frac{p(\varepsilon) p(\lambda)}{1 + \frac{100 - c}{c} \exp \left\{ -2 \left[ \frac{x - \varepsilon}{\lambda} \right] \ln \left( \frac{100 - c}{c} \right) \right\} } \right] + \frac{\overline{r_s}}{r_s} \right] \, d\varepsilon \, d\lambda.
$$

(6)

Thus, the mean rate-intensity curve is not influenced by the shape of the probability distributions for $r_s$ and $r_{max}$ and depends only on the means of these distributions. Integration over sloping-saturating units (described by Equation 3) yields a similar result for the weighting constant $\gamma$. The mean firing rate $\overline{r}$ is now completely deterministic, that is, no randomly-distributed parameters remain that might otherwise have to be obtained by curve-fitting to data.

The mean saturation rate, $\overline{r_{max}}$ was chosen in the following manner. Pooled saturation rate values from fiber characteristics supplied by Evans (see Fig. 3 in Evans and Palmer [1980]) and Liberman (1978, Fig. 17) were plotted vs. CF, smoothed (Cleveland, 1979), and interpolated between points (Figure 22) using a cubic spline routine (SigmaPlot). Although relatively constant at about 200/sec for frequencies less than 1 kHz, the plot of $\overline{r_{max}}$ dips idiosyncratically at 2.5 kHz, and rises dramatically to peak at about 10 kHz (two previously unnoted trends). Despite the variation of saturation rate with CF, however, simulations with different values of $\overline{r_{max}}$ indicated that dynamic range measures varied minimally with $\overline{r_{max}}$. Therefore, mean saturation rate $\overline{r_{sat}}$, was determined from the average of the pooled data of Evans (see Evans and Palmer, 1980, Fig. 3) and Liberman (1978, Fig. 17). Since saturation rate is much less
Table 5. Evidence for independence between neural fiber parameters within a spontaneous rate group.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$r_s$</th>
<th>$r_{\text{max}}$</th>
<th>$\varepsilon$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{max}}$</td>
<td>Various sources(^a)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Various sources(^{ab})</td>
<td>Evans &amp; Palmer, 1980(^c)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Evans &amp; Palmer, 1980(^d)</td>
<td>Various sources(^a)</td>
<td>Evans &amp; Palmer, 1980(^c)</td>
<td>-</td>
</tr>
<tr>
<td>$CF$</td>
<td>Liberman, 1978; (^f)</td>
<td>See Figure 22</td>
<td>Evans &amp; Palmer, 1980; Evans &amp; Geisler, Deng, &amp; Greenberg, 1985</td>
<td>1980(^d)</td>
</tr>
<tr>
<td>Kim &amp; Molnar, 1979</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Refers to inspection of Liberman, 1978; Kim and Molnar, 1979; Evans and Palmer, 1980; Geisler, Deng, and Greenberg, 1985; and others from supplied data or published graphs. \(^b\) This simplifying assumption may break down for other species. \(^c\) From raw data for high-$r_s$ units (Figure 22). There was no apparent correlation between $r_{\text{max}}$ and $\varepsilon$ for $0 < \varepsilon < 40$ dB SPL; other data for high spontaneous rate fibers (Liberman, 1978), showing a narrower range of thresholds (0-20 dB), nonetheless showed no visible correlation. Data for low-$r_s$ units were also available (Liberman, 1978; Evans and Palmer, 1980), but the sample sizes were too small for any reliable indication of trend; independence was therefore assumed outright for these other $r_s$ groups. \(^d\) We found a slight negative correlation (-.18) between $\lambda$ and $r_s$ within the high-$r_s$ fibers of Evans and Palmer (1980), which accounts for less than 4% of the variance, so that $\lambda$ and $r_s$ can still be treated as mutually independent (Figure 21). For the low- and medium-$r_s$ groups there was no suggestion of a correlation (perhaps due to the small sample size). There may be a between-group dependence of average dynamic range on unit spontaneous rate (Schalk and Sachs, 1980); the overall lower fiber dynamic ranges found by Schalk and Sachs may reflect a different criterion for dynamic range. \(^e\) High-$r_s$ units; Figure 20. A similar independence of $\varepsilon$ and $\lambda$ was also found for both low- and medium-$r_s$ groups, once sloping-saturating units were eliminated. \(^f\) Saturation rate is not independent of CF (Figure 22), but we assumed independence for computational simplicity since no other parameter is CF-dependent. \(^g\) Figure 20. Independence confirmed elsewhere; in computing discriminability limits at CF for 229 fibers, Delgutte (1987) defined the "10-dB dynamic range" as the dB range over which the difference limen dropped below 10 dB. This range showed no trend against CF regardless of $r$. 

Figure 20. (Left) Dynamic ranges of cat afferents vs. characteristic frequency (raw data of Evans and Palmer, 1980) for high-r, units only; 215 values. The variability is approximately uniform across CFs, and the smoothed trace varies only 5 dB on either side of 40 dB. (Right) Auditory fiber dynamic ranges plotted versus their thresholds for high-r, units (r>15/sec; raw data supplied by Palmer from Evans and Palmer, 1980). The smoothed curve (Cleveland routine, 1979) indicates the mean trend of the data and shows no systematic deviations from the horizontal, suggesting no correlation between a fiber's threshold and its dynamic range. Note the very high thresholds of some units. These units are not likely to be sloping-saturating fibers, as such units have much lower spontaneous rates (Liberman, 1988).

Figure 21. Fiber dynamic range vs. spontaneous rate. The high-r, data of Evans and Palmer (1980; r>15/sec) was pooled and the scatterplot smoothed with the Cleveland (1979) routine. Fiber dynamic range (in dB) appears to monotonically decrease with increase in r, from a mean of about 50 dB at 10 spikes/sec to 37 dB at 100 spikes/sec. This trend was double-checked by taking average values of dynamic range at each r, in the pooled data and then smoothing yet again. The same trend was found. No correlation was found between r, and A for lower r, values (1/sec<r,<15/sec); dynamic ranges for yet lower-r, fibers could not be tested, as it was not possible to discern whether fibers had such low spontaneous firing rates, from the supplied data.
ambiguous in definition than is threshold (see Chapter 6), the accuracy of this approach is thus limited only by the remaining source of ambiguity, potential across-cat variations in maximum rates. When pooled, these saturation values fit well to a Gaussian with mean of 200 spikes/sec (Figure 22). The corresponding smoothed trace for low-$r$, units (from Evans and Palmer (1980): $0<r<15$/sec; from Liberman (1978), set at $0<r<20$/sec) closely parallels the high-$r$, trace (Figure 22), suggesting 200 spikes/sec as a reasonable value for units of all spontaneous rates. (See Table 4 for mean spontaneous rates $r_{max}$.) Because there was no evidence that $r_{max}$ varied with spontaneous rate, the same value of $r_{max}$ was assumed for all spontaneous rate groups.

Equation 6 was integrated numerically (step sizes $\Delta \epsilon, \Delta \lambda = 1$ dB), where $\lambda - 4\sigma_{\lambda} < \lambda < \lambda + 4\sigma_{\lambda}$ for $\lambda - 4\sigma_{\lambda} > 0$. For $\lambda = 40$ dB these limits of 2 and 78 dB contain over 98% of the probability in the density function. Where $\lambda - 4\sigma_{\lambda} < 0$, the integration limits were $\lambda - 2\sigma_{\lambda} < \lambda < \lambda + 4\sigma_{\lambda}$ and the probability density was renormalised so that its lower limit was $\lambda - 2\sigma_{\lambda}$. (A discontinuity occurs for $\lambda = 0$, and $\lambda < 0$ is meaningless.) The integral over $\epsilon$ can be solved analytically (Appendix A) for a wedge-shaped $p(\epsilon)$ (Figure 17) which was used as a double-check by summing over $\lambda$ alone. At the 1 dB step size, there was no noticeable difference in plotted rates or intensity difference limens (see below) between the purely numerical double-summation, and the single-sum using the analytical integral over threshold $\epsilon$.

**COMPUTING JUST-DETECTABLE INTENSITY-INCREMENTS FROM THE MEAN RESPONSE-INTENSITY FUNCTION**

An $x$-dB tone initiates activity in the auditory channel corresponding to the point of maximal stimulation along the basilar membrane. Neural firing is variable; this spike count (over counting time $t$) varies across stimulus presentations, giving a distribution of neural counts. When stimulus intensity is increased by $\Delta x$ dB, the increment is assumed to shift this distribution along the decision axis by a constant amount, without changing its variance. Assuming a normal distribution gives the standard Signal Detection model (Figure 23), in which an observer's criterion is placed along the decision axis (counts) to maximize the payoff for responding.

The measure of sensitivity associated with Signal Detection Theory, $d'$, involves the difference in mean neural count between $x$-dB stimulus presentations, and $x + \Delta x$ dB stimulus presentations, specifically,
Figure 22. Maximum (=saturation) discharge rates \( r_{\text{max}} \) (Upper) Pooled data of Evans and Palmer (1980, fig. 3) and Liberman (1978, fig. 17) on maximum discharge rates in high-\( r_s \) units, plotted vs. the fiber's characteristic frequency. The data points were smoothed (Cleveland (1979) routine) and the resulting values interpolated with a cubic spline (SigmaPlot). Although relatively constant at about 200/sec for CF<1 kHz, the plot dips idiosyncratically at 2.5 kHz, and rises dramatically to peak at about 10 kHz. The dashed line is a smoothed trace for low-\( r_s \) units (Evans and Palmer: \( 0<r_s<15/\text{sec} \); Liberman: \( 0<r_s<20/\text{sec} \)). Note the correspondence between the two traces. (Middle) Fibers' maximum rates vs. their thresholds. Triangles are data of Liberman (1978) from all spontaneous-rate groups; open circles and solid circles are high- and low-spontaneous-rate units, respectively, from Evans and Palmer (1980). (Lower) A histogram of maximum rates pooled over all CFs from the raw data of Liberman (1978) and Palmer (Evans and Palmer, 1980). The solid line is a Gaussian distribution \( N(\mu, \sigma)=N(201.8,38.5) \); the dashed line is an extreme-value distribution \( E(\mu, \sigma) \) (see caption to Figure 14) with mean of 194.3/sec and a standard deviation of 35.6/sec. Although the Gaussian gives a slightly better fit, the extreme-value distribution gives relatively lower probabilities for the lower maximum rates and relatively higher probabilities for the higher maximum rates, in accord with the trend in \( r_{\text{max}} \), toward higher values in the range of \( \text{CF}=3-30 \) kHz. The closeness of the two calculated means suggests that 200 spikes/sec is a reasonable mean for \( r_{\text{max}} \).
Figure 23. Signal Detection Theory. The listener makes an observation along a decision axis and decides whether it actually arises from a signal presentation, using "C", an internal criterion point to distinguish between the internal distributions for noise (N) and signal-plus-noise (SN) (here, the evoked responses x and x+Δx).

Figure 24. Plots of the receiver operating characteristic (ROCs) for the equal-variance assumption (for d' = 1, the criterion assumed for our calculations) and the equivalent ROC obtained using the unequal-variance case (see text). The curves overlap so closely that they cannot be labelled. These curves plot false alarm rate p(Y/N) (probability of voting “yes” given the noise distribution) vs. hit rate p(Y/SN) (probability of voting “yes” given the signal-plus-noise distribution) in an imagined psychophysical experiment. The sensitivity index d' is held constant, hence their other name, isosensitivity curves (see Macmillan and Creelman, 1991). Courtesy Dr. Bruce Schneider.
\[
\frac{d'}{\sigma_{\hat{N}[x]}} = \frac{\hat{N}[x + \Delta x] - \hat{N}[x]}{\sigma_{\hat{N}[x]}}
\]

(7)

where \(\sigma_{\hat{N}[x]}\) is the standard deviation of the [normal] distribution of neural counts, \(\hat{N}[x]\), evoked by the \(x\)-dB tone is presented. For a pool of \(m\) fibers

\[
\bar{N}[x] = m \cdot t \cdot r(x)
\]

(8)

where \(t\) is the counting time. The ratio \(\rho\) of the mean neural count to its variance has been shown to be about 1.5 over a wide dB range for a counting time \(t = 51.2\) msec (Teich and Khanna, 1985). (More recent data suggests that \(\rho\) is not constant [e.g. Delgutte, 1987], but a constant \(\rho\) is reasonable for present purposes. Such variability data can be modelled as resulting from a dead-time Poisson process.)

The counting time \(t = 51.2\) msec is close to the tone durations used in the studies referred to in the Tables and Figures. This \(\rho\) also appeared to be independent of tone frequency, unit threshold, and unit spontaneous rate for units of medium- or high-rate. Following Hellman and Hellman (1990), if \(\hat{N}[x]\) is the single-fiber mean count for fiber number \(i\), and \(\rho\) is the same for all units,

\[
\sigma_{\hat{N}_i[x]}^2 = \frac{N_i[x]}{\rho}
\]

(9)

since for [assumed] independent units, the variance of the collected spike count is just the sum of the individual unit's variances, so that for \(m\) such units

\[
\sigma_{\hat{N}[x]}^2 = \sum_{i=1}^{m} \frac{1}{\rho} \sum_{i=1}^{m} N_i[x].
\]

(10)

Note that by the Central Limit Theorem the pooled counts will have a normal distribution given a large enough \(m\). The expected value for the sum of neural counts over \(m\) units is the mean count for \(m\) units, multiplied by \(m\); the expected value of the variance of the summed neural count is

\[
E\left[\sigma_{\hat{N}[x]}^2\right] = \left(\frac{mt}{\rho}\right) r(x).
\]

(11)

Using Equations (7, 8, 11) and replacing the mean count for \(x + \Delta x\) with the first 2 terms of its Taylor series gives
where \( r'(x) \) is the first derivative of \( r(x) \) so that

\[
d' = \frac{m \cdot t \cdot (\Delta x) \cdot r'(x)}{\sqrt{\left( \frac{m \cdot t}{\rho} \right) r(x)}}
\]

(12)

\[
\Delta x \ (dB) = \frac{d'}{r'(x)} \sqrt{\frac{r(x)}{m \cdot \rho \cdot t}}.
\]

(13)

Hellman and Hellman (1990) set their original equation in terms of sound pressure (i.e., in Newtons/meter\(^2\)), but an equation in decibels \( x \) lessens the error created by omitting the remaining terms of the Taylor series\(^{12}\). (Thanks to Bruce Schneider for pointing this out.) In retrospect, Equation 13 can be obtained through Siebert (1965), who developed the first theoretical measure of discriminability for neural pools; see Appendix B.

We assume that repetitions of the stimulus to a single fiber produces normally distributed spike counts. This assumption is reasonable for a reasonable accumulation during the counting time \( t \), but low-weight units firing near their thresholds may not satisfy this assumption individually or when spikes are pooled. Nonetheless computed dynamic ranges are not likely to change because the overlapping of threshold ranges over \( r \), groups assures that plenty of the more numerous high- and mid-weight units are already well-driven, such that the sum of counts is normally distributed.

For pooled counts we also make the signal-detection assumption \( \sigma_{N(x+\Delta x)} = \sigma_{N(x)} \); the true relation is

\[
\frac{\sigma_{N(x+\Delta x)}}{\sigma_{N(x)}} = \sqrt{\frac{r(x+\Delta x)}{r(x)}}.
\]

(14)

To assess the fairness of \( \sigma_{N(x+\Delta x)} = \sigma_{N(x)} \), we calculated Equation 14 for a single channel covering \( 1 \leq x \leq 110 \) dB SPL. We started with an intensity \( x \) and (1) computed the mean rate at \( x \), (2) computed the just noticeable increment in intensity \( \Delta x \) at \( x \), (3) computed the mean rate for \( x + \Delta x \), and (4) evaluated the ratio shown above which should be 1.0 given the equal variance assumption. These ratios were all under 1.03. Deviations in this range have little effect on computed threshold, as we found from an examination of the receiver operating curves (ROCs) for the equal-variance assumption (for \( d' = 1 \), the criterion assumed for our calculations) and the equivalent ROC obtained using the unequal-variance case (Figure 24).
To infer a practical dynamic range for the neural pool, we must establish some criterion based on actual performance. We note that for the longer 8 kHz tone, actual jnds are 5 dB at 43 dB SPL (1.5 sec tone; Rosenzweig, 1946) and 3.89 dB at 60 dB SL (Raab and Ades, 1946). More recent work shows lower jnds for such long tones at 1 kHz (Ostreich, Strominger, and Neff, 1977) implying lower jnds at 8 kHz also, but this may be offset by the rise in jnd as tones shorten (man; de Boer, 1986; Carlyon and Moore, 1984; Henning, 1971) towards the 40-50 msec used in neurophysiology, and below. Thus 3 dB seems an appropriate limit (see also Fay, 1988; Igarashi, Cranford, Allen, and Alford, 1979 [3.64 dB at 75 dB SPL, for 10 kHz]).

THE IDEAL OBSERVER

The following derivation is courtesy of Prof. Bruce Schneider. The algebra was checked by Nizami.

Siebert (1965) had already formulated the discriminability afforded by a neural population acting as an ideal computer, within the limits of the assumptions that he worked under. It is however not necessary to make all the assumptions that Siebert made (see Appendix B). The following pages derive the equation for the discriminability limen (jnd) of an ideal computer that follows a slightly different set of assumptions, ones that may be more realistic.

The ideal observer uses all of the available information to arrive at a decision with maximal efficiency; no real observer's performance is superior. Here an ideal observer sees the neural counts coming from fibers serving the critical band. We assume that within an $r$ group, each unit differs only in their threshold and dynamic range, sharing common values of $r_{max}$ and $r_e$ (the averages presented in Table 4) so to simplify the computations pursuant to the following algebra. For a fiber $i$ in a pool of $m$ neurons the average spike count during a $t$ second stimulus of intensity $x$ is

\[
\overline{N}_i[x] = tr_i(e_i, \lambda_i, r_{max}, r_e, x).
\] (15)

Let the spike count on any particular stimulus presentation, $N_i[x]$, be normally distributed with mean and variance

\[
\mu_i = \overline{N}_i[x], \quad \sigma_i^2 = \sigma_{N_i}^2[x].
\] (16)

We assume that the variance of $N_i[x]$ is the same as the variance of $N_i[x + \Delta x]$. A stimulus $x$ gives the $m$-
unit vector \( \{N_1, N_2, \ldots, N_m\} \). Now consider the ratio of the likelihood that \( \{N_1, N_2, \ldots, N_m\} \) came from a signal \( x + \Delta x \) dB to the likelihood that it came from \( x \) dB, given that we know the signal is \( x \) dB. For \( m \) mutually independent units, each of these likelihoods is the multiple of the respective probability densities, giving

\[
\frac{\prod_{i=1}^{m} \left( \frac{- \left( N_i - \bar{N}_i(x+\Delta x) \right)^2}{2\sigma_{N_i}^2(x)} \right)}{(2\pi)^{m/2} \prod_{i=1}^{m} \sigma_{N_i}^2(x)} = e^{- \sum_{i=1}^{m} \frac{(N_i(x+\Delta x) - \bar{N}_i(x))^2}{2\sigma_{N_i}^2(x)}} \frac{\prod_{i=1}^{m} \left( \frac{\left( N_i(x+\Delta x) - \bar{N}_i(x) \right)}{\sigma_{N_i}^2(x)} \right)}{(2\pi)^{m/2} \prod_{i=1}^{m} \sigma_{N_i}^2(x)}.
\]

Given the random variable \( N_i \), the maximum likelihood estimator (MLE) is the term involving \( N_i \), that is,

\[
\sum_{i=1}^{m} \left( \frac{N_i(x+\Delta x) - \bar{N}_i(x)}{\sigma_{N_i}^2(x)} \right) = \sum_{i=1}^{m} \left( \frac{r_i(x+\Delta x)}{r_i(x)} \right) N_i, \quad \text{since } (\text{Equation 9}) \quad \sigma_{N_i}^2(x) = \frac{\frac{t r_i}{t r_i}}{\rho}.
\]

The expected value of the MLE for \( m \) units presented with intensity \( x \) dB for \( t \) seconds is

\[
(\Delta x) \rho \sum_{i=1}^{m} \left( \frac{r_i'(x+\Delta x)}{r_i(x)} \right) E[N_i(x)] = (\Delta x) \rho t \sum_{i=1}^{m} \left( \frac{r_i'(x+\Delta x)}{r_i(x)} \right) E[N_i(x)]
\]

within each \( r_i \) group and by using the same steps as for Equation (19), the expected value of the MLE with intensity \( x + \Delta x \) dB is

\[
(\Delta x) \rho t \sum_{i=1}^{m} \left( \frac{r_i'(x+\Delta x)}{r_i(x)} \right) \frac{r_i'(x+\Delta x)}{r_i(x+\Delta x)}.
\]

The separation between the average values of the maximum likelihood estimates for \( x + \Delta x \), and \( x \), is then

\[
(\Delta x) \rho t \sum_{i=1}^{m} \left( \frac{r_i'(x+\Delta x)}{r_i(x)} \right) \left[ r_i(x+\Delta x) - r_i(x) \right] \frac{r_i(x+\Delta x)}{r_i(x)}
\]

\[
= (\Delta x)^2 \rho t \sum_{i=1}^{m} \left( \frac{r_i'(x+\Delta x)}{r_i(x)} \right)^2.
\]
because \[ r_i(c, \varepsilon_i, \lambda_i, r_{max}, r_s, x + \Delta x) - r_i(c, \varepsilon_i, \lambda_i, r_{max}, r_s, x) \] approximates \( r_i \cdot \Delta x \) as \( \Delta x \to 0 \) (definition of the derivative of \( r_i(c, \varepsilon_i, \lambda_i, r_{max}, r_s, x) \)). The variance of the MLE for stimulus \( x \) is now

\[
E \left[ \left( \sum_{i=1}^{m} \frac{(\Delta x)r'_i(c, \lambda_i, r_{max}, r_s, x)}{r_i(c, \lambda_i, r_{max}, r_s, x)} N_i[x] \right)^2 \right] - \left( E \left[ \sum_{i=1}^{m} \frac{(\Delta x)r'_i(c, \lambda_i, r_{max}, r_s, x)}{r_i(c, \lambda_i, r_{max}, r_s, x)} N_i[x] \right] \right)^2
\]

\[
= \sum_{i=1}^{m} \frac{(\Delta x)^2 \rho^2 r'_i(c, \lambda_i, r_{max}, r_s, x)}{r_i^2(c, \lambda_i, r_{max}, r_s, x)} \left( E[N_i^2[x]] - (E[N_i[x]])^2 \right)
\]

\[
= \sum_{i=1}^{m} \frac{(\Delta x)^2 \rho^2 r'_i(c, \lambda_i, r_{max}, r_s, x)}{r_i(c, \lambda_i, r_{max}, r_s, x)} \sigma_i^2[x]
\]

\[
= \sum_{i=1}^{m} \frac{(\Delta x)^2 \rho^2 r'_i(c, \lambda_i, r_{max}, r_s, x)}{r_i(c, \lambda_i, r_{max}, r_s, x)}
\]

\[
\text{The } d' \text{ is the difference of the means of the two MLEs divided by the standard deviation, so that}
\]

\[
d' = \frac{(\Delta x)^2 \rho t \sum_{i=1}^{m} \frac{r'_i(c, \lambda_i, r_{max}, r_s, x)}{r_i(c, \lambda_i, r_{max}, r_s, x)}}{\sqrt{(\Delta x)^2 \rho t \sum_{i=1}^{m} \frac{r'_i(c, \lambda_i, r_{max}, r_s, x)}{r_i(c, \lambda_i, r_{max}, r_s, x)}}}
\]

\[
[d']^2 = (\Delta x)^2 \rho t \sum_{i=1}^{m} \frac{r'_i(c, \lambda_i, r_{max}, r_s, x)}{r_i(c, \lambda_i, r_{max}, r_s, x)}
\]

For a single unit \( (m=1) \) the \( d' \)s of Equations 12 and 23 are the same. For \( m>1 \), the optimal detector (Equation 23) computes the squared \( d' \)s of each unit separately and then sums these to obtain the overall \( d' \). In contrast, Equation 12, not based on an MLE, first pools the neural counts and then computes \( d' \) (see Footnote 14). To obtain the average of \( (d')^2 \) for the optimal detector, we evaluate over all possible sets of size \( m \).

Thus,

\[
E((d')^2) = m(\Delta x)^2 \rho t \cdot E \left( \frac{r'_i(c, \lambda_i, r_{max}, r_s, x)}{r_i(c, \lambda_i, r_{max}, r_s, x)} \right)
\]

\[
= m(\Delta x)^2 \rho t \int_{\lambda = \hat{\lambda} - 4\sigma_{\lambda}}^{\hat{\lambda} + 4\sigma_{\lambda}} \left( \int_{\epsilon = \epsilon_L}^{\epsilon_H} \frac{r'_i(c, \lambda, r_{max}, r_s, x)}{r_i(c, \lambda, r_{max}, r_s, x)} p(\epsilon) d\epsilon \right) p(\lambda) d\lambda
\]
so that the intensity increment for \(E[(d')^2] = 1\) is given by

\[
\Delta x \ (dB) = \frac{1}{\int_{\lambda - \lambda - 4 \sigma_1}^{\lambda + 4 \sigma_1} \int_{e = e^2}^{r_{i/2}(e, \lambda, r_m, r_s, x)} p(e) \, de \ p(\lambda) \, d\lambda}.
\]  

Dividing \(p(e)\) into 12 rectangular regions (according to the histogram underlying Figure 17) allows the inner integral to be determined analytically, and the outer integral can be obtained numerically. The jnd for the MLE is still inversely proportional to the square root of the pool size, the duration of the stimulus, and the value of the mean to variance ratio, as it was for the \(\Delta x\) (Equation 13) for pooled counts.
NOTES

4. The single-unit firing equation is required to have a sigmoidal shape when plotted in coordinates of \( \{ \text{dB SPL, firing rate} \} \). Probability functions, the integrals of probability density functions \((= \text{probability distributions})\), have this property. The logistic density function is symmetric like the Gaussian but has more probability in the tails. Its integral, the logistic probability function, has the general form

\[
P(z) = \frac{1}{1 + e^{-\frac{(x+\mu)}{\sigma}}}.
\]

Being a probability function, this equation has a maximum value of unity and a minimum of zero (at \( z = -\infty \)). If we substitute in place of \( P(z) \) another measure which should have a maximum value of unity, the normalised value of the driven firing rate \((r-r_s)\) of a single auditory nerve fiber, we obtain

\[
y(\varepsilon, \lambda, x) = \frac{r-r_s}{r_{\max} - r_s} = \frac{1}{1 + e^{-\frac{(x+\mu)}{\sigma}}}
\]

where the variable \( z \) is replaced by \( x \), the applied sound pressure level in decibels SPL. The parameter \( \mu \) is defined such that when \( \mu = -\lambda \), \( y = 1/2 \). That is, when the sigmoid \( y(\varepsilon, \lambda, x) \) is at half of its maximum value (maximum=1), the corresponding value of the independent variable \( x \) must be exactly half-way between the point at which the sigmoid departs from zero, and the point at which it reaches its maximum value (unity). This corresponds to half-way between a fiber’s threshold and the upper limit of its possible dynamic range. Hence (in decibels) \( \mu = - (\varepsilon + [\lambda/2]) \), giving the full expression for normalised, driven rate \( y \),

\[
y(\varepsilon, \lambda, x) = \frac{1}{1 + e^{-\frac{(x-\varepsilon-\lambda)}{2}}}.
\]

The \( \sigma \) is the standard deviation of the logistic probability function (above), which can be obtained by noting that at the fiber threshold \( x = \varepsilon \), the lower limit of the fiber’s actual dynamic range,

\[
y(\varepsilon, \lambda, x) = \frac{1}{1 + e^{-\frac{1}{2}}}.
\]

At the upper limit of the fiber’s actual dynamic range, the saturation point of the normalised driven rate \( y \), we find \( x = \varepsilon + \lambda \), and

\[
y(\varepsilon, \lambda, x) = \frac{1}{1 + e^{-\frac{1}{2}}}.
\]
To solve for $\sigma$, let us assume that threshold $\epsilon$ is defined such that $(r-r_\epsilon) = 0.02 (r_{\text{max}}-r_\epsilon)$ so that $\sigma=\lambda/2\ln 49$. Assuming that $(r-r_\epsilon) = 0.98 (r_{\text{max}}-r_\epsilon)$ at $x=\epsilon+\lambda$ we again obtain $\sigma=\lambda/2\ln 49$. In general, a unique solution for $\sigma$ requires that for an assumed firing rate of $\beta'(r_{\text{max}}-r_\epsilon)$ at $x=\epsilon$ and an assumed firing rate of $\beta'(r_{\text{max}}-r_\epsilon)$ at $x=\epsilon+\lambda$, we must have $\beta'=1-\beta$. The general form of the firing-rate function is now

$$r(\epsilon, \lambda, x) = (r_{\text{max}}-r_\epsilon) \cdot y(\epsilon, \lambda, x) + r_s = (r_{\text{max}}-r_\epsilon) \left\{ \frac{1}{\epsilon + \frac{\lambda}{2}} \right\} + r_s$$

where $k=2\ln 49$ when the fiber’s dynamic range $\lambda$ is defined as the range of dB from 2% to 98% of its maximum driven rate. Other percentages will change $k$ accordingly. The equation above fills all the requirements of a rate-level function. Rate $r$ approaches its maximum $r_{\text{max}}$ as pressure becomes infinite, and approaches $r_s$ as pressure drops to zero (i.e. $x\rightarrow \infty$). As $x=\epsilon$, $(r-r_\epsilon)\rightarrow 0.02 (r_{\text{max}}-r_\epsilon)$. The verity of this limit can be examined by two representative examples. For a common assumed saturation rate $r_{\epsilon}=200$ spiked/sec (see Figure 22), we first examine a high-spontaneous-rate unit, $r_s=68.4$ spiked/sec. At $x=E$, $r=73.4$ spiked/sec. For a low-spontaneous-rate unit, say, $r_{\epsilon}=1$ spike/sec, then at $x=E$, $r=5.8$ spikes/sec. Although the error, percentage-wise, rises as $r_{\epsilon}$ falls, in neither case presented here is the error extreme.

5. Sampling error is a real danger; the available data samples never exceeded 1% of the neural populations. In the cat’s optic nerve a 5% sample introduced a 50% error (Donovan, 1967). One potential source of bias may be the decrease in fiber diameter with higher CF (Arnesen and Osen, 1978, and Alving and Cowan, 1971; noted in Warr, 1982), suggesting less likely electrode impalement with rise in CF. Also, pooling of data is unavoidable, as no single laboratory has published enough data to allow specification of probability density functions for threshold. Such pooling might produce a threshold variance larger than that of any individual animal (Palmer and Evans, 1982), as was found when recording from the pigeon (Richter, Heynert, and Klinke, 1995). We will show later that overall dynamic range of a fiber pool increases with threshold variance, hence the dynamic range derived from the mean rate-intensity function is likely to be inflated.

6. Note: not the 7 mm mentioned in Nizami and Schneider (1997), which was a misprint. From Greenwood (196)

The critical band $\Delta f=22.9(1 + 0.006046f)$

in the cat, where $f$ is frequency in Hertz, giving $\Delta f=1,131$ Hz at 8 kHz. From Liberman (1982),

$$f=0.456 \left( 10^{0.021(100-d)^{0.80}} - 0.80 \right)$$

where $d$ is % distance beyond the base of the basilar membrane,

so that $d=100 - \frac{1}{0.021} \log_{10} \left( \frac{0.8}{0.456} \right)$

where here $f$ is CF in kHz. Hence for the 8 kHz place, $d=39.83\%$ which corresponds to 9.957 mm from the base for an assumed basilar length of 25 mm (Greenwood, 1990). Assuming a symmetrical critical band with the frequency boundaries implied by Greenwood’s equation above, we obtain 10.315 mm and 9.601 mm as the upper and lower boundaries of the critical band (as distances from the base), for a critical band of 0.714 mm total width. To check the compatibility
of the two formulae used above, we first note that a half-critical-band is then 0.714/2 = 0.357 mm wide. Adding 0.357 mm to each side of 9.9575 mm (again assuming symmetry), this corresponds to percentage distances of 41.26% and 38.402%, or frequencies of 7.443 kHz and 8.600 kHz, for a critical-band-width of 1,156 Hz. Thus the two expressions are compatible for the critical band at 8 kHz.

7. A number of psychophysical studies suggest that the perception of sounds depends upon whether the energy in the stimulus is centered within a single limited frequency region (critical band) or spread across two or more such frequency regions. For example, if all of the energy in a complex signal (i.e. more than one component frequency) is contained within a critical band, the overall loudness of the sound depends solely on the total acoustic power (see Scharf, 1970, and Zwicker, Flottorp, and Stevens, 1957). That is, \( L_T = F(\Omega) \), where \( L_T \) is global loudness, and \( F \) is the power in the individual components of the complex sound. When the components of the complex sound are widely separated in frequency so that they fall into different critical bands, the loudness of the complex is the sum of the loudnesses within each critical band. That is, \( L_T = \Sigma L_i \), and where \( L_i \) is the loudness in critical band \( i \) (Fletcher and Munson, 1933; Marks, 1978; Schneider, 1988). Thus the nervous system, at some stage, must partition the pool of neural units serving the basilar membrane into different frequency regions. Therefore it is reasonable to limit the pool of neurons involved in the first stage of processing to those that fall (at most) within a critical band.

8. We will presently assume a CF response that fills the entire critical band. However, the slopes for actual fiber rate-level curves change systematically depending upon the relationship of the frequency of the tonal stimulus to the unit's CF (Sachs and Abbas, 1974). The mean rate-intensity function is most valid for the length of the basilar membrane for which fibers are responding to tones at their CF. This can be evaluated with neural recordings, and consists, conservatively, of about 10 inner hair cells at 8 kHz (see Winslow and Sachs, 1988), that is, roughly 240 fibers. In the absence of such neural recordings, estimates of the number of inner hair cells corresponding to CF can be obtained by the method of Kim and Parham (1991) by assuming that this number corresponds to a just-noticeable-difference in frequency. These authors noted that the relative frequency difference \( (\Delta f)/f \) for the cat (Elliott, Stein, and Harrison, 1960) can be translated into a basilar membrane distance ratio \( (\Delta y)/d \) and then, working according to Footnote 6, and using averaged values of Liberman's (1982) IHC packing density (those of Keithley and Schreiber, 1987, Fig. 4), we can obtain the number of IHCs corresponding to \( \Delta f \).

9. According to Bremaud (1988, p.98) and using very similar notation to his own: for a set of random variables \( X = \{X_1, \ldots, X_n\} \) (sometimes called a random vector), let \( G: \mathbb{R}^n \rightarrow \mathbb{R} \) be a function such that

\[
\int \cdots \int G(x_1, \ldots, x_n) f_X(x_1, \ldots, x_n) \, dx_1 \cdots dx_n
\]

has a meaning, where \( f_X \) is a non-negative function \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) which is the probability density of \( X \) (Bremaud, p.97), satisfying

\[
\int \cdots \int f_X(x_1, \ldots, x_n) \, dx_1 \cdots dx_n = 1.
\]

The expectation of the random variable \( G(X_1, \ldots, X_n) \), denoted by \( \mathbb{E}[G(X_1, \ldots, X_n)] \), is then defined by

\[
\mathbb{E}[G(X_1, \ldots, X_n)] = \int \cdots \int G(x_1, \ldots, x_n) f_X(x_1, \ldots, x_n) \, dx_1 \cdots dx_n.
\]

Furthermore, for independent random variables \( \{X_1, \ldots, X_n\} \), "such that each \( X_i \) admits a density [probability density function] \( f_i \),"
In other words, the value of the random variable \( G(X_1, \ldots, X_n) \), which can be the function of several random variables called either Equation 1 or Equation 3 in the text (what Bremaud calls “functionals of random vectors”), is

\[
f_X(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i)
\]

(Bremaud, p.105). The expectation value of the random variable \( G(X_1, \ldots, X_n) \), which can be the function of several random variables called either Equation 1 or Equation 3 in the text (what Bremaud calls “functionals of random vectors”), is

\[
E[G(X_1, \ldots, X_n)] = \int \cdots \int G(x_1, \ldots, x_n) \cdot \prod_{i=1}^{n} f_i(x_i) \cdot dx_1 \ldots dx_n
\]

which is the general form from which the mean rate function (Equation 5) derives. We now let the function \( G \) be the single-fiber firing-rate \( r \equiv r(x; \lambda, r_{\max}, r) \) such that \( \{X_1, \ldots, X_n\} = \{\varepsilon, \lambda, r_{\max}, r\} \) where the independent and non-stochastic variable \( x \) is applied sound pressure in dB SPL. The ensemble of single units in the summation pool is now completely characterised by the set of functions \( r_j = r_j(x; \varepsilon, \lambda, r_{\max}, r_{\ell_j}) \) where \( j \) represents one unit of the ensemble, and the infinite limits in the integration can safely be replaced by the actual limits of each distribution (for any given distribution, the integral over any other limits give zero as a multiplier such that only one non-zero multiple-integral is left over).

10. Some modellers prefer to express afferent’s firing rates as normalised driven rates with maxima of 1 and minima of 0. It is easily shown (proof omitted) that the normalised mean-driven-rate-function is the same as the mean function of the normalised-driven-rate, i.e.

\[
\bar{r} - r_s
\]

\[
\bar{r}_{\max} - r_s
\]

11. The precision of the computed rates had to be the best possible, considering restrictions on computer time (early trials used an IBM 386-20 clone). The step sizes for threshold and dynamic range, \( \Delta \varepsilon \) and \( \Delta \lambda \), were set equal for all trials, as there was no reason to do otherwise. Although step sizes of 5 dB produced notably different results from those of 1 dB (from visual inspection of plots), there was no noticeable difference between early results obtained with step sizes of 1 dB, 0.5 dB, and 0.2 dB.

12. The apparent mean-to-variance ratio \( \rho \) is not fixed, but may depend on the counting time \( \tau \), so that \( \rho > 1 \) for 0.1 msec < \( \tau \leq 400 \) msec. It then drops in a power-law fashion for \( \tau > 400 \) msec (Teich, Johnson, Kumar, and Turcott, 1990; note that they plot not \( \rho \) but \( \rho^{-1} \), their “Fano factor”).

13. To fill in the blanks: if \( \Delta x \) is the just detectable increment in level at level \( x \), then the number of spikes produced at level \( x + \Delta x \) and counted in a counting time \( \tau \) is

\[
N(x + \Delta x) = N(x) + \frac{dN}{dx} \Delta x + \frac{1}{2} \frac{d^2N}{dx^2} (\Delta x)^2 + O((\Delta x)^3)
\]

where \( O((\Delta x)^3) \) refers to an infinity of terms, starting with third order and progressing to the next integer higher in power with each successive term. These terms get successively smaller in magnitude. To a first approximation, the change in neural count with a level increment \( \Delta x \) on top of a level \( x \) is then
\[ \Delta(N(x)) = N(x + \Delta x) - N(x) = \frac{dN}{dx} \cdot \Delta x. \]

Hellman and Hellman (1990) now fall back on Signal Detection Theory (Figure 23), introducing the unitless detectability index \( d' \) into their assumption that

\[ \Delta(N(x)) = d' \cdot \sigma(x) \]

where \( d' \) is the distance between the means of the count distributions at intensity \( x \) and \( x + \Delta x \), in units of the standard deviation \( \sigma(x) \) at level \( x \). Hence

\[ d' \cdot \sigma(x) = \frac{dN}{dx} \cdot \Delta x \]

so that \( \Delta x = d' \frac{\sqrt{N(x)}}{\sqrt{\rho \frac{dN(x)}{dx}}} \)

which is the intensity increment \( \Delta x \) corresponding to a change \( \Delta(N(x)) \) in mean spike count.

14. Equations 12 and 23 can also be used to determine the relative advantage of the ideal observer over an observer that simply pools the responses from \( m \) units before making a decision. Observe that the square root of

\[
\int_{\lambda = \lambda - 4 \sigma_\lambda}^{\lambda + 4 \sigma_\lambda} \left( \int_{\varepsilon = \varepsilon_L}^{\varepsilon_H} \frac{r_i^2(\varepsilon, \lambda, \lambda_{r_{\max}}, r, x)}{r_i(\varepsilon, \lambda, \lambda_{r_{\max}}, r, x)} p(\varepsilon) d\varepsilon \right) p(\lambda) d\lambda \]

is the ratio of the \( d' \) of the ideal observer to the \( d' \) obtained from a simple pooling mechanism.
CHAPTER 3

ACTUAL DYNAMIC-RANGE COMPUTATIONS
Preliminary results of the computations have already appeared in print (Nizami and Schneider, 1996) and the complete results have been recently published both as a paper (Nizami and Schneider, 1997) and as an abstract that extended part of that paper (Nizami, 1998). We thank R.V. Harrison for comments on an earlier version of the published manuscript.

AUDITORY DISCRIMINABILITY FROM THE OVERALL MEAN RATE-INTENSITY FUNCTION

One channel

In computing the functions in Figure 25 we have pooled the different spontaneous rate groups into a single processing channel according to Equation 4 using the parameter values given in Table 4. The lower and upper panels respectively show how the mean rate and the size of the jnd change with sound pressure level. In order to compare the results of this simulation with others, we define the effective dynamic range of a pool of neurons as the dB difference between the upper and lower intensities producing a jnd of 3 dB. (As Figure 25 shows, choosing higher or lower dB limits, e.g. 5 or 2 dB, has only a marginal effect on the computed value of effective dynamic range.) The lower and upper 3-dB limits for the single channel model shown in Figure 25 are 7 and 89 dB SPL. This yields an effective dynamic range of 82 dB, which is 42 dB wider than the average unit's dynamic range. Repeating the exercises of Figure 25 assuming that all fibers had the same mean dynamic range revealed a disadvantage of only 1 dB compared to the case of a higher mean dynamic range for sloping-saturating units.

Of course, intensity-increment thresholds shown in Figure 25 assume no additional sources of noise other than those observed in 8th-nerve auditory neurons, that is, that the auditory system is a perfect discriminator once neural signals have departed the periphery. Additional internal noise added after this stage would increase the size of \( \Delta x \), the just-noticeable-difference in intensity, at all levels, and it seems inevitable that further error will creep in between ear and behaviour.

We also examined the effects of (a) a two-fold change in the standard deviation \( \sigma_r \) of the probability density function for dynamic range, for all \( r_s \) groups (Figure 26), and (b) a 20-dB change in the mean dynamic range for all \( r_s \) groups (Figure 27). Figure 26 shows that the jnd is relatively unaffected by substantial changes in \( \sigma_r \). Changing the mean dynamic range by 20 dB (Figure 27) reduces or raises overall dynamic range by only about 7 dB (overall ranges of 2-77 dB SPL and 12-101 dB SPL, respectively). It is highly unlike that the neural pool has a mean dynamic range of 60 dB, and even so, a single channel still falls short of the cat's full auditory range.

In Figure 28 we assumed that the standard deviation of the threshold distribution was either halved or doubled. To double the standard deviation while keeping the mean constant, we double each threshold and subtract the mean threshold \( \bar{e} \) by creating a new parameter \( 2 \bar{e} - \bar{e} \). This transformed parameter has
Figure 25. The overall mean response-intensity function (Equation 6; lower plot) and the intensity-increment function (Equation 13; upper plot) for the single channel model. See Table 4 for parameter values.

Figure 26. Overall mean response-intensity functions (lower plot) and intensity-increment functions (upper plot) when the standard deviations of the dynamic ranges are: 1) halved, 2) the same as in Table 4, or 3) doubled. See Table 4 for other parameter values.
a threshold histogram that is broader and flatter than before. Doubling gives a very large range (-15 to 107 dB) but requires more than 25% of the low-\(r\) fibers to have \(\varepsilon > 70\) dB, and more than 25% of the high-\(r\) fibers to have \(\varepsilon > 50\) dB, both experimentally unlikely. Changing the mean of the threshold distribution, \(z\), by an amount \(\pm \Delta \varepsilon\) dB assuming a constant \(\sigma_z\), will simply shift the functions in Figure 25 by \(\pm \Delta \varepsilon\), so that the effective dynamic range will not change\textsuperscript{15}.

Figures 26-28 reveal the best ways to increase the dynamic range of a single channel: (a) increase the mean dynamic range; (b) increase the threshold standard deviation. In particular, increasing the threshold standard deviation is surprisingly effective; doubling its value increases the effective dynamic range from 82 to 122 dB. (This effect was surprising; one might think that the best way to increase the ensemble dynamic range was to increase the fiber mean dynamic range.) A greater separation in mean thresholds across the 4 fiber subgroups will extend the effective dynamic range, but there is no experimental support for a larger separation\textsuperscript{16}.

Two channels

The single channel model, which weights each fiber equally (Equation 4), cannot attain a dynamic range as large as that exhibited by the cat. Because high-\(r\) units are such a large percentage of the neural pool, their contribution overrides that of the fewer high-threshold, low-\(r\) units. A number of experimenters (Munson and Gardner, 1950; Zwislocki, Pirodda, and Rubin, 1959; Yates, Winter, and Robertson, 1990; Zeng, Turner, and Relkin, 1991; Greenwood, 1993) have proposed two intensity channels to account for a variety of experimental results\textsuperscript{17}. Moreover, intensity discrimination under forward masking implies (Salvi, Saunders, Ahroon, Shivapuja, and Areholt, 1986) that high-\(r\) fibers form the lower intensity channel with the low-spontaneous-rate fibers forming the upper channel.

We pursued this scheme, with low-\(r\), fibers, including sloping-saturating fibers, as an upper channel separate from other units. The lower channel’s dynamic range limits (Figure 29) are 6 and 76 dB SPL. The upper channel’s range limits are 18 and 92 dB SPL. The two channels combined have a dynamic range of 86 dB that is only 4 dB larger than the single channel model. Based on recovery rates in forward masking experiments (Delgutte, 1987) some investigators have suggested a grouping of low- and medium-\(r\), units into a single channel. The effective dynamic range of such a channel is from 9 to 86 dB SPL. Note that the upper limit of such a channel is 3 dB lower than the single-channel model shown in Figure 25. Thus, having two separate channels can only slightly extend the dynamic range, given our assumptions about how these fiber populations differ in their threshold and dynamic range characteristics.

Dynamic range may depend on spontaneous-rate group, as noted by Schalk and Sachs (1980), who found lower average dynamic ranges and narrower dynamic range distributions than others did. We thus
Figure 27. Overall mean response-intensity functions (lower plot) and intensity-increment functions (upper plot) when the mean dynamic ranges are: 1) decreased by 20 dB, 2) the same as in Table 4, or 3) increased by 20 dB. See Table 4 for other parameter values.

Figure 28. Overall mean response-intensity functions (lower plot) and intensity-increment functions (upper plot) when the standard deviations of the thresholds are: 1) halved, 2) the same as in Table 4, or 3) doubled. See Table 4 for other parameter values.
tried a dynamic range density \( N(18, 4.78) \) for high-\( r \), units, \( N(25, 4.78) \) for mid-\( r \), units, and \( N(40.2, 9.57) \) for low-\( r \), units, with all other factors remaining the same. With these parameters a single channel model has a dynamic range of 78 dB (3-81 dB SPL), 4 dB less than under regular assumptions (see Figure 25). The upper limit of the channel actually drops by 11 dB. If we assume that low-\( r \), units form a separate upper channel whereas all other units form a lower channel, the overall dynamic range is 89 dB (3-92 dB SPL), which is 3 dB larger (at the lower end) than the two channel model in Figure 29. Most importantly, the overlap between the 3 dB ranges of the upper and lower channels now shrinks from 58 to 52 decibels, making each channel more distinct. (Remember that by pooling raw data for high-\( r \), units (Table 4), we had probably overestimated the variance \( \sigma_e \) of \( p(e) \) and thus previously overstated the overlap between the high-\( r \), and low-\( r \), channels\(^{18} \).)

Are sloping-saturating units special? Treating each fiber group as a separate channel (low-\( r \), mid-\( r \), high-\( r \), and sloping-saturating) gives 3 dB limits of 7-92 dB, which only extends the upper limit of dynamic range by 3 dB compared to the single channel of Figure 25.

**DISCRIMINABILITY FOR THE IDEAL OBSERVER**

Our model is unique in three ways. First, we do not assume an "ideal computer" (Siebert, 1965) for intensity coding as others did (Siebert, 1965; Viemeister, 1983, 1988; Delgutte, 1987; Winslow and Sachs, 1988). Second, we employ a new rate-level equation, including dynamic range as a parameter and giving fitted neural-firing characteristics very close to those observed (Table 1). Third, we compute a statistical mean rate-intensity function and an intensity discrimination function, based on known distributions of fiber properties, to account for the aggregate behavior of a neural population under the reasonable assumption of mutual independence between neural characteristics.

For Siebert's (1965) "ideal observer", maximum likelihood was applied to optimize the way information was combined across 8th-nerve units (Appendix B). The auditory computer was assumed to know the distribution of responses in every fiber for each pair of intensities, \( x \) and \( x + \Delta x \), i.e. knowledge of each fiber's rate-intensity curve, as well as \( x \) and \( x + \Delta x \). Siebert showed that if interspike intervals follow a gamma distribution, and if it possesses a certain scalar property, then the optimum decision rule is based on a weighted sum of the spike counts from each unit in the population. If spike count in each unit at every intensity level is normally distributed, the optimum decision rule in a two-interval, forced-choice situation is to compute a sum of weighted neural counts for each interval, choosing the interval of highest counts.

The parameter values used in this simulation are identical to those used in the single channel model above. Figure 30 shows the jnd for 938 channels (one critical band) for the model optimised under our assumptions of Chapter 2 (Equation 25). Overall range is extended from about 82 dB (Figure 25) to 101 dB,
Figure 29. Overall mean response-intensity functions (lower plot) and intensity-increment functions (upper plot) for two independent channels. See Table 4 for parameter values.

Figure 30. Intensity-increment functions for 1 channel (limits: 7-89 dB SPL), 7 channels (limits: 5-101 dB SPL), and 938 channels (limits: 3-104 dB SPL).
with almost all of the gain occurring at the high intensities. Unlike our single pooling node, however, each unit in the 938 channel model is weighted proportionally to its first derivative and inversely to its variance, implying an exceedingly complex network. The additional computational cost might well be worth the resources used, given the range extension, but fewer channels will produce nearly equivalent results. We created a 7-channel model by further subdividing each of our 4 r, groups. Each of the low spontaneous rate groups (see Equation 4) were subdivided into two subgroups, those with thresholds < 60 dB SPL and those with thresholds > 60 dB SPL. The medium-r, group was left alone but the high-r, group was divided into two groups (thresholds < 10 dB SPL and thresholds > 10 dB SPL), producing seven channels. These seven channels, optimally combined (Figure 30), produced an effective dynamic range nearly as large as that provided by 938 channels. Interestingly, assuming $\sigma_z = 0$ in Equation (25) has a negligible effect on the maximum likelihood estimate of $\Delta x$, giving a dynamic range of 6-101 vs. 5-101 dB SPL for 7 channels, and 4-102 dB SPL compared to 3-104 dB SPL for the 938 channel function (see Figure 31).

DISCUSSION

The validity of the assumptions

This work reports a solution for the intensity encoding ability of a neural ensemble innervating a given position on the basilar membrane. An advantage of this model over previous summation models (Goldstein, 1974; Howes, 1974; McGill and Goldberg, 1968; Lachs, Al-Shaikh, Bi, Saia, and Teich, 1984) is comprehensiveness; our method constructs the mean response from our best estimates of the sampling distributions of the two major parameters of individual rate-intensity curves, namely dynamic range and threshold. It is unlikely that the specific forms of the equations used for these distribution functions are crucial, as long as they provide a reasonable description of the data.

Our model's conclusions depend, of course, on the validity of our assumptions. We assumed that intensity encoding depends only on the total number of spikes occurring during a summation period. This, of course, ignores the fact that the rate of firing to a stimulus declines rapidly within the first few msec of stimulus presentation. If this adaptation pattern changes with intensity it could provide additional information that would aid in intensity discrimination. We also assume that spike timing does not contribute to intensity discrimination and that all units in the neural pool have rate-intensity functions that are characteristic of a unit driven at its CF, although slopes of rate-level plots change systematically depending on the relation of the stimulus tone's frequency to unit CF (Sachs and Abbas, 1974). Finally, dynamic range will depend both on the size of the neural pool, the duration of the summation period (assumed here to be 40-50 msec, but that may be as large as the energy integration time of 100-200 msec), and on the mean-to-variance ratio (see Chapter 2). However, the effect of a change in $\rho, t, \text{ or } m$ on overall dynamic
Figure 31. Intensity-increment functions for a 938-channel ideal computer (Equation 25) when the standard deviation of the probability density function for dynamic range is 0 in one case and 9.57 dB in the other.
range is relatively small; for example, doubling either \( \rho \), \( t \), or \( m \) changes the single channel range from 7-89 dB SPL to 6-93 dB SPL.

**Extending dynamic range by means of gain control**

The mammalian auditory system appears to have effective gain controls. One such mechanism is the middle-ear reflex, active at high SPLs. The middle-ear reflex does not come into effect until about 80 dB SPL or more (in humans [Moller, 1984]; in cats, at 80 dB re human SL [Eliasson and Gisselsson, 1955]). In cats, effective sound attenuation by the middle-ear muscles has been studied by direct stimulation of the relevant brainstem nuclei, with at best 5 dB of attenuation at 7 kHz (Teig, 1973; anesthesia). The effective reflex reduction of input sound, stimulated by contralateral tones, is best below 2 kHz and increases in effect with decrease in frequency (Moller, 1984). High-frequency tones will have perhaps a 5 dB attenuation, corresponding to a 5 dB drop in the ear's required neural dynamic range. Because the middle-ear reflex mimics an attenuation of the sound, the form of our single-unit equations does not change, and because the mean rate-intensity function approximates a linear sum of single-unit rates, neither does the computed discriminability.

**Structure of the olivocochlear gain-control system**

Another means of gain control is neural feedback. Accumulated evidence suggests that the superior olivary nucei in the pons is the first point of mutual interaction for neural impulses from both ears (Irvine review, 1986). Fex's (1962) pathway-mapping studies revealed that the crossed and uncrossed olivo-cochlear bundles (OCB) was a feedback system of the cochlea. This pathway uses the cochlear nucleus (CN) and the superior olive, and descending fibers from the inferior colliculus (Figure 32). Fex (1962) noted that the cat's efferents were sensitive to sound, and that their activity is inhibitory to primary afferents. In cats with one ear destroyed, crossed olivocochlear units could be activated both by electrical stimulation of the ipsilateral ear and sound in the contralateral ear.

Based on efferent projection, the superior olivary complex has been divided into medial and lateral zones (not specific nuclei; Guinan, Warr, and Norris, 1983). The MOC and LOC zones involve the periolivary nuclei, whose organisation is not well known, hence the more appropriate present use of a zonal classification (Irvine review, 1986). The periolivary cells lateral to the MSO (medial superior olive) obtain afferent innervation largely from the ipsilateral CN, and those medial to the MSO receive afferents largely from the contralateral CN (ibid.). Periolivary cells can project ipsilaterally or bilaterally to the cochlear nucleus (CN) or the inferior colliculus (IC), or bilaterally to the cochlea (the so-called MOC and LOC neurons).
Figure 32. (Overleaf) Possible or confirmed pathways of the medial olivo-cochlear feedback system in the cat. The portion below the horizontal dashed line is a transverse frontal section, midway dorso-ventral, in the brainstem. All connections are established by pathway tracing, not inferred from neurotransmitter studies. Branches from main pathways are all believed to be axon collaterals. Key:

COCB: crossed olivo-cochlear bundle
DAS: dorsal acoustic stria
DCN: dorsal cochlear nucleus
DMPO: dorso-medial periolivary nucleus
IAS: intermediate acoustic stria
IC: inferior colliculus
IHC: inner hair cell
LL: lateral lemniscus
LOC: lateral olivo-cochlear efferents
LSO: lateral superior olive
MNTB: medial nucleus of the trapezoid body
MOC: medial olivo-cochlear efferents
MSO: medial superior olive
OHC: outer hair cell
TB: trapezoid bundle
VCA: vestibulo-cochlear anastomosis
VCN: ventral cochlear nucleus
VNLL: ventral nucleus of the lateral lemniscus
VNTB: ventral nucleus of the trapezoid body

Upper graphic: Identified or suspected efferent (centrifugal) pathways (solid lines) of the medial olivo-cochlear (MOC) loop, for nuclei of the right side of the cat’s brainstem (primary source: Adams, 1983). Dashed lines are afferent projections. Projections from the left brainstem nuclei to the right (ipsilateral) side have been omitted but can be inferred from the illustration. Specific notes: 1. The thicknesses of lines emanating from the medial periolivary nuclei (DMPO, MNTB, VNTB) indicate the relative numbers projecting contralaterally vs. those projecting ipsilaterally. 2. The DCN is artificially elevated from the VCN for illustrative convenience. Intrinsic connections between DCN and VCN are from Oliver, Potashner, Jones, and Morest (1983). 3. There are no direct MOC projections terminating in the cochlear nucleus (Kane, 1976; Elverland, 1977). 4. We have omitted a few kinds of rare, unusual synapses at the cochlea (Ginzberg and Morest, 1984) that occur in addition to the morphology illustrated (Spoendlin, 1970; Ginzberg and Morest, 1984). Additional feedback control is postulated for the reciprocal synapses seen at the OHC afferent in humans, which might involve two-way transmission at the nerve terminal (Nadol, 1990). 5. The question mark indicates that the pathway from the ipsilateral IC to the DCN is only inferred (from Oliver, Potashner, Jones, and Morest [1983]), as is the separateness of the projections from the contralateral IC and from both LL (Kane, 1976). Other sources used in this illustration: Spangler, Cant, Henkel, Farley, and Warr (1987), Warr (1978). Labeling of a single efferent neuron all the way from the contralateral MOC area to the ipsilateral cochlea has been achieved by Liberaman and Brown (1986).

Lower graphic: Afferent (centripetal) pathways (solid lines) that have the potential to be involved with the medial olivo-cochlear (MOC) feedback system, originating from the right side of the cat’s head (primary source: this is a partial schematic of an illustration by Warr [1982]). Dashed lines are efferents. Whether or not all these projections are involved in MOC feedback is still not understood (Warr, 1982). For graphic simplicity, the lateral trapezoid bundle has been omitted, which projects from the VCN to the lateral lemnisci and inferior colliculi both ipsilaterally and contralaterally (Warr, 1972). The ventral cochlear nucleus (VCN) has been drawn in sagittal section for clarity (after Warr, 1982). Specific note: Both sets of peripheral afferents (from IHCs and OHCs) project to the cochlear nucleus (CN) (Leake-Jones and Snyder, 1982; Jones, Morest, Oliver, and Potashner, 1984). Afferents project to the anterior and posterior VCN as well as to the DCN (Osen, 1970; Fekete, Rouiller, Liberaman, and Ryugo, 1982; Leake and Snyder, 1989). Whether these afferents do project to the granular cell layer (Oliver, Potashner, Jones, and Morest, 1983) or do not (Osen, 1970; Leake and Snyder, 1989) is unresolved.
The myelinated medial olivocochlear (MOC) system projects to the outer hair cells (74% project contralaterally) and the unmyelinated lateral olivocochlear (LOC) system synapses with the radial fibers which innervate the inner hair cells (only 9% contralaterally; Guinan, Warr, and Norris, 1983). Later summaries suggest that 25% of LOC afferents are crossed over to the contralateral side (Guinan and Gifford, 1988). Painstaking tracing of a small number of fibers shows that MOC efferents do not make synaptic contact with inner hair cells (Liberman and Brown, 1986).

Action of olivocochlear gain control

Non-muscular gain control systems in the ear involve a rapid-response, frequency-specific local amplifier at the organ of Corti (the motile outer hair cell [OHC] itself) and a relatively slow response loop that presumably regulates long-term changes in auditory sensitivity (Dulon and Schact, 1992), involving the MOC (medial olivocochlear) efferents that innervate the OHCs. Anesthetised cats have been the principle source of information on this system, and in order to compensate for the presumed effects of anesthesia (see Klinke and Galley, 1974), as well as the delay in activation due to tones too short to be self-influencing, shocks are applied to the contralaterally-projecting crossed olivocochlear bundle (COCB) at the midline of the brainstem. Shocks give a suppression effect equivalent to as much as a 24 dB reduction in external sound intensity (Desmedt and Monaco, 1961; Desmedt, 1962; Wiederhold, 1970; Gifford and Guinan, 1983), equivalent to a lateral shift of the rate-level function to higher SPLs (Wiederhold, 1970). Gifford and Guinan (1983) found the same sort of parallel shift for almost all fibers from a local area of the basilar membrane (50 msec tone, 1 kHz place), a finding of particular relevance to our own computations, which also apply to a small region. All r groups show these shifts, whose magnitudes vary from unit to unit and generally agree with those of Wiederhold (1970). These rate-level shifts are essentially parallel, with little change of slope, for brief CF tones (Wiederhold, 1970; Gifford and Guinan, 1983; Guinan and Gifford, 1988), implying that the equations describing the rate-intensity data (Equations 1 and 3), and the mean rate-intensity functions produced from them (from Equation 5), need not be altered, at least to a first approximation. Computed mean rate functions would simply shift uprange or downrange by an amount determined by the level of activation of the MOC system. Efferent MOC fibers are found at all CFs and show a variety of thresholds at each CF (within individual cats; Liberman and Brown, 1986). Such a pattern is expected if the MOC system is to act smoothly and progressively as the auditory afferents are themselves progressively activated with rise in sound level, aided by a useful efferent dynamic range of at least 50 dB (Liberman, 1988).

Our computations have focused on the dynamic range of sound intensity encoded by a narrow portion of the organ of Corti. Such a portion may have efferent control; the similarity of the contralateral distribution of MOC efferents to the effectiveness of midline COCB stimulation on single-fiber threshold shifts (Wiederhold, 1970) suggests an OHC effect on these shifts (Guinan, Warr, and Norris, 1984). Finally, efferent gain control is not a transient response. Efferent firing continues even with tones 10 minutes long, despite some adaptation in firing rate (Liberman and Brown,
Therefore it is possible that efferent stimulation could produce a rightward shift in the rate-intensity curves of a population of neurons, equivalent to an rise in neural thresholds. Assuming that all fiber thresholds shift upwards by 20 dB due to gain controls pushes the 3 dB limit of the upper channel to 112 dB, giving a substantially improved overall dynamic range, but still well below human limits and probably below those of the cat. One solution to this problem is that much greater shifting may occur in the unanesthetised animal\textsuperscript{20,21}.

Other models that computed dynamic range

After Siebert (1965), Winslow and Sachs (1988) have used a maximum likelihood approach to study intensity coding. Their rate-intensity function contained 3 parameters (the equivalents of $r_s$, $r_{\text{max}}$, and $e$) whose values were varied in their calculations. Uniform distributions were assumed, with limits from the literature (Liberman, 1978; Sachs, Winslow, and Kozikowski, 1986). Because variation in any of these three parameters does not appreciably affect the dynamic range of a unit, their model does not take into account the observed variability in fiber dynamic range. We attempted to introduce dynamic range, $\lambda$, explicitly into the Sachs and Abbas (1974) equation, the progenitor of Winslow and Sachs (1988) but found this approach intractable. Further, the Winslow and Sachs equation contains two threshold-related terms, one of which must be fitted for each unit, and whose distributions are unknown, unlike our single and double-logistic functions, for which parameter distributions can be reasonably inferred\textsuperscript{22}.

Delgutte (1987) also used optimum processing, first determining mean rate-intensity functions for each of the 3 $r_s$ groups by averaging individual fibers after they had been shifted so that they were all centered at the point at which the first derivative of each fiber’s rate-intensity curve was maximal. The only random variable was the intensity level at which each individual fiber achieved its maximum discriminability. This is effectively equivalent to assuming that all fibers in an $r_s$ group had the same rate-intensity function, but differed only in threshold. Delgutte did not try altering the means and variances of the distribution of his one stochastic parameter, and also ignored variations in fiber dynamic range. Delgutte’s work was extensive and is described more fully in Appendix C.

Viemeister (1988) used a variation of Siebert’s optimisation, for the discriminability afforded by fiber populations of 10 and 50 units firing at their CFs in the 6-14 kHz frequency band. Like the present approach, Viemeister assumes Liberman’s (1978) spontaneous-rate classes and thresholds, and Evans’ and Palmer’s (1980) dynamic ranges. He also assumes mutual independence of fiber firing rates, and Gaussian fiber-spike-count distributions for any given fiber at a given SPL (after Teich and Khanna, 1985), with $\rho$=2 (rather than 1.5). Neural ensemble performance was judged by using the $d'$ measure of signal detection theory for each fiber; the $d'$s were weighted, by the proportions of fiber types in spontaneous-rate groups,
in an optimal pooling scheme. The resulting jnds are similar to those of Winslow and Sachs (1988); there is a strong dip centered at 30 dB SL, followed by a steady increase with rise in dB. Under the optimisation scheme, the jnd above 40 dB SL is largely due to low-r, fibers. Omitted was any special treatment for sloping-saturating units, or for any possible correlation between threshold and dynamic range among low-r, fibers, a simplification made here and elsewhere (i.e. Winslow and Sachs, 1988).

Discriminability limens computed based on optimisation schemes (Delgutte, 1987; Viemeister, 1988; Winslow and Sachs, 1988) have not followed Weber’s Law, which calls for a constant jnd. Figures 25 to 29 do show a relatively flat jnd, due presumably to not assuming an optimal combination across spontaneous-rate groups. Although clever manipulation of the pooled firing might give Weber’s law (see Delgutte, 1987), a physiological basis for arbitrary weighting has not been found. Nor has any physiological basis been found for Siebert-type optimisation, hence our present assumption of simple proportional weighting (Equation 4). Our attitudes reflect Viemeister’s (1988) note that “by arbitrarily altering the weights of the various types of fibers we can account for the psychophysical data or, for that matter, any Weber function whose thresholds are at least as high as those predicted under optimal combination. This, however, seems not particularly informative and, without some reasonable physiological constraints, would be little more than an exercise in curve-fitting”.

In comparing our results with others’, some common features emerge. First, all computations give discriminability at intermediate SPLs that is exaggerated compared to real performance. Secondly, all show, even with an optimum computation, that the ear’s entire dynamic range cannot be credited to a single channel of restricted basilar length.

**Rate-intensity functions, and possible correlations between their parameters**

A number of equations adequately describe the behavior of 8th-nerve fibers (Goldstein, 1974; Sachs and Abbas, 1974; Sachs, 1974; Lachs, Al-Shaikh, Bi, Saia, and Teich, 1984; Sachs, Winslow, and Sokolowski, 1989; Yates, Winter, and Robertson, 1990 [Guinea pig]). Sachs’ and Abbas’ (1974) equation, as updated in Sachs, Winslow, and Sokolowski (1989), well describes neural firing, including sloping-saturation, and relates this firing to basilar membrane displacement. However, there is no parameter that can be directly related to fiber dynamic range, which would make it hard to incorporate dynamic range distribution into our mean-rate model. To describe the gamut of rate-level shapes in the Guinea pig, others (Yates, Winter, and Robertson, 1990) used an equation similar to that of Sachs and Abbas (1974) and of Sachs, Winslow, and Sokolowski (1989). Advantageously, they involve the outer hair cell amplifier as an active feedback loop causing basilar nonlinearity, but it is hard to incorporate dynamic range variations into their equation23.

Our neural references have pertained to the cat unless stated otherwise. Inevitably we find species
differences. We assumed independence among the four parameters \( r, r_{\text{max}}, e, \) and \( A \) (Table 5), but for the Guinea pig there may be correlations between \( r \) and \( e \), and between \( r \) and \( r_{\text{max}} \) (Yates, 1991). However, the number of low-\( r \) units reported (2 units) was small, and the use of fibers representing a large range of frequencies leaves the possibility of a lower correlation over a smaller basilar membrane patch. Also, the sigmoidal curve loses its round edges with higher thresholds, producing a linear plot vs. dB SPL (Winter, Robertson, and Yates, 1990). Such units had not saturated at the highest levels tested (110 dB SPL). The cat could also have a high-threshold population (noted in Rhode and Smith, 1985), which would yield a greater dynamic range for a fiber pool if sufficiently populous, or if separately processed. When the correlations and populations noted above are confirmed and quantified our equations can easily be adjusted, and the computations redone.

It had already been suggested that a small number of units with very large dynamic ranges may account for the missing dynamic range at the upper limit of hearing, when an “ideal computer” is assumed (Viemeister, 1983). Earlier calculations showed that, in principle, a small auditory sub-ensemble could potentially fulfil this role (Hellman and Hellman, 1975; Viemeister, 1983). This ensemble would be a second channel, as suggested by Viemeister’s calculational requirement that saturated channels be ignored, a requirement common to any multi-channel model of auditory dynamic range.

**The role of sloping-saturating units**

Sloping-saturating units did not profoundly extend overall dynamic range. Clearly these units make a difference only when relatively isolated, and not as much as desired, due to both the relatively weak influence of the dynamic range distribution, and the considerable overlap of the threshold distributions. To emphasize, we recomputed the upper channel (low-\( r \), units) assuming the sigmoidal equation used for other units (Equation 1) and the same dynamic range distribution as for high-\( r \), units, so that threshold was the crucial factor. By thus repeating the simulations of Figures 25 and 29, the advantage over a single channel (Figure 25) was now only 2 dB, not much different from the small 4 dB we discovered.

We might ask what happens to dynamic range if it is based on single neurons and not pooled response, which, as we have seen, lowers the jnd according to the square root of the size of the neural population (Equation 13). Sloping-saturating units are commonly said to have much larger dynamic ranges than sigmoidal units, with dynamic range typically defined as “from the just subthreshold point in spontaneous activity up to the first level at which the discharge had saturated” (e.g. Palmer and Evans, 1979). Thanks to the method shown in Chapter 2 (Equation 13, \( m=1, \bar{r}(x)=r(x) \)), it is now possible to calculate the discriminability \( \Delta x \) for pure tones offered by a single auditory afferent fiber under the Signal Detection Theory assumptions that a) repetitions of the stimulus produce normally distributed spike counts over a
counting time \( \tau \) (the stimulus duration, not any suspected sensory "integration time"), b) the variance of these distributions is the same at the sound level \( x + \Delta x \) as at the level \( x \), for small \( \Delta x \). All else that is needed are the fiber's rate-level curve, an assumed counting time (here, 51 msec), and the mean-to-variance ratio \( \rho \) for the Gaussian distribution (here, assumed constant for all fibers, at \( \rho = 1.5 \)). Hence discriminability was computed for both the actual rate-level data points, and for the fitted curve \( r(x) \) (Equation 1 [for sigmoidal] or 3 [for sloping-saturating]). We employed 13 sloping-saturating fibers (1 from Wiederhold, 1970, 6 from Sachs and Abbas, 1974, 5 from Palmer and Evans, 1979, and 1 from Delgutte, 1987) and 13 sigmoidal fibers (1 from Wiederhold, 1970, 7 from Sachs and Abbas, 1974, 2 from Liberman, 1978, 1 from Delgutte, 1987, 1 from Sachs, Winslow, and Sokolowski, 1989, and 1 from Winslow and Sachs, 1988). These generally U-shaped curves (see Figures 33 and 34) vary from fiber to fiber and have a lowest point no lower than (at best) about 2.5 dB, offering discriminability much less than that afforded by pooled spikes from a fiber population. This demonstrates clearly the advantage of pooling. These lowest points can have similar magnitudes for both sloping-saturating fibers and sigmoidal fibers, which presumably occurs because the slopes of the steeper portions of the sloping-saturating fibers are similar to the steep slopes of the sigmoidal fibers. (Note the dependence of \( \Delta x \) on slope \( r(x) \) [Appendix D]; greater slope represents greater responsiveness to small changes in stimulus level, and vice versa.)

Useful encoding range was estimated first by differentiating the jnd equation for the fitted curve \( r(x) \), setting this derivative equal to zero, and solving for \( x \). Although \( x \) may have more than one value (see Appendix D), the value of \( x \) corresponding to the minimum in the U-shaped jnd-curve can be chosen without ambiguity with the help of Figures 33-34. This \( x \) in dB SPL is then substituted back into the jnd equation (Equation 13, \( m = 1 \), \( \tau(x) = r(x) \)). A criterion number of decibels, based on visual inspection of the jnd plots, is then added to this minimum value. This jnd corresponds to two points on the jnd curve computed from \( r(x) \); we assumed that the difference between their \( x \)-values in dB gives the useful dynamic range (see Appendix D for the algebra and numerical method). For criteria of 2, 4, and 6 dB, the average ranges were \{14.06, 15.86\}, \{18.90, 21.47\}, and \{22.32, 25.43\} for the 13 sigmoidal and 13 sloping-saturating units, respectively. The Student-t statistic reveals no significant difference between the numbers within each of these pairs, for a one-tailed test with acceptance of the null hypothesis (no difference; \( D = 0 \)) for \( t < t_{0.05} \) (Table 6). When two more sigmoidal units were evaluated (1 from Liberman, 1978, and one from Schalk and Sachs, 1980) and their computed dynamic ranges counted into the average dynamic ranges, the latter were \{13.13, 15.86\}, \{17.64, 21.47\}, and \{20.81, 25.43\} for the 15 sigmoidal and 13 sloping-saturating units, respectively. T-tests show a significant difference between the numbers within each of these pairs at \( p < .05 \), but significance disappears if the difference between the numbers is allowed to be 1 dB rather than none at all. That is, the numbers within each of these pairs do not significantly differ from each other by more than
Figure 33a. The discriminability of individual sigmoidally-saturating auditory-nerve fibers whose rate-level plots were taken from Fig. 2 (right-hand-side) of Sachs and Abbas (1974). Each fiber's plot is identified by its CF.
Figure 33b. The discriminability of individual sigmoidally-saturating auditory-nerve fibers whose rate-level plots were taken from a variety of sources. CF=0.28 is from Sachs and Abbas (1974, Fig. 2, right-hand side); CF=1.4 (all three) are from Liberman (1978, Fig. 16); CF=8.41 is from Delgutte (1987, Fig. 3a); CF=9.22 is from Sachs, Winslow, and Sokolowski (1989, Fig. 1); CF=13.7 (upper) is from Schalk and Sachs (1980, Fig. 4); CF=13.7 (lower) is from Winslow and Sachs (1988, Fig. 3a); and CF=16.5 is from Wiederhold (1970, Fig. 3a).
As long as the same criterion is applied to all fibers, and the criterion is sensible, then changing to a different criterion that follows these principles should not produce reversals in the overall pattern of results. For closer analysis, each of 6 sloping-saturating units, those of Figure 34a (CFs = 0.32, 1.76, 4.30, 7.74, 14.5, and 20.5 kHz [Sachs and Abbas, 1974, Fig. 2]), were paired according to similar CF with 6 sigmoidal units taken from the same study, those of Figure 33a (CFs = 0.32, 1.76, 4.30, 7.74, 14.5, and 24.5 kHz). The similar CFs assure nearby basilar locations for the units of each pair. The t-tests were repeated for this sample of 6 units of each kind. For a hypothesised difference of zero, the criteria of 2, 4, and 6 dB gave Student-t's of 0.53194, 0.58570, and 0.63581, respectively. These are less than $t_{0.05}=1.812$ for 10 degrees of freedom (Mendenhall, 1975, Table 4), not rejecting the hypothesised difference of zero. The jnd curves are juxtaposed in Figure 35 where the lack of difference can be visually appreciated.

For fiber discriminability curves that broaden out from the bottom, such as those seen here, the dynamic range assigned to each fiber depends on the arbitrary criterion number of dB added to the minimum point on the fiber's discriminability curve. This method of estimating fiber dynamic range is thus a circular argument; a higher criterion sets a higher dynamic range. Now, by restricting the method to only comparisons of one fiber's range to another's, this circularity appears to be removed. However, the different criteria employed (2, 4, or 6 dB) are still arbitrary choices and are different measurements, which can produce different results. Why choose 4 dB, say, instead of the 10 db used by Delgutte (1987; see below)? Inspection of the discriminability curves, and comparison to each corresponding rate-level curve, shows that these criteria represent the assumption that useful encoding of small changes in stimulus level can only be accomplished using the initial, steeply sloping portion of the rate-level function of the sloping-saturating fiber (unlike the inherent assumption in Delgutte). Because discriminability depends partly on rate and partly on slope (Appendix D), all that we have proven here is that the slopes of the initial, steeply sloping portion of the rate-level function of the sloping-saturating fiber do not differ significantly from those of sigmoidal fibers. The present exercise still has value, however, because it presents an alternate (and perhaps more reasonable) viewpoint about what constitutes true dynamic range.

All this evidence suggests that sloping-saturating units do not offer an extended dynamic range compared to sigmoidal units, and hence that sloping-saturating units are not a solution to the "dynamic range problem". When Evans and Palmer (1980) published the only study specifically devoted to dynamic range, they noted that fibers with spontaneous rates <15 spikes/sec had a mean dynamic range of 50 dB and that other fibers showed a mean range of 41 dB, a difference in ranges significant at $p<.005$ in a Student's t-test (cat). For 121 fibers from a single cat, 9% were said to have dynamic ranges exceeding 60 dB (Palmer and Evans, 1979). Their definition of dynamic range (see above) is much more generous than ours. The present
data suggests that all fibers have much smaller dynamic ranges than reported in the two latter papers, and that estimates of 60 dB are fanciful artefacts of a generous imagination that exaggerates the dynamic ranges of all auditory afferents.

Clearly, an operational definition of afferent dynamic range is needed, especially given that we cannot otherwise infer the discriminability offered by the pooling of auditory afferent firing. We discuss this issue further in Chapter 6. In the present Chapter we assumed these published dynamic ranges in our computations. These ranges appear in the equations for single-unit firing rate (Equations 1 and 3). Should these range have been much smaller? No, because the parameter \( \lambda \) representing fiber dynamic range appears to follow a definition that is effectively equivalent to the observational criterion of Palmer and Evans (1979) and other authors. That is, \( \lambda \) itself is larger than what we here infer to be the useful dynamic range of the fiber, and is used the way it is only because it helps to provide a good fit for the firing-rate equations. Indeed, as long as it is used in a manner consistent with the literature, it need not represent the fiber's inferred useful dynamic range.

The conclusion that sloping-saturating units offer no greater useful dynamic range than sigmoidal units is unlikely to change if we employ a floating value of \( \rho \) for low-\( r \), units (see Chapter 7). Indeed, making \( \rho \) lower as fiber threshold is approached actually increases the slope of the lower-intensity branch of the U-shaped jnd curve (Figure 34c). This reduces the single-fiber dynamic range estimated by the present method, perhaps by several dB, so that the average dynamic ranges offered by sloping-saturating units, for given criteria, are likely to not differ from those offered by sigmoidal units. Indeed, some sloping-saturating units seem likely to show lower dynamic ranges than those offered by some sigmoidal units.

There are already estimates of single-fiber dynamic range that are based on jnd curves. In the method above, we measured the width of the jnd plot at a dB level relative to its minimum point, a level that varies from fiber to fiber. Delgutte (1987) instead measured widths at a fixed discriminability of 10 dB, using actual 2IFC measurements of fiber discriminability. Examination of his two display plots, as well as our own, suggests that this width is measured too high up on the jnd curve to be a true discriminability measure. But it is easy to see why Delgutte chose a relatively high, fixed value for his dynamic range criterion. The lowest point (best discriminability) of his plots of single-fiber discriminability varied from less than 1 dB to over 10 dB, with the scatterplot apparently centered at about 3 dB. Thus measuring widths of jnd plots at a fixed level of (say) 3 dB would have assigned no dynamic range at all to many fibers, and large ranges to some of those whose jnd plots had lowest-points that were very low. The 10-dB range was clearly chosen to avoid the former problem, but it emphasised the latter problem. In view of the U-shape of the fiber-jnd plots, the 10 dB criterion should generally assign widest dynamic ranges to fibers whose discriminability plots have the lowest values of best discriminability. This appears true; close comparison of Delgutte's Figure 3A (best-discriminability values) to his Figure 3B (10-dB dynamic range) suggests that
Figure 34a. Similar to Figure 33, but for sloping-saturating fibers whose rate-level plots were taken from Fig. 2 (left-hand-side) of Sachs and Abbas (1974). The rate-level plots were fitted to the double-logistic function (Equation 3). These units have CFs similar to those of Figure 33a and therefore represent nearby positions on the basilar membrane, suggesting that differences between sloping-saturating behaviour and sigmoidal behaviour are not correlates of basilar-membrane locus.
Figure 34b. Similar to Figure 33, but for sloping-saturating fibers from various sources, fitted to the double-logistic function (Equation 3). The six panels on the left-hand side which have no CF designation are derived from rate-level plots taken from Palmer and Evans (1979, Fig. 2). CF=8.13 is from Delgutte (1987, Fig. 2b) and CF=5.26 is from Wiederhold (1970, Fig. 4a).
Figure 34c. Jnd curves for the sloping-saturating fibers whose rate-level plots were taken from Fig. 2 (left-hand-side) of Sachs and Abbas (1974). Unlike Figure 34a, the mean-to-variance ratio $\rho$ was adjusted relative to fiber threshold $\varepsilon$. Thus $\rho=0.1$ for $\varepsilon \leq 5$ dB; $\rho=0.5$ for $5 < \varepsilon \leq 10$ dB; $\rho=0.75$ for $10 < \varepsilon \leq 20$ dB; and $\rho=1.5$ for all other $\varepsilon$. The rate-level plots were fitted to the double-logistic function (Equation 3).
they are roughly mirror-images around a horizontal line (the abcissa, which for both illustrations was fiber CF).

A second reason for Delgutte's choice of a 10-dB criterion may have been conformity. The dynamic ranges he assigned to his fibers have a grand average (eyeballed from his Fig. 3b) in the range of 30-35 dB, with some fibers having dynamic ranges over 50 dB and very few having dynamic ranges below 20 dB. These predictions agree with the trends in cat dynamic range reported by Evans and Palmer (1980) and others.

In this Chapter we have used a fixed level of discriminability (3 dB) to estimate the overall dynamic range offered by a neural pool, under the assumption of equal weighting of fiber counts at the putative counting center. However, it appears inappropriate to estimate useful dynamic range for a single neuron according to the width of the discriminability plot at a fixed level of discriminability (Delgutte's method). Delgutte's inherent assumption is that fibers contribute to overall perceptual discriminability in proportion to the lowness of their best-discriminability value. Thus fibers with a best-discriminability that is relatively high might be considered useless for purposes of discrimination. But there appears no a priori reason for Delgutte's assumption; for example, if only a fiber of relatively poor discriminability had a firing-rate threshold low enough to respond to a very quiet sound, then this particular fiber would have to be the one employed for coding changes in stimulus intensity. That is, units with higher best-discriminability values might be used decisively if other units are already firing at saturation and are thus unable to signal changes in stimulus level. Our own method for single fibers (above) does not make Delgutte's assumption, which was equivalent to a weighting, but instead assumes that each fiber can potentially be called upon for its powers of discriminability. Like Delgutte's assumption, this too is an arbitrary one.

Would low-spontaneous-rate neurons have any advantage as an upper channel?

The role of the low-\(r\) units as an upper channel is suggested by the superior encoding ability for speech syllables presented at 70 dB SPL in a noise masker, as compared to that of high-\(r\) units (cat; Silkes and Geisler, 1991). This is what we would expect if a lower channel saturated by about 50-60 dB SPL and was hence not well able to code for 70 dB SPL speech sounds. Such a lower channel's output would presumably be ignored above the crossover point for the two channels.

Other evidence suggests the superiority of low-\(r\) units in realistic listening conditions. Inspection of fiber rate-level curves to CF tones in both long-term continuous noise and simultaneously-gated noise shows a greater firing rate to low-level tones, a drop in maximum rates, and a shift of the rate-level curve to higher SPLs (Costalupes, Young, and Gibson, 1984). These effects increase with noise level, causing a progressive compression in the fiber's range of available firing rate. This compression gets stronger with
Table 6. Values of the Student-t statistic for comparison of the average dynamic ranges of sigmoidal vs. sloping-saturating fibers. These were compared to the tabulated values of Student's $t$ found in Mendenhall (1975). The statistic is

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{s} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $\bar{y}_1 =$ the larger of the two averages being compared,

$\bar{y}_1 =$ the smaller of the two averages being compared,

$D_0$ is the hypothesised difference of the two averages,

$s$ is the $\sqrt{\text{sample variance}}$ which is

$$s^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{j2} - \bar{y}_2)^2}{n_1 + n_2 - 2}.$$ 

The number of fibers of each fiber type are $i$ and $j$. The $t$-test assumes that the individual fiber dynamic ranges are samples from a Gaussian distribution. A Gaussian can roughly fit the dynamic range data that we employ (Figure 18) in computing the dynamic range of a neural pool, that of Evans and Palmer (1980). Although their observational criterion for dynamic range (Palmer and Evans, 1979) inflates fiber dynamic range (see text), we will nonetheless assume that the dynamic ranges arising from our own measurement technique also follow a Gaussian. For 13 sigmoidal units and 13 sloping-saturating units:

<table>
<thead>
<tr>
<th>Criterion increment, in dB</th>
<th>$D_0$</th>
<th>$D_0$</th>
<th>$D_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>+2</td>
<td>1.42493</td>
<td>0.63475</td>
<td>-0.15543</td>
</tr>
<tr>
<td>+4</td>
<td>1.50484</td>
<td>0.92039</td>
<td>0.33594</td>
</tr>
<tr>
<td>+6</td>
<td>1.54323</td>
<td>1.04756</td>
<td>0.55188</td>
</tr>
</tbody>
</table>

When two more sigmoidal units are included (see text),

<table>
<thead>
<tr>
<th>Criterion increment, in dB</th>
<th>$D_0$</th>
<th>$D_0$</th>
<th>$D_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>+2</td>
<td>1.98542</td>
<td>1.25688</td>
<td>0.52834</td>
</tr>
<tr>
<td>+4</td>
<td>2.05878</td>
<td>1.52132</td>
<td>0.98385</td>
</tr>
<tr>
<td>+6</td>
<td>2.09551</td>
<td>1.64166</td>
<td>1.18782</td>
</tr>
</tbody>
</table>
Figure 35. A comparison of jnds calculated from functions fitted to digitised data of Sachs and Abbas (1974). The solid curves are sigmoidal units (from Figure 33a) and the dashed curves are sloping-saturating units (from Figure 34a). Each unit is identified by its characteristic frequency (CF).
rise in fiber spontaneous rate, leading Costalupes et al. to suggest that low-spontaneous-rate units may be the main purveyors of intensity information at high SPLs. A similar shift in fiber operating range under the long-term influence of a masker tone near the probe CF was found in the chinchilla, whose spontaneous-rate groupings are similar to those of cats (Sinex and Havey, 1986). Tone-masked thresholds rose as spontaneous rates dropped, an effect confirmed for fibers of similar CF but different r, giving a potential advantage of 10 dB or so to low-r, units, even in a small area of basilar membrane for which our computations apply. Low-r, units are also better encoders for a stimulus that is closer to real speech than are pure tones, an amplitude-modulated narrow-band sound (125 Hz wide, 2 kHz center frequency). Higher envelope modulation was replicated in the firing rates of low-r, afferents as compared to high-r, units (Wang and Sachs, 1992).

Thus there appears to be some evidence that low-spontaneous-rate units are advantageous to hearing. It seems reasonable to postulate that this advantage can only be enjoyed if these units are computationally isolated in a second stream of auditory intensity processing, i.e. a second channel. There may be a neural basis for the advantages shared by low-r, units; there is a greater representation of low-r, units as compared to high-r, units, due to differing branching patterns (Fekete, Rouiller, Liberman, and Ryugo, 1984) and different numbers of terminal boutons (Ryugo and Rouiller, 1988), in certain portions of the cochlear nucleus, the only terminus of primary afferents (Osen, 1970, and others noted there; Fekete, Rouiller, Liberman, and Ryugo, 1984).

Possible dynamic ranges of auditory channels in man

How does the cat neural population within the critical band- and hence the inferred discriminability limen - compare to that of man? Comparison of critical-band width at the same frequency may not be as appropriate as comparison of bandwidth at equivalent percentage-distances along the respective basilar membranes (Greenwood, 1991). The number of fibers in a human critical band at mid-to-high CFs in man can be calculated by noting that the number of neurons per inner-hair-cell (IHC) is roughly constant at 10.5 (Nadol, 1990), that the ratio IHCs/mm is approximately constant at 86.35 (average over many cochleas; Ulelova, Voldrich, and Janisch, 1987) and that the ratio [mm of basilar membrane/critical band] is about 0.53 on the Greenwood assumption of equal cochlear lengths per critical band (for the narrower estimates of the critical band; see Greenwood, 1991). This argument gives 481.4 afferents per critical band in the middle range of human frequencies, about half the number we employ for the cat at 8 kHz. The human numbers will shift the computed jnds upward by $\sqrt{2}$ of their present values, reducing the encoding range of each channel. Because the change is multiplicative, the plotted jnd curves themselves will not simply shift upwards by a constant amount; their shape will also change. The actual frequency that corresponds to the 8 kHz place in the cat can be computed from Greenwood (1990) by calculating the proportional distance from the apex of the organ of Corti for 8 kHz in the cat (Greenwood's Fig. 2) and substituting this distance into the comparable equation for man (Greenwood's
Equation 1). The result is 2,862 Hz. JND measurements are available for monaural 30 msec stimuli centered at a nearby frequency, 3 kHz (Buus, 1990). Assuming that a lower estimate of the human critical band is the equivalent rectangular bandwidth (ERB), then the equation of Moore and Glasberg (1983) gives 365 Hz. The Buus jnds for frozen noise (identical sounds) of 150 Hz bandwidth are 3.44, 2.69, and 1.15 dB for 30, 60, and 90 dB SPL pedestals, and 2.99, 2.56, and 0.75 dB for a 500 Hz bandwidth. (JNDs for non-frozen [random] noise were 0.3-1.5 dB higher.) The 365 Hz bandwidth would presumably produce intermediate jnds; let’s choose 1 dB at 90 dB SPL and assume that if the computed cat jnd at 90 dB SPL falls above this limit, after correcting our computations for the difference in population size, then human discriminability also fails. Under this criterion, a single channel fails beyond 66 dB SPL; shifting all mean thresholds upwards by 20 dB still gives 1-dB limits of only 38-86 dB SPL for a single channel. A low-r, upper channel (composed of 82 units) has an even lower limit of 62 dB SPL, due to the drop in population size in the critical band as compared to the cat.

**SUMMARY OF RESULTS**

Loudness change is perceived over a far greater range than the discrimination range of single afferents (the "dynamic range problem"; Evans, 1981). However, earlier neural count models had not fully considered how variability in the dynamic ranges, thresholds, spontaneous rates, and saturation rates of these afferents may affect overall dynamic range. In the present model all four of these characteristics appeared in a logistic rate-intensity function that fits well to sigmoidally-firing cat neurons; a double-logistic fits well to sloping-saturating units. These equations were averaged statistically over parameter distributions for each of 3 spontaneous-rate groups. An average over these groups, weighted by relative group size, was used to compute discriminability for a patch of basilar membrane. The upper limit on dynamic range for a patch one critical band wide at 8 kHz was 89 dB SPL. A two-channel model extended the upper limit by only 3 dB and splitting the population into four channels did not improve this limit. Seven channels, optimally combined, provided a dynamic range nearly equal to that of the ideal auditory computer that treats each fiber as a separate channel. Moreover, the contribution of sloping-saturating units to discriminability was not as profound as popularly expected. Changing the variances of fiber dynamic ranges did not strongly affect overall dynamic range in any of these computations. In sum, when the distributions of neural parameters are accounted for, local pooling can greatly mitigate the "dynamic range problem".

It is unlikely that uncertainty in the estimated values of the neural parameters is responsible for the failure of a single-channel model. In estimating overall dynamic range we chose the highest likely means and variances for the two essential distributions \( p(\epsilon) \) and \( p/(\Delta) \). Different measurements give notably different values for threshold (Liberman, 1978; Geisler, Deng, and Greenberg, 1985; Rhode and Smith, 1985) and dynamic range (Evans and Palmer, 1980; Schalk and Sachs, 1980). Indeed, a single-channel model employing more conservative estimates of fiber dynamic range and threshold variation (after Schalk and
Sachs, 1980) produces an overall dynamic range that is considerably smaller that the cat’s actual range. Splitting the fibers into two channels failed to extend overall dynamic range to the behavioural limit, regardless of whether the upper channel was composed only of low-r, units, or both low- and mid-r, fibers, implying that threshold distributions (at least for upper channels) must be wider, or that a mechanism must exist to compensate for their narrowness. Compensation could come from an active gain control involving the efferents to the outer hair cells (Duron and Schacht, 1992), which if functioning in active, alert, and perhaps attentive animals, might be able to extend overall dynamic range by shifting the thresholds of all the units in a pool. An improved overall range is thus obtained without having to assume a computationally complex ideal observer.
NOTES

15. Changing \( \varepsilon \) by \( \pm \Delta \varepsilon \) causes an identical shift in the mean rate plot. To see this, replace \( \varepsilon \) by \( \varepsilon + \Delta \varepsilon = \beta \), a leftward shift along the \( \varepsilon \) dimension, such that \( p(\varepsilon + \Delta \varepsilon) = p(\beta) \) and \( d(\varepsilon + \Delta \varepsilon) = d\beta = d\varepsilon \). Substituting into Equation 1 produces \( r(\varepsilon, \beta, \lambda, r_{\text{max}}, r_t, x-\Delta \varepsilon) \) which with all appropriate substitutions made into Equation 6 gives

\[
\frac{r(x-\Delta \varepsilon)}{r(x)} = \frac{\beta_n}{\beta_c} = \frac{r_{\text{max}} - r_t}{1 + 100 - c \exp \left\{ -2 \left( \frac{x - \beta}{c} \right) \ln \left( \frac{100 - c}{c} \right) \right\} + r_t} p(\beta) p(\lambda) d\beta d\lambda.
\]

The mean rate has shifted leftward along the \( x \)-axis; the right-hand-side is the old mean rate (Equation 6), since \( \varepsilon \) and \( \beta \) are both independent of \( x \) and the shape of the threshold distribution has not changed. Repeat these steps for \( \varepsilon - \Delta \varepsilon \).

16. Although our computations show that the jnd curve must eventually turn upward at the upper limit of auditory dynamic range, few studies have used sufficiently high SPLs to reveal this feature. JNDS for the 1 kHz tone do turn upward above 95 dB SPL, although they are still smaller at 120 dB SPL than at 70 dB SPL (Viereimier, 1988). This upturn is inevitable, as the number of unsaturated nerve fibers available to signal intensity changes must eventually decline.

17. Greenwood (1961a) observed a slight drop in masked thresholds when the total power in the critical band exceeded 50 dB, a phenomenon that he attributed to the sudden excitation of a new population of hair cells. Greenwood (1993) also re-analysed Viemeister’s (1983) notched-noise experiment and found a similar (but less sharp) transition for a narrowband noise. Munson and Gardner (1950), and Zwislocki, Pirodda, and Rubin (1959), found that the threshold of a 1-kHz pure tone, preceded by a 1-kHz masker, increased with masker intensity, except in the region between 50 and 80 dB, where increases in masker intensity had little effect on threshold. Munson and Gardner (1950) suggested that a transition to a second population of nerve fibers was responsible for this plateau in the otherwise steadily-increasing masking function. Indeed, Yates, Winter, and Robertson (1990) have suggested such a multichannel model for the guinea pig. Yates et al. concluded that “the full range of loudness seems, in the guinea pig at least, to be coded by two or more sets of fibers, each with a different dynamic range. Simple summation of the responses of a set of fibers would not lead to the psychoacoustically-observed loudness functions, but a relatively simple gating scheme under which the more sensitive set of fibers was gradually switched out as stimulus intensity was increased and other, less sensitive, fibers began to respond, could well do so”.

18. These channels may be even further apart. For low spontaneous rate units (\( r_t < 0.5 \) sec), experimental mean-to-variance ratios are not available below 20 dB above each fiber’s threshold (Teich and Khatana, 1985, Fig. 7), so as a trial run we approximated this effect by using a \( \rho \) that depends on the stimulus level with respect to the minimum value of the threshold density function \( p(\varepsilon) \), i.e. the ensemble of low-\( r_t \) units was treated as if being a single (average) unit of variable \( \rho \). For a fiber ensemble with lowest thresholds at 15 dB SPL, the following ratios were extrapolated: 0.1 at 20.5 dB; 0.5 at 28 dB and 30 dB; and 0.75 at 40 dB SPL. The resulting differences in \( \Delta \varepsilon \) are not notable for the useful middle range of \( \Delta x \), except that the lower boundary of this useful range steepens and hence moves upward by 3-5 dB, increasing the separation between lower and upper auditory channels.

19. How do our single-channel results compare to whole-nerve models when the latter are restricted to a limited basilar region such as that serving a single critical band? Starting with Goldstein (1974), we find that in his model, using the simpler of two possible across-basilar-membrane neural excitation functions (the one that takes lesser account of the contributions of the long tails of the tuning curves of high-CF fibers (cf. Kiang and Moxon, 1974), gives a saturation that starts at about 50-60 dB SPL and is essentially complete by 90 dB SPL for the model’s tonal frequency (5.83 kHz). In Howes’ equations (1974), fiber recruitment with spread in CF appears to be a prerequisite for the continued increase in summed firing rate (1974, Fig. 2), and appears as the basilar membrane neural fiber density...
displacement was modelled as having three distinct regions; higher fiber thresholds represent higher thresholds on this curve. The rate-intensity functions form a continuum showing a systematic change in slope with increase in threshold, from flat-saturating to sloping-saturating to the shallowest functions, straight (linear in dB). As threshold rises, fiber dynamic range also rises (although the behaviour of the *useful* dynamic range is not clear; see the above section on sloping-saturating fibers). The use by Sachs, Winslow, and Sokolowski (1989) of a higher power-function exponent in the sub-equation for basilar-membrane displacement is said to eliminate the straight function, in the cat (Yates, 1990). Like the Sachs *et al.* equation, however, the probability distributions for some non-constant curve-fitted parameters are not presently available. Straight rate-intensity functions can be incorporated into our model if necessary, using a separate, straight-line equation incorporating probability density functions for fiber threshold $\epsilon$ and fiber dynamic range $\lambda$, and by estimating the mean spontaneous rate. We also make some assumptions about the levels at which $r_{\text{max}}$ occurs and what is its mean value.

24. These comments apply to tones presented in quiet, the case considered here and in all the mean-rate computations discussed above. Sloping-saturating units respond differently in notched noise, having dynamic ranges shifted to higher SPLs (cat, tone at unit CF, 5 fibers, one-octave notch, Palmer and Evans, 1982; see also Palmer and Evans, 1979). Milder shifts are seen in other auditory fibers, and the majority (62%) showed no change in firing rate in notched noise (ibid.). Recall (Chapter 1) that Viemeister (1983) used a notched noise to show the unexpectedly large dynamic range of a limited portion of the basilar membrane, and we must assume that the encoding abilities of sloping-saturating units contributed to this range. Hence Viemeister’s results may have derived from more than just the encoding ability of the basilar patch as seen in the absence of noise.

25. Bandlimited noise is not really an appropriate comparison for our model, since the amplitudes of the many components are chosen randomly. Hence frozen noise, which is just a single bandlimited noise repeatedly presented, is also not an appropriate comparison. The only appropriate comparison is a sound which stimulates each frequency within the critical band at the same amplitude.
CHAPTER 4

EXPERIMENTAL METHODS, AND DATA ANALYSIS
EXPERIMENTAL METHODS

Subjects

Testing was rigorous and time-consuming, so only 3 subjects (BAK, LN, and JWC [two males and one female, respectively; average age = 32]) provided forward-masked detection thresholds. One of the subjects (LN, male, age 35) was the first author, one (BAK, male, age 23) was a paid undergraduate, and one (JWC, female, age 38) was a research assistant. Previous listening experience in this laboratory varied, respectively, from none (LN), to moderate (BAK), to extensive (JWC). Only 2 subjects (average age=23.5) provided forward-masked intensity-increment thresholds. JW (male, age 26), a paid lab technician, had moderate previous listening experience in this laboratory and ZS (male, age 21), a paid undergraduate, had none. All subjects had normal audiometric thresholds according to a commonly accepted standard (ANSI S3.6-1969). All subjects were treated according to American Psychological Association guidelines as published periodically in J. Acoust. Soc. Am..

Stimuli

A 2 kHz, 200 msec (excluding ramps) forward-masker was followed in one of the two forced-choice intervals (detection thresholds) or both of the two intervals (intensity-increment thresholds) by a 2 kHz tone pip (Figure 36). In Experiment 1 the masker was fixed at 97 dB SPL; in Experiment 2 it was always 20 dB lower than the pedestal level of the tone pip. The tone pip was formed by multiplying a 2 kHz pure tone carrier by a Gaussian amplitude envelope (maximum amplitude =1) whose standard deviation was 0.5 msec. The reported pip intensity is the root-mean-square sound pressure level (dB SPL) of the carrier. The 200 msec masker was formed by adding together 201 of these tone pips spaced 1 msec apart following the procedure used in Schneider et al. (1994) to create the 200 msec 2 kHz tone whose on and off ramping was determined by the Gaussian envelope of its individual pips. Now, in most publications the probe pip delay is defined as the interval between the offset of the masker and the onset of the probe, or the interval between the offsets of the masker and the probe pip. Furthermore, the many previous studies of forward masking tended to use much longer probe stimuli, so that post-masker behaviour could not be described as a function of specific post-masker times. Thus behaviours at two different masker-pip time gaps $t$, regardless of how $t$ was specified, could only be compared when these $t$ were separated by more than the duration of the probe, so that there was no overlap of probe profiles in time. Here, although the Gaussian ramping of both the pip and masker makes it difficult to specify the interval between the offset of the masker and the onset of the probe, $t$ can be defined as extending from the center of the last Gaussian envelope in the masker to the center of the pip. This interval was always an integer multiple of the period of the 2 kHz carrier, giving zero phase shift between masker and tone and thus eliminating a potential detection cue to which the human ear is quite
Figure 36. The stimulus and the probe pip's energy spectrum. A 97 dB SPL, 2.5 Hz tone, 200 msec long (excluding ramps), was followed by a 2 Hz tone pip having the same Gaussian ramping (standard deviation=1/2 msec). Standard deviation, 1.5 msec. The stimulus and the probe pip's relative spectral density of the tone pip.
sensitive (Rajcan, 1976). The pip’s amplitude is down to 14% by $\pm 2\sigma$ of the Gaussian envelope’s peak and 1% by $\pm 3\sigma$, which should minimise any small amount of post-masker perceptual adaptation that might be caused by the pip itself.

Another advantage of Gaussian shaping is that it provides the smallest amount of spectral spread for a given tone-pip duration (Gabor, 1946; see Schneider et al., 1994). Moreover, the relative spectral density of the tone pip shows only a single relatively narrow lobe (Figure 36). It is generally recognised that many non-Gaussian ramping schemes (e.g., linear ramps, Hanning and/or Hamming windows) produce multiple side-lobes in the relative spectral density, spreading energy over a wide frequency range. Such energy, which is spectrally remote from the carrier frequency, could be used as a detection cue. Wightman (1971) has illustrated this spread for 2 kHz tone pips with linear ramps (Figure 37).

All stimuli were generated as digitised voltage values at a sampling rate of 20 kHz, using software kindly supplied by Prof. Bruce Schneider. These were stored as files on diskettes until needed. When used, the voltages were converted to analog form using a 16-bit digital-to-analog converter (TMS320C25 digital signal processing system).

**Apparatus**

Subjects sat in a double-walled sound-attenuating chamber wearing a TDH-49 headset (sound to the right ear only, red lead, serial # A17522), with the input from a Tucker-Davis HB5 headphone amplifier. Harmonics appeared at 4, 6, and 8 kHz, at 50 dB down; however, anything higher than 6 kHz is beyond the range of the headphones. The headphone was calibrated just before the experiment (Bruel and Kjaer 2209 sound level meter, 1 inch microphone in a 6 cc coupler) and headphone voltage calibration was done weekly. A Hewlett-Packard 350D passive attenuator (always set at 40 dB of attenuation) was also used for sounds below 40 dB SPL. Maximum output to the right earpiece was 102.3 dB SPL and all subsequent reported sound levels were rounded down by 0.3 dB to achieve whole numbers25.

**Procedure**

Subjects performed two-interval forced choices (2IFCs). A 1.5 sec delay separated each of the two intervals in each forced-choice. The literature suggests that 1 sec is sufficient recovery time for typical forward-maskers (Samoilova, 1959, 20 msec probes; Weber and Moore, 1981, 5 msec probes). In experiments determining forward-masked probe-detection thresholds, the masking situation most like our own employed a 70 dB SPL pure-tone forward-masker, lower than the present level, but, at 300 msec, longer in duration (Fastl, 1979). Now, the shift in the pure-tone probe’s threshold induced by pure-tone maskers of 90 dB SPL (Kidd and Feth, 1982, 20 msec probe) or bandpass-noise maskers of 50 dB SPL/Hz (Carlyon,
1988, 10 msec probe) generally increase substantially with increase in masker duration, at least up to 300 msec. Thus we may assume that the effects of Fastl’s 300 msec maskers were similar to our own. Fastl’s probe was also the most similar to our own, with 1 msec Gaussian ramps, a 2 msec plateau level, and frequencies of 3.5, 4, or 5.5 kHz. Under these conditions, psychophysical recovery of nonmasked threshold was almost complete by only 200 msec post-masker (Fastl, 1979). Thus an inter-stimulus interval of 1.5 sec should allow substantial recovery from the 97 dB SPL forward-masker. An inter-stimulus interval of 1 second also appears adequate for recovery of the whole-nerve action potential (CAP) in the cat (Abbas, 1984), in which the CAP is a neural correlate of the brief Gaussian pip.

In forward-masked threshold detection, one interval was a comparison condition in which the masker preceded a pip of constant level. The other interval contained the same masker and time gap, but no pip. The order of these 2 intervals was randomised on each forced-choice, and the subject’s task was to identify the interval containing the pip, with their choice indicated as right or wrong by a visual signal following each discrimination. Establishment of forward-masked intensity-increment thresholds differed slightly. In one interval of the 2IFC task, a 2 kHz, 200 msec (excluding ramps) forward-masker, was followed by a 2 kHz tone pip having the same Gaussian ramping (σ=0.5 msec) but 20 dB higher in level (Figure 38). This was the comparison condition containing the comparison pip of constant level (the pedestal level). The other interval contained the same masker and time gap, but a pip of higher level created by simply increasing the level of the pedestal tone. The order of these 2 intervals was randomised on each forced-choice, and the subject’s task was to identify the interval containing the pip of higher intensity, with their choice being indicated as correct or incorrect by a visual signal following each discrimination.

The pairs of forced-choices were given in blocks of 100 with both pip intensity and masker-pip time gap remaining constant within each block (the method of constant levels). Over 100 such self-paced choices the subject produced a percentage-correct score recorded as p(c). Although p(c) cannot be greater than 1, it can be less than 0.5 if the stimulus is presented at such a low level that correctly choosing the stimulus interval is a matter of chance. To avoid any performance drop due to fatigue or attention loss (Zwislocki et al., 1958), 10-20 min. breaks were offered between blocks. No subject was tested until ready; much longer breaks were given if requested, and subjects were free to request.

The order of time gaps was irregular. When each subject started at a new time gap, a new series of pip levels was chosen that was customised to each subject based on their performance at nearby time gaps. (For the first few time-gaps, some trial-and-error was required.) The experimenter first chose a pip level just above that level at which p(c) starts to drop below 1.0. Over successive blocks the pip level was lowered monotonically until the subject started making errors (p(c)<1.0) and then further until performance approached chance (p(c)=0.5). Such monotonic intensity changes proved easier for the participants to follow
Figure 37. (Wightman, 1970, Fig. A-4) Examples of the effects of ramping duration on the relative energy spectrum of a 2 kHz tone, for a linear ramping (i.e. a trapezoidal tone envelope). The rise time is the time from envelope-amplitude=0 to envelope-amplitude=constant. Total tone duration was not mentioned.

Figure 38. Forward-masked increment-detection. Examples of the forward masker and tone pip for the 3 msec masker-pip time-gap. The forward masker has a 200 msec plateau and is a 2 kHz tone with the same Gaussian ramping as the pip (standard deviation =0.5 msec). For clarity of illustration the central 185 msec of the masker is not shown.
than the sudden large relative changes in level that occur when pip level is changed randomly from block to block over a range of dB. Randomising intensity across blocks was tried initially and resulted in poorer performance. (In particular, in the pip-detection task (Experiment 1), the great widths found for the psychometric function at early t meant that random changes in pip level could be profound and hard to cope with.) Thus, to make the detection task easier, subjects were told to use any listening cue that worked. Commentary from the subjects was encouraged in order to check for changes in listening criterion or problems in performance, and the subjects were encouraged to confirm their impressions with each other.

Each day, enough data was gathered to assemble a psychometric function (percent correct vs. intensity level). The number of pip levels required for each psychometric function varied from 5-14 (median=8) and depended on the width (in dB) of the psychometric function, determined by trial-and-error over the first 1 or 2 days' testing at each time gap. In order to maintain a good estimate of where p(c)=0.75 (the threshold) was found, a minimum of 5 different pip levels was employed each day. Experience showed that these levels need not be separated by less than 1/4 dB. In order to maintain a high density of points on each psychometric function, even when the function was quite broad, the separation between adjacent pip levels never exceeded 2 dB. Thus for a psychometric function spanning 10 dB, for example, there would be 5 pip levels each 2 dB apart, plus other levels in-between at the experimenter's discretion, which were never closer than 1/4 dB to any other level employed.

On successive days, psychometric functions were determined for a particular time gap until the subject's performance failed to show an improvement of more than a few percentage points relative to the previous day. Experience showed that a total of 3-5 experiment days were required, depending on time gap. The p(c) scores from the last 2 days at each time gap were averaged at the common pip levels to get the final psychometric function. In forward-masked threshold detection, p(c)=0.75 indicated the pip level giving the detection threshold (see below). In establishing intensity-increment thresholds, p(c)=0.75 indicated the pip level giving the intensity-increment threshold (see below). For nonmasked intensity-increment thresholds, the pip pedestal was present in one interval of the 2IFC task and the incremented pip appeared in the other. Nonmasked pip thresholds were determined at the end of the experiment; in the randomised 2IFC task, one of the intervals was silent, the other contained a pip located 390 msec into the interval.

Performance

The subjects had about 20 hours practise before data collection started, and p(c) values were accepted only when they stabilised, requiring 3-5 days work for each t. Due to the long hours required (2.5-8 per day; median=4), the test days were at the subjects' convenience, averaging 2-3 per week. Experience suggested that short-term learning started to reverse after 7 consecutive days absence, as evident by the larger
number of days required before their psychometric function stabilised. After finishing a particular time gap, subjects were sometimes absent for as much as several weeks. If so, when beginning at a new time gap, the subjects were allowed to repeat individual blocks, or even a whole day's listening, until they recovered their previous performance levels.

For forward-masked detection thresholds, the number of pip levels required for an intensity-increment threshold varied from 5-14 (median=8). The number of pip levels required for a forward-masked intensity-increment threshold varied from 7-10 and was typically 8 or 9. In both these experiments, the number of stimulus levels used depended on the width (in dB) of the psychometric function, as successive pip levels were never separated by less than 1/4 or more than 2 dB.

We are confident that the hours of daily listening did not affect the data, for the following reasons. As noted above, the subjects were carefully monitored. Fatigue was not the only reason that a subject might be terminated for the day. Verbal expressions of disgruntlement or boredom were taken seriously, and if necessary, the subject was terminated for the day regardless of their apparent progress. Furthermore, eight-hour days were infrequent, and continuous inspection of the data suggested no difference between performance on long days relative to performance on short ones.

To avoid any performance drop due to fatigue or attention loss (Zwislocki, Maire, Feldman, and Rubin, 1958), 10-20 min. breaks were offered between sets and no subject ran until ready. Subject performance was monitored closely; experience confirmed that discerning threshold fine structure with changes on the order of 10 dB or less requires intense concentration and long sessions, as reported elsewhere (Kemp, 1979). All subjects maintained high morale despite the aggregate of hours (535 for BAK, >1,000 for LN; ~300 per subject for both subjects in the intensity-increment-detection task). Subjects were aware of all p(c)'s and were encouraged to compete for highest p(c) (lowest thresholds). The subjects were free to offer self-evaluation. They soon became adept at self-assessment, recognising those episodes when performance was sub-standard due to fatigue or attention lapses, even before seeing their scores. In such cases the subjects themselves always insisted on being re-tested, and were always re-tested, with the same block, after a suitable break. This typically occurred once daily. Nonmasked-pip thresholds were determined last of all. For pip-detection thresholds, one of the intervals in the 2IFC task was silent, the other contained a pip located 390 msec into the interval. For intensity-increment thresholds, one of the intervals in the 2IFC task contained a pip located 390 msec into the interval, and the other interval contained a pip of higher level, also located 390 msec into the interval. Although any maskers were absent, the subjects were so well practised that they had no trouble differentiating the two 2IFC test intervals. (Nonetheless, a new listening criterion had to be established.) Throughout the threshold-detection task, no TTS (temporary threshold shift) or long-term threshold shifts were seen; such would have appeared as noticeably lower
scores, or been noted by self-report. TTS had been looked for but not found when using a comparable masker at a fixed \( t = 10 \) msec (100 msec, 2 kHz, 100 dB SL; Ehmer and Ehmer, 1969), and no TTS was mentioned in comparable studies (Weber and Moore, 1981; Zwicker, 1984; Scharf and Buus, 1986; Carlyon, 1988). The masker in the increment-detection task was never strong enough to conceivably cause TTS.

**DATA ANALYSIS**

One reason for using the very painstaking method of constant levels is that current studies using adaptive tracking of thresholds tend to express measurement variability, for each subject, as the standard error of repeated measurements of the observed quantity (here, the detection threshold). This gives large error bars. True estimates of detection thresholds are further avoided when ignoring the possibility of real differences in performance between individuals and instead averaging within-subject averages across subjects, at which point associated error is often not reported at all. The present method allows small error bars, and data was not initially averaged across subjects.

The psychometric threshold was estimated using probit analysis (McKee, Klein, and Teller, 1985) in which \( p(c) \) scores are fitted to the cumulative normal distribution,

\[
p(c | w) = \frac{1}{2} + \frac{1}{2} \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{w} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx
\]

\[
= \left[ \frac{1}{2} + \frac{1}{2} \left( \frac{1}{2} + \frac{1}{2} \text{erf} \left( \frac{w-\mu}{\sigma\sqrt{2}} \right) \right) \right]
\]

(26)

so that \( 0.5 \leq p(c | w) \leq 1 \); multiplying by 100 gives actual percentages. \( \text{Erf} \) refers to the error function found in published tables and software packages (e.g. GrafTool); care must be taken, because not all tables define \( \text{erf} \) in exactly the same way. Here \( x \) represents the stimulus intensity continuum. Within this continuum lies \( \mu \) in dB SPL that corresponds to \( p(c | w) = 0.75 \), so that \( \mu \) is the pip-detection threshold in forward-masked detection thresholds, and \( [\mu \text{ minus the level of the comparison pip}] \) is the intensity-increment threshold for forward-masked intensity-increment thresholds. The term \( \sigma \) varies as the width of the psychometric function and hence represents the inverse of the resolution of the detection threshold or the intensity-increment threshold, as the case may be.

In probit analysis, the values of \( \mu \) and \( \sigma \) are selected that provide the best fit of Equation 26. Confidence limits called "fiducial limits" can be placed on these parameters (see Finney, 1971), thus giving confidence limits on \( \mu \). Fiducial limits of 95%, 99%, and 99.9% were computed for detection thresholds \( \mu \),
and fiducial limits of 95% were computed for intensity-increment thresholds, \( \mu \) minus the level of the comparison pip. The error bars actually illustrated for each detection threshold or intensity-increment threshold are the 95% fiducial limits at \( p(c|w) = 0.75 \) (McKee, Klein, and Teller, 1985, Equation 4); these are asymmetrical with the lower fiducial limit extending further below the plotted threshold than the upper fiducial limit extends above it (ibid.)\(^7\). The present fiducial limits are well-behaved, as \( n \) (the number of trials per \( p(c) \)) \( \geq 100 \) at all pip levels. These bars shorten as the number of trials increases (McKee et al., 1985), and they lengthen when the number of trials at each level and the number of pip levels are both fixed, but the psychometric function becomes shallower. Thus as the psychometric function widens and its slope decreases, the number of different pip levels used must increase to compensate for this loss of precision, given the same number of forced-choice trials at each level. Although this increase is theoretically better when at higher relative pip levels, equally-spaced pips were used in the hope of attaining a fairer estimate of the \( p(c) \) curve's standard deviation, \( \sigma \). (We thank Prof. Doug Creelman for reminding us of probit analysis.)

When the 99 or 99.9% confidence limits for two different thresholds (or intensity-increment thresholds) do not overlap, we can be reasonably certain that the thresholds (or intensity-increment thresholds) are reliably different from one another. That is, changes in the plotted pattern of the thresholds (or intensity-increment thresholds) are not caused by their inherent variability.
25. Thanks to the use of a 16-bit computer register for storing voltage values, there are at most $2^{16}/2 = 32,768$ voltage steps in the $+$ or $-$ voltage values. At 60.5 dB down from maximum available SPL, corresponding to about 42 dB SPL, there are 30.93 steps above the voltage floor (voltage=0). This must be rounded to a whole-number of steps, which is thus 31, corresponding to a voltage to the headphones of 60.48 dB down. This represents an acceptable deviation from 60.5 dB down, since a difference of 0.02 dB is well below what any human can discriminate. Similarly, at 60 dB down (roughly 62 dB SPL), there are 32.77 steps, rounded off to 33, giving an actual SPL of 59.94 dB down. This is also acceptable, but does not allow differences of more than 0.5 dB to be used by the experimenter. For a difference of 1 dB, i.e. between 60 dB down and 61 dB down, there are only 3.5 actual voltage steps, compared to (for example) 9.6 steps between settings of 46.25 and 46.75 dB down (a difference of only 0.5 dB). Thus, below 49 dB SPL the actual sound levels expressed at the headphones, and the differences between nearby actual sound levels, start to deviate substantially from the small changes (0.5 dB or less) that are desired. Hence the use of the attenuator, which allows much higher voltage levels to be sent to the headphones, by way of the attenuator, where they are diminished (with an ignoraibly small associated voltage loss). The maximum voltage at the headphones is thus (very slightly less than) the maximum amplifier level of 4.879 V.

26. Learning effects were paramount despite differences in the subjects’ previous listening experiences (none [subject LN], some [BAK], or very extensive [JWC]). At each new time-gap $t$, improvement of as much as 10 dB occurred from first to second day’s trials, with successively lesser improvements over the following 1 or 2 days for a given $t$, as was found for detection of longer tones (Zwislocki, Maire, Feldman, and Rubin, 1958; Lukazewski and Elliot, 1962). There was a daily warm-up improvement in $p(c)$ as expected from long-tone thresholds (Loeb and Dickson, 1961); $p(c)$ for the first, and sometimes second, daily set of 100 could consistently be improved by retesting at day’s end, unless the subject reported fatigue. Occasional within-day retesting with multiple pip levels showed that daily detection performance asymptoted, but continued to improve (Zwislocki, Maire, Feldman, and Rubin, 1958; Loeb and Dickson, 1961; Lukazewski and Elliot, 1962) over successive days. As $t$ changed from week to week, there was an imperceptible decline in the range of improvement in $p(c)$ over successive days at a given $t$, an improvement seen elsewhere with another constant-level 2IFC task (Whitmore, Ermey, and Williams, 1968) involving comparable time (months) and practise (thousands of trials). Superimposed on this trend, learning had to re-occur at each new $t$, as found in experiments where a detected tonal frequency was changed (Zwislocki, Maire, Feldman, and Rubin, 1958). Thus the learning effects experienced here reflect all those noted over a variety of much earlier studies, despite the fact that not all those studies employed 2IFC.

27. $P(c)$’s fitted to a rescaled cumulative normal distribution extending from $p(c)=0.5$ to $p(c)=1.0$ such that $p(c)=0.75$ represents a Z-score (units of standard deviation $\sigma$) of 0. $P(c)$’s are weighted inversely by their intrinsic binomial variability, the standard error $(p(c)(1-p(c))/n)^{1/2}$, where $n$ is the number of trials per $p(c)$, of a proportion for random samples of size $n$ drawn from a binomial distribution. See McKee, Klein, and Teller (1985).
CHAPTER 5

EXPERIMENT 1: FORWARD-MASKED PIP-DETECTION THRESHOLDS
We thank Professors Fan-Gang Zeng and William Rhode for their comments on a manuscript that was derived from parts of this chapter. Early results were reported in Nizami and Schneider (1994a).

ADDITIONAL NOTE ON METHODS

Methods were as described in Chapter 4, with the exceptions noted below.

Data analysis

Differences in the thresholds inferred at different post-masker times may be significant if, for pairs of thresholds ½ msec apart (our best resolution), the later point of the pair shows a lower fiducial limit that reaches to a higher dB SPL than the upper fiducial limit of its earlier neighbour. There are 80 such possible differences for a 0.5 msec increment in \( t \) over a total duration of 40 msec. When using 99.9% or even 99% fiducial limits, it is quite unlikely that we would observe such reversals by chance alone.

RESULTS

Nonmasked pip-detection thresholds

The average of the absolute pip thresholds was 25.2 dB SPL, which reflects superior performance considering the average age of 33 for BAK (threshold=22.6 dB SPL), JWC (threshold=25.3 dB SPL) and LN (threshold=27.8 dB SPL). Because of the pip's bandwidth (Figure 36), \( \geq 3 \) critical bands at 10 dB down (Scharf and Buus, 1986, Table 14.4), it may be more appropriate to compare to thresholds for filtered clicks; subjectively, the present pip sounded like a bubbling or clapping partway between click and tone, as expected from its shortness (Burck, Kotowski, and Lichte, 1935; Bentzen, 1950; Miskolczy-Fodor, 1953). But there are problems with defining click intensity (Yost and Klein, 1979; Stapells, Picton, and Smith, 1982). Absolute thresholds for very brief Gaussian pips are not found in the literature. However a 2 kHz non-Gaussian pip with 0.6 msec ramps, 2 msec long between the half-maximum points, had a mean threshold of 19.6 dB "with respect to threshold at 1 sec duration" (Hempstock, Bryan, and Tempest, 1964) that was unstated. The latter is assumed to be the 1.0 dB SPL binaural free-field threshold plus an assumed 8 dB correction to monaural headphone listening (Scharf and Buus, 1986; Table 14.1) plus a 0.9 dB adjustment for TDH-49 headphones (Michael and Bienvenue, 1977), thus eventually giving a comparison threshold of 29.5 dB SPL for subjects aged 18-30.

Absolute thresholds are also known for a Hanning-windowed 2 kHz tone of 5 msec total length, for three subjects of unstated ages (Kohlrausch, Püschel, and Alphei, 1992). These thresholds lay between 17 and 22 dB SPL, as we might have expected from our own subjects if tested with these effectively longer (and therefore more energetic) tones. Thus, despite a lack of perfect comparisons, our subject's thresholds are quite good, and we can now approach the general results armed with some confidence in the subjects'
abilities.

**Forward-masked pip-detection thresholds**

We found a steep initial drop in pip-detection thresholds (Figures 39, 40, and 41), from about 75 dB SPL at \( t=0 \) msec to about 55 dB SPL at about \( t=6 \) msec. Periodically, thresholds increased. These reversals could occur over a \( \frac{1}{2} \) to 1 msec period such that previous studies, which sampled the recovery curve much less densely, would not have been able to reveal this pattern. The true threshold pattern may be even finer than shown; thresholds drop to a sideways dislocation at 3.5-4 msec (for LN, at 4.5-5 msec). This dislocation appears (see JWC) to be a compensation between a rapid drop followed by a rapid increase, a hidden spike-shaped rise implying undersampling of the phenomenon. Now in all but 2 of the 14 cases in which the 99.9% fiducial limits indicated a significant increase in threshold with a \( \frac{1}{2} \) msec increase in masker-pip gap, there was also an increase in the slope of the psychometric function. This increase (corresponding to a decrease in \( \sigma \), see below) was significant for \( n=14 \) at \( p<.0132 \) using a two-tail sign test (\( 2 \times p<.0066 \)) on the differences between pairs of \( \sigma \)'s (tables of cumulative binomial probability [e.g. Pagano, 1998, p.520]). These were the \( \sigma \)'s for the psychometric functions for each of the thresholds that dip and then rise \( \frac{1}{2} \) msec later.

The actual psychometric functions are shown in Figure 42 for the postmasker times showing significant threshold rises, as marked by the asterisks in Figures 39-41. The psychometric functions overlap little if at all. Of course, some overlap would be expected on the psychometric functions' infinitely long tails, which were omitted for clearer inspection. For subject LN, removing the visual clutter caused by the data points (Figure 42d) reveals separation between the psychometric functions even at 26 msec post-masker, far beyond the 99% and 99.9% significance limits indicated earlier by asterisks. Thus the threshold oscillation appears surprisingly long-lived. *(Note well: In the context of the present results, the words “oscillation” and “periodicity” refer only to nonmonotonicities, and are not meant to imply actual periodicity.)*

**The standard deviation \( \sigma \) of the psychometric functions**

Lest the duration of the threshold oscillation be assumed inflated, consider another property of the psychometric function, its slope. The actual psychometric function for each threshold point has a slope, the derivative of Equation 26, which is a Gaussian function of stimulus intensity, with standard deviation \( \sigma \). These \( \sigma \)'s clearly oscillate, for a considerable time post-masker (Figure 43). The significance of this oscillation will be dealt with shortly.
Figure 39. Subject LN. Forward-masked thresholds with 95% fiducial limits. Asterisks indicate t for thresholds significantly higher than those 0.5 msec earlier: ***, based on 99.9% fiducial limits; **, based on 99% fiducial limits. These significant increases occur despite the general downward trend of recovery.
Figure 40. Subject BAK. Forward-masked thresholds with 95% fiducial limits. Asterisks indicate thresholds significantly higher than those 0.5 msec earlier: ***, based on 99.9% fiducial limits; **, based on 99% fiducial limits. These significant increases occur despite the general downward trend of recovery.
Figure 41. Subject JWC. Forward-masked thresholds with 95% fiducial limits. Asterisks indicate $t$ for thresholds significantly higher than those 0.5 msec earlier: ***, based on 99.9% fiducial limits; **, based on 99% fiducial limits. These significant increases occur despite the general downward trend of recovery.
DISCUSSION

Regarding forward-masked thresholds

Averaging of the thresholds for the sake of comparison to the literature (Figure 44) removed much of our threshold fine-structure due to its lack of precise alignment across subjects, a lack implying that an equipment flaw was not the source of the pattern. The average trend of the subjects' pip-detection thresholds compares well with the literature; mean thresholds are lower than obtained elsewhere (Weber and Moore, 1981; Zwicker, 1984) for comparable stimuli. The pip sounded like a bubbling or clapping partway between click and tone, perhaps providing a "pitch cue" to differentiate it from the masker (see Terry and Moore, 1977), and may thus have been easier to detect than a slightly longer tone which would have greater tonality (Scharf and Buus, 1986). This may explain the subjects' superior detection performance.

There is only one published study that employed a 2 kHz probe over a fine enough time scale to provide enough detail to compare to our own results. Kohlrausch, Püschel, and Alphei (1992) employed an 80 dB SPL frozen noise with a rectangular envelope to forward-mask a 2 kHz tone shaped by a Hanning window (a cosinusoid; see Blackman and Tukey, 1958, p.14). The 5 msec total duration of the tone means a lower resolution than presently achieved. This lower resolution may explain the lack of apparent oscillation in their data. Also not apparent is the precision of their data, given that no error bars were presented. However, the general trend of their thresholds mirrors our own results, showing a very steep recovery curve over the first 5 msec post-masker, followed by a relatively sharp transition to a much shallower, linear descent over 5-20 msec post-masker. Also apparent are individual differences in their subjects' recovery curves, with one subject showing consistently lower thresholds beyond 5 msec post-masker, as evident here for BAK as compared to LN and JWC (Figure 44).

Kohlrausch et al. also modelled the decline in forward-masked thresholds. Below we will offer several models for our data, 2 mathematical, one conceptual. Neither of our mathematical models deal with the full range of the data from \( t=0 \) to \( t=40 \) msec, so it may be appropriate to mention Kohlrausch's approach, which covers that range. Briefly, the sound produces a response that is half-wave rectified and low-pass filtered at 1 kHz in a single frequency-tuned channel, mimicking the critical band and inner hair cell. This signal is the input to a model in which a sound range of 100 dB is mapped to a model output of 100 dB in such a way that a 1 dB change in signal level corresponds to a 1 dB change in model output. This difference, when it occurred, was taken as detection threshold. The first of these models was a logarithmic compressive transformation that could be followed by a low-pass filter with its own time constant, that is, an integrator. This model provided threshold plots containing a single sharp rise about 5 dB high and 3 msec wide superimposed on threshold-level curves that fell either too linearly or too quickly. As an alternative model, then, Kohlrausch et al. proposed processing loops, each having a different time constant, and each feeding-back on itself to provide a gain control. Five such circuits were arranged in series, each with a
Figure 42. (Next 5 pages) The subjects' actual performance, percentage correct $p(c)$ vs. the level of the Gaussian tone pip, for neighbourhoods on the recovery curve that show sudden increases in threshold. The psychometric functions were plotted using Grafool's built-in error function $\text{erf}(x)$ whose parameters $\{\mu, \sigma\}$ were taken from the cumulative Gaussian fitted by the probit analysis routine (Equation 26). The curve for $p(c)$ was then computed for level $x$ expressed in dB SPL, by noting that actual percentages $p(c)$ obey

$$p(c) = 100 \left[ \frac{1}{2} \left( 1 - \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\frac{x - \mu}{\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \right) \right], \quad 50 \leq p(c) \leq 100$$

(from Equation 26) and by letting

$$z = \frac{x - \mu}{\sigma}, \quad \text{so that} \quad dx = \frac{dz}{\sigma}$$

and so that

$$\frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(z)^2}{2}} \, dz = \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{x^2}{2}} e^{-\frac{z^2}{2}} \, dz + \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{x^2}{2}} e^{-\frac{z^2}{2}} \, dz + \frac{w - \mu}{\sigma} / \sqrt{2} \int_{0}^{\frac{x^2}{2}} e^{-\frac{z^2}{2}} \, dz.$$

The integral in the line above can be solved by letting

$$v = \frac{z}{\sqrt{2}}, \quad \text{so that} \quad dv = \frac{dz}{\sqrt{2}} \quad \text{hence} \quad \frac{w - \mu}{\sigma \sqrt{2}} \int_{0}^{\frac{x^2}{2}} e^{-\frac{z^2}{2}} \, dz = \frac{1}{2} \int_{0}^{\frac{x^2}{2}} e^{-\frac{z^2}{2}} \, dv = \frac{1}{2} \text{erf} \left( \frac{w - \mu}{\sigma \sqrt{2}} \right).$$

Thus we finally obtain

$$p(c) = \left[ \frac{1}{2} + \frac{1}{2} \left( \frac{1}{2} + \frac{1}{2} \text{erf} \left( \frac{w - \mu}{\sigma \sqrt{2}} \right) \right) \right].$$

(A cautionary note: the error function $\text{erf}$ may be defined in different ways in different software and handbooks.) The following 5 pages show the subjects' $p(c)$ scores as described by this equation. The first 2 pages show $p(c)$ scores for BAK and JWC; the following 3 pages show $p(c)$ scores for LN.
Figure 42a. Psychometric functions for BAK and JWC near t=6 msec and t=10 msec. See also Figure 42b.
Figure 42b. Psychometric functions for BAK and JWC near t=6 msec and t=10 msec. Same data as Figure 42a, but for the sake of clarity only 2 of the psychometric functions from Figure 42a are shown here.
Figure 42c. Psychometric functions for LN near t=6, 10, 14, 16, 18, 21, and 26 msec.
Figure 42d. Psychometric functions for LN near t=6, 10, 14, 16, 18, 21, and 26 msec. Same data as Figure 42c, but the data points have been removed for the sake of clarity.
Figure 42e. Psychometric functions for LN near t=6, 7.5, 10, and 14 msec. Same data as Figure 42c, but with the inclusion of the data for the rise in threshold at t=7-7.5 msec, which was not presented earlier.
different integration time, followed finally by a 200 msec integrator. This time the output more closely followed the pattern of the actual forward-masked thresholds, including the dip and subsequent rise just before the end of the masker, as seen in their data.

Thus there are potential alternatives to models such as ours. Although the latter model of Kohlrausch et al. mimics the behaviour of afferent fibers, the feedback loops beg for physiological correlates. No feedback is known that involves the periphery alone, or the afferent fibers alone; the situation is more complex (see for example Figure 32). Furthermore the model does not predict the oscillations found here, presumably because ringing was not accounted for explicitly. Our own model deals with the ringing of neural response without attempting to model the transformations performed by the inner hair cell, and produces an initial rise in threshold at the end of the masker, in the zone in which pip and masker do not fully overlap. Note that Kohlrausch et al. defined masker-pip time gap as from masker’s end to pip’s end, which was not the definition used for our own data, so that t as it is used in this chapter is 2.5 msec earlier than the time gaps referred to by Kohlrausch et al.. Respecting this difference in definition reveals an intriguing difference in the patterns of forward-masked thresholds that occurs in the zone of physical overlap of masker and pip. In our own data, the region of overlap (t=0-1.5 msec) produces a temporary rise in pip thresholds, whereas thresholds are already steadily declining during the last 5 msec of masker-pip overlap in Kohlrausch et al., whose thresholds instead rise and then fall over the last 10 msec in which the pip lies entirely within the masker. We speculate that this difference in behaviour must be related to the difference in maskers, in our case a pure tone, in the other case a frozen noise low-passed below 5 kHz. Because further comment would be pure speculation, we let the matter rest.

There is another study that may seem compatible to our own, due to the use of relatively brief tone probes (7 msec at 0.5, 1, and 4 kHz) and relatively intense forward maskers (white noise, 90 dB SPL), that of Elliott (1962). However, her data are not comparable to our own, for two reasons. First, her method differed significantly. The probe was always presented in the second interval of the two-interval task, and the subjects established threshold by adjusting the pip’s level until the difference in pitch between masker and masker-plus-tone was just discernible. Presumably the use of a noise masker created significant differences in pitch, making this method practical. However, this method differs greatly from the present one (for example, subjects did not make forced choices), and because Elliott’s subjects may be detecting a level difference only in a roundabout fashion, we hesitate to compare her data to our own. The second reason for ignoring her data is that the published graphs were so small that actual data points could not be discerned unambiguously.

**Regarding recovery times**

Recovery from an intense forward-masker must reflect both recovery from neural fatigue as well
Figure 43. The standard deviations $\sigma$ of the fitted psychometric function, the cumulative normal (Equation 26). The least-squares-fitted straight lines extrapolate the post-masker duration required for $\sigma$ to return to its value for the nonmasked pip. As $t$ decreases towards 1 msec and $\sigma$ increases, "threshold" becomes less well-defined. At 9 msec for BAK, for example, $\sigma=7.52$ dB corresponds to a psychometric-curve width (say, $\pm 2\sigma$) of 30.1 dB. (With a Gaussian envelope, amplitude is down to 14% by $\pm 2\sigma$ of the Gaussian envelope's peak and 1% by $\pm 3\sigma$) This $\pm 2\sigma$ width does not even fall below 10 dB until 14 msec for JWC, 17.5 msec for BAK and 22 msec for LN. For nonmasked pips the $\pm 2\sigma$ widths are 6.56, 7.32, and 9.56 dB respectively.
as recovery from neural ringing. These components are revealed in the differences in the times required for recovery from oscillations in \( \sigma \), and total recovery of behavioural sensitivity. We would expect oscillations in \( \sigma \) to cease once ringing was over, given that ringing is the presumed cause of the oscillation (as modelled below). The average recovery time of the standard deviation \( \sigma \) of the psychometric functions was 31 msec (Figure 43). Behavioural thresholds, however, should not fully recover until the neurons are fully refreshed from their sound exposure. The lowest realistic estimates of threshold recovery times were \( t = 69.2, 63.1 \), and 78.2 msec (LN, BAK, and JWC respectively), which were inferred from fitted straight lines (least squares, \( t \geq 13 \) msec). The slopes of these lines are recovery rates of -0.38, -0.39, and -0.36 dB/msec, numbers that are remarkably consistent (Figure 45). Although this remarkable consistency might not have been revealed without this curvefitting exercise, the inferred recovery times are undoubtedly underestimates.

Not shown on Figures 39-41 were detection thresholds at 70 msec, obtained for LN and BAK. These thresholds were 34.05 and 29.9 dB SPL, respectively, and appear in Figure 46. Noted on this figure are threshold-recovery times that were more realistic than those obtained from the straight-line extrapolations of Figure 45, but that were less precise due to extrapolation of threshold’s asymptotic decay. Recovery times resulted from fitting \( A \) and \( B \) in the function \( \text{Threshold (dB SPL)} = A/(B + \ln t) \) (see Xu, Payne, and Nelson, 1995), for \( t \geq 6 \) msec. Setting \( A/(B + \ln t) = \text{Threshold [nonmasked]} \) then gave \( t = 434, 475, \) and 1,206 msec (BAK, LN, JWC). The latter estimates are compatible with the literature for similar stimuli (\( t \geq 400 \) msec; Weber and Moore, 1981), times undoubtedly dependent on recovery of neural firing rates when tones are used as maskers and probes (Smith, 1977; Harris and Dallos, 1979). In between the recovery times estimated above are the minimum recovery times for a 5 msec Hanning-windowed 2 kHz pip following an 80 dB SPL frozen noise, which we can graphically estimate as 140 msec (Fig. 2 of Kohlrausch, Püschel, and Alphei, 1992).

Our probe tones are shorter than these others, so one objection is that our brief pip provokes only one or a few spikes in each neuron. This coordinated firing is indicated by the compound action potential (CAP) that is indeed provoked by Gaussian pips of comparable brevity to our own (Hoke, 1974). But the CAP’s usual reported amplitude, that of the \( N_1 \) component, shows forward-masking of brief tone bursts by tone maskers (cats, in Abbas and Gorga, 1981, and Dolan, Teas, and Walton, 1983; chinchillas, in Relkin, Doucet, and Sterns, 1995; Mongolian gerbils, in Cacace and Smith, 1986; both chinchillas and gerbils, in Relkin and Smith, 1991) with the same kind of general dependence on masker frequency and intensity, and on probe frequency, as seen for firing rates and psychophysical thresholds. (There is an exception: that CAP threshold continues to grow without saturation as masker level increases, whereas threshold for an observable change in single-nerve firing saturates [Relkin and Smith, 1991].) Thus, by analogy, pip-threshold recovery times are likely to result from recovery from neural fatigue\(^{11,12} \). (The term “neural fatigue” may refer, physiologically, to an inability to respond due to a long-term stimulation (seconds,
Figure 44. Comparison of the forward-masked thresholds to the literature. a. A scatterplot of all the subjects' thresholds. b. Forward-masked thresholds for 2 of the 3 subjects of Kohlrausch, Püschel, and Alphei (their Fig. 2 of 1992). Sound levels are with respect to 0 dB=97 dB SPL. The masker turns off at END=300. c. Averaged forward-masked thresholds vs. others obtained using long maskers and short tones, all stimuli at 2 kHz. All time gaps are from the end of the masker to the middle of the probe. Conditions of masker duration (msec), masker amplitude (dB SPL), and tone duration (msec) were: 400, 82, 5 (Weber and Moore, 1981); 100, 102, 5 (Ehmer and Ehmer, 1969). For the latter data, we assumed 0 dB SL=12 dB SPL and monaural listening.
minutes, or even hours), and can be related to temporal threshold shift psychophysically. This possible meaning of "neural fatigue" is not the one used throughout the present text, where "neural fatigue" is used to refer to very-short-term changes only.

The same conclusion can be reached by other means. Our own probe sounded somewhat click-like. If it is a neural response that is slow to recover from a click, and that is measured using a click probe, then a similar effect should appear even if basilar motion is believed irrelevant, as with cochlear implants, placed in individuals whose inner hair cells are believed to be non-functional. In implantees, non-overlapping neural responses can be assured by using, as either masker or probe, a pulse with a very brief square positive phase preceding a similar negative phase. When thresholds were then obtained by varying probe-pulse width (~100 µsec; Blamey and Dooley, 1993), a cochlear implantee showed a recovery congruent to the present data (Figure 47). Threshold recovery curves in cochlear implantees (Shannon, 1990) have shown that forward masking occurs even without involvement of the synapse between the inner hair cell (IHC) and its afferent, and with an asymptotic recovery comparable to longer times seen in normals (e.g. Champlin and Wright, 1993). Also notable was that our subjects (Figures 39-41) showed a square rise 1 msec wide, centered at 6.5-7 msec (undersampled in JWC). The cochlear patient showed a sudden peak of similar relative profoundness at t=6 msec, for trials involving either single or quadruple masking pulses, a threshold feature which apparently does not need operable hair cells. Thus the ringing is a property of the neurons themselves. The similarity of the subject curves and implantee curve in Figure 47 suggests that neural fatigue accounts for most of the general upward shift in masked threshold.

In Chapter 4 we raised the issue of what constituted an adequate inter-stimulus interval between the two intervals of the two-interval forced-choice task that the subject performs. The recovery times extrapolated above suggest that the 1.5 sec inter-stimulus interval was more than adequate for recovery of absolute pip threshold. It might be thought that such an after-the-fact assertion does not follow from the experiment. But an inadequate inter-stimulus interval would have reduced responsiveness to the tone, lowering \( p(c) \) scores so that detection-thresholds would have risen. Thus the entire recovery curve would have risen up relative to the proper one, so that extrapolated recovery times would have been higher than appropriate. It seems, then, that any reasonably-extrapolated recovery time that falls below the length of the inter-stimulus interval is, by itself, evidence of an adequate inter-stimulus interval. Of course, we still have to justify the inter-stimulus interval before the experiment; we cannot simply choose one arbitrarily.

We end this section by asking about possible age or experience effects. The thresholds of BAK, the youngest subject, were lower than, and recovered more quickly than, those of the other 2 subjects. Indeed the extrapolated recovery times follow the order of the subject's ages. Thus, an age effect is apparent, although it does not seem to affect the threshold oscillations. Indeed, the extrapolated recovery times from the oscillation (Figure 43) were in reverse order to the subject's ages, although the time differences are so
Figure 45. Least-squares fits of straight lines to subjective thresholds for $t \geq 13$ msec.
small, and the extrapolation method so crude, that there may be no statistical significance to this latter correlation. Also, it would be rash to speculate about age effects given the small sample size and relatively limited age range. As for the subject's prior listening experience, given the extensive training involved in this experiment, it is unlikely that previous experience would make a difference. And, indeed, no experience effects were noted.

A unique experimental problem at very early recovery times

The rise in thresholds over $t=0-2$ msec caused a unique experimental problem. At these levels, the masker amplitude "rolls off" so that the pip is superimposed on a declining masker level. Because the masker and probe are in phase (separated by 0.5 msec = 1 wavelength), the probe level can be increased until the level of masker-plus-probe equals the plateau level of the masker. The result is a masker just 1 or more wavelengths (0.5 msec) longer. For $t=0.5$ or $t=1$ or $t=1.5$, this sound cannot be distinguished from the masker alone, the comparison sound presented in the other interval of the 2AFC task. Hence performance should drop towards chance at such probe levels. This did happen. At higher probe levels, the probe once again becomes more and more distinguishable, as a blip of increasing loudness at the "end" of the masker. Conversely, as the level of masker-plus-probe drops (due to decreasing probe level), the pip again becomes increasingly distinguishable, then decreasingly distinguishable as probe level approaches masker level such that the target condition cannot be distinguished from the comparison condition (masker presented alone). It is the latter detection threshold that is reported over $0\leq t\leq 1.5$ in Figures 39-41. The experimentalist was careful to avoid testing the "upper" detection threshold - probably an intensity-increment threshold - and little data was recorded for this circumstance.

The phenomena discussed in this paragraph do not seem to represent the "confusion effect" reported by Neff (1986), which involves narrow-band noise as the forward-masker that precedes a probe tone.

Alternative explanations for the threshold oscillations

Regardless of the possible roles of physical sound overlap, or of sensory overlap caused by ringing, we can eliminate some alternative hypotheses for the threshold oscillations in Figures 39-41. Threshold oscillation might represent interference between the probe pip and evoked otoacoustic emissions, producing a lightly damped wave in the ear canal. But OAE amplitudes are rarely above 20 dB SPL (Kemp, 1978; Zurek, 1981); a typical evoked emission of about 10 dB SPL (Veuillet, Collef, and Duclaux, 1991), if added to the pip, would not allow even a 1 dB change in pip level until the latter dropped below 30 dB SPL. For a 20 dB SPL emission, the latter limit is 38 dB SPL, corresponding to $t\geq 20$ msec. Alternatively, oscillation in threshold and $\sigma$ could occur by interference of the tone pip with basilar vibration caused by long-ringing subharmonics (Dallos, 1973) generated from the forward masker. Generally, these undertones appear to be
Figure 46. The fit of the function $\text{Threshold (dB SPL)} = A/(B + \ln(t))$ to pip-detection thresholds for $2.5 \leq t \leq 40$ msec. Full perceptual recovery time is taken as the point where the fitted curve intersects the dashed horizontal line representing nonmasked pip threshold.
30-70 dB below the stimulating tone, whereas undertones strong enough to cause a notable change in threshold here would have to be ≤30 dB below the masker level. Furthermore, the masking pattern caused by the 2 kHz forward-masker itself would extend upwards in frequency (Scharf and Buus, 1986, Fig. 14.33), not overlapping with the undertones, which would thus have been hearable and distinct.

MODEL

Forward-masked thresholds as due to neural ringing

Loss of behavioural sensitivity to the probe after cessation of the forward masker may be due to two factors. First, a drop in firing rates may occur due to neural fatigue. Secondly, the well-known frequency specificity of auditory afferents (Kiang, Watanabe, Thomas, and Clark, 1965) makes each neuron a filter, and filtering should cause the sound-provoked neural response to persist for some time after the sound terminates ("neural ringing"). Of course, the current wisdom is that the neuron itself has no tuning property, tuning being determined by the basilar membrane. Regardless of the source of the tuning, it is generally true that the response of a filter at any time T is a function of the input at time T plus a weighted response to all previous input. Therefore, ringing is ubiquitous.

Now if the probe appears during a period of significant ringing, behavioural thresholds should rise, even in the absence of neural fatigue, because the probe-induced neural activity may be masked by neural ringing. Despite many publications, the contribution of neural ringing has not been accurately determined over its expected post-masker durations, because earlier work used probe stimuli that were too long, producing probe-detection thresholds that were too far apart in time.

There may be some concern about the term "neural ringing". In particular, it might be assumed that any oscillation in the response is certainly due to the residual ringing of the basilar membrane (BM). We employ the term "neural ringing" because of a paucity of evidence for unusually long ringing of the BM itself. Only one paper seems to show this ringing, by way of the cochlear microphonic (Echeverria and Robles, 1983). Given that the BM is only one component of the auditory filter, we thought it was conceivable that neurons might ring even after basilar membrane ringing has ended. Admittedly, this claim is contentious.

Nonetheless, the unexpected pattern of the forward-masked thresholds can be predicted by a neural model involving two different types of primary auditory afferents, those with a short ringing time, and those with an unexpectedly long ringing time. Neural ringing is evident in post-stimulus-time histograms (PSTHs) that plot the frequency with which neural spikes occur in contiguous fixed time periods following a sound. Because any area of the basilar membrane can function as a bandpass filter (Scharf and Buus, 1986), a neuron servicing this area will continue to fire after the sound terminates (ringing). PSTHs in the cat reveal two types of neural ringing provoked by a click (Pfeiffer and Kim, 1972). Type I neurons (93% of afferents)
Figure 47. (Upper) Averaged forward-masked thresholds (LN, BAK, JWC) vs. those of a cochlear implantee. The vertical scales are normalised since there is no known conversion from electrical to acoustical dB (Shannon, 1990). (Lower) The average threshold shifts caused by the masker (LN, BAK, JWC) vs. those of a brief sound pulse preceded by an identical pulse. The sound pulse was a damped wave roughly 3 msec long produced when an exponential electrical pulse was passed through a telephone (Chistovich and Ivanova, 1959).
and Type II neurons (7% of afferents) both show a damped decay in PSTH amplitude, but in Type I units, ringing times decrease from at least 12 msec for neurons whose peak sensitivity (characteristic frequency or CF) is 0.47 kHz, to about 4 msec for 2 kHz, to 1 msec for >10 kHz (Kiang, Watanabe, Thomas, and Clark, 1965)\textsuperscript{37}. Type II neurons, however, can ring for much longer times, for instance lasting 8 msec or more at 15 kHz (Kiang, Watanabe, Thomas, and Clark, p55). The PSTHs show a modulated envelope not seen spontaneously in quiet, whose duration and number of peaks both increased with click intensity over 55-95 dB SPL for CFs ≤ 2 kHz (Pfeiffer and Kim, 1972). The height of the histograms also increased, indicating greater probability that a spike would be fired within a given counting bin. For CF=2 kHz, ringing could last 15 msec after clicks at 95 dB SPL, and we may imagine longer ringing with greater energy at CF=2 kHz, i.e. a rapidly ramped intense tone such as ours. For instance, after 80 dB SPL pure tones that were 25 msec long, subsequent ringing can last as long as 20 msec at CF=1.9-2.5 kHz in afferents having no other unusual features (Rhode and Smith, 1985, cat; Rhode, personal communication). In such units ringing times even down to 20 dB SPL could be longer than expected, and even those fibers that were normal-ringers at low SPLs could ring for several msec more at high levels. Thus a 97 dB SPL tone might well produce ringing beyond 20 msec. We must be careful, however, because the apparent extension of ringing time with increase in masker level could be artefactual. At lower masker levels, the lower PSTH amplitudes may ring for just as long as for the higher amplitudes seen for more intense maskers, but the later portion of the PSTH may be lost in neural noise.

Note that in today's parlance, "Type I" and "Type II" neurons refer to two anatomical classes of neurons that innervate the inner and outer hair cells, respectively, and that no one has recorded from Type II cells (M. Chris Brown, personal communication). Here we do not intend that Type I cells refer to those innervating inner hair cells and Type II cells as innervating outer hair cells. The nomenclature "Type I" and "Type II" was that of Pfeiffer and Kim (1972) who, in everyone's hindsight, had presumably recorded from IHC afferents only.

Careful inspection of the literature shows that Type II behaviour is not restricted to the cat\textsuperscript{38}.

Long ringing also appears coincidental to the neural phase-locking seen at low CFs in the cat, in which spike firing synchronises with the sound-pressure waveform; unexpectedly long click-provoked ringing can last at least 8 msec even at 15 kHz (Kiang, Watanabe, Thomas, and Clark, 1965, p55), which is far beyond the CFs showing phase-locking.

Our forward-masked thresholds can be attributed to Pfeiffer's and Kim's Type I and II behaviour. The neuron's behaviour as a filter is commonly characterised by its response to a click. The PSTH generated by condensation clicks is generally shifted by ½ a period relative to that generated by rarefaction clicks, so that when spike counts generated by condensation and rarefaction clicks are displayed on the same plot (Pfeiffer and Kim, 1972), and if the plot is then normalised, the resulting PSTH looks like an amplitude-
modulated tone whose frequency is the neuron's CF. The modulation envelope approximates the impulse response, the response to an infinitely narrow square pulse. Figures 48a and 48b depict respectively the hypothetical impulse responses of Type I and Type II units tuned to 2 kHz. The impulse response of a pool of such units is the normalised compound PSTH of the sums of neural counts (Figure 48c) and therefore is the simple average of the impulse response functions of the pool's individual units.

We assumed that the output of this filter was first squared and then integrated over a period of the filter's CF (here, 2 kHz) to obtain the running average of the power passing through, computed every ½ msec. We then plotted a unitless decibel measure \( dB_R(t) \), the ratio of running average power to its maximum value:

\[
 dB_R(t) = 10 \cdot \log_{10} \frac{P_R(t)}{\max[P_R(t)]},
\]

where \( P_R(t) = \frac{1}{0.0005} \int_{t-0.0005}^{t} y_f^2(\tau) d\tau, \quad y_f(t) = \int_{0}^{\infty} y_e(\tau)[g(t-\tau) + s(t-\tau)] d\tau \)  \( (27) \)

where \( g(t) + s(t) \) is the filter's input for masker \( g(t) \), probe \( s(t) \).

During the masking tone, average power is constant, giving the term \( \max[P_R(t)] \). After termination, several bumps or near-plateaux appear (Figure 48d). Presenting a 87 dB SPL tone pip at \( t=0 \) produces a sudden power increase (Figure 48e) whereas at \( t=1.5 \) a response is not noticeable (Figure 48f) because the pip is added to the right edge of the first plateau. Thus for detection of a local intensity increment the pip would have to be more intense at 1.5 msec than at 0 msec. At \( t=2.5 \) msec (Figure 48g), a notable response appears again. A second minor plateau occurs between 5 and 7.5 msec with similar results for a pip at 68 dB SPL (Figures 48h,i,j). Generally, a greater pip level will be needed for threshold if the response to the pip occurs at the right edge of a plateau in the filter's power curve, so that multiple plateaus produce multiple nonmonotonicities in recovery.

We can predict post-masker thresholds by assuming that threshold is reached when the output of the filter, rather than remaining at a steady state or decreasing (see Figure 48d), suddenly increases by a fixed increment (Figures 48e or 48g). Now, the separation between the threshold curves shown for such increments of \( \frac{1}{2} \) dB and 1 dB (Figure 48k) is greater when the tone pip is presented in the middle of a plateau in the relative-power curve than when it occurs just to the right of the edge of the plateau. Thus, more of the energy passed by the neural filter would be needed to jump from the lower threshold plot to the upper threshold plot for a tone pip on the plateau of the relative-power curve. Assuming a monotonic correspondence between increments in tone-pip energy and the resulting increments in neural power, the
Figure 48. Model. Hypothetical PSTM at a characteristic frequency (CF) of 2 kHz (after Pfeiffer and Kim, 1972) for a) a Type I afferent (response $y_I$), b) a Type II afferent (response $y_{II}$), where

$$y_I(t) = [\cos(2\pi2000t)] \cdot 6.16397 \times 10^7 t^{1.75} e^{-2.500t}$$

$$y_{II}(t) = \cos(2\pi2000t) \left[ 6.16397 \times 10^7 t^{1.75} e^{-2.500t} + 1.90783 \times 10^{50} t^{19.5} e^{-2.500t} + 9.09724 \times 10^{43} t^{19.5} e^{-1.250t} + 3.71906 \times 10^{49} t^{49.5} e^{-1.875t} \right]$$

and where c) the hypothetical impulse response $y_C$ of a pooling of Type I and II responses is

$$y_C(t) = 0.93 y_I(t) + 0.07 y_{II}(t), \text{ for } t \text{ expressed as seconds.}$$

The equations were chosen for convenience only. The units for $y_I$ and $y_{II}$ are spike counts, whereas $y_C$ is a hypothetical impulse response function. The parameter values in $y_I$ and $y_{II}$ were somewhat arbitrary, due to insufficient numbers of neural recordings; many different combinations of values might provide the same impulse response functions. d) The relative power output of the filter specified by the above equations and Equation 27, near the end of the 97 dB SPL, 2 kHz masker. e) Power output of the same filter to the masker plus an 87 dB SPL tone pip presented at $t=0$, f) at $t=1.5$, g) at $t=2.5$ msec. The tone pip produces a negligible increase at $t=1.5$ msec because the response to it occurs when the response to the masker is just beginning its decline. h) The filter's response to the masker plus an 68-dB tone pip at $t=5.5$, i) at $t=8$, j) at $t=10$ msec. k) Predicted thresholds, assumed to occur when the output of the filter produced either a 0.5 dB increase in power, or a 1 dB increase in power.
Type I

Weighted impulse response

Time gap, t (msec)

Type II

Weighted impulse response
Relative power (dB)

$\text{pip} = 87 \text{ dB SPL}$

$\text{pip} = 68 \text{ dB SPL}$

$\text{t (msec)}$

$\text{no pip}$

$\text{pip} = 87 \text{ dB SPL}$

$\text{pip} = 68 \text{ dB SPL}$

$\text{pip} = 68 \text{ dB SPL}$

$\text{t (msec)}$

$\text{pip} = 68 \text{ dB SPL}$

$\text{Predicted threshold (dB SPL)}$

$1 \text{ dB criterion}$

$0.5 \text{ dB criterion}$

$\text{Time gap, t (msec)}$
result is a psychometric function with a shallower slope in which the threshold is at a local minimum on the threshold plot. Similarly, a steeper slope is predicted when the tone pip is positioned to the right of the edge of a plateau in the relative-power curve, i.e. in which pip threshold will be at a local maximum in Figure 48k. The slope of each threshold’s psychometric function is inversely related to the parameter $\sigma$, with $\sigma$ the standard deviation of the Gaussian function whose integral is fitted to the data points (Equation 26). A small $\sigma$ indicates a steep slope, and vice versa. Closer inspection of the data in the neighbourhoods of $t=1, 6$, and $10$ msec revealed that $\sigma$ drops as threshold rises, and starts to rise again as threshold starts to drop (Figure 49).

Figures 48d-j ignore neural fatigue, which should otherwise raise the thresholds in Figure 48k. Furthermore the peaks in the model pattern may not align with those of Figures 39-41 due to such things as species differences; for instance, the $2$ kHz place in the cat is not at the same relative position on the basilar membrane as in man (Greenwood, 1990) and so may display mechanical dissimilarities. But our results imply Type II fibers in man, and with ringing times similar to those in the cat. Certainly Type I fibers should exist in humans. In man the $4$ msec observed ringing time (Carlyon, 1988; Shailer and Moore, 1987) at $2$ kHz is expected from the width of the critical band, a filter at least as narrow as that of the cat (Fay, 1988), and the majority “Type I” ringing times in the cat agree with the inverse of the auditory filter width (Carlyon, 1988) in man (see Moore and Glasberg, 1983). Clearly, Type I ringing times in man resemble those of the cat. Such estimates were probably not sensitive to Type II ringing, perhaps because of its small contribution to the amplitude of pooled ringing beyond $t=4$ msec. The present data suggest that recovery from a forward masker is affected by neural ringing up until about $25-30$ msec post-masker, far beyond the limits previously assumed.

**Uses of the ringing model: recovery times for forward-masked clicks**

An interpretation of threshold elevation in terms of ringing should find especial application in any situation in which forward-masking might be caused by ringing alone. Whereas the present model cannot provide a complete account of the recovery function for long-duration forward-maskers, which induce neural fatigue, it should be able to account for the recovery from very short maskers that presumably produce little, if any, neural fatigue. Indeed, long, fatiguing tones like our $200$ msec masker are not essential for forward-masking; even clicks can reduce the neural response to similar clicks (rats; Moller, 1970) and relatively intense clicks can produce psychophysical threshold shifts of up to $55$ dB (Chistovich and Ivanova, 1959; Raab, 1961). Threshold shift is lesser and recovery time shorter when the intense masker click is reduced by $15$ dB (Raab, 1961), as expected from the apparent intensity-dependent increase in PSTH amplitudes and ringing times of Type II neurons (Pfeiffer and Kim, 1972; Rhode and Smith, 1985) and as thus predicted by a ringing model. Indeed, clicks spread their energy across a broad frequency band regardless of their
amplitude, so forward masking of clicks by clicks must have another cause besides neural fatigue. This cause becomes evident from the different orders of magnitude of post-masker durations required for recovery from oscillation and recovery from fatigue. Chistovich and Ivanova (1959) determined thresholds for a click-like sound forward masked by the same sound. Figure 46 plots their masked thresholds as a function of time along with the average of our three subjects. Note that nearly the same elevation in masked threshold is produced at short masker-probe gaps, but the recovery function when the masker is a click-like sound is much more rapid than when the masker is a 200 msec pure tone. The recovery of pulse threshold to its nonmasked value (Chistovich and Ivanova, 1959) took 28 msec, in general agreement with the duration of activity in long-ringing neurons and with the 31 msec average recovery time of the standard deviation $\sigma$ of the psychometric functions (Figure 43).

This agreement may be only serendipitous. Recovery of click threshold from forward-masking by a click shows a clear dependence on forward-masker level for the two such levels employed (Raab, 1961). For the more intense of the two masker levels, a forward-masker 85 dB above the click's absolute detection threshold, at least 100 msec appears to be required for recovery. Thus the argument in the above paragraph is misleading. The intensity-encoding response to the very brief pip might be not a firing rate, but the coordinated initial neural firing that produces the $N_1$ potential. If this burst from the afferent population is the stimulus-intensity cue, and if the forward-masker causes a discoordination of the neural firing contributing to this burst, as suggested by Relkin and Smith (1991), then the decline in $N_1$ amplitude will correspond to a rise in detection thresholds. The discoordination of firing, caused by forward-masking, should be most profound at the probe's CF, so that off-CF fibers will make a proportionally greater contribution to $N_1$ since they suffer less from this discoordinating effect (Relkin and Smith, 1991). Thus spread-of-excitation along the basilar membrane becomes important even for very brief stimuli. But spread-of-excitation is restricted by the use of a Gaussian pip having a narrow frequency spectrum (Figure 36), such that this effect would not be as important as for (say) forward-masking of a click by a tone.

Even a click might cause some discoordination in firing, such that a click could forward-mask a click. Because a click passed through a speaker or headphone is still relatively broad-band, spread-of-excitation will not help a click as well as a tone. Indeed, when forward-masker and probe energies are equated, a Gaussian pip following a tonal masker might conceivably produce the same recovery curve as a click following a click masker. Figure 46 suggests that this is not impossible.

Limitations of the model: the fit to the data

It might be objected that the nonmonotonicities in the model recovery do not at all resemble the nonmonotonicities in the data. This is a valid point that reveals a weakness of the model. Unfortunately both the neural and psychophysical data are not available in enough bulk or fine detail to allow fine tuning of the
model. Furthermore individual neurons differ idiosyncratically in the number of oscillations seen, and their separation in time (see Pfeiffer and Kim (1972) and Rhode and Smith (1985)). However, no alternative models presented themselves.

We also note certain objections to fitting the model to data produced using a pure tone. If neural ringing resulted directly from basilar ringing, such basilar ringing would cause each basilar location to ring at its own characteristic frequency (CF), which would scramble the phase responses of any perceptual event that was integrating activity over some basilar region. Although this appears true, the scrambled phase responses might average out in such a way as to preserve useful information at the CF of the applied tone. The objection also applies to the driven neural response that occurs while the tone is present. It is common knowledge that increasing the tone’s intensity broadens the tone-provoked pattern of physical excitation of the basilar membrane. So how is tonality preserved, especially for tones that are ramped in such a way as to produce multiple side-lobes in the relative energy spectrum (e.g. Figure 37)?

There is a third issue regarding how the data might result from the firing of an afferent population. Detection thresholds are usually assumed to mirror the most sensitive of the afferents in the assumed population. Thus the normal-ringing neurons, which appear to be so relatively populous, might determine detection thresholds unless the long-ringing neurons were the most sensitive. Within our model this would be true for the first few milliseconds, beyond which normal ringing ceases so that the activity provoked by the long-ringing neurons can still interfere with detection. The degree of this interference cannot be estimated given that the relevant data on neural thresholds is lacking. However, as noted elsewhere, long-ringing neurons do not appear to be restricted to one spontaneous-rate group (i.e. to one distribution of thresholds).

Limitations of the model: the threshold shift in afferents vs. the threshold shift in psychophysics

The findings of Relkin and Turner (1988) appear to counteract our model. The actual degree of threshold shift found in behavioural work such as our own is not replicated in primary afferents, so that much of forward masking would seem to be central rather than peripheral. Relkin and Turner (1988) examined the number of spikes produced by a neuron stimulated at its CF by adjusting tone intensity up or down in response to neural counts, eventually converging on the probe level that in theory would give \( p(c)=0.66 \). This procedure allowed individual neurons to provide growth-of-masking functions in which postmasker threshold was plotted vs. masker level, under fixed timing conditions. Postmasker thresholds tended to rise as masker level rose from 0 to 60 dB SPL, but appeared to level off between 60 dB SPL and the highest level employed, 80 dB SPL. The upward shift in probe-detection threshold caused by the masker was typically less than 20 dB, much less than that seen in psychophysical studies. Thus, the response properties of individual neurons cannot explain the full degree of forward-masking, suggesting a substantial central
component. However, we have no idea how much pooling of neural responses might occur, nor do we know whether the total transformation between periphery and perception is linear or otherwise. There is little doubt, however, that peripheral forward-masking occurs. This masking may be indicative of a serial chain of fatigue events, one for each succeeding synapse. Probe duration and rise-time also affect the degree of forward-masking (Turner, Relkin, and Doucet, 1994), so that rise-time in particular may play an important role in correlating forward-masking data between physiology and psychophysics.

For very brief stimuli, study of the compound action potential may point the way towards an explanation for the observed differences between psychophysical and neural forward-masking. Relkin and Turner (1988) had dealt with rate responses to 25 msec tones. Utilising the same stimulus conditions, and the same species (chinchilla), Relkin and Smith (1991) performed a similar study with the exception that the tone and masker were at 5 kHz. This time it was the compound action potential that was subject to the 2IFC procedure described above, giving growth-of-masking functions for the N1 potentials, which were recorded during the first 12.5 msec of the probe tone. The functions were straight lines when masked threshold in dB SPL was plotted vs. masker level in dB SPL, showing no inclination towards saturation even at a masker level of 90 dB SPL. The growth-of-masking function was seen to parallel those known for human observers for a nearby frequency and a probe of similar duration. Thus psychophysical forward-masking was reflected in a population response, the N1, a finding that was difficult to explain “since the CAP is not a variable which is directly transmitted to the central nervous system” (Relkin and Smith, p.137). But it is the coordinated neural response reflected by the CAP that may be transmitted to the central nervous system19.

We have shown that a minority of nerve fibers can produce peaks in the post-masker threshold curve regardless of the absolute population size, by assuming (caption to Figure 48) that all of the Type I and II fibers respectively produce stereotyped responses, that is, there is zero variance across-fibers within each group. Of course, some variance must actually occur.

The literature suggests that adaptation of neural firing must be ubiquitous to synaptic transmission, implying that models focusing on properties of 8th-nerve afferents may be simplistic. However, we were fascinated by the unusual long-ringing neurons noted by Pfeiffer and Kim, and others, and perceived a putative effect of their activity. Such activity might very well occur at higher points in the chain of transmission.

**Limitations of the model: regarding the rise in threshold over t=0-2 msec**

The earliest post-masker nonmonotonicity is the profound rise in forward-masked threshold over 0-2 msec. Now, primary auditory afferents act as bandpass filters and continue to respond (ring) even after the stimulating sound has been discontinued, so that threshold rise within a few milliseconds of the cessation of a masker could reflect difficulty in distinguishing probe-induced activity from masker-induced ringing.
This ringing, as a property of the filter, should be independent of the masker’s length or intensity, and hence not dependent on the degree of neural fatigue as usually seen (e.g. Smith, 1977, and Harris and Dallos, 1979). For example, when a 125 msec bandpass noise preceded (but never physically overlapped with) a click probe, threshold still rose for \(0.1 \leq t \leq 1\) msec even for two different noise levels (32 or 12 dB SPL per cycle; Penner, 1974). Similarly, noise of equal levels but very different durations (from 25 to 625 msec) can produce almost the same amount of forward-masking for clicks over \(0.1 \leq t \leq 1\) msec (Penner, 1974).

Carlyon (1988) appears to have paraphrased this point. He proposed that longer maskers should not produce higher thresholds than shorter maskers, for the same masker level, when masker ringing overlaps with the probe; such masking is “simultaneous” rather than “forward”. Carlyon went even further, assuming that probe thresholds with longer maskers should be lower (Carlyon, 1988, p.70) based on the so-called “overshoot effect” established by others. This was what Carlyon found over at least \(t=10\) msec or so for 250 Hz probes (broadband masker) but only over 1 msec at best for 2 kHz probes. For this effect, only the region of profound ringing amplitude seems to matter, considering the estimated 19 msec of ringing at 250 Hz (Carlyon, 1988). This extends the results of Duifhuis (1973) who used a shaped tone-burst of only 10 cycles, much shorter than the present probe, forward-masked by an identical burst of 40 dB SL, much weaker than the present masker. At 1, 2, and 4 kHz the probe-detection thresholds dip at \(t=0\), rise to about \(t=2.5\) msec, then drop. Duifhuis appears to define time-gap as peak-to-peak so that physical overlap of masker and probe occurred for 5 msec at 2 kHz, giving a threshold rise only for regions of heavy masker-probe overlap and not the complete 2 kHz ringing time of about 4 msec (Shailer and Moore, 1987; Carlyon, 1988).

Duifhuis (1973) proposed that the dip in threshold at \(t=0\) represents detection of a just-noticeable-difference (jnd), given that masker and probe can physically overlap. We computed the jnds over \(t=0-2\) msec (Table 7) after necessarily making assumptions about the integration time involved (see Appendix E). The present jnds at 0 msec vary from about 0.5 to 0.7 dB for an integration time of \(\pm 0.5\) msec (= one standard deviation of the Gaussian envelope = one period of the 2 kHz tone) to 0.3-0.4 dB for an integration time of \(\pm 1.5\) msec. These jnds are better than those of about 0.66 dB and 1.3 dB found for a much weaker 5 msec masker and 5 msec pip (Duifhuis, 1973), suggesting that an overshoot effect was present in our experiment but not in that of Duifhuis because his forward-masker was so short. In Chapter 5 we establish nonmasked intensity-increment thresholds for the Gaussian probe pips used here, for pip levels up to 90 dB SPL. For these nonmasked pips there is no need to specify an integration time. To briefly jump ahead: the discriminability limens averaged about 1.25 dB at 90 dB SPL for two subjects. Now for comparison to the present jnds, let us assume a moderate integration time of \(\pm 1\) msec, which includes almost all of the pip’s energy. Then only for \(0 \leq t \leq 0.5\) msec, a region of heavy masker-probe overlap, do the pip thresholds represent jnds that fall below the jnds for nonmasked pips. Beyond this region, then, we may not be detecting a jnd.
Table 7. Pip-detection thresholds over $0 \leq t \leq 2$ msec as just-noticeable-differences in level, computed according to the method in Appendix E. The term $n$ refers to the number of standard deviations of the pip ($\sigma=0.5$ msec) that were used as the limits of integration of pip energy. Although three decimal places are reported, the actual accuracy is probably much less, being dependent on the accuracy with which the masker and pip levels are reported.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Time gap, $t$ (msec)</th>
<th>$n=1$</th>
<th>$n=2$</th>
<th>$n=3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LN</td>
<td>0</td>
<td>0.716</td>
<td>0.558</td>
<td>0.427</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>2.377</td>
<td>1.50</td>
<td>1.033</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>2.848</td>
<td>0.914</td>
<td>0.438</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>16.818</td>
<td>5.01</td>
<td>1.319</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>34.136</td>
<td>12.50</td>
<td>1.950</td>
</tr>
<tr>
<td>BAK</td>
<td>0</td>
<td>0.588</td>
<td>0.457</td>
<td>0.349</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.736</td>
<td>1.07</td>
<td>0.728</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>5.251</td>
<td>1.924</td>
<td>0.968</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>15.532</td>
<td>4.227</td>
<td>1.046</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>33.289</td>
<td>11.76</td>
<td>1.680</td>
</tr>
<tr>
<td>JWC</td>
<td>0</td>
<td>0.501</td>
<td>0.389</td>
<td>0.296</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.268</td>
<td>0.771</td>
<td>0.518</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>5.239</td>
<td>1.92</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>13.190</td>
<td>3.01</td>
<td>0.676</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>33.624</td>
<td>12.06</td>
<td>1.783</td>
</tr>
</tbody>
</table>
was not sufficient data for BAK and JWC around f = 1.4 kHz.

Figures 4.9. Close-ups showing forward-masking thresholds (from Figures 3.9:}

\[ \text{Time gap (msec)} \]

\[ \text{dB (arbitrary scale)} \]

\[ \text{Threshold} \]

\[ \text{Threshold} \]

\[ \text{Threshold} \]

\[ \text{Threshold} \]

\[ \text{Threshold} \]

\[ \text{Threshold} \]

\[ \text{Threshold} \]

\[ \text{Threshold} \]
Beyond a few milliseconds post-masker, the jnd's become enormous; but the auditory system is still suffering from post-masker fatigue. Thus some other listening cue becomes appropriate in the 2IFC task in order to just discriminate the interval containing the pip from that containing the masker alone. Subjective impressions bear this out; Penner's subjects found that when the masker was close to the pip, it could be detected as a change in masker quality, whereas with greater separation, the pip appeared discretely. BAK, JWC, and LN noted a similar and strong listening dichotomy that reversed near \( t=1-1.5 \) msec.

It is also interesting to note that almost the same amount of either forward- or backward-masking was seen for clicks over \( 0.1 \leq t \leq 1 \) msec (Penner, 1974), which also agrees with Duifhuis' (1973) proposal that profound backward masking, and some part of the early component of forward masking, have the same cause: overlap of pip-provoked ringing with the masker, and vice versa. But this does not explain the rise over \( 0 < t \leq 2 \) msec that is not found in all backward masking (e.g. studies cited by Duifhuis; Dolan and Small, 1984) and hence is not expected in all forward-masking either.

It appears that processes happening below \( t=2 \) msec differ from those that occur later. We had modelled ringing as the cause of the early threshold rise, which agrees with the data of Penner (1976). But the contribution of ringing is hard to judge; when there is physical overlap of masker roll-off with the probe, only the regime of overlap seems to matter. Because this region includes the few msec over which ringing is profound, and bearing in mind that Penner's ringing effect lasted only 1 msec, it may be that only the profound initial ringing amplitude may have made any difference in these experiments. Thus we might assume that further effects of ringing do not make an appreciable contribution to the average degree of masker-induced fatigue suffered by the auditory system.

**SUMMARY OF RESULTS**

Subjective detection thresholds for a short probe tone in quiet can be assumed to follow the lowest neural detection thresholds, those of neurons of high-spontaneous-rate. This capability should be temporarily disabled by an intense forward-masker, so that neurons of lower spontaneous rate (and higher threshold) mediate some short period of post-masker detection. The transition between these neural channels should be marked by a cusp in the plot of probe-detection threshold vs. post-masker duration. But none appeared for a 2 kHz Gaussian-shaped probe (standard deviation=0.5 msec), centered at intervals of as little as half a millisecond over 0-30 msec following a 200 msec, 97 dB SPL, 2 kHz tone. Surprisingly, there were small, sudden rises and falls superimposed on each subject's generally smooth recovery. Even more obvious were nonmonotonicities in the standard deviation of the cumulative normal fitted to each threshold's psychometric function.

The increase at \( t=1 \) msec, which has been seen elsewhere, might be viewed as a just-noticeable-
difference. The subsequent observed threshold oscillation is predicted if the observer's decision about the presence or absence of the pip depends on pooling of the spikes from two types of primary afferent neurons, a majority of neurons whose ringing times depend on the width of the auditory filter, and a hitherto mysterious minority which ring for much longer. Pooled ringing also explains the counterintuitive discovery that as thresholds rise, there is a drop in the decibel range over which pip identification varies from random to perfect at a given post-masker time. Thus sensitivity to changes in the pip's intensity are greater at recovery times showing reduced sensitivity to the pip's presence.

Instead of imagining two neural channels (populations) with different recovery rates, we might have instead considered a small group of primary afferents, each recovering at a different rate and each starting its recovery at a slightly different post-masker delay (with a rate distribution whose variance would perhaps be on the order of a few milliseconds). These different recovery processes could conceivably produce nonmonotonic recovery functions. However, such a model would have all the same problems of accounting for the actual time course of the data as the present model. It is a problem to specify just how many neurons would be required with what spectrum of different recovery times, and how that particular population size could be justified in terms of the critical band, and of known properties of neural convergence.

For those who would have preferred more data, the results should perhaps be viewed in terms of costs and benefits. This experiment had great costs, in particular the time and commensurate dedication demanded of the subjects. The benefits are the fine detail that cannot be achieved by any other means, and the first psychometric functions for forward-masked tones. The discovered thresholds were statistically meaningful for each of several subjects, but due to the costs, it was not possible to repeat the experiment for any one subject. Thus, although it would be fascinating to examine the stability of the threshold curves, the repeatability of the results within a subject must be left to later investigation. The issue of test-retest reliability seems to be answered by the consistent smallness of the error bars seen across each subject's own detection thresholds. For example the 99.9% error bars indicate that 99.9% of the time the true threshold value will be found within this interval. When two temporally adjacent sets of 99.9% intervals do not overlap, we can be quite confident that the thresholds differ. Graphical comparison of the psychometric functions when there is no overlap indicates quite clearly that there is no overlap. Although we did not actually retest at a particular level we are confident that were we to do so, we would obtain the same results.
NOTES

29. Although probit analysis allows estimation of fiducial limits, the Weibull function (Relkin and Pelli, 1987) may be preferable as a psychometric function in 2IFC (McKee, personal communication). The fitting of subject’s \( p(c) \) scores to a psychometric function was repeated using the cumulative Weibull probability, the probability function corrected for a minimum value of 0.5,

\[
p(c) = 1 - (1 - \gamma) e^{-\left(\frac{c}{a}\right)^p}
\]

where \( \gamma \) is the false-alarm rate \( p(0) \) (see Pelli, 1985), set to 0.5 for 2AFC experiments. Threshold was assumed as \( p(c) = 0.75 \). The width of the fitted curve (in dB) was estimated as the difference between stimulus levels computed to produce \( p(c) = 0.55 \) and \( p(c) = 0.95 \). These plots respectively showed the same patterns as the masked thresholds and standard deviations \( \sigma \) from the probit analysis, the only exception being that the Weibull thresholds were consistently about 1/4-1/2 dB higher. This process was repeated for the psychophysical functions for intensity-increment thresholds (Chapter 5), with similar results.

30. The thresholds quoted by others may vary from each other and from Figure 44 due to differences in masker length and frequency, masker and probe ramping, probe length and frequency, and definition of masker-pip time gap. The masker’s spectrum also matters; narrow-band noise and sinusoids give different degrees of masking for the brief 2 kHz probe, producing recovery curves that can cross over each other (Weber and Moore, 1981). Hence comparison of masked threshold is tricky, especially since different ramping envelopes cause differing degrees of frequency splatter into nearby critical bands, which may increase loudness.

31. Comparing masking using masker and probe in one ear versus masker in one and probe in the other lead Luscher and Zwischlocki (1949) to conclude that forward masking is a monaural and hence peripheral process. Luscher and Zwischlocki also deduced that recovery from a forward masker was the reversal of a peripheral adaptation process and Carlyon (1988) concluded that recovery is at least partially the reversal of adaptation. Smiarowski and Carhart (1975) state that their forward masking trials also support this interpretation. Forward-masked detection thresholds increase monotonically with rise in masker level (Zwicker, 1984; Fastl, 1979; Carlyon, 1988) implying dependence on degree of neural fatigue. This intensity-dependence can be seen clearly in forward-masked peripheral afferents (Smith, 1977; Harris and Dallos, 1979), further supporting the view of recovery as reversal of peripheral adaptation.

32. We can be sure that a notable \( N_1 \) potential happens for a 2 kHz pip, for the following reasons. In awake Guinea pigs given tone pips at 2 kHz or its harmonics (3 cycles rise/fall time), there was a significant elevation in the detection thresholds for CAP as frequencies dropped (Cazals, Aran, and Erre, 1983). Presumably the lower frequency splatter associated with less abrupt ramping lowers the CAP amplitude, by reducing the population of units responding. CAP amplitude was not affected by extending the intensity plateau of the pips from 1 to 30 msec, suggesting that the transient period of the stimulus is all that matters. In man, awake (but sometimes sedated) subjects also produce a CAP, in response to clicks (e.g. Yoshie, 1968). Brief none-click sounds also evoke an \( N_1 \) in man; these were sinusoids (500 Hz or harmonics of it) damped to give a half-sinusoidal pulse with a damped tail (Elberling and Salomon, 1971). Finally, for a stimulus similar to the present probe, human subjects produce a curve of \( N_1 \) vs. level which is similar in shape to that of other mammals (Gaussian pips of 3 msec total duration at 2,140 Hz; Hoke, 1974). Generally the CAP is produced, at least at low stimulus levels, by nerve fibers local to the CF of the stimulating tone (as noted by Ozdamar and Dallos, 1976, and Abbas and Gorga, 1981). Thus the fiber firing produced by longer tones and the CAP for short pips should originate from the same fibers.
33. Very brief square current pulses can be sent to a single electrode of a cochlear implant. A pulse with a positive phase preceding a negative phase can give a neural firing histogram narrower than 1 msec (Javel, Tong, Shepherd, and Clark, 1987), presumably because the evoked basilar motion from the compression pulse opposes that provoked by the rarefaction pulse (Moller, 1970) when the two pulses are correctly separated in time (see Pfeiffer and Kim, 1972). This leaves only the initial basilar movement (no ringing) and its response, so that when biphasic pulses are used as both masker and probe only msec apart, the resulting neural responses will not overlap in time.

34. The use of patients as subjects restricts practice time (Shannon, 1990) so that threshold fine structure might be lost by avoiding potential improvements across-days, that could reduce data jitter. To further confuse things, error bars are typically unstated, and the maskers tended to be only moderately loud (Shannon, 1990). It is thus not clear whether actual threshold oscillation occurred for Blamey and Dooley’s patient.

35. Is middle-ear activity responsible for the masked-threshold microstructure? Stimuli other than sound can evoke middle-ear reflexes (Djupesland, 1965) and anticipation of the loud forward-masker, or the pressure of the headphones on the jaw, might activate the stapedius reflex quasi-continuously. The masker appears to be energetic enough to activate the stapedius reflex (see Gelfand, Silman, and Silverman, 1981) and lengthening the masker beyond 200 msec would not likely reduce its reflex threshold (Djupesland and Zwischlocki, 1971) suggesting a reflex latency <200 msec and perhaps much lower (Perlman and Case, 1939). Hence oscillatory features should not represent jiggle due to reflex turn-on. The test pip thresholds themselves are likely too weak to evoke the reflex (Gelfand, Silman, and Silverman, 1981). The overall threshold pattern seen in the text occurs with much lower-intensity maskers (Duijshuis, 1973) that should not be strong enough to provoke middle-ear reflexes, suggesting that such reflexes are not critical to the existence and pattern of auditory recovery. If we assume a reflex turn-off latency comparable to turn-on time, then the middle-ear reflex is again unlikely to be involved.

36. Oscillations are hard to find in psychophysical data. Undersampled oscillations may manifest as sudden blips or changes of slope in the recovery curve. Recovery from 80 dB noise maskers shows a notable slope change at about 7.5 msec for the 2 kHz, 5 msec test tone (Zwicker, 1984), and a slope change is barely discernable for the 8 kHz probe at 80 and 60 dB masker levels for the 200 msec masker. These small changes are hard to evaluate due to lack of stated error bars. We also suspect that individual data may have shown more profound oscillations, which would be lost through the custom of across-subject averaging, due to the subtle differences in time structure across subjects (compare Figures 39-41 with Figure 44c).

37. The damped sinusoidal decay in PSTH amplitude, which illustrates ringing, can be seen even for much higher frequencies, when paired clicks are used to produce interference patterns in basilar motion (Moller, 1970; rats). Normal ringing (rats) varies from 1 msec at 5.4 kHz to 0.3 msec at 30.5 kHz, becoming progressively shorter with rise in CF (as expected, since the auditory filter tends to widen as CF increases).

38. Clicks of 128 dB SPL can “frequently” provoke ringing of at least 15 msec (CF=1 kHz) in Caiman crocodilus, a reptile, with a PSTH which appears to be modulated (Klinke and Pause, 1980, Fig. 6), compared to 4 msec of presumably “normal” ringing at 0.8 kHz. Ringing can persist for at least 8 msec at only 1.0 kHz in the chinchilla (Salvi, Henderson, and Hamernik, 1979, Fig. 6) and at 1.10 kHz, phase-locked click-provoked ringing persists for about 10 msec in the Guinea pig (Versnel, Prijs, Schoonhoven, 1997; high-SR unit, Fig. 2c). Such ringing lasted about 4 msec for a 6.84 kHz high-

r unit (ibid.), longer than expected if Guinea pig critical bands are comparable to those of the cat. Lingering response also shows in the cochlear microphonic (CM); times for the damped decay of the CM to baseline (Guinea pig) were about 8 msec at 2 kHz to 15 msec at 0.8 kHz and damped oscillations of 18 msec were seen (Echeverria and Robles, 1983, Fig. 2).
39. Relkin and Smith (1991) note that the probe-provoked CAP is a special cue since it represents a coordinated neural firing that defines the masker-probe time gap. If this onset response is reduced, due to a discoordination of firing caused by forward-masking at the probe's CF, then off-CF fibers make a proportionally greater contribution to $N_i$ since they suffer less from this discoordinating effect (ibid.). Thus spread-of-excitation becomes important even for very brief stimuli that provoke little more than an initial burst of neural firing. But spread-of-excitation is restricted by the use of a Gaussian pip having a narrow frequency spectrum, as done here. Thus we might expect the amplitude of $N_i$, and thus the growth-of-masking function, to differ for Gaussian pips. Further examination is required to confirm or deny this speculation.
CHAPTER 6

EXPERIMENT 2: FORWARD-MASKED JUST-DETECTABLE INTENSITY-INCREMENTS
Preliminary results of this experiment appeared as an abstract (Nizami and Schneider, 1994b). Since that time a variety of evidence has appeared that casts doubt on Zeng et al.'s interpretation of the “mid-level hump” in terms of differential recovery rates. We note this material in this chapter’s Discussion and propose an alternative model for the mid-level hump in light of present results. We also thank Dr. Fan-Gang Zeng for comments on an earlier write-up of some of the material in this chapter, and for the data shown in Figure 59.

RECAP OF METHODS

The methods were as described in Chapter 4, and are briefly recapitulated below.

Subjects

Two male subjects were run (ages 22 and 23 at start of the study), one a technician in another laboratory, one an undergraduate.

Stimuli and procedure

Subjects performed two-interval forced choices (2IFCs). Each interval lasted 500 msec, with the masker starting 100 msec into the interval and the probe following the masker by up to 80 msec, leaving almost 120 msec of quiet at the end of each interval. A 1.5 sec delay separated the two intervals. In both of the two test intervals on each forced-choice, a 2 kHz, 200 msec (excluding ramps) forward-masker was followed by a 2 kHz tone pip. One interval was a comparison condition in which the masker preceded a comparison pip of constant level (the pedestal level), always set 20 dB higher than the masker. The other interval contained the same masker and time gap, but the pip was incremented so as to be slightly more intense than the pedestal level. The order of these 2 intervals was randomised on each forced-choice, and the subject’s task was to identify the interval containing the pip of higher intensity, with their choice being indicated as correct or incorrect by a visual signal following each discrimination. Over 100 such self-paced choices the subject produced a percentage-correct score recorded as 0.5≤p(c)≤1.0 (method of constant levels), with p(c)=0.75 indicating the pip level giving the intensity-increment threshold. Pip intensity-increment thresholds were also measured when the masker was absent.

The order in which the masker levels were chosen was irregular. The subjects provided complete psychometric functions at both 3 msec and 80 msec delays before moving on to the next masker level, in which the process was repeated. Obtaining nonmasked intensity-increment thresholds required some retraining, to familiarise with the different sound; in order not to break the subject’s training, these were all obtained last. The number of pip levels required for an intensity-increment threshold varied from 7-10 and
was typically 8 or 9, depending on the width of the psychometric function, as successive pip levels were never separated by less than 1/4 or more than 2 dB. To avoid any performance drop due to fatigue or attention loss (Zwislocki et al., 1958), 10-20 min. breaks were offered between sets and no subject ran until ready. 

RESULTS AND DISCUSSION

Nonmasked pip-detection thresholds

The absolute pip thresholds were 28.50 dB SPL (JW) and 26.62 dB SPL (ZS), which compare well to the average of 25.2 dB SPL found for 3 other university subjects in Experiment 1 (Chapter 5). Thus the lowest pip pedestal, 30 dB SPL, was just several dB above threshold, as anticipated (Chapter 1).

The detected intensity-increments

Figures 50 (subject JW) and 51 (subject ZS) show the detected intensity-increments for the 3 msec masker-pip time gap, for the 80 msec time gap, and for the nonmasked pip.

For both subjects, a general trend becomes immediately apparent for all three time gaps. The size of the intensity increment rises up as masker level approaches 50 dB SPL, corresponding to a pip pedestal level of 70 dB SPL. It then falls quickly to its lowest values, for the 70 dB SPL forward masker (90 dB SPL pip pedestal). Evidently, there is no consistent difference between performance at the three time gaps, which was surprising; there is a clear time-dependence for forward-masking involving the kind of maskers and probes used by Zeng et al. (e.g. Plack et al., 1995).

The lack of time-dependence is counterintuitive; the 80 msec time gap allows greater perceptual recovery from the masker (Figures 39-41), which the literature suggests is caused (at least partly) by a recovery of neural responsivity. This in turn should allow greater neural response to a given tone, which should allow a lower intensity-increment threshold. (This argument assumes a constant neural increment for successful discrimination.) Certainly there is a clear time-dependence for forward-masking involving the kind of maskers and probes used by Zeng et al. (e.g. Plack et al., 1995).

According to the present interpretation of the two-channel Zeng et al. model, the peaked pattern of the just-detectable intensity-increments should not have appeared at all under the present experimental design (Figure 9). The peaked pattern also fails to support the referential-coding model of Plack et al. (Figure 11), which predicts a mid-level hump for the nonmasked pip but no hump at 3 msec. Thus the results do not support either of the competing theories that motivated this study, inasmuch as the perceptual processes affecting a brief Gaussian pip can be assumed to reflect the perceptual processes affecting the 25 msec tones used by Zeng et al. (1991) and Plack et al. (1995). In lieu of their models, an alternate model will be presently shortly that may explain the behaviour of the Gaussian pip under certain qualifying assumptions.
Figure 50. Tone-pip just-detectable intensity-increments for subject JW. The comparison pip, before multiplication by the Gaussian envelope, is always 20 dB higher than the forward masker.
Figure 51. Tone-pip just-detectable intensity-increments for subject ZS.
There is, however, one trend in the just-detectable intensity-increments that suggests that the forward-masker was used as a reference. The curve of just-detectable intensity-increments was generally lower for the 3 msec forward-masked condition than for the nonmasked pips, for both JW and ZS. Both subjects’ better discrimination at 3 msec might have been due to the closeness of the pip to the masker’s termination, which may have provided a reference sound against which changes in pip level were more clearly heard. The clapping or bubbling sensations provided by both masker roll-off and by the pip itself can indeed interact to provide a discrimination cue, as suggested by the subjects’ comments and the experimenter’s personal experience in Experiment 1 (Chapter 5).

The trend to lower just-detectable intensity-increments for \( t=3 \) msec than for \( t=\infty \) shows an exception that actually reinforces the concept of masker as reference. For both JW and ZS, the error bars of the just-detectable intensity-increment for the nonmasked 30 dB SPL pip did not overlap with those of the masked pip at \( t=3 \) msec, whose just-detectable intensity-increments were higher. For ZS the error bars were non-overlapping even when the fiducial limits were assessed at 98% instead of 95%, and there is little doubt that with further trials JW could have achieved the same. The 10 dB SPL masker itself was inaudible (JW) or sometimes barely audible (ZS), so that the difference of about 0.4 dB in the just-detectable intensity-increments may reflect the breakdown of the masker’s “referential” role once it becomes inaudible. Apparently subjects can learn to use the masker as a reference with sufficient practise at the 2IFC discrimination task. Nonetheless, referential coding, if present, was not strong enough to reduce the mid-level hump evident for all three time conditions.

**The standard deviation \( \sigma \) of the psychometric functions**

Of longstanding interest is the question of how an intensity-increment threshold relates to its underlying psychometric function. The intensity-increment thresholds seen in Figures 50 and 51 were inferred from cumulative normal curves (Equation 26). The standard deviations \( \sigma \) of these normals replicate the intensity-dependence seen for the just-detectable intensity-increments (Figure 52). Figures 53 and 54 show some of these cumulative normal curves, the actual psychometric functions for detection of the intensity-increment.

The subject with the lower just-detectable intensity-increments (ZS) also had lower \( \sigma \)’s overall. Furthermore, ZS showed \( \sigma \)’s for the 3 msec delay that were consistently lower than for the nonmasked pip, the same pattern seen in just-detectable intensity-increments. JW’s standard deviations followed suit with one exception, for the 30 dB SPL masker - 50 dB SPL tone condition. The size of \( \sigma \) clearly peaks for ZS at the 50 dB SPL masker level, but for JW there is not a clear peak, but instead a broad rise. Thus the true peak in \( \sigma \) may lie in-between the 40 and 50 dB SPL masker levels, implying the same for the just-detectable
Figure 52. The standard deviations $\sigma$ of the cumulative normal distribution which was fitted to the subjects' $p(c)$ scores, for JW and ZS.
intensity-increments. Thus the sharp peak seen in each subject's just-detectable intensity-increments might reflect an even higher just-detectable intensity-increment at a lower pedestal level.

The \(\sigma\)'s not only parallel the just-detectable intensity-increments, but the two kinds of plots are very close together for ZS and only slightly further apart for JW (Figures 55 and 56). Sigma appears to be just short of being a just-detectable intensity-increment. The correlation of these two quantities becomes clear when the two data plots seen in each frame of Figures 55 and 56 are replotted as a single plot in each corresponding frame of Figures 57 and 58, in which the \(\sigma\) for a given stimulus condition is plotted versus the corresponding just-detectable intensity-increment for that same stimulus condition. This relation can be crudely quantified by assuming that \(\sigma\) is a multiple of the just-detectable intensity-increment. The just-detectable intensity-increment must become vanishingly small as the pips presented in each of the two intervals of the 2IFC task approach equality; the corresponding psychometric function for increment detection, which is the integral of a Gaussian distribution (Equation 26), must steepen as \(\sigma\), the standard deviation of the Gaussian, approaches zero. Thus the straight lines shown in Figures 57 and 58 must intercept both the abcissa and ordinate at zero.

The fitted slopes of these lines are independent of the values of the correlation coefficients \(r^2\), whose calculation does not require curve-fitting, and whose high values indicate a strong correlation between just-detectable intensity-increment and the standard deviation \(\sigma\) of the psychometric function. Note that strong individual differences in performance, as evident in the just-detectable intensity-increments (Figures 50 and 51), also appear in Figures 57 and 58. Subject JW shows a \(\sigma\) which is 0.66-0.80 of the just-detectable intensity-increment, whereas for ZS these numbers are 0.91-0.95. This difference may imply real physiological differences between what constitutes a just-detectable intensity-increment for each subject. The values of \(r^2\) were slightly higher for ZS, which, rather than implying physiological differences, may imply greater facility at the 2IFC task.

Some might object that the congruency of \(\sigma\) to the just-detectable intensity-increment might disappear if that threshold was determined by \(p(c)=0.95\), say, instead of \(p(c)=0.75\) as done here. But it is hard to imagine such a situation. Of course, the relation between \(\sigma\) and the just-detectable intensity-increment might no longer be linear, and the correlation coefficient might drop accordingly. Also, a higher \(p(c)\) would mean a higher just-detectable intensity-increment, such that, for each psychometric function, \(\sigma\) (a constant for each established psychometric function) would be even smaller than the just-detectable intensity-increment than found here for \(p(c)=0.75\).

Comparison of the just-detectable intensity-increments to Zeng's measure (1994)

There are two obvious questions about the behaviour of the just-detectable intensity-increments.
Figure 53. The psychometric functions (Equation 26) for intensity-increment detection by subject JW. The forward-masker levels were (left column) 10 dB SPL (pip = 30 dB SPL), (middle column) 50 dB SPL (pip = 70 dB SPL) and (right column) 70 dB SPL (pip = 90 dB SPL). The time gaps were (upper) 3 msec, (middle) 80 msec, (lower) ∞ (i.e. no masker).
Figure 54. The psychometric functions (Equation 26) for intensity-increment detection by subject ZS. The forward-masker levels were (left column) 10 dB SPL (pip = 30 dB SPL), (middle column) 50 dB SPL (pip = 70 dB SPL) and (right column) 70 dB SPL (pip = 90 dB SPL). The time gaps were (upper) 3 msec, (middle) 80 msec, (lower) = (i.e. no masker).
First, why do they follow the peaked pattern? Secondly, why did the presence of the masker apparently not affect this pattern?

To begin an answer to the first question, we must return to procedure. In the 2IFC task the subject produces scores representing the percentage of times, on average, that they correctly perceive a difference in intensity between a tone and a more intense tone. Steadily raising the size of the tone's increment results in a steady improvement in the probability that the larger of two internal responses corresponds to the interval containing the incremented tone, and not just the pedestal. Because tone-pip intensity is the only stimulus characteristic that is manipulated by the experimenter when masker and pedestal are fixed, it is reasonable to infer that perceived differences in tone-pip intensity correspond to differences in perceived intensity, that is, differences in loudness. The effects of forward-masking on probe-tone loudness had been examined by Zeng (1994), employing the same stimuli used in the original discovery of the mid-level hump (Zeng et al., 1991) with the exception that the masker was also a 1 kHz tone, not a noise. This time however the tone in the other interval of the 2IFC task was a nonmasked standard. The level of the forward-masked tone was adjusted by the subject to match its loudness to that of the nonmasked standard. This level was found to equal that of the standard at 30 dB SPL but progressively dropped below that of the standard as the standard increased to 70 dB SPL. The matching level then progressively rose again, until the two probes were again equal in loudness at 90 dB SPL. The difference in level between the forward-masked tone and the nonmasked standard amounted to as much as 10-16 dB at mid-levels of the standard (and zero at 30 and 90 dB SPL). For each loudness match, the width of the psychometric function from the 50% to the 76% points was taken as the variability of the judged loudness. For each loudness match, the width of the psychometric function from the 50% to the 76% points was plotted. This plot showed a mid-level rise (Figure 59).

Zeng employed \( p(c)=0.76 \), whereas the present intensity-increment thresholds employ \( p(c)=0.75 \). Thus Zeng's plotted quantity should be nearly equivalent to a present just-detectable intensity-increment. The numbers should be negligibly different if the effect of forward-masking is the same in the two experiments, but Zeng's values are much larger (Figure 59). This might be credited to a less conservative method (adaptive tracking) or less rigorous training of the subjects, but the difference is compelling. Thus the forward-masker used here, despite its other possible effects, may have been ineffectual at provoking the kind of perceptual changes responsible for the mid-level hump seen by Zeng et al. (1991). This discrepancy is compatible with the literature (Zeng et al., 1991; Zeng and Turner, 1992; Plack et al., 1995), which shows that the shifts that forward-masking causes for the just-detectable intensity-increments are larger than the excursions in magnitude seen in Figures 50 and 51. Thus the mid-level hump seen in the latter figures must arise from a different process than that which causes the mid-level hump in a Zeng et al. (1991) experiment².
Figure 55. The just-detectable intensity-increments (from Figure 50) and the standard deviations \( \sigma \) of the fitted psychometric function (Figure 52), for subject JW and for each stimulus condition.
Figure 56. The just-detectable intensity-increments (from Figure 51) and the standard deviations $\sigma$ of the fitted psychometric function (Figure 52), for subject ZS and for each stimulus condition.
Such a process may be the loudness enhancement or decrement found by Elmasian, Galambos, and Bernheim (1980).

Zeng's (1994) experiment has been explained in order to show that two sets of experimental results can be compared. From this comparison, it is evident that the perceptual activity provoked by the brief Gaussian pips used here must be very different from that provoked by the forward-masked 25 msec tones used by Zeng and his contemporaries. Because of its briefness, the tone pip used here probably caused only an initial neural burst (see Footnote 32), reflected in a compound action potential (CAP). The component of CAP that is usually reported is \( N_1 \). The behaviour of \( N_1 \) can mimic that of firing rates and psychophysical thresholds. The \( N_1 \) for brief tone bursts can be forward-masked by tones (Abbas and Gorga, 1981, and Dolan et al., 1983; cats) with the same general kind of dependence on masker frequency and intensity, and on probe frequency, as seen for tone-provoked neural firing rates (Smith, 1977, gerbils; Harris and Dallos, 1979, chinchillas) and for psychophysical detection thresholds (e.g. Fastl, 1979; Zwicker, 1984).

Before attempting to answer the second question posed at the start of this section, we must ask whether the Zeng et al. and Plack et al. models would be expected to apply under the physiological response represented by \( N_1 \). For a stimulus similar to the present probe, human subjects produce a curve of \( N_1 \) vs. level that is similar in shape to that of other mammals (Gaussian pips of 3 msec total duration at 2,140 Hz; Hoke, 1974). No data could be found on forward-masking-provoked shifting of the \( N_1 \) curve in man. However, in Mongolian gerbils the \( N_1 \) resulting from rapidly-ramped 6 kHz probes tends to drop under forward-masking, so that when \( N_1 \) amplitude is plotted against probe level, the resulting curve shifts up-level while maintaining its general shape (Cacace and Smith, 1986). (Recall that the Gaussian envelope of the present experimental probe produces a localised power spectrum (Figure 36) so that the resulting \( N_1 \) is best compared to the \( N_1 \) resulting from the onset of rapidly-ramped pure tones, which should originate from a narrower patch of basilar membrane than the \( N_1 \) for clicks.) Changes in the rate-of-growth of \( N_1 \) are thus (roughly) preserved, so that the pattern of just-detectable intensity-increments should also shift up-level under forward-masking, just like the recovery-rate model for the rate-level curves for 25 msec tone bursts. Because \( N_1 \) is subject to forward-masking, it must also show recovery, introducing a time-dependence in post-masker \( N_1 \) amplitude (see Murnane, Prieve, and Relkin, 1998). If this time-dependence mimics that known for the recovery of longer tones, the Zeng et al. recovery-rate model should hold. The referential-coding model should also still apply, given that no new circumstances have arisen that would remove its applicability.

Another point of information must be settled before attempting an answer the second question posed above, and that is whether the present forward-masker was actually effective in shifting the curves of \( N_1 \)-vs.-probe level. Recall the present experimental conditions: a 200 msec, 2 kHz masker precedes a probe of the
Figure 57. The just-detectable intensity-increments plotted vs. the standard deviations $\sigma$ of the relevant psychometric functions (subject JW), for each respective stimulus condition. Also shown are the correlation coefficients, the least-squares-fitted straight lines, and their slopes.
Figure 58. The just-detectable intensity-increments plotted vs. the standard deviations $\sigma$ of the relevant psychometric functions (subject ZS), for each respective stimulus condition. Also shown are the correlation coefficients, the least-squares-fitted straight lines, and their slopes.
same frequency. For equal-frequency conditions a forward-masker only 50 msec long, of only 40 dB SPL, was found to decrement $N_1$ amplitudes for probes as intense as 80 dB SPL (Cacace and Smith, 1986; 6 kHz, Mongolian gerbil). Thus if $N_1$ is the crucial variable, the masker used here should have had some effect upon it. The same had been found for the cat; the $N_1$ for a rapidly-ramped 10 msec probe of 41 dB SPL at 8 kHz can be decremented by an 8 kHz forward-masker of only 20 dB SPL and 100 msec duration, for time gaps up to perhaps 50 msec (Abbas and Gorga, 1981). The same was found true for the same frequency conditions for 200 msec maskers and fixed time gaps of 5 or 20 msec. Thus it appears that the forward-masker should have affected the physiological response to the Gaussian probe.

To now answer the second question: it is unclear why the forward-masker did not affect the just-detectable intensity-increments, given that the forward-masker should have been effective. Indeed, it is hard to believe that a forward-masker as intense as 70 dB SPL would not cause some shifting of the $N_1$-vs.-probe level curves, up towards higher stimulus levels. This would cause differences in magnitude between the forward-masked intensity-increment threshold and the nonmasked intensity-increment threshold, in either model (Figures 9 and 11). But these differences were absent. The just-detectable intensity-increments of Figures 50 and 51, besides presenting novel data, leave little doubt that brief pips do not follow all the same rules as longer tones. The following model attempts to complete an answer to the first question that was posed at the start of this section, that is, why the just-detectable intensity-increments follow the peaked pattern.

**MODEL**

The nonmasked just-detectable intensity-increments show the same, peaked pattern as those of the forward-masked just-detectable intensity-increments (Figures 50 and 51). We might instead have expected nonmasked 2 kHz pips to produce a gently declining curve with rise in pedestal level, the "near-miss to Weber's law" as found for much longer, gently-ramped, nonmasked 2 kHz tones (Figure 60; Jesteadt et al., 1977). Interestingly, the latter pattern is predicted neither by the Zeng et al. recovery-rate model (Figure 9) nor the Plack et al. referential-coding model (Figure 11). Also, using tones with plateau levels lasting only 1 msec, Carlyon and Moore (1984) showed much lower just-detectable intensity-increments than seen in Figures 50 and 51, for frequencies nearby 2 kHz and with 5 msec onset and offset ramps. The behaviour of the Gaussian pip appears to be unique for a tone-pip.

No published just-detectable intensity-increments could be found for brief Gaussian pips. But as the pip was indeed quite brief (±2 standard deviations spans 2 msec) and, in timbre, halfway between a tone and a click, it is not inappropriate to compare the just-detectable intensity-increments to those for acoustic clicks. The just-detectable intensity-increments of Figures 50 and 51 indeed follow those of nonmasked
Figure 59. Just-detectable intensity-increments for the nonmasked pip, vs. the variability of the loudness matches from Zeng (1994). For Zeng's data, the abcissa represents the level of the forward-masked probe tone; for the present data, the abcissa represents the pedestal level of the Gaussian pip.

Figure 60. The just-detectable intensity-increments for nonmasked pips (from Figures 50 and 51), compared to those for 2 kHz tones of 500 msec duration (Jesteadt et al., 1977).
click stimuli (Figure 61), which peaked at mid-levels (Raab and Taub, 1969), as did intensity-decrements measured using a comparison click and a variable click that approached the comparison from below (Avakyan and Radionova, 1963). The common patterns suggest a common neural representation.

**Background: just-detectable intensity-increments and \( N_i \) variability**

\( N_i \) had already been used to explain the just-detectable intensity-increments for clicks. Taub and Raab (1969) obtained \( N_i \) vs. click level for a Guinea pig. Amplitude histograms of \( N_i \) for click levels 5 dB apart in this animal's lower intensity range turned out to be unimodal and approximately Gaussian. The standard deviations of \( N_i \) were estimated over a similar intensity range for two other animals, giving 6-7 data points for each animal. These standard deviations rose with level and then started to decline, showing a maximum near the level corresponding to about 38 dB SL for a human listener, in the region of maximal Weber fractions for humans (Raab and Taub, 1969). Because the Weber fractions obtained from the \( N_i \) also peaked at about 38-48 dB SL relative to a human listener's click-detection threshold, Taub and Raab concluded that the variability in \( N_i \), represented by its standard deviation, accounted for the mid-level rise in human just-detectable intensity-increments for clicks (Raab and Taub, 1969).

This point was reinforced by an unusual effect that occurred when a click was presented on a continuous background of white noise. As noise level rose to 65 dB SPL, there was a substantial drop in the highest values of the listeners' just-detectable intensity-increments, and the peak in the pattern of just-detectable intensity-increments that had been seen for the nonmasked clicks moved up-level and flattened out. In animals absolute thresholds for \( N_i \) rose and amplitudes declined under the same sort of noise (Taub and Raab, 1969, Guinea pigs; see also Frishkopf, 1956, for cats), which presumably happened because noise-evoked firing reduces a neuron's ability to respond to a concurrent click, taking some neurons "off-line". Histograms of these \( N_i \) amplitudes in the Guinea pig revealed a drop in variability, that is, a smaller standard deviation of the \( N_i \) distributions. Taub and Raab concluded that human just-detectable intensity-increments reflected the variability in neural response that was represented by the standard deviation of the distribution of \( N_i \). In the following sections we shall briefly review two models that relate the mean \( N_i \) amplitude to the standard deviation of the distribution of \( N_i \). The later of these models implies that two neural populations are needed to explain the ear's dynamic range. Despite the increasing doubtfullness of the Zeng *et al.* (1991) model of just-detectable intensity-increments as due to two neural channels of differing recovery rates (Plack *et al.*, 1995), it is still possible to interpret the present results in terms of two neural populations.
Figure 61. The just-detectable intensity-increments for nonmasked pips (from Figures 50 and 51), compared to data for nonmasked clicks. The Weber fractions of Raab and Taub (1969) were converted to just-detectable intensity-increments from the tables in Taub (1969). The reference levels used to compute dB SL presumably differed in each cited study, so that the horizontal positions of the respective curves may differ from what is shown here if presented on a dB SPL scale. For instance, the 0.3 msec click used by Avakyan and Radionova for their intensity-decrement thresholds contains more energy than the 0.1 msec click used by Raab and Taub when both stimuli have the same amplitude before filtering. Assuming equal filtering, the SLs of Avakyan and Radionova were probably computed with respect to a lower sound pressure level. Thus their data would shift leftwards relative to that of Raab and Taub, so that the peaks of the respective curves would be in closer apposition. N.B.: The Weber fractions of Avakyan and Radionova (1963) as shown in Raab and Taub (1969, Fig. 1; from Taub, 1969, Fig. 5) are not plotted correctly, due to an apparent miscalculation by Taub which we could only replicate by using an incorrect formula (omitted) which inflates the Weber fractions of Avakyan and Radionova (1963).
N₁ in terms of two neural populations

The Gaussian pip's just-detectable intensity-increments mimic those for clicks. Hence in order to model the just-detectable intensity-increments of Figures 50 and 51, it may be useful to start with perhaps the first model that related the click-provoked N₁ to its own variability, that of Frishkopf (1956). Frishkopf had no click-discriminability data for man. He noted however that when N₁ was plotted vs. click level, the N₁ in cats showed a notable change of slope in the middle range of click intensities, the "knee" since replicated by many others. In order to account for this curve, Frishkopf first assumed N units, whose thresholds fluctuate from instant to instant. Thus a neuron's threshold was defined not by a discrete stimulus level, but by its probability of falling within a given stimulus interval - that is, each neuron's threshold was described by some stationary probability density function D(S) assumed identical for all N neurons. Although the N units fluctuate with mutual independence, each unit would have the same cumulated probability p(S) of firing in response to a stimulus of intensity S, given by the integral of D(S) from minus infinity to S. This stimulus would fire a number of units n(S) that would vary from presentation to presentation with a variance \( \sigma^2 \) thanks to the fluctuation in each neuron's threshold. Successive responses of a neuron (to fire or not to fire) were assumed mutually independent, so that discharge history had no effect.

Frishkopf then stated the probability for a given S that n units would respond out of N, a binomial statistic, from which followed the mean number of units firing at a given intensity, \( \bar{n}(S) = Np(S) \), and the variance associated with this mean. This variance \( \sigma^2 \) is zero for \( p=0 \) (no units firing) and \( p=1 \) (all units firing); in-between is an inverted parabola with maximum variance at \( p=\frac{1}{2} \). Assuming a constant response amplitude called r for each neuron in a neural population, that population's total N₁ amplitude on a given stimulus presentation was \( A(S) = rn(S) \) and its mean amplitude was \( \bar{A}(S) = r\bar{n}(S) \). From this followed an expression for the variance of \( A \). This \( \sigma^2 A(S) \) was a smooth symmetrical curve that rose from zero, peaked where \( A \) is steepest, and returned to zero.

Frishkopf then assumed that there were other sources of N₁ amplitude variability, which took on positive and negative values \( A_0 \) with equal probability so that total instantaneous amplitude \( A = A + A_0 \), \( \bar{A}(S) = \bar{A}(S) \), and \( \sigma^2 A(S) = \sigma^2 A(S) + \sigma^2 \) where \( \sigma^2 \) was constant. Frishkopf measured mean N₁ amplitude and its standard deviation in several cats. Maximum N₁ amplitude, \( A_{max} = rN \), was estimated from the plateau of mean N₁ amplitude, and \( p(S) = \bar{A}(S)/A_{max} \) was also inferred from mean N₁ amplitude. The free parameters \( \sigma_0 \) and \( N \) were then found by fitting \( \sigma_\tau \) to the measured standard deviations. The values of \( \sigma_0 \) could also be estimated in the absence of the stimulus, and agreed reasonably well with the fitted values. \( N \) was chosen to match the measured standard deviation to \( \sigma_\tau \) at the sound level where \( p(S) = \frac{1}{2} \). With these choices, \( \sigma_\tau \) fit well to the actual standard deviation of the N₁ over the lower range of click levels, but a single-population
model predicted a drop-off at 35 dB SL that was not seen, leading to the conclusion that two neural populations, one "sensitive" and the other "insensitive", were needed. For a single neural pool \( A \) rises in sigmoidal fashion, so to describe an \( N_i \) curve with a knee, two populations were required which overlap barely, if at all, in Frishkopf's model. Thus there would be a drop in \( \sigma_{N_i} \) at the knee in the \( N_i \) curve, and \( \sigma_{N_i} \) was highest at the steepest points in the plot of \( N_i \) amplitude (Frishkopf, 1956, Fig. 26). Frishkopf's actual equations are presented in Appendix F.

Frishkopf's theory was parsimonious but in hindsight some of its physiological parameters were not realistic. For example, the cat has roughly 50,000 primary afferents (Dannhof and Bruns, 1993) but Frishkopf's model predicts only 300 or so units responding to a click at 35 dB above detection threshold \( (N=360, 320 \text{ or } 175; \text{three cats}) \). Granted, there is a substantial breadth of fiber thresholds, even within a cat's critical band (Chapter 2); but such numbers appear very low for a sound that still has a very wide frequency spectrum even after filtering through a speaker or headphones. Frishkopf's model also cannot convincingly describe the second channel's properties.

A two-channel model of discriminability might still be appropriate, but it is not Frishkopf's model. As noted, the Gaussian pip's just-detectable intensity-increments mimic those for clicks, for which Radionova (1963) explained the mid-level hump as follows. Radionova assumed that the \( N_i \) arose from the simultaneous discharge of many neurons. Each neuron's own threshold was believed to obey some probability distribution which, for a given stimulus level \( S \), would cause the neuron to fire on some occasions but not on others (after Frishkopf, 1956). Thus a population of \( N \) neurons would be described by \( N \) threshold distributions, one for each neuron.

Radionova then assumed that in the 0.5 msec of the response to a click, the nerves could produce only one spike each that contributed to \( N_i \), such that mean \( N_i \) amplitude at level \( S \) was proportional to the number of units responding above their thresholds. This number was the total assumed population size \( N \) multiplied by the integral of the probability density function for the \( N \) mean neural thresholds, taken from minus infinity up to \( S \). The variability shown by the \( N_i \) over repeated presentations of the same stimulus level was due to a fluctuation in the number of neurons achieving threshold. Radionova proposed that if each neuron contributed equally to \( N_i \) amplitude, the variance \( \sigma_{N_i}^2 \) reflected the number of neurons whose thresholds fell within some small interval centered around the given stimulus level. Radionova then boldly assumed that, effectively, this interval was infinitely thin, so that \( \sigma_{N_i}^2 \) reflected the neural threshold density in thresholds/dB. Integrating this quantity from \(-\infty\) up to the level of the applied stimulus would give a number that was proportional to the total number of active neurons and hence proportional, in turn, to the amplitude of \( N_i \).

Radionova had measured the mean and variance of \( N_i \) in cats, finding that \( \sigma_{N_i}^2 \) followed a bimodal
plot for 12 click levels between 5 and 115 dB SL. She then computed $\int \sigma_{ni}^2 dS$ from $-\infty$ to $S$, plotted this quantity and $N_i$ for 12 click levels $S$, and found congruence of the two plots up to 65 dB SPL. Frishkopf (1956) had noted that when $N_i$ for the cat was plotted vs. click level, it showed a notable change of slope in the middle range of click intensities, the "knee" since replicated by many others. Radionova's plots of total active neurons and of $N_i$ both showed a mid-range knee, above which the mean $N_i$ rose more quickly than the total responding population. Of course, Radionova might have employed $\sigma_{ni}$ instead of $\sigma_{ni}^2$. The only apparent justification for using $\sigma_{ni}^2$ was the success with which its definite integral imitated $N_i$.

The bimodal nature of $\sigma_{ni}^2$ was compatible with a threshold density composed of two unimodal neural-threshold distributions with different mean values. Because the mean $N_i$ does not decrease over 5-115 dB SL, Radionova reasoned that neurons continued to achieve threshold at least over this range, so that the threshold density function(s) extended at least this far. When the two distributions barely overlap, the good discriminability afforded by continuing recruitment of the lower-threshold units is exhausted, and large numbers of higher-threshold units are not yet accessed. (This is similar to the Zeng et al. (1991) model for tones, but without their upward shifting of mean-rate curves with forward-masking (Figure 7).) Thus the knee in $N_i$ was due to a slowdown in the rate of growth of the active neural population, caused by insufficient overlap of the two threshold distributions.

Radionova's model and Frishkopf's model have certain commonalities. Both are forced to assume two neural pools with little overlap, effectively two separate channels, to account for the wide auditory dynamic range reflected in the non-decreasing amplitude of $N_i$. Both accept that each and every neuron's threshold is probabilistic rather than fixed, and neither party specify the form of the relevant probability density function. But Frishkopf assumed that this function was the very same for all neurons, perhaps the reason that his theory produced very small estimates of pool size $N$. Indeed, the requirement for an identical $D(S)$ for all neurons would force $D(S)$ to be very wide, with a standard deviation on the order of tens of decibels, in order to achieve a reasonable dynamic range for a neural population. The literature suggests that the density functions governing single-neuron threshold, whatever they are, cannot be this wide. Radionova's model instead gives a distribution for mean thresholds, one for each of the neurons in a neural pool.

We must ask whether the Frishkopf and Radionova models can actually specify the neural populations underlying intensity encoding for clicks. Frishkopf's model was adjusted so that variability of neural response, represented by the total variance $\sigma_{n}^2$, reached its maximum at the steepest points in the $N_i$ curve and its lowest at the bottom, the knee, and the saturation level. Radionova's model makes assumptions that result in the same pattern. The plots of $\sigma_{ni}^2$ vs. click level made by Frishkopf are somewhat jagged; several peaks may appear, suggesting several underlying populations. It may not be possible to clarify this
pattern within the context of either model. Radionova's approach requires many measurements of $\sigma_{wi}^2$, needed to obtain a smooth enough plot to be able to fit a smooth curve composed of two unimodal probability distributions. Frishkopf's model allows the underlying probability density functions for neural threshold to be determined only when they are non-overlapping, and only when the number of distributions is known. Thus although these intriguing models offer the possibility of two or more channels at work, they do not offer a way of unambiguously identifying these channels.

A counter-intuitive conclusion, and its resolution

This brief historical review seemed appropriate because the just-detectable intensity-increments of Figures 50 and 51 followed the same pattern as did the just-detectable intensity-increments (Raab and Taub, 1969) and just-detectable intensity-decrements (Avakyan and Radionova, 1963) of nonmasked clicks. Before proceeding further, let us briefly recap the model. Figure 62 shows the interrelationships of the four relevant quantities: 1) the variance of the pip-provoked spike count (proportional to the variance of $N_i$), 2) the number of neurons that are actively contributing to $N_i$ (the driven neurons), 3) the inferred density of neural thresholds (Radionova), and 4) the inferred just-detectable intensity-increments (after Radionova). The upper illustration shows the variability of spike count vs. the growth in size of the population of driven neurons, after both Frishkopf (1956) and Radionova (1963). Because $N_i$ amplitude is assumed proportional to the number of neurons responding to the stimulus, and because each neuron is assumed to provide an equal contribution to $N_i$ amplitude (perhaps a single spike), the variance in $N_i$ amplitude is thus proportional to the variance in spike count. This variance is pictured as having two disjoint distributions. The middle illustration shows the presumed density of neural thresholds, a density assumed proportional to the variance of $N_i$ (after Radionova, 1963). The two threshold distributions overlap little if at all, causing the temporary slowdown of the number of neurons firing in response to the stimulus. The lower illustration shows the distributions of neural thresholds, and the computed rise in the just-detectable intensity-increment, found at levels in-between the mean values of the two threshold distributions. The gap between distributions corresponds to a temporary slowdown in the growth of the number of neurons firing in response to the stimulus, which causes the knee in the $N_i$ curve. The pattern of just-detectable intensity-increments has been extrapolated at very low and very high sound levels where it presumably rises with the inevitable decline in the rate of growth in the number of responding neurons.

Longer stimuli do not show the mid-level peak in just-detectable intensity-increments (e.g. Jesteadt et al., 1977). Radionova suggested that the difference was due to the different kinds of neural activity that are available to indicate intensity. Longer sounds provoke spike trains that can indicate intensity by the firing rate of individual neurons, that is, spikes per some unit of counting time. Very short sounds do not
Figure 62. Model. (Upper frame) The variability of neural counts vs. the growth in size of the population of driven neurons, after Frishkopf (1956) and Radionova (1963). Since $N_1$ amplitude is proportional to the number of active neurons, and since each neuron is assumed to provide a single spike, the the variance in $N_1$ amplitude is thus proportional to the variance in neural counts. (Middle) The presumed density of neural thresholds, assumed proportional to the variance of $N_1$ (after Radionova, 1963), vs. the growth in size of the population of driven neurons. (Lower) The intensity-increment threshold (after Radionova, 1963, with extension to very low and very high sound levels) vs. the presumed density of neural thresholds.
allow this cue. For these stimuli, the neuron may ring for much longer than the duration of the stimulus, but the number of emitted spikes that contribute to the compound action potential is a few, or even just one. Hence inspection of firing rate, the number of spikes fired over some counting time, would not provide a basis for identifying stimulus intensity or changes in stimulus intensity. Indeed, due to the neuron's ringing, such a code would be misleading. It might suggest a continuing stimulus when the actual stimulus is already turned off. It might also provide an unreliable indicator of stimulus intensity, an indicator that might not increase monotonically as stimulus level rises.

Instead of depending upon firing rate, the auditory system must utilise some aspect of the total number of stimulus-provoked spikes, and hence some aspect of the size of the responding neural population. The present model proposes that this crucial quantity is the rate of change of the size of the responding neural population, that is, the derivative of the function describing the number of driven neurons. In Radionova's model, this derivative is the density of neural thresholds. As threshold density drops, so does discriminability, so that the just-detectable intensity-increment rises. A drop in threshold density corresponds to a drop in count variance. Both theoretically (Frishkopf) and experimentally (Radionova), this corresponds to a drop in $\sigma_n^2$ (Figure 62).

But this model is incompatible with the data. In Guinea pigs, the distribution of $N_1$ amplitudes had a smaller standard deviation under simultaneous white-noise masking (Taub and Raab, 1969). Under the same conditions, human just-detectable intensity-increments fell (Raab and Taub, 1969), implying that the size of the just-detectable intensity-increment is correlated with variability in $N_1$. This result is in accord with signal detection theory, in which detectability $d'$ will rise with a drop in variance (all other terms remaining constant), giving better discriminability. Thus the Frishkopf-Radionova model is not compatible with signal detection theory. A comparison of the signal-detection interpretation of the present mid-level hump with the Radionova (1963) model suggests that there is no good reason to reject the signal detection model, and that the Radionova model is incorrect. Rather than clutter the discussion with the technical details, these details are relegated to Appendix G.

**Firing rate vs. rate of population growth: strategies for encoding intensity change**

The apparent incorrectness of Radionova's mathematical assumptions, as demonstrated in the Appendix, does not force the general principles of her model to be wrong. Very brief stimuli may nonetheless be encoded differently from longer stimuli, and if so the crucial quantity for the discriminability of very brief stimuli is still the rate-of-change of the size of the responding neural population. The relevant population must be much smaller for Gaussian pips than for clicks, given that Gaussian pips have a much narrower frequency spectrum (Figure 36). Nonetheless the just-detectable intensity-increments for the
Gaussian pip are of similar magnitude to those for clicks (Figure 61), implying that it is not the total size of the active population that determines intensity-increment detection for very brief stimuli, but rather the rate of population growth. Now, rate is independent of the size of the active population and dependent only on the number of free fibers, which should still differ greatly for clicks vs. pips. Thus the crucial factor may be the population growth relative to the total neural population, a weighted growth rate\textsuperscript{45}.

Thus what we have called population encoding is best referred to as rate-of-growth encoding. It would be useful to quantify this growth in order to infer the resultant discriminability, perhaps in the manner employed by Radionova (1963). Unfortunately, it is still not well understood how population size changes with change in stimulus level. It has been assumed that the rise in number of driven neurons follows the intensity function of Frishkopf (1956, Fig. 26) (Figure 62). But it appears that any plot of $N_1$ that shows an outright plateau cannot be used to infer the contributing neural population, unless the signal-detection model of discrimination is to be deemed irrelevant in this circumstance (Appendix G).

There is a second objection to the inference of population growth from the amplitude of $N_1$. The mean amplitude of $N_1$ is accepted as a broad indicator of the number of active units, but it is biased towards fibers of higher CF. Basilar excitation patterns tend to move up-frequency, as demonstrated by multiple nerve-fiber recordings in cats (e.g. Pfeiffer and Kim, 1975; Evans, 1981) and patterns of simultaneous masking in humans (e.g. Scharf and Buus, 1986, Fig. 14.33). Hence $N_1$ shows a steep rise at higher SPLs as the excitation pattern spreads (e.g. Ozdamar and Dallos, 1976, Fig. 2b). The extent of this bias can be seen for the 7 kHz tone in the Mongolian gerbil; the steep rise usually seen in $N_1$ amplitude is reduced to an actual decline when fibers above 10 kHz (for a species whose frequency range extends to perhaps 40 kHz) are subjected to high-pass noise, which eliminates their contribution to $N_1$. This high-frequency bias in $N_1$ can be countered somewhat by using narrow-band stimuli such as a Gaussian pip. Thanks to its narrow spectrum, the pip should provoke only the local frequency components of $N_1$; such local components do exist, as noted by Ozdamar and Dallos (1976) and Abbas and Gorgia (1981) and as recently reiterated by the frequency specificity of forward-masking of $N_1$ (Cacace and Smith, 1986). The discovery of a mid-level hump for Gaussian pips, without substantial lateral spread of the basilar excitation pattern, reinforces the usefulness of two channels in accounting for the mid-level rise (Figure 62).

Very brief tones, whose ramping is perform short, may also employ population encoding and therefore produce similar mid-level humps in the just-detectable intensity-increment as seen for clicks and Gaussian pips. A mid-level rise happens for intensity decrements as well as intensity increments (Figure 61). But there is an apparent exception; the brief tones used by Florentine (1986) at the relatively low frequencies of 250 Hz and 1 kHz do not provide a mid-level hump, presumably because of spread of the basilar excitation profile up towards higher frequencies, accessing new afferents so that the rate of population
growth remains high. It is instructive to compare discriminability limens for brief tones with pedestals at 40, 65, and 85 dB SPL and frequencies of 250 Hz, 1 kHz, and 8 kHz (Florentine, 1986). Generally, these increment thresholds fell as tones lengthened from 2 msec to 2 sec (=intensity-plateau durations; rise/fall times =1 msec). However, closer inspection reveals that, for the 8 kHz tone and the shortest durations, just-detectable intensity-increments were highest at the intermediate intensity level, 65 dB SPL. Indeed, the magnitudes of this just-detectable intensity-increment and that found for the 85 dB SPL tone, and their mutual difference, resembled those of subject JW (Figure 50). (Florentine's just-detectable intensity-increments, at least for 250 Hz and 1 kHz, were otherwise consistent with the magnitudes of intensity-decrement thresholds known for square-gated tones of similar lengths, as brief as 2 msec (Henning, 1971; see Florentine, 1986, Fig. 5).)

As tones lengthen, a firing-rate code becomes available, perhaps even for tones whose plateau level is only a few milliseconds long. The improvement in discriminability as tones lengthen implies that rates provide better discriminability cues than the very brief bursts of neural firing reflected in the $N_1$. As tone level rises, fibers that are already firing above their spontaneous rates will increase their rate further, and those that are dormant, due to higher thresholds, will start to come on-line. Nonetheless, close inspection of the literature reveals that a mid-level hump is possible even for longer tones, at higher frequencies (Jesteadt et al., 1977; Long and Cullen, 1985). For example, when Carlyon and Moore (1984) examined nonmasked just-detectable intensity-increments at 6.5 and 8 kHz, a mid-level elevation was seen that increased when the tone was presented simultaneously with bandpass noise centered at the tone's frequency. These tones were 26 msec long, plus 5 msec ramps. The discriminability pattern was little changed by masking the ramping of the tone, using noise bands presented before and after. This is expected given that the length of the ramps should eliminate the $N_1$ potential (Goldstein and Kiang, 1958) and, by inference, any cue that might have been present due to the frequency splatter caused by the ramping itself. Also, the 5 msec ramps of Carlyon and Moore should have allowed a fairly good localisation of tonal energy. But the use of high tonal frequencies restricts the usual up-frequency spread of basilar excitation patterns; the mid-level rise in the just-detectable intensity-increment implies that rate coding involves integration of evoked spikes over some range of CFs which is fairly broad$^{46}$.

Although we do not know the precise timing conditions under which a rate-of-growth code would give way to a firing-rate code, we can assume that rate-of-growth codes must fail if $N_1$ is absent. How rapid does ramping have to be to preserve $N_1$? The most authoritative study may be that of Zaboeva (1966). Like Goldstein and Kiang, she employed white noise, with a total duration of 20 msec and an amplitude $U_f=U_0(1-\exp(-t/\tau))$. Increasing $\tau$ from 0.2 to 16 msec for a plateau level of 30 dB SPL causes $N_1$ to drop quickly. On average (12 cats), $N_1$ vanishes by $\tau=9.8$ msec. $N_1$ threshold rises monotonically as $\tau$ decreases (average, 8
cats); increasing \( \tau \) from 0.2 to 6 msec causes a threshold rise of 16 dB. For fixed values of \( U_0, N_1 \), amplitude increases monotonically as \( \tau \) drops from 4 to 0.2 msec (average, 8 cats). Under this change of \( \tau \), \( N_1 \) amplitude rises for fixed above-threshold amplitudes \( U_0 \). The consistency of these results is broken only by the observation that decreasing \( \tau \) from 4 to 0.2 msec did not alter \( N_1 \) amplitude for stimuli of 5 or 10 dB SL.

In sum, a rate code must apparently take over for approximately \( \tau > 10 \) msec for white noise of 30 dB SL, regardless of the length of the plateau in the stimulus' amplitude. We would expect that this \( \tau \) shrinks as sound level drops, and conversely that \( \tau \) increases as sound level rises. Thus for white noise, at least, a rate-of-growth code could work for a relatively large range of \( \tau \), given this exponential ramping. These \( \tau \)'s should represent auditory upper limits; as tones have a much narrower spectrum, they should produce a smaller \( N_1 \) for a given sound level, and thus the relevant \( \tau \)'s will probably be much smaller.

**Mid-level humps: the role of spike pooling**

A gathering of peripheral output must happen in order to allow counting of the spikes provoked by very brief sounds, nonmasked or otherwise, so that population-based encoding can occur. Spike pooling is thus a possible link between psychophysical and neural behaviour. It may also be a factor common to intensity encoding for nonmasked brief stimuli, and forward-masked longer stimuli. As for rate-based encoding, a mid-level hump is found in *backward* masking (Plack et al., 1995), in which the probe precedes the masker. Backward masking of 25 msec tones implies a central locus in which pooling of neural activity occurs over some integration time whose length must be reflected in the time-dependence of both forward- and backward-masking (see Plack et al., 1995). This dependence appears to be consistent for both kinds of masking, as would be expected. A central pooling also explains the existence of contralateral forward- and backward-masking for threshold detection (Elliott, 1962) and for just-detectable intensity-increments (Plack et al., 1995), in which masker and tone are presented respectively to the different ears.

Suggestions of central pooling appear elsewhere. An attempt to resolve the origin of the mid-level hump in just-detectable intensity-increments had come from Zeng and Shannon (1995) using a 1 kHz masker in place of their earlier noise masker. For two patients with implants that stimulate the auditory nerve, there was a mid-level rise in forward-masked just-detectable intensity-increments, as well as reasonable nonmasked just-detectable intensity-increments. For patients with implants that stimulate the cochlear nucleus, weak forward maskers produced only an elevation in just-detectable intensity-increments but no hump. Zeng and Shannon concluded that the midlevel hump is not due to the first afferent synapse, implying a gathering of spikes at the cochlear nucleus or higher. Evidence from contralateral masking suggests higher (Schlauch, Clement, Ries, and DiGiovanni, 1999).

The final component of the present model can now be put in place. Spikes from responding neurons
must be gathered somewhere to provide total counts, so that intensity differences can provoke count differences. Whatever the peripheral intensity encoding, the afferent response can be emulated by neurons of the cochlear nucleus, then sent to the various nuclei of the brainstem, with what is believed to be the first interaction of left- and right-ear inputs occurs (Irvine, 1986). We may assume that these inputs are collected by collector neurons, in which left-ear input can cause fatigue and hence reduced response of the collector to right-ear inputs, and vice versa, allowing contralateral masking. Thus although pooling might occur at the cochlear nucleus, if there is a comparison point that is common to all nonsimultaneous masking it must lie above the cochlear nucleus. Of course, comparison need not occur in just one place; close inspection of the literature reveals that contralateral forward maskers cause a small mid-level rise not as profound as that for ipsilateral maskers (Zeng and Shannon, 1995). This contralateral-ipsilaterial difference apparently depends on the time gap between masker and probe (Plack et al., 1995), so that contralateral masking actually causes greater shifts in the just-detectable intensity-increments for time gaps of 12.5-50 msec. The same phenomenon was seen in backward masking (Plack et al., 1995), implying that contralateral and ipsilateral masking may be mediated at different loci.

Summary of the model

The purpose of this study was to test two theories of intensity encoding, the Zeng et al. (1991) model of two recovering channels and the Plack et al. (1995) model of referential-coding. But there was no obvious difference between the just-detectable intensity-increments for the three time gaps. Thus, in an earlier section, two questions were posed about the behaviour of the just-detectable intensity-increments. First, why the peaked pattern? Secondly, why did the forward-masker apparently not affect this pattern?

The second question was answered, as best as possible, in that same earlier section. An attempt was made to answer the first question by noting the similarity of just-detectable intensity-increments for the Gaussian pip to those of clicks (Avakyan and Radionova, 1963; Raab and Taub, 1969). Longer stimuli do not show the mid-level peak in just-detectable intensity-increments, leading Radionova (1963) to suggest that although the intensity of longer stimuli might be indicated by the firing rate of individual neurons, very brief stimuli (such as clicks) do not allow this cue. Now, the measured physiological response to clicks is the amplitude of the N1 potential. In Radionova's (1963) model, the variance of N1 reflected the neural threshold density, with mean N1 amplitude being proportional to the number of units responding above their thresholds. This population is distributed into two groups; the knee in N1 is caused by the lack of overlap of these distributions, which results in a slowdown of the rate-of-growth of the active population. This rate-of-change is just the derivative of the mean N1 amplitude, i.e. the density of neural thresholds. As density drops, spike-count variance drops, but just-detectable intensity-increments rise.
Although it now appears unlikely that measured $N$ variance reflects the density of neural thresholds (Appendix G), as Radionova assumed, this by itself does not rule out two population densities that somewhat overlap. Two overlapping populations could produce mid-level humps as seen for clicks and Gaussian pips, if the rate of population growth is the basis for encoding intensity change for very brief tones with brief ramping. Due to limited spectral spread, Gaussian pips access far fewer neurons than clicks. Thus the crucial factor in discriminability may be population growth relative to the accessible population size. In order that population-based encoding can occur, peripheral output must be gathered to allow counting of the spikes provoked by very brief sounds, allowing intensity differences to produce count differences.

If we assume that the forward-masker was effective, we have a test of the Radionova model. To explain: forward-masking causes shifting of the curve showing $N$ amplitude vs. level, and we might expect that this represents shifting of the neural thresholds for response to a brief stimulus. This in turn corresponds to shifting of Radionova’s two fiber-threshold distributions. Thus for $t=3$ msec, the pip should always access the same fiber population in the same channel, giving a constant just-detectable intensity-increment for all pip levels. For $t=80$ msec, we might assume as per Zeng et al. (1991) that the upper population corresponds to low-spontaneous-rate neurons that recover from forward-masking more slowly than the presumed high-spontaneous-rate neurons of the lower population. If so, the perceptual gap already predicted for nonmasked pips under a Radionova model would be even more profound.

But discriminability is similar for all three time gaps (3 msec, 80 msec, and $\infty$). Thus, if the forward-masker affects discriminability, the two-population Radionova model must fail. Alternatively, the forward-masker was ineffectual. We are left with the conclusion that either all three models (Zeng et al., Plack et al., and Radionova) are inapplicable, or that the forward-masker had no effect on the brief Gaussian pip.

**SUMMARY OF RESULTS**

When an intense forward-masker precedes a 25 msec tone by 100 msec, the tone’s just-detectable intensity-increments are inflated at moderate pedestal levels (Zeng et al., 1991). The present work investigated two models that account for this inflation, the Zeng et al. (1991) recovery-rate model and the Plack et al. (1995) referential-coding model. Intensity-increment thresholds were obtained at three masker-pip delays (3 msec, 80 msec, and $\infty$) for a very brief Gaussian-shaped pip of 2 kHz forward-masked by a long 2 kHz tone. A constant pip/masker intensity ratio of 20 dB was employed. The literature implies that the forward-masker affects the encoding of the pip, as it should for the 25 msec, more gently ramped tones used elsewhere (Zeng et al., 1991; Plack et al., 1995). Thus the pattern of the just-detectable intensity-increments should have differed for the three different time gaps employed.

A mid-level elevation was seen in the just-detectable intensity-increment for both the forward-
masked and non-masked Gaussian pips. If the forward-masker had been effective, the peaked pattern would support neither the Zeng et al. (1991) model nor the Plack et al. (1995) model. If the forward-masker was not effective, only the referential-coding model would have worked. The similar discriminability for the Gaussian pips and nonmasked clicks suggested a common explanation. A reduction in the rate of growth in the number of neurons activated by a very brief nonmasked stimulus would reduce sensitivity to stimulus increments, creating the mid-level hump. In this regard Radionova (1963) had suggested a transition between two neural populations of different average thresholds, channels that act to extend auditory dynamic range. Although these two channels might still be differentiated by the neurons' spontaneous rates, there was no specification of different recovery rates, unlike the Zeng et al. model. Like the other two models, the two-population model fails if the forward-masker is effective. Thus an effective forward-masker implies support for none of the three models, and an ineffective forward-masker implies support for both the referential-coding model - should that model apply to such brief stimuli - and the two-channel model of Radionova (1963). We have already seen that two neural populations of different mean threshold have the potential to substantially extend auditory dynamic range when firing-rate encoding is employed (Chapter 3). The present experiment suggests the possible involvement of such a dynamic-range-extending mechanism, for very brief sounds that access rate-of-population-growth encoding.

Although the encoding of just-detectable intensity-increments must depend on central pooling of spikes for both very brief nonmasked stimuli and forward-masked longer tones, the present results, combined with the literature, suggest that these stimuli may experience two different intensity-encoding processes.

The method of constant levels allowed psychometric functions to be assembled for the intensity-increment detection. Assumed to underly each psychometric function is a Gaussian probability density whose standard deviations closely follow the just-detectable intensity-increments. Indeed, when standard deviations are plotted against just-detectable intensity-increments, high correlations and linear relationships are found.
NOTES

40. The importance of procedure is sometimes forgotten. For example, substantial training effects were found in attaining click intensity-decrement thresholds by an adjustment method (Avakyan and Radionova, 1963), so much so that the subjects' first 50 trials at each level were not counted towards the intensity-decrement threshold. These authors, who acted as subjects, noted the difficulty they had in making discriminations, especially at low levels. A profound improvement in just-detectable intensity-increments occurred also in four cats for click trains at a level of 44 dB SPL (Saunders, 1969). Our own method of constant levels may allow the subject more time to minimise bias in their listening criterion. Nonetheless, substantial day-to-day improvements in performance were seen even when the subjects had become well-acustomed to the task. In further agreement with the literature, both subjects agreed that discrimination was especially difficult at low pip levels, including absolute threshold detection of the lone pip.

41. Plack, Carlyon, and Viemeister (1995) found the mid-level hump for a 50 dB SPL pedestal, for both forward and backward maskers, but noted a drop in the just-detectable intensity-increment as the 20 msec maskers and tones were brought together from 100 msec separation (the time-gap commonly used by Zeng, Turner, and Relkin (1991) and others) until they were half-overlapping. In another example, Zeng and Turner (1992) extended the original Zeng et al. study by using the same stimuli as Zeng et al., a 100 msec long, 90 dB SPL narrowband noise centered at 1 kHz, preceding by 100 msec a 1 kHz probe, 25 msec long. Four masker-probe time gaps were used (50, 100, 200, and 400 msec) and the increase in the increment-threshold, compared to nonmasked controls, was progressively lower with greater time gap.

42. Two cautions must be noted. First, pitch cues are suspected to assist the subject in forward-masked pip detection, in which the masker appears in both intervals of the 2IFC task but the pip to be detected appears in only one of the intervals (e.g. Terry and Moore, 1977). Pitch cues might thus improve the detectability of Zeng's forward-masked probe, increasing the probe's apparent loudness. A second concern is that the facilitation caused by the masker may not be consistent with time gap. Carlyon and Beveridge (1993) reasoned that adaptation caused by forward-masking should, on average, reduce the neural response to the more intense tone (the incremented tone) in the two-interval forced-choice procedure, making it sound more like the nonmasked comparison (pedestal) tone. Conversely, discrimination should be easier when the masker precedes the comparison tone, increasing the perceived difference between pedestal and pedestal-plus-increment. For a 1 kHz tone, Carlyon and Beveridge utilised a time gap of 170 msec, and three pedestal levels (low, medium, and high). A second study employed a 15 msec gap, only one pedestal level (about 60 dB SPL), and a reduced masker (70.5 dB SPL). The just-detectable intensity-increments for the 170 msec time gap ran opposite to prediction, such that the masker increased the loudness of the forward-masked tone. For the 15 msec gap, however, the forward-masked just-detectable intensity-increments behaved as predicted. This time-dependence suggests that Zeng's results may apply to only one of the forward-masked time gaps (3 vs. 80 msec) employed here.

43. The behaviour of clicks is not consistent over all publications. Penner and Viemeister (1973) obtained Weber fractions (ΔI/I) for clicks, with both broadband and high-pass filtering and with or without background noise. If these Weber fractions show a rise at moderate click levels, so should the discriminability limen expressed in dB (see the formulae in Appendix E). Penner and Viemeister found no profound mid-level hump and no effect of background noise. However, their Weber fractions for nonmasked broadband clicks (ibid., Figure 2) are inconsistent; two of the three subjects show a steep initial drop in (ΔI/I) at 0 dB SL, followed by a more gentle decline (perhaps even a plateau) having very large bumps. The third subject however shows no steep decline, but a progressive, bumpy rise towards the highest click levels used (59 dB SL). Their Weber fractions are also inconsistent within-subjects. Left-ear Weber fractions sometimes differed substantially from right-ear Weber fractions, so that the two plots sometimes appearing to progress in different directions (up or down) with change in SL. The absence of error bars makes the
significance of these differences hard to assess. Because of these problems we hesitate to model these particular results.

44. Rather than computing the just-detectable intensity-increments, Taub and Raab (1969) compared the $N_i$ amplitudes evoked by clicks in each interval of a 2IFC task, one interval containing the comparison click and the other the incremented click. When the interval containing the incremented click also gave the higher of the two $N_i$ potentials, a correct vote was recorded, and the level of the incremented click was adjusted in a staircase procedure according to these votes in order to arrive at a 75% correct vote. It was assumed that for neighbouring click intensities, the distributions for $N_i$ had different means but the same standard deviation. Thus, the act of specifying some probability (e.g. 75%) that the interval chosen as having the more intense stimulus did indeed contain that stimulus, corresponded to specifying the separation between the means of the distributions of $N_i$ evoked by repeated presentation of those two stimuli. This separation was taken as an intensity-increment limen.

45. We might be tempted to graphically illustrate the growth in population using basilar excitation patterns (after Whitfield, 1967). In these models, driven firing rates of neurons (firing rate minus spontaneous rate) are plotted as vertical bars along an axis indicating basilar location, under simplifying assumptions such as a constant distribution of neurons along the basilar membrane (after Ozdamar and Dallos (1976) and others). However, clicks or very brief pips can be expected to provoke an all-or-nothing response from a single neuron; the neuron fires above its spontaneous rate, or does not fire at all. Granted, the neuron will ring after the stimulus disappears (see for example Pfeiffer and Kim (1972) for clicks). But only the initial burst, perhaps only a single spike, is expected to contribute to $N_i$. Thus, if we were to concern ourselves only with that portion of the response contributing to $N_i$, then the responses to very brief pips would have to be illustrated by vertical bars of a constant height. Such bars would form comb-like patterns due to lateral spread of excitation across the basilar membrane. These illustrations would have little utility as a tool for understanding just-detectable intensity-increments for very brief stimuli. Clicks, for example, must stimulate a very great portion of the basilar membrane unless they are very weak, in which case neurons at the ends of the membrane, those of very low or very high CFs, have thresholds too high to be fired by the click. Thus clicks of even moderate levels will already have provoked a response from most of the neurons having relatively low thresholds. Spread of excitation across the basilar membrane cannot help explain the mid-level hump for Gaussian pips either, since these have a very narrow spectrum (Figure 36) and the growth in $N_i$ is presumably due to accessing of neurons with higher thresholds, not to lateral spread.

46. To continue this principle: When this excitation pattern is restricted, such that there is a smaller range of free frequencies to expand into, rate cues are lost due to eventual saturation of the firing rates of the firing neurons. New cues do not become available until a higher neural channel is accessed. Bandpass (notched) noise takes useful neurons off-line, exacerbating this effect. Notched noise has been used as an investigative tool in order to minimise off-frequency cues, but the role of notched noise masking is still unclear; it has opposite effects on the just-detectable intensity-increments for nonmasked (Carlyon and Moore, 1984) vs. forward- or backward-masked sounds (Plack and Viemeister, 1992), even at frequencies as low as 1 kHz (e.g. Schneider and Parker, 1987). Such apparent contradictions may reflect detection cues other than neural response profiles, so that attempts to incorporate effects of notched noise in a two-channel model of discriminability may not be prudent.

47. Regardless of where neural fatigue may occur, it appears to go hand-in-hand with forward masking. Forward-masked detection thresholds increase monotonically with rise in masker level (Zwicker, 1984; Fastl, 1979; Carlyon, 1988), an intensity-dependence that can be seen clearly in forward-masked peripheral afferents (Smith, 1977; Harris and Dallos, 1979). Comparing masking using masker and probe in one ear versus masker in one and probe in the other lead Lüscher and Zwischlocki (1949) to conclude that forward
masking is a monaural and hence peripheral process. Lüscher and Zwischlocki also deduced that recovery from a forward masker was the reversal of a peripheral adaptation process. Results of more recent forward-masking studies agree at least partially with this interpretation (Smialowski and Carhart, 1975; Carlyon, 1988). Of course, it is not clear how “peripheral” this effect must be; here “peripheral” might well be any level in the auditory system that does not have binaural interaction.

48. If we assume that any basilar motion that might be caused by electrical stimulation is irrelevant, then the recovery seen in implant wearers (Figure 47) implies not basilar motion but processes further along the auditory pathway. In such deaf patients the outer hair cells are presumed destroyed. If masking is not cochlear it can then occur at the cochlear nucleus (Kaltenbach, Meleca, Falzarano, Myers, and Simpson, 1993; Shore, 1995) the common destination of the afferents from the inner hair cells (e.g. Fekete, Rouiller, Liberman, and Ryugo, 1984; Osen, 1970), although effects occur which are not seen in primary afferents and which are blamed on nuclear architectonics (Kaltenbach, Meleca, Falzarano, Myers, and Simpson, 1993; Shore, 1995). All these findings imply that forward masking may occur not just in primary afferents (Smith, 1977; Harris and Dallos, 1979) but anywhere where neural response can suffer fatigue.
CHAPTER 7

FURTHER DISCUSSION, AND FINAL SUMMARY OF RESULTS
In this section we explore issues that are germane to the question of the role of channels in auditory dynamic range, but that were not necessary to describe the results from earlier chapters.

**CAN OUR COMPUTATIONS (CHAPTERS 2 AND 3) BE GENERALISED TO OTHER SPECIES?**

**THE ROLE OF DESIGN**

Although we focused our attention on the cat, the computational methods of Chapter 2 might be applicable to mammalian hearing in general. Such a broadened scope would bring useful new results. Below we review what is common, and not common, in the characteristics that describe the firing-rate behaviour of peripheral afferents. We also discuss the problems involved in measuring these characteristics. For the sake of simplicity, we discuss only pure-tone stimulation of fibers, the most commonly found case. Stimulation by noise need not guarantee the same characteristics, due to the much larger number of component frequencies in the noise.

**Common features of fiber properties across species**

An ubiquitous feature of mammalian audition is that spontaneous rates of single afferents follow two distinct distributions, one for low-spontaneous-rates that is roughly exponentially declining in its probability density, one for higher spontaneous rates that is roughly Gaussian in shape. These density functions have mean values (of spontaneous rate) that are of the same orders of magnitude across species, when mutually comparing the macaque (Nomoto, Suga, and Katsuki, 1964), the Guinea pig (Evans, 1972, and Manley and Robertson, 1976), the squirrel monkey (Geisler, Rhode, and Kennedy, 1974), the cat (Liberman, 1978, and Sachs, Woolf, and Sinnott, 1980), the chinchilla (Borg, Engstrom, Linde, and Marklund, 1988, and Relkin and Doucet, 1991), and the rat (El Barbary, 1991a). In contrast stand the pigeon and the blackbird, with no spontaneous rates below 2/sec, and only a single spontaneous-rate distribution (Sachs, Woolf, and Sinnott, 1980).

Then there is threshold. The criteria for threshold may differ from study to study (as will be discussed), and sample sizes are always relatively quite small. Bearing in mind these concerns, low-spontaneous-rate and high-spontaneous-rate fibers may appear to differ in mean threshold within each species (Guinea pig, Evans, 1972; cat, Liberman, 1978; chinchilla, Relkin and Pelli, 1987, and Borg, Engstrom, Linde, and Marklund, 1988; rat, El Barbary, 1991b; Mongolian gerbil, Ohlemiller, Echteler, and Siegel, 1991). Indeed, there appears to be a strong correlation between threshold and spontaneous rate at a given CF for the pigeon (Smolders, Ding-Pfennigdorff, and Klinke, 1995), which does not disagree with the strategy of evaluating low-spontaneous-rate and high-spontaneous-rate units as separate channels. To obtain accurate computations of dynamic range, this correlation would have to be taken into account arithmetically, perhaps within a 1-channel model.

Consider also the maximum firing rates. The Gaussian distribution of maximum firing rates $r_{\text{max}}$ (Figure 22) is seen elsewhere, for the squirrel monkey (fig. 7 of Rose, Hind, Anderson, and Brugge, 1971) and the pigeon (Sachs, Woolf, and Sinnott, 1980). A change in saturation rates with CF (Figure 22), showing a rise in the region of $\text{CF}=10$ kHz,
also appears for the Mongolian gerbil (Ohlemiller, Echteler, and Siegel, 1991).

Having mentioned three properties of all fibers (spontaneous rate, threshold, and saturation rate), we should bear in mind that not all fibers show the same sort of rate-level behaviour. Those units that are a minority in this respect, the sloping-saturating fibers, appear for other species as well as cats, which include both mammals and birds (gerbil, Ohlemiller, Echteler, and Siegel, 1991; Guinea pig, Winter, Robertson, and Yates, 1990; pigeon, Temchin and Moushegian, 1992; blackbird, Sachs, Woolf, and Sinnott, 1980). This reference list is incomplete, given that evidence for sloping-saturating units in a certain species has sometimes been found by more than one laboratory. Furthermore there are probably other species showing sloping-saturation, in reports that were not noticed. It has been suggested that all sloping-saturating units in mammals have their trademark “bend” at the same sound pressure level (Richter, Heynert, and Klinke, 1995), implying a mechanism in basilar-membrane mechanics, but this did not happen for the pigeon (Richter, Heynert, and Klinke, 1995). This is also clearly not the case for the units displayed for the cat by Sachs and Abbas (1974).

There are a variety of other cross-species similarities that, as yet, are not as global as those noted above. For example, a roughly Gaussian distribution for thresholds, as we sometimes employed in our simulations for the cat (Figure 17), has been noted for the macaque monkey (Katsuki, Suga, and Kanno, 1962). Strangely, this distribution splits into two apparent Gaussians for CFs below 9 kHz, with the mean of one distribution being 40 dB higher than the mean of the other. This higher distribution could well be a separate population of sloping-saturating fibers; Nomoto et al. were not clear about this.

As another example, the roughly Gaussian distribution of dynamic ranges in sigmoidally-responding fibers (Figure 18) has also been noted in the pigeon (Richter, Heynert, and Klinke, 1995) with a mean value similar to that of those of the cat. Also, the pigeon shows three types of fiber firing-rate functions, the sigmoidal, sloping-saturating, and straight (Richter, Heynert, and Klinke, 1995), just as found in the Guinea pig (Winter, Robertson, and Yates, 1990).

Despite gaps in our knowledge, it appears that fiber characteristics are similarly distributed across-species. In particular, spontaneous rate is sufficiently well recorded in the literature that we may infer a common distribution across species, and it is not unthinkable that maximum rates are common also. These are the two characteristics for which measurement is fairly unambiguous. But the same cannot be said of threshold and dynamic range, as we note below. Nonetheless, the similarity of afferent neural rate-level properties across-species suggests a common architecture for stimulus-intensity encoding at the basilar membrane.

**Differences in fiber properties across species: due to problems of definition?**

Despite the across-species similarities noted above, there are many apparent differences from study to study in the properties of afferent firing-rate behaviour. These differences are found within as well as
across species. As demonstrated by the computations in Chapter 3, different distributions of fiber properties make differences in computed overall dynamic range. It is, however, very hard to tell whether these differences in fiber properties are real, in view of the many different ways in which firing characteristics have been defined. We explore this issue below. In an attempt to avoid ambiguity, we have sometimes quoted from the original papers.

Problems of defining threshold

An auditory afferent’s threshold has usually been defined in terms of firing rate when stimulated at the stimulus frequency to which the fiber shows its lowest comparative threshold - the characteristic frequency (CF), sometimes referred to as best frequency (BF). Thus fiber threshold is the minimum point on the frequency tuning curve (FTC), the plot of fiber threshold vs. tonal frequency, which is drawn by subjecting the fiber to a slow frequency sweep. This plot depends on a firing-rate criterion for threshold, and the known sharpness of neural tuning suggests that very fine steps in frequency are needed to establish the fiber’s true CF. Hence the accuracy with which the FTC is determined is a limiting factor on the accuracy of CF and thus on the accuracy of threshold at CF. The limit of the accuracy of the assigned CF is set by the size of the inner-hair-cell itself, to which the afferents connect. Each IHC takes up a certain length of the basilar membrane. Because frequencies themselves can be infinitely fine, each IHC must best respond not to one frequency, but to a range of frequencies within which the number of stimulating frequencies is limited only by the frequency-generating equipment, and by the outer- and middle-ear’s ability to faithfully reproduce those frequencies.

Thus measurement of threshold cannot be separated from establishment of the FTC. All the same, threshold measures are usually stated in terms of firing rate at CF. Independently of the issue of the accuracy of the assigned CF, establishing threshold is a problem. When defined in terms of firing rate at CF, threshold is not defined consistently across studies, even for the same species. For the pigeon, for example, threshold has been defined as a 50% increase over the fiber’s mean spontaneous firing rate (Sachs, Young, and Lewis, 1974) but alternatively as a 25% increase over the fiber’s mean spontaneous firing rate (Richter, Heynert, and Klinke, 1995). These spontaneous rates can be established with minimum variance by averaging over relatively long time intervals (say, 1 minute). For the Mongolian gerbil, however, threshold was “the smallest stimulus amplitude that yielded at least 2 more spikes in a 50-ms window with the stimulus present than in an identical window during which there was no stimulus” (Ohlemiller, Echteler, and Siegel, 1991). Note that this would seem to correspond to an increase of 40 spikes per second - a relatively enormous change!

To see the differences that dissimilar threshold definitions have made for a single species, consider the subject of our computations, the cat. In their oft-cited paper, Sachs and Abbas (1974) defined threshold pressure
as "the sound pressure level which produces a rate which exceeds the spontaneous rate by an amount equal to 10% of the difference between saturation and spontaneous rates" (Sachs and Abbas, 1974, p. 1846). This has the potential for giving vastly different thresholds for units with similar spontaneous rates but very different saturation rates. For example, for two fibers with the same \( r_s = 1/sec \) and saturation rates of either 101/sec or 301/sec, the rate changes for the threshold criterion are 10 and 30 spikes/sec, respectively.

Liberman's (1978) measure of threshold involved subtracting the number of spikes occurring within the typical counting time of 50 msec after the tone burst with the number occurring during the burst itself. This difference was then compared to a criterion value in such a way that the measured threshold "is within 10 dB of the tone level required to elicit a firing-rate increase of 10 spikes/sec...regardless of the unit's rate of spontaneous discharge". What this appears to mean is that a unit with \( r_s = 1/sec \) can have its threshold decided by the same 10 spikes/sec change in firing rate that may occur for a unit with \( r_s = 100/sec \). For the first unit, this constant change of 10/sec represents a rate change of hundreds of percent; for the second unit, this change is only 10% of resting rate.

Geisler, Deng, and Greenberg (1985) pointed out this potential conflict between thresholds determined by percentage changes in firing rates and those determined by absolute differences in actual firing rate. In an effort to get to the root of the problem, Geisler's group used three different ways of quantifying threshold. One was a threshold-tracking program like that used by Liberman (1978) but with some differences. The routine sampled the unit's spontaneous rate over 20 separate periods of 250 msec each. This allowed a mean value and a standard deviation to be assigned to the spontaneous activity of the fiber. The threshold criterion was then set at 1 spike greater than the sum of the mean and standard deviation obtained earlier. The SPL of the stimulus tone was then incremented in 2 dB steps until a count crossing level was found. These "tracking thresholds" were found to be reliable, within a few dB of each other, over repeated determinations.

Geisler, Deng, and Greenberg (1985) also defined what they called a "3\( \sigma \) threshold". This involved calculating the mean and standard deviation of the time-interval between spikes, in quiet. The threshold was then the stimulus pressure that produced a mean interval that was three standard deviations less than the mean interval during spontaneous firing. As a final definition of threshold, a "shape threshold" was defined based on the shapes of rate-level curves whose higher-intensity portions were monotonically increasing (the majority of fibers studied under this criterion). It was found that the differences between the 3\( \sigma \) threshold and the shape threshold "rarely exceeded 5 dB" for the small sample of fibers used. The tracking threshold technique gave thresholds about 5 dB higher than those from shape-threshold measurements.

Liberman's (1978) measurements indicated that thresholds of low-spontaneous-rate fibers averaged 20 dB higher than thresholds of high-spontaneous-rate fibers. Geisler et al.'s measurements, on the other hand, showed a difference of at best 5 dB. Yet other definitions of threshold have been used by other laboratories with yet differing results. In the Guinea pig, threshold occurred for "an increase in discharge rate of the RMS of (20 spikes/sec and 20% of the spontaneous rate). For fibres with spontaneous rates less than 30 this
approximates to a 20 spikes/s increase while for spontaneous rates greater than 100 it approximates to 20% of spontaneous" (Winter, Robertson, and Yates, 1990). The inconsistency of this criterion is self-evident; the criterion increase in neural counts at "threshold" is clearly much closer to the fiber's spontaneous rate for high-spontaneous-rate fibers than for low-spontaneous-rate fibers. Such a criterion would only make sense if low-spontaneous-rate fibers had rate-level curves that started out with very shallow slopes (compared to those of high-spontaneous-rate units), such that the effective threshold of the fiber, defined here as the bottom limit of its useful dynamic range, occurred well above the point at which its firing rate first deviated from its spontaneous rate. Such a rate-level function would be that of a sloping-saturating unit, turned upside-down and back-to-front. No such rate-level functions have yet been noted. For low-spontaneous-rate units the firing-rate increase at threshold is relatively enormous; consider the percentage change in rate represented by an increase of 20 spikes/sec in a fiber with a spontaneous rate of 1 spike/sec. Compare this to the relatively fine criteria used by Geisler et al., a single spike in a 250 msec period corresponding to a change of only 4 spikes/sec. Hence the criterion rate changes at threshold of Winter et al. appear closer to those of Sachs and Abbas (1974) and Liberman (1978) than the finer gradations of response used by Geisler et al.:

We should consider one possible solution to the threshold problem, offered by the fast computer algorithm tested by Relkin and Pelli (1987, chinchilla). Relkin and Pelli employed a two-interval forced-choice task (see Figure 5) to construct a neurometric function, the neural equivalent of a psychometric function. A probe tone 25 msec long was presented in one of two intervals, with a masking tone presented for 100 msec at the beginnings of both intervals. If the number of spikes counted with the tone present exceeded the number counted with the tone absent, the tone was registered as being correctly detected. Threshold was then defined as the tone intensity required for 61% correct detection, read from a plot of percentage correct vs. sound intensity. These thresholds agreed well with those found in the same experiment using an up-down convergence procedure in which stimulus intensity is changed in increasingly narrow steps until some criterion minimum intensity change (e.g. 0.5 dB) is reached. Although this method seems to solve the problem of defining threshold, it has not yet produced enough data to assign threshold distributions to the different spontaneous-rate groups.

There is obviously disagreement over what constitutes an appropriate measure of threshold. Attempts to get around this problem may only serve to confuse the issue of whether there are real correlations between fiber characteristics. For example, one alternative is to work with relative thresholds, thresholds adjusted relative to some other measurable phenomenon, perhaps some other measurement of threshold. In measuring fiber thresholds in the Guinea pig, Winter, Robertson, and Yates (1990) normalised the fiber thresholds, measured across several animals at the fiber CFs, by subtracting the threshold for the CAP (compound action potential) measured at the same CF for each animal. This CAP threshold is thus used
as an "average" threshold because the CAP represents the coordinated initial neural response of the IHC afferents. For instance, if a unit had a threshold at CF of 20 dB SPL, and the CAP threshold at that CF was 25 dB SPL, then the relative threshold for that unit would be -5 dB. This scale was used as the ordinate in plotting the unit's spontaneous rate vs. dB re CAP threshold. This ordinate varied from -20 to +20 dB, a range of 40 dB. A strong correlation was thus found between spontaneous rate and normalised thresholds. Contrast this result to Ohlemiller, Echteler, and Siegel (1991), who found a strikingly similar plot across-animals for the Mongolian gerbil, without adjusting thresholds by CAP values (ibid., Fig. 2). Both studies used somewhat similar simple counting criteria to establish thresholds. Both data plots' abscissas covered a range of 40 dB.

Problems of defining dynamic range

Different ways of defining threshold imply different measures of dynamic range, which at its broadest is the difference between SPL at threshold and at saturation. We may count ourselves fortunate that saturation rate has an unambiguous meaning and that saturation rates, although measured by different criteria from study to study, probably do not differ much according to method.

However, identifying dynamic range is still a problem. For the pigeon, dynamic range was defined as extending "between the CF-threshold of the fibre and the intensity for 80% of the saturation discharge rate" (Richter, Heynert, and Klinke, 1995). For the Mongolian gerbil, however, dynamic range was "the difference of sound-pressure levels yielding a driven rate change equal to 0.9xR_{MD}" where \( R_{MD} = r_{max} - r_s \) (Ohlemiller, Echteler, and Siegel, 1991). This vague description might be equivalent to an earlier measure used with the Guinea pig, wherein a fiber's dynamic range was "measured between 10% and 90% of the total spike rate change, i.e., between spontaneous and saturation rates" (Harrison, 1981). In phrasing the dynamic range problem, Evans himself referred to dynamic range as extending "from the just sub-threshold point in spontaneous activity up to the first level at which the discharge rate had saturated" (Evans, 1981, p. 721, a rather ambiguous definition requiring a certain amount of judgement. This nonetheless appears to be the broadest possible dynamic range that could be assigned to a fiber, likely equalling or exceeding the ranges produced by the other definitions mentioned above. Within Evans' definition and within the intensity levels used in his neural recordings (typically ≤100 dB SPL), some of his fibers showing sloping saturation could not be assigned dynamic ranges, only minimum estimated dynamic ranges (the lines with arrows attached, seen in Figure 3). Evans found that most units had a mean dynamic range of 41 dB, the value that we assume for most auditory afferents in our computations (Table 4). However 9% had ranges over 60 dB and 5% over 70 dB, and some had not saturated at the highest levels employed, despite having thresholds apparently no higher than 60 dB SPL (Evans, 1981; Figure 3). But the useful dynamic range of a sloping-saturating fiber
(Chapter 3; Figures 33-35) is much less than the dynamic range estimated in the manner employed by Evans and all others, because these definitions place the upper limit of dynamic range well past the point at which the slope of the rate-level plot changes from steep to shallower. Thus the average useful range of these fibers must be much less than 40 dB. Assuming useful dynamic ranges of 60-70 dB is clearly inappropriate.

It is transparent that none of the experimentalists mentioned above has examined whether their definition of dynamic range is operationally valid, that is, whether the assumed dynamic ranges of neurons actually correspond to some criterion of useful discriminability. The only person to have apparently done this is Delgutte (1987), who examined neural rates for stimulation at CF in 229 fibers from 17 cats, using a 2IFC technique apparently identical to that of Relkin and Pelli, 1987 (and of Relkin and Turner, 1988; see Chapter 5). The resulting neurometric functions were used to estimate difference limens vs. sound level for individual nerve fibers. The resulting U-shaped plots are similar to those of Figures 33-35. Delgutte defined the width of each of his plots, at a difference limen of 10 dB, as the “10-dB dynamic range” (in decibels) for each fiber. These ranges appear quite generous, given that they are estimated well up on each plot. These ranges varied from 20 to 50 dB, well within the ranges stated by Evans and Palmer (1980), but with no units having dynamic ranges >60 dB, unlike Evans and Palmer (1980). Thus Evans and Palmer, working with the same species, appear to estimate an even greater dynamic range than Delgutte.

By Delgutte’s 10-dB criterion, fibers of medium spontaneous rate (defined therein according to Liberman (1978) as 0.5<r<18 spikes/sec) had the largest dynamic ranges, clearly higher than the band of lowest dynamic ranges, those of high-r, units. Delgutte provided a rationale for this by noting that the mid- and low-r, units appear to be the ones whose steepest portion of the rate-level curve is widest (in dB). Delgutte also plotted the sound level Lmin corresponding to the minimum point on each fiber’s difference-limen plot. The collection of these Lmin can be displayed in histograms of occurrence rate vs. tone level, collected separately for each of the Liberman (1978) spontaneous-rate classes. Tone level (abscissa) was expressed as a relative scale in which 0 dB corresponded to roughly 10 dB SPL for 1<CF<10 kHz. The mean of Lmin for low-r, units was 30 dB, that of mid-r, units was 23 dB, and that of high-r, units was 12 dB. These differences between these mean values resembles that between mean firing thresholds according to the criterion of Liberman (1978). This correlation might be thought of as evidence that Liberman’s firing-threshold criterion is most appropriate. However, Delgutte’s Lmin is not a firing-threshold criterion per se, but perhaps a criterion for the midpoint of a fiber’s useful dynamic range (assuming symmetrical difference-limen curves for each fiber).

Figures 33-35 make it clear that the dynamic ranges assumed by others are exaggerations of the useful dynamic range, and probably very substantial exaggerations. This does not mean that the possible usefulness of two computationally-separated intensity-encoding channels (Chapter 3) has been mis-
estimated. The dynamic ranges we employ are the estimates of Evans and Palmer, which are overestimates of the useful dynamic range for any class of fiber with which they dealt. However, our dynamic range term \( \lambda \) (see Figure 2) is a very generous estimate, probably very close to the Evans and Palmer estimates of dynamic range for fibers with sigmoidal rate-level responses. Comparison is more difficult for sloping-saturating units, given that we employ a function with two dynamic ranges (\( \lambda_1 \) and \( \lambda_2 \); see Equation 3) to describe the rate-level functions of sloping-saturating units. Nonetheless, we assumed the average dynamic ranges of Evans and Palmer (see Table 4) for \( \lambda_1 \). (“Straight” rate-level functions were discovered later and for other species. They defy definition of dynamic range in terms of saturation rates because of their high thresholds, not far from maximum employed stimulus levels.)

We note that even the psychophysical auditory dynamic range is not uniquely defined (see Chapter 1). In one definition, dynamic range spans detection threshold to sensation threshold; in another definition, dynamic range was the contiguous intensity range over which intensity-increment thresholds were small (about 1 dB or less). But the former spans 0-120 dB SPL, whereas the latter is unknown because we have no jnds for 105-120 dB SPL or above, even, we suspect, for the well-worked 1 kHz tone. This suggests experiments exploring discriminability at high SPLs. However, the very conservative attitudes of ethics committees may restrain or prevent such explorations, in which case this issue will not be resolved.

**Actual across-species differences**

Some across-species differences in threshold and dynamic range may derive only from differing ways of measuring these fiber characteristics from study to study. There are some characteristics of fiber firing that appear independent of such measurement problems. We refer to the different percentages found for the three types of rate-level function (sigmoidal, sloping-saturating, and straight). Rate is obtained by the counting of spikes within a commonly used interval, usually 40-50 msec, starting at a common starting point (about 10 msec after the stimulus is turned on). Repeated trials are employed for the same neuron, at least 10 presentations at a given sound level, often many more. In the pigeon, 37% of units sampled were sigmoidal, and 50% were sloping-saturating (Richter, Heynert, and Klinke, 1995), the rest being “straight”. In the Guinea pig, 42% were sloping-saturating, 48% were sigmoidal, and 10% were straight (Winter, Robertson, and Yates, 1990). In the cat, only 8% or so of fibers were sloping-saturating, and no “straight” rate-intensity functions were noted (Liberman, 1988). The distribution of these different fiber types in a given animal may well be CF-dependent, and the number of units sampled may be relatively small; for example, Winter, Robertson, and Yates (1990, Table 2) described only 89 units, taken from 10 animals. Hence all these percentages are not firm. It is not inconceivable that interanimal differences within a species might also be significant.

More problematic is the issue of how many of these rate-level types belong in each of the recognised
spontaneous-rate classifications. Liberman (1988, cat) found sloping-saturating units only within the low-spontaneous-rate fibers ($r_t<0.5$ spikes/sec). However, Winter, Robertson and Yates (1990, Guinea pig) found sloping-saturating units in all spontaneous-rate classes. Because spontaneous rate is measured relatively unambiguously, involving spike counts over periods of time that are long compared to the mean interspike intervals, one cannot blame this apparent difference on measurement criteria. The situation differs when relating fiber rate-level type to threshold, because threshold criteria are not uniform across studies (see above).

**Summary of problems in defining fiber properties**

Differences in measurement criteria may represent differences of attitude; measures that are convenient for the experimentalist may offer little sense of rigour to the theorist. However, the cacophony of different criteria for measuring threshold and dynamic range may be causing the perception of false relations between characteristic properties of neural activity, obscuring real relations. We cannot let ourselves be trapped into thinking that appropriate definitions of threshold and dynamic range are somehow unreachable. Although there may in fact be unrecognised theoretical reasons why such terms cannot be defined without bias, the proliferation of differing criteria are a nuisance.

Can a threshold criterion be recommended among those presently used? The answer is yes. The 2IFC method of Relkin and Pelli has the advantage of compatibility with a well-known model of discriminability, Signal Detection Theory, and their neural intensity-increment thresholds predicted by the method of constant levels agreed closely with those obtained through another, common psychophysical method, up-down convergence on the intensity-increment threshold. The application of psychophysical procedures to nerve fibers provides a feeling of union between neural and psychophysical measurement. Furthermore, two-interval comparisons are a component of theoretical jnd measures such as that of Siebert (1965) and Hellman and Hellman (1990). Psychophysical discriminability computed using 2IFC can now be compared to discriminability limens computed in the same way from neural performance at above-threshold levels (Delgutte, 1987; Relkin and Turner, 1988; *et alii*). We note that comparison of two or more intervals is inherent to Signal Detection Theory, although that approach was invented much later than 2IFC, which dates back at least as far as Montgomery (1935).

**Summary of this Section**

Until the relations between fiber thresholds, rate-level types, and dynamic ranges can be established beyond reasonable doubt, it will not be possible to understand stimulus-intensity processing by the basilar membrane. In Chapter 3 we assumed overlapping channels based on Liberman's (1978) criterion for fiber
thresholds. If these criteria turn out to be misleading, then other approaches are called for. For example, in a completely different approach, Winter, Robertson, and Yates (1990) suggested that low-r, units form a separate channel to other units from 5-100 dB SPL (or higher), whose purpose is frequency coding by basilar membrane locus, a sort of parallel channel rather than the present model of a sequential, overlapping channel. Such a role is suggested elsewhere by the rate-place profile, the plot of single-fiber firing rate vs.fiber CF, for low-spontaneous-rate units. In response to a tone a clear peak is apparent even up to a tone intensity of 87 dB SPL (Shofner and Sachs, 1986; tone frequency=1.5 kHz). In contrast, rate-place profiles for other r, groups saturated at high SPLs, so that tone frequency presumably cannot be represented by local ensembles of those particular neurons. A greater frequency specificity of the rate-place profiles of low-r, units was also found by Kim, Chang, and Sirianni (1990).

WHERE AND HOW MIGHT TWO AUDITORY CHANNELS (CHAPTERS 3 AND 6) INTERACT?

In Chapter 3 we explored the extension of overall dynamic range that might be offered by a second channel when encoding of intensity and intensity change was assumed to involve firing rates. In Chapter 6 we proposed two intensity-encoding channels to account for the discriminability of Gaussian pips that provoke a population response, not a firing rate. Although we assumed in both chapters that the two channels were the high- and low-spontaneous-rate units, there seems to be no guarantee that the two different encoding methods represent the same neural pathways. For example, rate-based encoding may involve the critical band (as modelled in Chapter 2), or less, whereas population-based encoding may involve the entire basilar membrane. At present, we will ignore this question of the peripheral component of the physiological pathways - we have not enough evidence to discuss this issue - and focus instead on where two auditory channels might come together to interact. The channels must interact, in order that saturation of firing rate in one channel does not overwhelm the brain with neural impulses, disguising the behaviour in another, less saturated channel. A switchover from one channel to the other may be part of this process, but it is difficult to determine the stage of auditory processing at which a switchover might occur. More than one processing locus might be involved. The firing rates in two channels might be processed in parallel up to some point, with the higher nervous system subsequently “attending” to one channel or another.

A possible locus of two-channel interaction

Assuming that fibers of differing spontaneous rate form channels, as done in Chapter 3, the logical place to start looking for a two-channel interaction is the first locus at which these fibers can interact through synaptic contact or mutual interneurons. As pictured in Figure 32, the cochlear nucleus (CN), a complex and only mildly understood body, is the only terminus for the primary afferents from the inner hair cells (Osen,
1970, confirming earlier studies; Fekete, Rouiller, Liberman, and Ryugo, 1984). Horseradish peroxidase tracer allows the entire course of the auditory afferent fiber to be described, revealing that the afferent fiber splits into two branches at the CN, an ascending branch that terminates in the anteroventral CN (AVCN) and a descending branch that leaves axon collaterals in the posteroventral CN (PVCN) and (in most cases) continues into the dorsal CN (Fekete, Rouiller, Liberman, and Ryugo, 1984, confirming the staining study of Osen, 1970; Leake-Jones and Snyder, 1982; Leake and Snyder, 1989).

This presents a possibility for two-channel interactions. Blackburn and Sachs (1990) imagine a differential weighting of fiber inputs to certain cells of the cochlear nucleus, resulting in "selective listening". This is one kind of two-channel interaction: activity in the second channel automatically suppresses activity in the first channel when some criterion is exceeded. This is akin to the retinal circuitry of various species, in which it is now acknowledged that a mutually suppressive interaction occurs between rods and cones at intermediate levels of light intensity (for the cat, see for example Rodieck and Rushton, 1976; Levine, Frishman, and Enroth-Cugell, 1987).

Footnote 1 introduced the idea that the suspected rod-cone break in visual recovery might have an auditory analogy. Similarly, visual gating might provide insights into auditory gating. There is a wealth of psychophysical and physiological evidence for rod-cone interaction based on a wide variety of experimental phenomena. Psychophysical work suggests such an interaction in determining absolute thresholds (Drum 1981a,b, 1982), the contribution of rods and cones to brightness (Palmer, 1976), and brightness increment thresholds (Buck and Pulos, 1987; Sharpe, Fach, Nordby, and Stockman, 1989). This interaction is likely to be neither totally inhibitory nor fully complementary (i.e. additive rod and cone responses), but somewhere in-between (Drum, 1982), and is not strictly a one-way effect of rods on cones, but may be reciprocal (e.g. Sugita, Itabashi, and Tasaki, 1988). A variety of evidence from different species suggests that this interaction occurs either through direct rod-cone contact or through intermediary retinal cells. Thresholds for ganglion cell response demonstrate a sub-additive distance-dependent interaction between rods and cones that is likely to be mediated by lateral interconnecting cells at the retina (goldfish; Shefner and Levine, 1977). A similar experiment on the cat's ganglion cells produced similar results (Levine, Frishman, and Enroth-Cugell, 1987). Rod-cone interaction can also be inferred in the human electroretinogram, which is recorded at the cornea and results from current distal to the layer formed by the retinal ganglion cells (Sandberg, Berson, and Effron, 1981). With the help of the anatomical search for possible retinal circuits (see Sterling, Freed, and Smith, 1988), it is now generally agreed that a variety of rod-cone interactions are likely to happen.

A similar interaction may occur in audition. One possible participant in such an interaction is the chopper cell. As well as terminals from primary afferents, possibly inhibitory nerve terminals are found on
chopper-type cells in the antero-ventral portion of the cochlear nucleus (AVCN; papers noted in Blackburn and Sachs, 1989). Chopper units are characterised by a peristimulus-time histogram showing a series of peaks whose number and amplitude increase with rise in SPL, as the interpeak interval monotonically shortens (Webster, 1977), although some choppers have only transient responses (Rhode and Smith, 1986). (A peristimulus-time histogram is a time histogram (number of event occurrences plotted vs. time) of the number of neural spikes that occur within successive counting bins of equal duration, while the stimulus is present.) The peaks in the peristimulus-time histogram are regularly spaced and are not related to the phase of the tone, and chopper cells are found for all CFs (Pfeiffer, 1966).

Afferents of all spontaneous rates contact a category of cell that is described by its shape, the stellate cell. These cells have chopper activity (papers noted in Wang and Sachs, 1992), producing a comb-like histogram of firing rate vs. time that is not phase-locked to the stimulus and hence appears unrelated to sound frequency (see also Webster, 1977). Blackburn and Sachs (1990) suggest that these cells may be the place for selective listening, a two-channel crossover. Blackburn and Sachs have not ruled out the possibility of a mutual suppression between the two known chopper-cell subgroups. Perhaps such an interaction is mediated by interneurons in a manner akin to rod-cone interaction. Similarities between various response properties of CN cells and those of primary afferents (see Blackburn and Sachs, 1989) imply that the “rods” and “cones” are still the high- and low-r, fibers, respectively.

It is still not certain that two intensity-encoding channels are necessary (see Chapter 6). Compared to fibers of the cochlear nerve, a large proportion of CN cells, more in the dorsal CN rather than in the ventral CN, have relatively large dynamic ranges. The mean dynamic range of 83 dB was credited to inhibitory side-bands activated by the notched-noise masker (500 msec tone in notch 1 octave wide; Palmer and Evans, 1982). Peripherally, notched noise creates a substantial extension of dynamic range only for sloping-saturating afferents (ibid.), and the small sample size (5 units) precludes generalising this result to the entire eighth nerve. The important point is that in certain listening conditions, the saturation of firing rates in primary afferent fibers does not preclude a larger dynamic response range at single cells further up the processing ladder. Further discussion of CN phenomena is beyond the scope of this thesis.

A potential model species for two-channel interaction: the moth

We have reviewed some suggestions that there could be a two-channel interaction at the cochlear nucleus in the cat. The complexity of this body’s interconnections, and the number of its different cell types, in terms of both morphology and physiology, conspire against any immediate understanding of its possible role in solving the problem of how a smooth transition in firing rate might occur between two auditory channels. The answer may instead lurk in an architectonically simpler auditory system, that of the moth.
Recall Figure 4, which showed how two firing-rate curves differing only in threshold could be summed to extend dynamic range. The moth has only two peripheral receptors, which are also the primary afferents, which are sensitive to ultrasound frequencies. Furthermore the moth is studied without anesthesia. This offers a great advantage over mammalian studies, because it is generally suspected that anesthesia affects mammalian hearing in unknown ways.

A simple model of overall auditory dynamic range?

Understanding moth hearing is indeed germane to understanding mammalian hearing; for instance, the moth’s auditory fibers display the same sort of firing-rate adaptation found in the cat and other mammals (see Perez and Coro, 1985). Furthermore, dynamic ranges of moth primary afferents are of the same order of magnitude as those in mammals. Each of the moth’s receptors has a useful dynamic range of 30 dB or so (Ascalapha odorata and Empyreuma pugione, Coro and Alonso, 1989; other moth species, Perez and Coro, 1985) covering 35-65 and 65-95 dB SPL respectively in one subject, well within the overall auditory intensity range of cats and humans. The two peripheral receptors, called A1 and A2, have overlapping response ranges, and the A1 receptor may actually decrease its firing rate above 65 dB SPL. Perez and Coro (1985) noted the dynamic range problem as it applied to the moth, and suggested that inhibitory interactions between the two afferents might happen in some species that they had studied, thus somehow extending dynamic range.

Second-order repeater neurons R1 and R2 have similar properties to A1 and A2, and both sets of units show an onset response that is higher than the steady-state activity (as in the cat). The third repeater cell, labelled R3, has a wider dynamic range than any other cell, one of 40-50 dB (Figure 63). This neuron has a smoothly monotonic rate-level function, and inferential evidence suggests that it receives input from both “A” neurons. (There is also a chopper interneuron that may receive input from both peripheral receptors [Coro and Alonso, 1989]). Other neuron types seen in the cat cochlear nucleus (both DCN and VCN) appear to have single-unit analogs in these moth species (Coro and Alonso, 1989) and in other moth species (Boyan and Fullard, 1986). Boyan and Fullard (1986) subjected Heliothis viriscens to short ultrasound bursts at test frequencies comparable to those used by Coro and Alonso (1989). They found similar primary afferents as found by Coro et al., “at least seven” subsequent interneurons in the pterothoracic ganglion, the apparent equivalent of the cochlear nucleus, and similar firing rate behaviour. Again, inferential evidence suggests a single interneuron of wide dynamic range that receives input from both peripheral “A” receptors. The smooth response of such an integrator neuron poses a solution to the problem of a non-monotonic total spike count when intensity is encoded using two neural channels. All these features offer a system that may be a useful model for auditory processing in mammals (Coro and Alonso, 1989).
Figure 63. Average numbers of spike counts within a 45 msec, 16 kHz pulse in an unanesthetised moth Ascalapha odorata (Fig. 2 [lower] of Coro and Alonso, 1989). The auditory receptors are labeled $A_1$ and $A_2$ and the their repeater neurons are labeled $RA_1$ and $RA_2$. The $RA$ neuron, a possible summator, has a higher dynamic range than either the $A_1$ or $A_2$ neuron.
Coro and Alonso (1989) also note the similarity of their RA neuron to the \( O_C \) interneuron in the PVCN of the cat (Rhode and Smith, 1986). The \( O_C \), an "onset-chopper" unit that shows 2-4 peristimulus-time response peaks followed by continued firing in response to the stimulus tone, may be a class of integrator neuron. \( O_C \) units have the largest dynamic ranges (average of 54 dB vs. 22 dB for primary afferents) and highest discharge rates in the CN, and their rate-level curves are monotonically increasing and are found at all CFs. Inferential evidence suggests that several neurons converge on each \( O_C \) unit (Rhode and Smith, 1986). The cat, rather than utilising just a single cell like the moth, may have a population of integrator neurons for intensity coding.

*A neural assembly for response speed and frequency resolution?*

The complexity of the cat's CN precludes any further statement. However, if mutual inhibition of high-\( r_s \) and low-\( r_s \) signals occurs at an interneuron, it helps solve another problem, that of speed of response. If switchover between two channels was regulated by neural systems in the pons (the forward-bulging portion of the brainstem), such as the lateral- or medial-olivocochlear system, a significant delay would occur due to synaptic delay times. Convergence on a single cell, however, would allow these changes to happen much faster.

Finally, one must ask why a population of integrative interneurons may be necessary in the cat but only a single integrator neuron in the moth. The answer lies in frequency resolution. The moth with two peripheral receptors cannot distinguish frequencies (papers noted in Coro and Alonso, 1989), and therefore convergence of two fibers' outputs onto a single second-order neuron does not remove any notion of frequency. This is not the case in cats or humans, which are sensitive to a wide frequency range. In mammals, however, sideband inhibition for cells of the CN can preserve the identities of neighbouring, simultaneously-presented frequencies (Greenwood and Maruyama, 1965; Whitfield, 1967).

In order to preserve this continuum of frequencies while encoding the intensity at any possible presented frequency, there must be a battery of intensity-integrator neurons. How many? The number of these neurons might equal the number of just-noticeable-differences in frequency, \( \Delta f \), such that the computational methods of Chapter 2 are useful only for one \( \Delta f \). These frequency difference limens can be provided by subjective discriminations, which can then be combined with basilar maps and fiber packing-density data to infer the number of peripheral neurons involved, and from that, the degree of neural convergence. The number of peripheral afferents for one \( \Delta f \) can also be implied from coupling basilar maps and fiber packing data with reasonable assumptions derived from rate-vs.-place behaviour of peripheral firing (e.g. Winslow and Sachs, 1988, p.175). It might also be possible to infer the peripheral population from application of Signal Detection Theory to population response profiles along the basilar membrane (as
done by Kim and Parham, 1991). Although psychophysical frequency-discrimination limens are available for some animals, we are not aware of any published comparison of the peripheral populations inferred from the purely neural, and the partly psychophysical, approaches.

**FINAL SUMMARY OF RESULTS**

Auditory stimuli provoke the activity of the primary afferent neurons attending the peripheral hearing organ, the Organ of Corti (sometimes referred to by its substructure, the basilar membrane). The basilar membrane’s afferents encode stimulus intensity and intensity changes, presumably (in most models) by firing rates measured over some brief counting window. But the ranges over which afferents can effectively encode changes in sound level as changes in firing rate are quite limited, as is the spread of afferent firing thresholds, according to data from anesthetised animals. In humans however there is a very broad contiguous range of sound levels over which subjects can adequately distinguish (discriminate) between two brief stimuli varied in intensity but not in frequency. Changes of a few decibels [dB] can be detected over at least 100 dB, whereas an auditory neuron’s tone-provoked firing rate will change, on average, only over 40 dB. And the latter estimate is optimistic. This discrepancy in ranges, known as the “dynamic range problem”, is the topic of the present work.

One computational model was followed by two experiments. The model provided the dynamic range afforded by pooling of the individual peripheral fiber’s neural spikes over a counting window at some proposed counting center. Previous attempts to compute dynamic range in this fashion only averaged over a few “representative” neurons (e.g. Siebert, 1965; Goldstein, 1974; Howes, 1974; Winslow and Sachs, 1988) and did not account comprehensively for the variability of neural characteristics such as dynamic range, spontaneous firing rate in quiet, maximum possible firing rate (saturation rate), and threshold stimulus intensity required to provoke firing above spontaneous rate. In the present work, a mathematical function of logistic form was devised in these neural characteristics appeared as four parameters. This function was shown to provide a good fit to the individual firing rates of nerve fibers. A double-logistic fitted well to a fiber sub-group whose plots of rate vs. level showed a sudden change of slope partway through what is commonly presumed to be their effective encoding range (“sloping-saturating” neurons).

Now by spontaneous rate, neurons fall into 3 groups whose average thresholds differ. For each group, firing-rate equations were averaged over four probability distribution functions, one for each of the fibers’ characteristics. This integration was done for the frequency band presumably serving loudness (the critical band). The 3 results were then averaged again by relative group size. An existing model relating discriminability to neural rate was then modified slightly, and used to compute dynamic range. This range, 82 decibels (dB), is an optimistic estimate that is much smaller than the suspected perceptual range.
The distributions of the four fiber characteristics were manipulated to examine their relative importance. The most notable changes in overall dynamic range arose from changes in the distribution of fiber threshold, not that of fiber dynamic range. Many other poolings were investigated; computationally isolating low-spontaneous-rate neurons, as a separate "channel", extends the upper limit by only 3 dB (to 92 dB SPL). Interestingly, when the dynamic range was maximised using maximum-likelihood methods, the usual method found in the literature, only 7 channels were required to provide almost the same range as offered by treating each fiber in the critical band as a separate, mathematically maximised channel. However, maximum-likelihood methods have never been physiologically substantiated, hence our focus on non-ideal computations that would probably require much less physiological sophistication.

When fibers of low-spontaneous-rate are computationally isolated as an upper channel, the combined dynamic range is 86 dB, only 4 dB greater than a single channel. When medium-spontaneous-rate units are included in the upper but not the lower channel, the dynamic range spans 9-86 dB SPL, the upper limit now being 3 dB lower than for a monolithic, single channel. Dynamic ranges may depend on spontaneous-rate group; assuming average dynamic ranges of 18, 25, and 40 dB for the high-, mid-, and low-spontaneous-rate fibers, respectively, reduces the dynamic range of a single-channel model to 79 dB. Isolating low-spontaneous-rate units as an upper channel extends overall range to 89 dB (3-92 dB SPL), which is 3 dB larger than noted for a two-channel model in the previous paragraph. These computations imply that computationally splitting the afferent population into two channels offers little advantage over a single channel. In the present computational scheme, pooling of neural spikes into two channels only has the potential to mitigate the "dynamic range problem" if the two channels operate in succession as stimulus intensity increases, with one group being a lower-intensity-encoding channel and one an higher-intensity-encoding channel. For this to happen, true neural properties would have to differ substantially from published estimates. Great separations would be required between the mean values of the threshold distributions assumed for high- vs. low-spontaneous-rate afferents, with the mean value for thresholds of medium-spontaneous-rate afferents commonly expected to fall somewhere in-between. The average dynamic range of high- and medium-spontaneous-rate afferents would also have to be smaller than presently assumed, and the average dynamic range of low-spontaneous-rate afferents would have to be higher. Thus fibers of higher spontaneous rates would effectively encode only the lower stimulus intensities (say, 10-50 dB SPL), and changes in those intensities, and the other fibers would have to effectively offer detection and discrimination for the rest of the behavioural dynamic range. Further exploration of the encoding abilities of pooled neural firing must await common agreement on meaningful measures of a neuron's threshold and dynamic range.

Another variation tried, using the average dynamic ranges noted just above, was to isolate high-,
mid-, and low-spontaneous-rate fibers as separate channels, with the low-spontaneous-rate fibers further split into two channels according to sigmoidal-saturating and sloping-saturating behaviour (giving four channels in total). Isolating sloping-saturating fibers in this fashion offers an overall range of 7-92 dB SPL. This is only a 3 dB advantage over the simple single-channel model. Also, when the dynamic ranges of individual sloping-saturating fibers were computed, these ranges did not prove notably superior to those of other fibers. These computations spoil the popular myth that sloping-saturating neurons solve the “dynamic range problem”. There are also doubts as to whether these neurons actually have substantially higher thresholds than other fibers.

To further investigate the possible contributions of a second channel, two experiments were done. Both utilised the knowledge that fiber firing rates in response to a brief probe tone will be systematically lowered after previous exposure to a conditioner containing the same stimulus frequency (forward-masking). The low-spontaneous-rate and high-spontaneous-rate populations of neurons have different post-masker recovery rates, and the low-spontaneous-rate neurons typically have thresholds that are higher than those of the more numerous high-spontaneous-rate neurons. Now, subjective detection thresholds for a short probe tone in quiet can be assumed to follow the lowest neural detection thresholds, those of neurons of high-spontaneous-rate. This capability should be temporarily disabled by an intense forward-masker, which should fully saturate the high-spontaneous-rate neurons because of their relatively low thresholds and limited dynamic ranges. Thus neurons of lower spontaneous rate (and higher threshold) mediate some short period of post-masker detection. Eventually the high-spontaneous-rate neurons will recover, ultimately determining the shape and asymptote of the plot of probe-detection threshold vs. time gap. The transition between these neural channels should be marked by a cusp in the plot of probe-detection threshold vs. post-masker duration. Thus, it might be possible to observe the auditory analog of the rod-cone break. Following an intense bleacher that saturates the rods, the curve describing recovering visual detection threshold for a small point of light shows a cusp midway through its course. The early part of the recovery curve reflects cone activity, whereas the later part reflects rod activity.

Because single-neuron recovery from a forward-masker can be substantial within a short duration post-masker, a probe-detection curve of high temporal resolution is required, but an unambiguous masker-probe time gap t demands rapid ramping of stimulus amplitude, introducing a spread in the stimulus’ Fourier frequency components. This spread was compensated for as best as theoretically possible, by shaping a 2 kHz probe tone with a Gaussian envelope. The resulting pip was positioned at ¼ msec intervals following the cessation of a 97 dB SPL, 200 msec masker, so that forward-masked thresholds were obtained at a much finer resolution than previously seen. Thresholds for probe-detection were established by the method of constant levels. This method paid off in ways that the much more popular and much faster staircase
procedures (adaptive tracking) might not have allowed, such as very small error bars, and the first psychophysical functions for forward-masked detection and discrimination. However, assurance of uniform subject performance requires repeated measurement, so that only one subject had the endurance to explore fine structure beyond $t=10$ msec in the first experiment.

Based on a two-channel model for intensity encoding, we expected to find a cusp in the behavioural threshold recovery function. None was observed. Thus, the masker may not have fully saturated the high-spontaneous-rate neurons, making it difficult to assess the separate contributions of the two populations of neurons. Alternatively, and more likely, the high-spontaneous-rate and low-spontaneous-rate neurons do not act as expected. Not previously seen, however, were periodic increases in threshold superimposed on the otherwise monotonic decline that agrees with the (somewhat sparse) literature. A rise in threshold at $t=1$ msec, which has been seen elsewhere, was followed by progressively smaller increases appearing as peaks in the plot of threshold vs. time. The oscillating detection threshold could be conceptually modelled as detection of a just-noticeable-difference (jnd) in a post-masker pulsatile neural noise, assigning a role to auditory Type II IHC afferents (in the nomenclature of Pfeiffer and Kim (1972)) that show exceptional pulsatile “ringing”. This model explained why the width of the psychophysical threshold function (in dB) appeared to oscillate out-of-phase to that of the thresholds.

The second experiment employed similar stimuli and procedures as the first one, so that only two subjects could be employed. Zeng, Turner, and Relkin (1991) had observed a mid-level rise (circa 40-60 dB SPL) in the magnitude of the just-detectable intensity-increment in a forward-masked probe tone, using stimuli similar in durations to those used by Relkin and Doucet (1991). In the latter study, an auditory probe’s masker-induced average relative neural threshold shift, in dB, was plotted vs. probe presentation time post-masker (Relkin and Doucet, 1991). The average relative increase of the probe’s detection threshold recovered to zero much more slowly for afferents with spontaneous rate less than or equal to 20 spikes/sec than for afferents with spontaneous rate exceeding 40 spikes/sec. Zeng et al. (1991) explained their result in terms of these differential recovery rates. It was hypothesised that under certain circumstances these differences caused the fibers of high spontaneous rate to act as a channel for the encoding of softer stimuli, and those of low spontaneous rate to encode for louder stimuli, as follows. The literature shows that when a forward-masker precedes a brief probe, the curve showing neural response vs. probe level shifts up-level. There is evidence that such shifting occurs for curves that plot afferent firing rate vs. level for 40 msec tones (Smith, 1977) as well as for curves expressing $N_t$ potential vs. level (Cacace and Smith, 1986). This shifting may persist even at 100 msec post-masker for low-spontaneous-rate neurons, so Zeng et al. chose that delay between masker offset and tone onset, so that high-spontaneous-rate units, but not low-spontaneous-rate units, should have recovered from the forward-masker. Intensity-increment thresholds for soft tones should
be relatively unaffected by the masker because the "soft" (high-spontaneous-rate) channel should have recovered from adaptation. For very high intensity tones, the "loud" (low-spontaneous-rate) channel has not completely recovered from adaptation, but the test pip is sufficiently above this channel’s threshold to permit good discrimination. Poor discrimination, however, should occur for intermediate-level pips. These are out of the lower channel’s range of good discriminability, and close to the "louder" channel’s threshold, which is presumably still elevated due to the lingering effects of the masker.

The "mid-level hump" found by Zeng et al. was for a 25 msec tone presented 100 msec after an intense masker. Others later found that this hump decreased in size as the time gap between masker and probe decreased, suggesting that masker-probe proximity allowed encoding of probe level by reference to the masker (Plack, Carlyon, and Viemeister (1995)). These recovery-rate and referential-coding models predict different changes in the size of the just-detectable intensity-increment, with changes in the masker-probe time gap. By employing three such time gaps (3 msec, 80 msec, and ∞) and a 20 dB pip/masker intensity ratio, both models could be tested simultaneously.

Thus the Zeng et al. model was directly tested in our second experiment, in which discriminability limens were measured for 2 kHz Gaussian pips succeeding 200 msec, 2 kHz maskers, set 20 dB lower in level. The just-detectable intensity-increments for the brief pip were independent of gap but differed by pedestal level, peaking at 70 dB SPL. To interpret in terms of the models, if the forward-masker was effective in shifting the stimulus-response curves (rate-vs.-level, or, here, N_t-amplitude-vs.-level), then both recovery-rate and referential-coding models fail. Only referential-coding is supported if the forward-masker was ineffectual (which is unlikely). Since the peaked discriminability pattern also occurs for clicks, a model was proposed that is analogous to that for clicks. That is, as brief tones shorten even further, the encoding of intensity change shifts from a dependence on firing rates to a dependence on the rate of growth of the number of responding afferents (= the rate of growth of the total stimulus-provoked spike count). Such rates are reflected by the slope of the plot of N_t.

In this model, a decrease in N_t slope, corresponding to a slower rate of growth in N_t, reflects the inflation of the intensity-increment threshold as probe levels approach 70 dB SPL. The slowdown in growth rate at intermediate probe levels can result from little or no overlap of two distributions of afferents with differing mean threshold (Radionova, 1963). Two such distributions would extend dynamic range for brief probes, as actually seen in the broad dynamic range of N_t. Now, in the above computations, we had already assumed two population distributions acting as channels, a concept applied to longer tones such as those of Zeng et al. These channels were the high- and low-spontaneous-rate neurons, as assumed in the Zeng et al. model of their mid-level hump. However, even Radionova’s model is contra-indicated by the lack of time-dependence in our just-detectable intensity-increments, since some time-dependence should have been
evident for efficacious masking.

Recall that the intensity-increment thresholds for brief pips were obtained by building psychometric functions. Presumed to underly these psychometric functions are Gaussian distributions, whose standard deviations $\sigma$ were plotted vs. the just-detectable intensity-increments for the two subjects and three conditions. The resulting six plots each revealed a linear relation. One subject's data showed jnd (just-detectable intensity-increment)=$1.28\sigma$, jnd=$1.52\sigma$, and jnd=$1.23\sigma$ respectively for the 3 msec, 80 msec, and nonmasked conditions. The other subject showed jnd=$1.1\sigma$, jnd=$1.08\sigma$, and jnd=$1.05\sigma$ for the same respective conditions. Thus, under present definitions and experimental conditions, there is a linear relation between the just-detectable intensity-increment and the parameter $\sigma$ of the fitted psychometric function.

Both of the present experiments were designed to look for signs of differential neural-recovery rates. The lack of the expected cusp in our own forward-masked pip-detection thresholds (and in detection thresholds from the similar but less detailed study of Kohlrausch, Püschel, and Alphei (1992)) seems to imply that the mid-level rise in discriminability seen by Zeng et al. (1991) does not depend on differential recovery rates of peripheral afferents, a much more recent conclusion reached by Zeng and others.

To conclude: the data revealed in this dissertation suggest that partitioning afferents by average threshold, and thus by spontaneous rate, does not explain auditory dynamic range for tones or brief forward-masked probes. What, then, is the solution to the "dynamic range problem"? This discrepancy between neural and behavioural dynamic range seemed to imply that afferent neurons could not be fully responsible for intensity encoding, an unlikely situation. The problem had to be a product of the way that dynamic range was studied, and indeed it resulted from too narrow a perspective. Ignored were other aspects of afferent behaviour beyond just the rate-level responses to brief stimuli presented in quiet. For instance, the forward-masked shifting of rate-level curves for tones (Smith, 1977), and for plots of $N_1$-amplitude-vs.-level (Cacace and Smith, 1986), suggests that neural adaptation provides a large portion of the ear's dynamic range in everyday listening, given that background stimuli are always present to act as a forward-masker. Another role of background stimuli may be to alter cochlear mechanics by way of olivo-cochlear feedback, which has the apparent capability to shift the afferents' operating range up- or down-level. Furthermore, shifting of neurons' dynamic range by forward-masking and olivocochlear feedback are probably not the same thing, since the former effect occurs naturally in anesthetised animals, whereas under anesthesia olivocochlear feedback has to be provoked by applied electrical stimulation of the olivocochlear bundle. That is, one effect is neural in origin, and the other is mechanical.

We close with a brief list of the "firsts" achieved in this thesis:

- the first use of the mean rate-intensity function that we are aware of;
- the first theoretical modelling of the discriminability offered by individual sloping-saturating units, and the discovery that their useful dynamic range is not likely to exceed that of the majority sigmoidal units;
- the discovery of the fine structure in early recovery of the auditory detection threshold;
- the first psychometric functions for forward-masked probe-detection thresholds and increment-detection thresholds;
- the first suggestion of a special role for the “Type II” inner-hair-cell afferents of Pfeiffer and Kim (1972);
- the discovery that the pattern of discriminability thresholds for Gaussian pips resembles that for clicks;
- the first quantitative comparison, that we know of, between the width of the psychometric function under forward-masking, and the size of the discriminability limen."
51. There is a trap set for the theorist who attempts to account for the concept of a “threshold” for driven (above-spontaneous-rate) firing when crafting rate-vs.-level equations. This concern which may explain the continuing difficulties that researchers have had in finding an equation that precisely describes the rate-level curve of a single neuron. This concern is that no equation can provide proper limiting conditions for both the case where driven activity ceases for sufficiently-low-intensity stimuli, as well as the case of an absence of driven activity in the complete absence of a stimulus. That is, for a threshold called $x_{\text{min}}$

$$r(x) \rightarrow r_s \text{ only as } x \rightarrow x_{\text{min}}$$

contradicts

$$r(x) \rightarrow r_s \text{ only as } x \rightarrow 0.$$  

Equivalently, no equation can include spontaneous rate as a sole remaining term in the limit of declining stimulus level, as done in many rate-level equations, and hope to exactly describe the [experimentally averaged] single-nerve firing rate in this limit. This argument is independent of the value of $r_s$ and does not depend on the random nature of a threshold. The italicised statements are broken only if $r(x_{\text{min}}) = r(0)$, that is, if driven firing occurs for any stimulus level. But neurons do not have infinite dynamic range, where for present purposes, dynamic range means that $r_s < r(x) < r_{\text{max}}$ (dynamic range being interpreted in such a way that it has its largest possible value). The statement $r(x_{\text{min}}) = r(0)$ cannot be true for a monotonic rate-level function such as characterises the experimentally-averaged spike count for an auditory primary afferent, taken over some fixed counting time. There exists some non-zero stimulus intensity below which the neuron is impervious to the stimulus. Since “threshold” is always used to refer to the lowest level at which the neuron starts to fire reliably above its spontaneous rate, threshold cannot be described precisely by a monotonic rate-level function. Neither can dynamic range, if dynamic range is defined partly in terms of threshold. Considering these theoretical restrictions, our logistic rate-level equation, with its upper and lower limits that do not quite match $r_{\text{max}}$ and $r_s$, may be the best compromise yet in describing auditory rate-level behaviour.

52. In the Introduction we introduced the concept of multiple cone mechanisms in visual response (Footnote 1). The different cone types are frequency-dependent so that a given single wavelength of light excites each type to a differing degree, because of differences in their light-absorbing pigments. It has been noted that these mechanisms may interact (DuCroz and Rushton, 1966; Eisner, 1986), so that rod-cone interaction may not be the only intensity-dependent interaction at the retina. To continue the analogy to hearing, interactions may occur for overlapping distributions of thresholds, for different spontaneous-rate groups or different types of rate-level functions.

53. The use of mean-rate functions to describe channel behaviour may pose a problem not found in whole-nerve spread-of-excitation models. In those models the stimulus intensity is reflected in the whole-nerve rate, which monotonically increases with rise in SPL. Thus loudness can arise by comparing the total rate to an internal standard, perhaps corresponding to spontaneous rate only. Comparing the loudnesses of two tones, then, consists of comparing their whole-nerve rates, automatically and without conscious control. However, when dealing with multiple channels, switching out internal attention from one channel to another, unconsciously and automatically, will not necessarily produce monotonic changes in firing rates; sudden large differences could occur. Two approaches can reduce the problem: either label the channels somehow so that rate differences are automatically compensated for, or combine channel outputs in some fashion that
produces monotonic rate change. Some sort of feedback loop might provide the latter scheme.

54. One might imagine that some kind of gating is possible at the IHCs, involving the gating-out of the activity of afferents from the inner hair cells by the efferent synapses, which are on the afferents themselves (see Spoendlin, 1975, cat). However, available evidence suggests that the LOC is not a dynamic-range-extending system. The LOC alone can be activated when the crossed OCB is cut and allowed to degenerate, and it affects N1 only on the stimulated side, a decrease equivalent to about a 6-7 dB drop in sound level (Desmedt and LaGrutta, 1963). A similar small effect was found by Sohmer (1966), who compared the effects of homolateral (ipsilateral LOC) vs. contralateral (crossed OCB) electrical stimulation on N1, finding that reduction of N1 was six times greater from contralateral neurons than from homolateral ones. These changes in effective SPL are more compatible with those caused by the middle-ear reflex than by the MOC system, implying that the LOC bundle may prove part of a startle system, a total auditory gating whose neural pathway is similar to that of the goldfish (see Lin and Faber, 1988). In this species, giant cells which are part of the afferent pathway will, with a single neural spike, activate inhibitory efferent nerves which have endings on the primary afferents, postsynaptic to the hair cell. This fast action (a latency of 6-8 msec) temporarily prevents auditory response by turning down the gain. Such a role for the cat LOC system cannot be discounted, since technical problems have forbid confirmed recordings from single LOC efferents (Liberman and Brown, 1986; Brown, personal communication). We may speculate that the system may not fire unless the animal is startled, a condition usually tested in awake animals.
CHAPTER 8

SUGGESTIONS FOR FURTHER WORK
IDEAS

Here we describe several studies suggested by (and relevant to) the computations and experiments already described. These studies regard 1) ways to improve and extend the dynamic range computations, 2) improved definitions for quantities used in those computations, 3) further exploration of detection thresholds under non-simultaneous masking, 4) further exploration of auditory discriminability, and 5) further exploration of the psychophysical functions describing detection and discrimination.

1a. Can the computations of Chapter 3 be improved or extended? The parameter $\rho$

In the computations of Chapter 3, it might be objected that certain choices of parameter values were inappropriate. For example, we accepted the dictum that the mean-to-variance ratio has value $\rho = 1.5$ for all units (Teich and Khanna, 1985). However, others have used higher values of $\rho$. For example, Viemeister (1988) chose $\rho = 2$ for his ideal computations of dynamic range, although he did not do neural recordings, and his choice of $\rho$ was left unexplained. Delgutte (1987) did do actual recordings, counting means and variances in response to 50 msec CF tones from 229 fibers from 17 cats. His data-analysis and conclusions differed from those of Teich and Khanna. First, the fibers were separated according to spontaneous-rate group by Liberman's (1978) classification (see Chapter 2), and then the averages of the spike-count variances across-fibers were plotted vs. the averages of the mean count, for each $r_i$ group. The curves for each $r_i$ group closely overlapped, but mean count increased more rapidly than variance, such that the variance-to-mean ratio appeared to asymptote. Thus the mean-to-variance ratio $\rho$ was always $> 1$, reaching at least 3 and possibly going higher.

We noted earlier (Chapter 3) that the jnd $\Delta x$ varies as the square root of $\rho$ and that consequently the change in computed dynamic range is small. But complacency may be dangerous. The proper choice of mean-to-variance ratio remains unclear, and may have a larger effect than expected if it is frequency-dependent, intensity-dependent, or dependent on fiber characteristics such as spontaneous rate. For example, for low-$r_i$ units, fiber mean-to-variance ratio may depend on stimulus level above fiber threshold, that is, relative stimulus level $x - \varepsilon$ (Teich and Khanna, 1985, Fig.7). By re-evaluating Equations 9-13 using a new integration variable $r_i/\rho$ we can eventually improve accuracy by using the same function $\rho(x, \varepsilon)$ for each possible fiber $i$ implied by the family of rate-level equations $r_i (c, \varepsilon, A, r_{max}, r_{sat}, x)$ (Equations 1 and 3), when $\rho(x, \varepsilon)$ is clearly quantified.

Thus the choice of $\rho$ is one area in which the accuracy of our dynamic-range computations might have to be improved. Figures 64 and 65 imply that the right $\rho$ might be crucial for dynamic-range computations involving very small populations. Here the neural intensity difference limens shown by Delgutte (1987, Fig. 2a) for two illustrative fibers, one sigmoidal and one sloping-saturating, are replotted.
Figure 64. The computed neural discriminability function $\Delta x$ (Equation 13) for a population of 1. This jnd is shown by the smooth curves. The joined symbols are jnds obtained by Delgute (1987) by applying a 2AFC procedure to neural firing, for a low-spontaneous-rate, high-threshold unit. Spontaneous rate was set to zero. a. The computed jnd for a mean-to-variance ratio $\rho=1.5$. b. The computed jnd for a mean-to-variance ratio which is set relative to the fiber's threshold. For $x-e < 5$ dB, $\rho=0.1$; for $5 < x-e \leq 10$ dB, $\rho=0.5$; for $10 < x-e \leq 20$ dB, $\rho=0.75$; $\rho=1.5$ for $20 < x-e$. c. The computed jnd using $\mu(r) = 5.94 \times 10^{-3} r - 6.24 \times 10^{-5} r^2 + 2.78 \times 10^{-7} r^3 - 4.08 \times 10^{-9} r^4$ for single-fiber (mean) rate $r$ in spikes/sec, fitted by least-squares to plots of variance vs. mean published by Delgute (1987). This fitted $\mu$ extrapolates to zero when neural firing rate is zero, as inferred from Delgute (1987) and Teich and Lachs (1985). d. The computed jnd when $\mu(r) = (5.94 \times 10^{-3} r - 6.24 \times 10^{-5} r^2 + 2.78 \times 10^{-7} r^3 - 4.08 \times 10^{-9} r^4)^{3.3}$. 
The fibers' published rate-level plots were digitised, then fitted to a logistic function (Equation 1) and a double-logistic (Equation 3), respectively. Fiber discriminability was then computed using the equation for the just-noticeable-difference (Equation 13) with population size $m$ set to unity. The value of the mean-to-variance ratio was first set to 1.5 as used for the computations in Chapter 3. The resulting curves for the theoretical jnds are narrower than the measured ones, and have notably higher minima for the sigmoidal fiber (Figure 64). Then a graduated set of discrete values $\rho<1.5$ were employed for sound levels less than 20 dB above fiber threshold. This produced a steeper increase in jnd at the lower limits of discriminability for the sloping-saturating unit, and no visible change for the sigmoidal unit. For an improved fit, Delgutte's plot of variance vs. mean (Delgutte, 1987, Fig. 6b) was fitted to a fourth-order polynomial. This polynomial, chosen only for its algebraic simplicity, gives $\rho$ as a function of firing rate $r$, which is the "mean" rate for a single unit as reported in the literature. This new $\rho$ was then substituted into the jnd equation, producing only a slight improvement in fit. Nonetheless the Delgutte measurements and the computed jnd curves can be brought into reasonable conjunction if $\rho$ is first phrased in terms of driven rate $(r-r_d)$ rather than firing rate $r$, and then raised to the exponent 2.5.

Even here caution is called for, because the psychophysical techniques applied to nerve fibers by Delgutte (1987) and Reklin and Turner (1988) must contain hidden assumptions about how neurons discriminate. Thus their data might well be a best resolution, beyond that which the auditory system actually obtains from the fiber. How peripheral response is actually used will be revealed only through further physiological investigation of the cochlear nucleus and higher centers.

1b. Can the computations of Chapter 3 be improved or extended? Mutually-dependent parameters

In Chapter 7 we noted that in some species the independence of single-fiber firing properties such as threshold and dynamic range cannot be guaranteed. Continuing the notation of Footnote 9: when dealing with non-independent variables, the mean rate-level function must deal with the joint probability distribution $f_x$ that describes the $n+1$ -dimensional distribution function for those $n$ variables. Thus

$$E[G(X_1,\ldots,X_n)] = \int \ldots \int G(x_1,\ldots,x_n) f_x(x_1,\ldots,x_n) \, dx_1,\ldots,dx_n. \quad (28)$$

gives the mean rate-level function for non-independent parameters $\epsilon$, $\lambda$, $r_{\text{max}}$, and $r_s$ when $G$ is the single-fiber firing-rate $r=r(x; \epsilon, \lambda, r_{\text{max}}, r_s)$ and $\{X_1,\ldots,X_n\} = \{\epsilon, \lambda, r_{\text{max}}, r_s\}$. 
Figure 65. The computed neural discriminability function $\Delta x$ (Equation 13) for a population of 1. This jnd is shown by the smooth curves. The joined symbols are jnds obtained by Delgutte (1987) by applying a 2AFC procedure to neural firing, for a high-spontaneous-rate, low-threshold unit. a. The computed jnd for a mean-to-variance ratio $\rho = 1.5$. b. The computed jnd for a mean-to-variance ratio which is set relative to the fiber's threshold. For $x < 5$ dB, $\rho = 0.1$; for $5 < x < 10$ dB, $\rho = 0.5$; for $10 < x < 20$ dB, $\rho = 0.75$; $\rho = 1.5$ for $20 < x$. c. The computed jnd using $\rho(r) = 5.94 \times 10^2 r - 6.24 \times 10^4 r^2 + 2.78 \times 10^6 r^3 - 4.08 \times 10^9 r^4$ for single-fiber [mean] rate $r$ in spikes/sec, fitted by least-squares to plots of variance vs. mean published by Delgutte (1987). This fitted $\rho$ extrapolates to zero when neural firing rate is zero, as inferred from Delgutte (1987) and Teich and Lachs (1985). d. The computed jnd when $\rho(r) = (5.94 \times 10^2 r - 6.24 \times 10^4 r^2 + 2.78 \times 10^6 r^3 - 4.08 \times 10^9 r^4)^{1/3}$, and (dashed lines) when $\rho(r) = (5.94 \times 10^2 r - 6.24 \times 10^4 r^2 + 2.78 \times 10^6 r^3 - 4.08 \times 10^9 r^4)^{1/3}$.
1c. Another use of the computational model for dynamic range

The mean-rate computation method of Chapter 2 may prove useful for any situation in which a rise in stimulus level is represented in the activity of many neurons. Certainly the “dynamic range problem" exists in other neural systems that are characterised by variety in neural thresholds and dynamic ranges. An example is the system that apparently regulates the wake-sleep cycle. The suprachiasmatic nucleus (SCN) in the rat, hamster, and human is a circadian pacemaker whose activity entrains (synchronises) to environmental light cycles. This luminance may vary over 10^4-10^5 lux (units of luminance) in a daily cycle. However, thresholds vary between light-sensitive SCN neurons in the rat, and similarly in the hamster, and are of the order of 0.1-1 lux, with saturation occurring at 1-3 logarithmic units above threshold (Groos and Meijer, 1985). This restricts the SCN to being sensitive to the luminance changes that occur at dawn and dusk (Meijer, Groos, and Rusak, 1986). Nonetheless the dynamic range of the system can be extended beyond that of any given cell by normalising and averaging the responses of several cells (ibid., Fig. 7). As the fitted rate-level curves for single cells are sigmoids much like those of Figure 13, then once sufficient sample sizes for neural parameters become available, the mean-rate method of Chapter 2 could be modified to describe the presumed ability of the SCN to signal a broad luminance range.

2. A study of threshold criteria for the rate-vs.-level response of single neurons

It is imperative to establish appropriate criteria for both neuronal threshold and dynamic range (Chapter 7). One way to do this is to test proposed definitions of threshold and dynamic range, comparing the predictions to some set standards of performance. For example, one standard to be met, when attempting to define “threshold", is that units of differing spontaneous activity must be treated consistently (Chapter 7). (In this respect, the method of Relkin and Pelli (1987) cannot be accepted until this freedom from bias is demonstrated.) One might test criteria by applying them to actual neural recordings, but in the absence of fresh recordings, the investigator must use either digitised rate-level curves, or simulate rate-level curves using an equation for single-nerve firing.

A standard for dynamic range might be set by assuming effective dynamic ranges using Figures 33-35, which can then be used to test existing measures of fiber dynamic range to see which, if any, come close. Note that, because useful dynamic range is not likely to start at a fiber’s threshold, a standard for the useful dynamic range need not involve whatever definition of threshold turns out to be best.

3. A study of backward-masked thresholds

The obvious control experiment for the present work, as used many times elsewhere, is a backward-
masking experiment, in which the probe precedes the masker (all other conditions remaining the same). Time was not available for such an experiment, and there were no reasons to suspect different results from those found elsewhere, but backward masking is still worthy of note. Psychophysical backward masking is mainly peripheral (Duifhuis, 1973) and appears to have 2 phases, an initial phase representing overlap between probe-induced ringing and the masker, and a slower phase of very gradual recovery attributed to central processes (Dolan and Small, 1984). Threshold elevation is absent previous to 30-40 msec pre-masker and is substantial only within about 4 msec (Scharf and Buus, review, 1986). Backward masking is largely absent previous to 10 msec pre-masker for a variety of masker and probe stimuli (Duifhuis, review, 1973) including long masker tones and short probe pips (Fastl, 1976). Only the initial phase of backward masking occurred even in an implant wearer (Blamey and Dooley, 1993) implying that the later component is weak.

The backward-masking data suggests that such a control experiment with the present stimuli would be useful given that neural fatigue would be mostly or completely absent, thus helping uncover the long oscillation pattern if present, because it would not be superimposed on a background threshold elevation caused by fatigue. Such oscillation could conceivably result from the probe pip itself, which then would have extended the oscillatory threshold pattern to that actually seen in Figures 39-41. If the oscillations are not present, we can feel sure that the pip does not provoke these oscillations by itself.

If there occurs an actual long-term oscillation in backward-masked pip threshold, beyond the 0-5 msec span (Dolan and Small, 1984) over which responses to masker and probe overlap, it may imply a central rather than peripheral source of all long-term pip-threshold oscillations, as was implied when the mid-level hump in discriminability was found for backward as well as forward masking (e.g. Plack, Carlyon, and Viemeister, 1995). That is, some sort of integrative process occurs at some level above the cochlea. Backward masking may indeed be influenced by oscillation, given that even clicks create backward masking of clicks (Raab, 1961; Chistovich and Ivanova, 1959), an effect presumably caused by ringing. This masked threshold drops steeply over t = -10 to t = 0 msec, more steeply than the detection thresholds of forward-masked clicks over a comparable time (Raab, 1961; Chistovich and Ivanova, 1959), and threshold shift persists at a low level for at least 100 msec thereafter (Chistovich and Ivanova, 1959).

4. Various discriminability studies

Another possibility for further work is to measure the intensity-increment threshold after lengthening the standard deviation σ of our Gaussian-shaped test pip, giving gentler ramping and hence a narrower frequency spectrum. We can do this without altering the pip's energy (which would complicate interpretation of the results), because maintaining the pip's level preserves the same total sound energy per pip (see Glossary) within some presumed sensory integration time >> σ. Changing σ allows us to investigate
whether restricting the spectral spread of the pip, and hence the number of new fibers recruited, does in fact change the discriminability limen within the short ramping durations that provoke the CAP. There is a problem here, in that two processes may be at work that produce the same effect. Reducing the rate of rise of the stimulus’ amplitude may reduce the synchrony of the initial neural bursts and hence reduce \( N_i \), whereas narrowing the spectral spread should reduce the size of the responding population for a given sound level, also reducing the amplitude of \( N_i \).

In Chapter 6, we proposed that lengthening a tone pip beyond 1 or 2 msec would eventually allow a rate code to be available for the auditory system’s use. A code based on population size would also be present due to the simultaneous initial firing of the responding population, whose size would be roughly represented by the amplitude of the \( N_i \) potential. This code might compete with the rate code until lengthening the pip’s ramping causes the \( N_i \) to disappear. We can study the effect of this putative competition by obtaining discriminability limens for tones with plateaus long enough to provide a rate code, say, plateaus of 5 msec or more. These tones would have Gaussian ramps (not Gaussian envelopes) whose \( \sigma \) can be lengthened as mentioned above, preserving the same energy in the ramping itself. By altering the ramping (of durations \( \leq 5 \) msec) while keeping the plateau duration constant, and vice versa, we can fathom the contribution of the \( N_i \) code and the rate code. Furthermore, we can in principle repeat the investigation by measuring \( N_i \) in awake humans (see the references in Footnote 32).

There are other discriminability studies that might be worth pursuing. As the pip intensity approaches very low pedestal levels from higher pedestal levels, the intensity-increment threshold should rise (Figure 62) even as variability falls. Generally, in the absence of forward maskers, increments get larger as low SPLs are approached for 1 kHz tones (and also white noise) of much greater length than the pips used here (Scharf and Buus, 1986). These longer tones had gentle ramping, long enough that \( N_i \) was not likely to be available as a possible aid to discriminability (see Footnote 32). This suggests a common very-low-intensity mechanism that serves both of the putative stimulus-intensity codes, the rate code and the population code.

The above suggestions lead to another possible study. As absolute threshold is approached, the detected increment should grow in size, as known experimentally (e.g. Jesteadt, Wier, and Green, 1977). As the pedestal level drops below absolute detection threshold, the detectable increment should equal the difference between the [below-threshold] pedestal level and the absolute threshold. Eventually, as the pedestal continues to drop, this jnd will approach a constant value, the value found in the complete absence of sound. In this particular model, then, the intensity-increment threshold for above-threshold pedestals should not approach infinity, but should instead not exceed some finite value, perhaps set by neural noise. Even if an infinite intensity-increment threshold is not seen in the limit as the pedestal approaches the
absolute detection threshold, absolute threshold itself may still be a perceptual discontinuity, with the just-detectable intensity-increments obtained from above-absolute-threshold pedestals approaching different values from those obtained from below-absolute-threshold pedestals, as the pedestal level approaches absolute threshold from opposite directions. We are unsure whether the just-detectable intensity-increment has been computed theoretically for this intensity regime, or whether this discontinuity hypothesis has yet been investigated. If so, it has most likely been done for tones, or white noise, having the 5-35 msec plateau levels that are typical of most published work. It may be illuminating to see what happens when using a brief Gaussian pip, which, as we have argued, presumably encodes intensity change through a spike-population code rather than a firing-rate code.

There also has been very little investigation of how the intensity-increment thresholds behave at pedestal levels of 100 dB SPL and higher. Apparently such investigations have been disallowed because of fears that repeated exposure, such as in 2IFC tasks, could cause auditory damage. These concerns may be exaggerated, and there may now be sufficient accumulated data on temporary threshold shifts so that experiments can be designed that pose no threat of damage to the subjects.

Even the results of Chapter 6 can be improved upon. A possibility raised in Chapter 6 is that the elevation in just-detectable intensity-increments seen in Figures 50 and 51 is higher than already evident, due to undersampling of the curve. Thus we do not know the precise position and shape of the peak in that curve. The curve must be investigated in finer detail by evaluating the just-detectable intensity-increment using many more pedestal levels in the range of 60-80 dB SPL, say, at 2 dB increments along the dB SPL axis.

5. What is the probability distribution function underlying either neural or psychophysical performance on 2IFC tasks?

Inspection of the psychophysical functions for detection and discrimination discovered in the two Experiments (e.g. Figures 42, 53, and 54) suggest that at the intensities below the level corresponding to \( p(c) = 0.75 \), detection and discrimination performance as measured by \( p(c) \) can drop off more quickly than expected for an underlying Gaussian distribution. This pattern is not rare in the present work, and can be seen for psychometric functions elsewhere in the literature. This pattern cannot be described by the integrals of symmetric probability density functions such as the Gaussian or Weibull distributions (see Footnote 29), but must reflect skewed density functions such as the extreme-value distribution (see caption to Figure 18). A variety of such distributions should be fitted to our psychometric data, whose sheer bulk - a total of 164 psychometric functions - provides a rare opportunity for examination. Neurometric functions, whenever they become available in sufficient numbers, should be similarly examined so that any common underlying
density function can be revealed.

Another feature of our psychometric functions that we have not yet remarked upon is the sometime appearance of a brief upward surge in $p(c)$ scores during the downturn mentioned above, usually at the second-lowest pip level used, just before the drop to chance performance. It is hard to tell whether this apparent change is a chance occurrence or whether it is due to some aspect of procedure, such as an actual improvement in perceptual power. Examination of fresh performances in the ranges of low relative pip levels corresponding to $0.5 \leq p(c) \leq 0.75$ might resolve this issue.
55. It is interesting that the oscillations in threshold (Figures 39-41) and in the standard deviation of the psychometric function for threshold (Figure 43) graphically resemble an interference pattern produced in a 2IFC task for detection of a gap in the middle of a sinusoid, with a continuous sinusoid as comparison (Shailer and Moore, 1987). When the gap starts at the end of a wavelength and is restarted with zero phase shift after an integer multiple of half-wavelengths $\frac{1}{2}f$, $p(c)$ scores oscillate at $\frac{1}{2}f$. When ringing during the gap meets the continuing sinusoid, it is either in phase or in antiphase, reducing or enhancing the gap with declining effect as the gap widens. This pattern is not as clear for $f=2$ kHz as for lower frequencies, where ringing persisted for at least 8 msec at 400 Hz; but some oscillation still occurs for 4 msec, the ringing duration inferred from the width of the auditory filter (Carlyon, 1988). Perhaps subsequent ringing would improve pip detection by providing a higher basilar oscillation to the probe, and hence explain the steep drop in early recovery beyond the limit of small jnds ($1 \leq t \leq 5$ msec). (For the neural equivalent of Moore and Glasberg's detection thresholds, see Moller, 1970, Fig. 1 [brief clicks]).
REFERENCES


Otolaryngologica, 71, 262-265.
Electroencephalography and Clinical Neurophysiology, 7, 399-406.


Kupperman, R. (1970) ON and OFF responses as measured in the cochlea of the Guinea pig. *Journal of the
Acoustical Society of America, 47, 518-524.


3, 455-472.


Palmer, A.R., and Evans, E.F. (1979) On the peripheral coding of the level of individual frequency components of complex sounds at high sound levels. In O. Creutzfeld, H. Scheich, and Chr. Schreiner (Eds.), *Hearing mechanisms and speech* (pp. 19-26). Heidelberg: Springer-Verlag.


Perez, M., and Coro, F. (1985) Physiological characteristics of the tympanic organ in noctuoid moths. II. Responses to 45 ms and 5 s acoustic stimuli. *Journal of Comparative Physiology, 156*, 689-696.


Acoustical Society of America, 92, 3097-3101.


Research, 82, 31-43.


Warr, W.B. (1972) Fiber degeneration following lesions in the multipolar and globular cell areas in the ventral cochlear nucleus of the cat. *Brain Research, 40*, 247-270.


Recall the mean-rate equation (Equation 6)

\[ \bar{r}(x) = \int_{e=e_k}^{e_{max}} \int_{\lambda=\lambda_{min}}^{e} \left[ \frac{r_{max} - r_s}{1 + \frac{100-c}{c} \exp \left\{ -2 \left( \frac{x-e}{\lambda} \right) \ln \left( \frac{100-c}{c} \right) \right\} } + r_s \right] p(e) p(\lambda) d\lambda d\lambda. \]  

(A1)

With a criterion \( c=2 \),

\[ \bar{r}(x) = \int_{e=e_k}^{e_{max}} \int_{\lambda=\lambda_{min}}^{e} \left[ \frac{r_{max} - r_s}{1 + 49e} \right] p(e) p(\lambda) d\lambda d\lambda. \]  

(A2)

**General solution for the integral involving threshold \( e \) only**

The indefinite integral over threshold alone is

\[ \int \frac{c-a e}{1+49Be} d\epsilon = \lambda \int \frac{c-a \lambda \ln \left( \frac{u-1}{49B} \right)}{k} du, B = e^{-\frac{k}{\lambda}}, u = 1 + 49Be^{\frac{k}{\lambda}}, \epsilon = \frac{\lambda}{k} \left( \frac{u-1}{49B} \right). \]  

(A3)

Here \( B \) can be treated as a constant for purposes of integration. Now the integral becomes

\[ \frac{\lambda}{k} \int \frac{c+\frac{a \lambda}{2} - ax - \frac{a \lambda}{k} \ln(u-1)}{u(u-1)} du = \ln(49B) = \frac{k}{2} \frac{kx}{\lambda}, \text{ also } u>1 \Rightarrow (u-1)>0 \]  

(A4)

which in turn becomes

\[ P \int \frac{1}{u(u-1)} du = Q \int \frac{\ln[u(u-1)]}{u(u-1)} du \text{ where } P = \frac{\lambda}{k} \left( c + \frac{a \lambda}{2} - ax \right), Q = a \left( \frac{\lambda}{k} \right)^2. \]  

(A5)

Here \( P \) and \( Q \) act as constants, and the integral simplifies to
and thus to

\[ P \ln \left( \frac{u-1}{u} \right) - \frac{Q}{2} \ln^2(u-1) + Q \int \frac{\ln(u-1)}{u} \, du = \frac{1}{u(u-1)} - \frac{1}{u} \]  \hspace{1cm} \text{(A6)}

\[
\frac{1}{u(u-1)} = \sum_{n=1}^{\infty} \frac{1}{u^n} \]  \hspace{1cm} \text{(A8)}

(multiply both sides by \(u-1\) for proof) so that

\[ \ln(u-1) = \int \frac{1}{u-1} \, du = \sum_{n=1}^{\infty} \int \frac{1}{u^n} \, du = \ln u - \sum_{n=1}^{\infty} \frac{1}{nu^n} \]  \hspace{1cm} \text{(A9)}

which gives

\[
\int \frac{\ln(u-1)}{u} \, du = \int \frac{\ln u}{u} \, du - \sum_{n=1}^{\infty} \int \frac{1}{nu^{n+1}} \, du = \frac{\ln^2 u}{2} + \sum_{n=1}^{\infty} \frac{u^{-n}}{n^2}. \]  \hspace{1cm} \text{(A10)}

Note that we have omitted the constant of integration that must appear in any indefinite (bound-less) integral.

Solution for the mean-rate equation

The mean-rate equation, which has stated bounds of integration, can now be solved. To continue,

\[
\int \frac{e^{-ae} \, de}{1+49Be} \]  \hspace{1cm} \text{(A11)}

\[
= P \left( \frac{k}{2} - \frac{kx}{\lambda} + \frac{k \in}{\lambda} \right) - P \ln \left( 1 + 49Be \frac{ke}{\lambda} \right) + \frac{Q}{2} \ln^2 \left( 1 + 49Be \frac{ke}{\lambda} \right) - \frac{Q}{2} \left( \frac{k}{2} - \frac{kx}{\lambda} + \frac{k \in}{\lambda} \right)^2 + Q \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + 49Be \frac{ke}{\lambda} \right)^{-n} \]

\[
= -\frac{a}{2} e^2 + c e - P \ln \left( 1 + 49Be \frac{ke}{\lambda} \right) + \frac{Q}{2} \ln^2 \left( 1 + 49Be \frac{ke}{\lambda} \right) + Q \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + 49Be \frac{ke}{\lambda} \right)^{-n} + c \left( \frac{\lambda}{2} - x \right)^2 + c \left( \frac{\lambda}{2} - x \right) \]

where again the constant of integration has been ignored. When integrating over definite bounds for \(e\), the last two...
terms act as constants and, along with the constant of integration, they will vanish. Thus

\[
\int_0^{55} \frac{c - ae}{ke} \, de = -\frac{a}{2} \cdot 55^2 + 55c - P \left[ \ln \left( 1 + 49Be^{\frac{55k}{k}} \right) - \ln \left( 1 + 49B \right) \right] + \frac{Q}{2} \left[ \ln^2 \left( 1 + 49Be^{\frac{55k}{k}} \right) - \ln^2 (1 + 49B) \right] + Q \sum_{n=1}^\infty \left[ \ln^2 \left( 1 + 49Be^{\frac{55k}{k}} \right) - \sum_{n=1}^\infty \frac{1}{n^2} \left( \ln (1 + 49B) \right)^n \right].
\]

To investigate further we let

\[
m_0 = k(55 - x)
\]
\[
m_1 = -kx
\]
\[
g = \frac{a}{k^2} \quad (= 6.256 \cdot 10^{-6})
\]
\[
p_1 = \frac{c - ax}{k} \quad (= 2.435 \cdot 10^{-5})
\]

and noting that \(-\left(55^2 \cdot a/2 + 55c = 1\right)\), we now have the slightly different form

\[
\int_0^{55} \frac{c - ae}{ke} \, de = 1 - \left[ p_1 + p_2 \lambda^3 \left( \ln \left( 1 + 49e^{\frac{m_0}{k}} \right) - \ln \left( 1 + 49e^{\frac{m_1}{k}} \right) \right) \right] + \frac{q}{2} \left[ \ln^2 \left( 1 + 49e^{\frac{m_0}{k}} \right) - \ln^2 \left( 1 + 49e^{\frac{m_1}{k}} \right) \right] + q \lambda^2 \left[ \ln^2 \left( 1 + 49e^{\frac{m_0}{k}} \right) - \ln^2 \left( 1 + 49e^{\frac{m_1}{k}} \right) \right] + q \lambda^2 \sum_{n=1}^\infty \frac{1}{n^2} \left[ \ln^2 \left( 1 + 49e^{\frac{m_0}{k}} \right) - \ln^2 \left( 1 + 49e^{\frac{m_1}{k}} \right) \right].
\]

If the probability density function for threshold \(\epsilon\) is shifted along the \(\epsilon\) dimension, the limits of integration will change from \(0 \leq \epsilon \leq 55\), to \(T_1 \leq \epsilon \leq T_2\) where \(T_2 = T_1 + 55\). In that case

\[
c = \frac{a}{2} \left( T_1 + T_2 \right) + \frac{1}{T_1 + T_2}, \quad m_0 = k(T_2 - x), \quad m_1 = k(T_1 - x).
\]
Assuming infinite limits for the dynamic range distribution \( p(\lambda) \) (bounds which must be reduced to non-zero limits in actual computations), the portion of the mean rate-level function which is integrated over \( \varepsilon \) and \( \lambda \) is now

\[
\frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(\varepsilon-\lambda)^2} d\varepsilon \int_{0}^{1+49e^{1/\lambda}} e^{\lambda} \left( \ln \left( 1 + 49e^{\frac{m_i}{\lambda}} \right) - \ln \left( 1 + 49e^{\frac{m_i}{\lambda}} \right) \right) d\lambda
\]

(A16)

Since all probability density functions must integrate to unity, the first term is just "1". The fourth term can be examined to establish a crude boundary on its value. We first put in more realistic boundaries of \((0,\infty)\) for \( \lambda \) i.e. we assume that the amount of accumulated probability below \( \lambda=0 \) is negligible compared to the rest. We then note that

\[
\left| \frac{q}{\sigma \sqrt{2\pi}} \int_{0}^{1+49e^{1/\lambda}} \lambda^2 e^{-\frac{1}{2\lambda^2}(\varepsilon-\lambda)^2} \left( \sum_{i=1}^{\infty} \frac{1}{n^2} \left( 1 + 49e^{\frac{m_i}{\lambda}} \right)^{-n} - \sum_{i=1}^{\infty} \frac{1}{n^2} \left( 1 + 49e^{\frac{m_i}{\lambda}} \right)^{-n} \right) d\lambda \right| < \frac{q}{\sigma \sqrt{2\pi}} \int_{0}^{1+49e^{1/\lambda}} \lambda^2 e^{-\frac{1}{2\lambda^2}(\varepsilon-\lambda)^2} \left( \sum_{i=1}^{\infty} \frac{1}{n^2} \left( 1 + 49e^{\frac{m_i}{\lambda}} \right)^{-n} - \sum_{i=1}^{\infty} \frac{1}{n^2} \left( 1 + 49e^{\frac{m_i}{\lambda}} \right)^{-n} \right) d\lambda
\]

(A17)

since generally

\[
\left| \int_{a}^{b} f(\lambda) d\lambda \right| \leq \int_{a}^{b} |f(\lambda)| d\lambda.
\]

(A18)

We also note that

\[
1 + 49e^{\frac{m_i}{\lambda}}, 1 + 49e^{\frac{m_i}{\lambda}} > 1 \text{ hence } \left( \frac{1}{n^2} \right) \left( 1 + 49e^{\frac{m_i}{\lambda}} \right)^{-n} < \left( \frac{1}{n^2} \right)
\]

(A19)

so that
\[
\left| \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + \frac{1}{49e^{\frac{\mu}{\lambda}}} \right)^n - \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + \frac{1}{49e^{\frac{\mu}{\lambda}}} \right)^{n-1} \right| < \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + \frac{1}{49e^{\frac{\mu}{\lambda}}} \right)^n \leq \sum_{n=1}^{\infty} \frac{1}{n^2} + \sum_{n=1}^{\infty} \frac{1}{n^2} = 2 \left( \frac{\pi^2}{6} \right) = \frac{\pi^2}{3}.
\]

Finally,
\[
\left| \frac{a}{\sigma \sqrt{2 \pi}} \int_{\lambda_0}^{\lambda} \frac{1}{2} \left( 1 + \frac{\lambda - \mu}{\sigma} \right)^{\frac{\lambda - \mu}{\sigma}} \left( \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + \frac{1}{49e^{\frac{\mu}{\lambda}}} \right)^n - \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + \frac{1}{49e^{\frac{\mu}{\lambda}}} \right)^{n-1} \right) d\lambda \right| 
\leq \frac{a}{\sigma \sqrt{2 \pi}} \int_{\lambda_0}^{\lambda} \frac{1}{2} \left( 1 + \frac{\lambda - \mu}{\sigma} \right)^{\frac{\lambda - \mu}{\sigma}} \left( \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + \frac{1}{49e^{\frac{\mu}{\lambda}}} \right)^n - \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + \frac{1}{49e^{\frac{\mu}{\lambda}}} \right)^{n-1} \right) d\lambda 
= \frac{a}{\sigma \sqrt{2 \pi}} \int_{\lambda_0}^{\lambda} \frac{1}{2} \left( 1 + \frac{\lambda - \mu}{\sigma} \right)^{\frac{\lambda - \mu}{\sigma}} \left( \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + \frac{1}{49e^{\frac{\mu}{\lambda}}} \right)^n - \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 1 + \frac{1}{49e^{\frac{\mu}{\lambda}}} \right)^{n-1} \right) d\lambda.
\]

For values \( \mu=40.2 \) dB and \( \sigma^2=9.57^2 \) the latter term in Equation A21 becomes 0.003518. Thus the contribution of this term to the mean rate is at best \( \left[ (r_{max} - r_i) \right] \cdot 0.003518 \) which for \( r_{max}=200/\)sec and \( r_i=68/\text{sec} \) (high-\( r_i \) units) gives 0.464/sec and which for \( r_{max}=200/\)sec and \( r_i=0.4/\text{sec} \) (low-\( r_i \) units; Table 4) gives 0.702/sec. In both cases, the contribution to mean firing rate is negligible.

The portion of the mean rate-level function involving integration over \( \varepsilon \) and \( \lambda \) is now solved:
\[
\bar{r}_{max} = \int \frac{1}{\varepsilon + \lambda} \frac{\lambda^{\frac{\lambda - \mu}{\sigma}}}{1 + 49e^{\frac{\mu}{\lambda}}} \left( \ln \left[ \frac{\lambda}{2} \left( 1 + 49e^{\frac{\mu}{\lambda}} \right) \left( 1 + 49e^{\frac{\mu}{\lambda}} \right) \right] - p_2 \right) - p_1 \lambda \right] d\lambda.
\]

Some constants used here are: \( k=2 \ln 49=7.78364, \ a=3.79 \cdot 10^4, \) and \( c=2.86 \cdot 10^{-2} \) (for \( \varepsilon \) limits of 0 and 55 dB SPL), \( q=a/k^2=6.256 \cdot 10^4, \) and \( \beta^2=\beta k^2=2.435 \cdot 10^{-5}. \)

We have ignored the small probability step in \( \varepsilon \) that occurs between -5 and 0 dB SPL which is due to a small number of fibers with very low thresholds. For this constant probability density \( c_2=0.003652, \)
\[
c_2 \int \frac{1}{1 + 49e^{\frac{-\varepsilon \lambda}{\kappa}}} d\varepsilon = c_2 \varepsilon - c_2 \frac{\lambda}{k} \ln \left( 1 + 49Be^{\frac{\varepsilon \lambda}{k}} \right)
\]
(where again the constant of integration has been omitted) so that the portion of the mean rate-level function involving integration over $\varepsilon$ and $\lambda$ is now

\[
\frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{c_2}{\sigma} e^{-\frac{1}{2\sigma^2} (\varepsilon - \varepsilon_0)^2} d\varepsilon \int_{-\infty}^{\infty} \frac{c_2}{\lambda} e^{-\frac{1}{2\lambda^2} (\lambda - \lambda_0)^2} d\lambda = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{c_2}{\lambda} e^{-\frac{1}{2\lambda^2} (\lambda - \lambda_0)^2} \left[ 5c_2 - c_2 \frac{\lambda}{k} \ln \left( \frac{1+49e^{\frac{-\lambda}{\lambda_0}}}{1+49e^{\frac{-\lambda_0}{\lambda}}} \right) \right] d\lambda
\]

\[
= \frac{c_2}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\lambda} e^{-\frac{1}{2\lambda^2} (\lambda_0 - \lambda)^2} \left[ 5 + \frac{\lambda}{k} \ln \left( \frac{1+49e^{\frac{-\lambda}{\lambda_0}}}{1+49e^{\frac{-\lambda_0}{\lambda}}} \right) \right] d\lambda.
\]

Note that for large values of $\varepsilon$ (in dB SPL)

\[
\left( \frac{1+49e^{\frac{-\varepsilon}{\lambda}}}{1+49e^{\frac{-\varepsilon_0}{\lambda}}} \right) \to 1 \text{ so that } \ln \left( \frac{1+49e^{\frac{-\varepsilon}{\lambda}}}{1+49e^{\frac{-\varepsilon_0}{\lambda}}} \right) \to 0
\]

hence

\[
\frac{c_2}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\lambda} e^{-\frac{1}{2\lambda^2} (\lambda_0 - \lambda)^2} \left[ 5 + \frac{\lambda}{k} \ln \left( \frac{1+49e^{\frac{-\lambda}{\lambda_0}}}{1+49e^{\frac{-\lambda_0}{\lambda}}} \right) \right] d\lambda \to \frac{5c_2}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\lambda} e^{-\frac{1}{2\lambda^2} (\lambda_0 - \lambda)^2} d\lambda = 5c_2.
\]

The entire derivation up to this point is courtesy of Dan Schnabel, M.A. (Math.) circa 1992.
APPENDIX B

SIEBERT'S (1965) "IDEAL COMPUTER"

Theoretical values for discriminability emanated from Siebert's (1965) concept of the "ideal computer", which lead to an expression for the Weber fraction, and an optimum-combination rule for combining [the values of] random variables. Siebert recognised that in certain respects action potentials are random, or "stochastic". This nature limits detection and discrimination. Siebert imagined a computer making the best discriminations that can be had from auditory-nerve activity, for example from the whole-nerve activities produced by two successive tones of differing intensity. As Siebert notes, the computer's "decisions would be optimum in the sense that, for a given reliability of performance, no device (including the brain) given only this input could operate with a smaller difference in intensity" (Siebert, 1965, p.207). Siebert accepts that there is no reason the brain should be optimal at psychophysical tasks.

His model makes the following assumptions. No phase-locking is desired; hence sound frequencies are restricted to above 5 kHz. The time interval between successive neural spikes is a random variable. Tones are long enough that transients can be ignored (i.e. minimum frequency splatter). Adaptation is slight enough to be ignored. (This is usually insured in the kind of neural recordings that we employ, by omitting the first 10 msec or so from the repetition-averaged record of neural firing.) Individual nerve fibers are statistically independent of each other in their firing patterns. Finally, there are assumed to be multiple units of interest at any given basilar locus.

A fiber's firing is represented by a distribution function $p_i$ for interspike time $\tau$. Under the assumption that changes in a stimulus' frequency or intensity will change only the scale of $\tau$, so that the shape of the distribution $p_i$ remains the same, a bound can be put on the Weber fraction $\Delta$ detectable by the ideal computer 75% of the time in a two-interval forced-choice task:

$$\Delta^2 = \frac{1}{s^2 T \sum_{s \tau} G_i \left( \frac{\delta r_i}{\delta s} \right)^2}$$  \hspace{1cm} (B1)

(Siebert's notation) where $r_i = I/\tau_i$ is the unit's average firing rate, $s$ and $f$ are the tone's intensity and frequency, $x$ is the fiber's position along the basilar membrane, and the counting index $i$ is the unit's identifier at $x$. The $G_i$ are dimensionless constants ($1 < G_i < 2$) such that

$$G_i = \frac{1}{\sigma_i^2}$$  \hspace{1cm} (B2)

where $\sigma_i^2$ is the variance of $q_i$, a probability distribution with mean =1 defined such that

$$p_i(x;s,f;x) = r_i(s,f,x)q_i[\tau r_i(s,f,x)].$$  \hspace{1cm} (B3)

If the $q_i$ are gamma functions then "an optimum decision procedure [in discriminating two sounds] is simply to compare weighted counts of the total number of events during each burst" (Siebert, p.210). The change in total rate, $\delta r$, is assumed small ($\delta r/\tau$) so that $\sigma^2$ at $r$ is the same as $\sigma^2$ at $r+\delta r$ for the decision variable for each unit. The decision variable for a population is the sum of the individual decision variables and is assumed Gaussian. Siebert's jnd is optimal (a best boundary) in the sense that it is based on an optimal statistical test to decide whether $r$ or $r+\delta r$ was presented first. Siebert's equation depends on the condition that $rT$>>1, which he assumes is true of most fibers for counting times $T$>100 msec.

Siebert's optimisation focuses on log-likelihood ratios rather than spike counts. Siebert (1965, Eqn. 4) optimises the jnd for an assumed $d'=1$ by first adding the single-unit log-likelihood ratio over all units. The detectability index $Q=(d')^2$, assumed equal to unity, is a function of these sums. Compare this to the Hellman and Hellman (1990) approach; in the 2IFC paradigm, the Hellman and Hellman equation (Equation 13) compares total ensemble counts in the two intervals. (A $Q$ obtained by comparing counts in one interval with counts in the other will
always be less than or equal to that found using log-likelihoods (Siebert). Our own version of the Hellman and Hellman (1990) equation for the jnd can be obtained through Siebert. Switching to our own notation (intensity=x) and assuming a single basilar membrane locus corresponding to a single frequency,

\[ \Delta x = \frac{1}{\left( \sum_i \frac{G_i}{r_i(x)} \left( \frac{\delta r_i}{\delta x} \right)^2 \right)^{1/2}} \]  

(B4)

About G, Siebert notes that "G, measures the effect of the stochastic behavior of the primary units; if the firing pattern were regular, \( \sigma^2 \) would be small, G, would be large, and the limiting value of the DL [difference limen] would be small" (Siebert, p.209). As G falls within the experimental bounds of the single-unit mean-to-variance ratio (Teich and Khanna, 1985), we assume that \( G_r = \rho \), so that

\[ \Delta x = \frac{1}{\left( \sum_i \frac{\rho_i}{r_i} \left( \frac{\delta r_i}{\delta x} \right)^2 \right)^{1/2}} \]  

(B5)

In the limit that all m units behave as a single average unit f (the mean rate-intensity function), such that total average firing rate is mx, and with an assumed constant \( \rho \) (as per Hellman and Hellman, 1990), then

\[ \Delta x = \frac{1}{\left( m \frac{\rho}{r} \left( \frac{\delta r}{\delta x} \right)^2 \right)^{1/2}} = \frac{\sqrt{r}}{\sqrt{m \rho} \frac{\delta r}{\delta x}} \]  

(B6)

This is the Hellman and Hellman equation (Equation 13) for the intensity difference limen when \( d \approx 1 \). Thus Equation B6 is a single term of the Siebert jnd, which removes its optimality, since we are first summing the neural response and then computing the jnd instead of forming a continuous running best-jnd computation, for a given stimulus level and over the entire fiber ensemble.

In Hellman and Hellman (1990), the response of interest is not spike count over the whole nerve, but over some population which is said to be "optimally selected" but not otherwise specified. The Hellmans' model assumes that counts across channels (here, fibers) are uncorrelated (after Johnson and Kiang, 1976) and hence that the variance of the firing sum is the sum of the fiber variances. Also assumed are steady-state conditions (as per the Teich and Khanna (1985) data) and a constant spike-count mean-to-variance ratio \( \rho \) across fibers. Total spike count is an unweighted sum across individual units; unlike Siebert, no specific assumption is made about the spike count distribution function. However, Hellman and Hellman represent the change in mean count \( \Delta N \) resulting from a jnd, as \( \Delta N = d' \alpha(I) \), where \( \alpha(I) \) is the standard deviation of the count at intensity I. This implies the signal detection theory assumption of Gaussian distributions for spike count, and a common \( \alpha(I) \) for each distribution of total counts (\( N(I) \) and \( N(I+\Delta I) \)). This in turn implies a relatively small jnd \( \Delta I \), as Siebert assumed. Hellman and Hellman do not require \( rT > 1 \), as does Siebert, but such a limit might affect the magnitude of the ignored terms of the Taylor series in spike count N (see Footnote 13).
APPENDIX C

A COMPARISON OF OUR MODEL OF AUDITORY NEURAL DISCRIMINABILITY TO DELGUTTE’S (1987) MODEL

Our model and Delgutte’s enjoy similar statistics. His fibers’ mean levels $L_{\text{min}}$ for high, mid, and low-spontaneous-rate units and our mean fiber thresholds $\bar{L}$ closely correspond. Regarding populations, the proportions for high-, mid-, and low-$r_f$ units again closely correspond despite the differences in threshold criteria. Beyond this point there are fundamental differences in the way that we and Delgutte go about quantifying the ear’s discriminability. We average a logistic function of four parameters, obtaining the mean firing-rate function of a fiber population as if it were a single fiber. Delgutte at first averages the experimental plots of the single-fiber difference limens. The Delgutte distribution function over threshold is for $L_{\text{max}}$; ours is for actual fiber threshold. Delgutte did not investigate the effects of altering the means and variances of the distribution functions of his one stochastic parameter $L_{\text{min}}$ presumably because he was most interested in summation activity across channels, not within them. We did manipulate these aspects of threshold and dynamic range, as these two parameters appear explicitly in our single-unit logistic function. We also assumed that 50% of low-$r_f$ units were sloping-saturating and assigned them a higher average dynamic range. Delgutte used a fiber rate-level function (Sachs, Winslow, and Sokolowski, 1989) that fits these fiber’s firing curves well, and hence may not have felt obliged to make such a subdivision.

Delgutte generates difference limens algebraically from the three average rate-level functions, assuming that the outputs of separate fibers are independent, random variables (after Johnson and Kiang, 1976). Like Delgutte, we assume a neural count distribution for a single unit at any given dB level, and a $d = 1$. But the wide CF range that Delgutte uses includes units with CFs $< 1 \text{ kHz}$, that might show phase-locking. We avoid any potential role of phase-locking by using much higher frequencies.

How the single channel is handled also differs. Under a single-channel model, the performances of the three $r_f$ groups were combined under an optimum-combination rule favoured by others. We simply added the percentage-weighted mean-rate functions for the $r_f$ groups (Equation 4) - not an optimisation scheme - and then calculated discriminability according to Equation 13, an equation which by itself appears equivalent to Siebert’s optimised just-noticeable-difference under our own assumptions about neural behaviour (see Appendix B). For Delgutte’s second case, the pure-tone stimulus, the 76 parallel channels are each narrower than the single critical band to which our computations are restricted. Delgutte pools performances of three $r_f$ groups within each frequency-defined channel using an optimum-combination rule, then pools across channels again using the same optimal rule. At this point, our approach and Delgutte’s diverge so much that comparing predicted difference limens is no longer appropriate.

In Delgutte’s single-channel model for the broadband noise, high-$r_f$ fibers provide most of the discrimination power at low intensities, mid-$r_f$ units are most important at somewhat higher levels, and low-$r_f$ fibers at the upper 20-30 dB. In our model, which is most applicable to one critical band in width, high-$r_f$ units dominate the mean rate-intensity function for single-channel computations. The difference is presumably due to the optimum-processor model used by Delgutte. Delgutte’s whole-nerve computation for the 1 kHz tone contains a wiggle, giving two plateaux; our critical-band model does not. This difference is presumably due to across-basilar-membrane contributions between adjacent, frequency-restricted channels. For his 1 kHz tone, high-$r_f$ units have the lowest difference limens, about 0.04 dB, and low-$r_f$ have the highest.
APPENDIX D

COMPUTING THE DYNAMIC RANGES OF SINGLE NERVE FIBERS

We infer single-nerve dynamic range from a smooth curve expressing the fiber’s discriminability limens. This jnd curve for single fibers is obtained by first fitting either the single-logistic (Equation 1) or the double-logistic (Equation 3) to sigmoidal or sloping-saturating fibers, respectively. To find the width of the jnd curve, we first find its lowest point, by letting the derivative equal zero, and solving for the corresponding sound pressure level \( x \) in dB SPL. Recall that (Equation 13)

\[
\Delta x \ (dB) = \frac{d'}{r'(x)} \sqrt{\frac{r(x)}{m \cdot \rho \cdot f}}. \tag{D1}
\]

for mean rate \( \bar{r} \). Now for a population size \( m=1 \), a single-unit firing rate \( r(x) \), and \( \rho \) and \( \tau \) assumed constant, letting

\[
\frac{d(\Delta x)}{dx} = 0 \tag{D2}
\]

results in

\[
\frac{1}{2} = \frac{r(x)r''(x)}{[r'(x)]^2} \text{ where } r'(x) = \frac{d[r(x)]}{dx}, \quad r''(x) = \frac{d[r'(x)]}{dx}. \tag{D3}
\]

Now for sigmoidal units, from Equation 1,

\[
r(x) = \frac{r_{\max} - r_s}{1 + 49^{\frac{-2(x - \alpha)}{\lambda}}} + r_s. \tag{D4}
\]

By letting

\[
\alpha = \frac{1}{1 + 49^{\frac{-2(x - \alpha)}{\lambda}}}, \tag{D5}
\]

we can solve Equation D3 after some algebra, giving

\[
\frac{1}{2} = \frac{(1 - 2\alpha)[\alpha(r_{\max} - r_s) + r_s]}{\alpha(1 - \alpha)(r_{\max} - r_s)}. \tag{D6}
\]

Now by allowing \( r_{\max} - r_s = \delta \), and rearranging terms, we obtain

\[
-\frac{3}{2} \delta \alpha^2 + \alpha \left( \frac{\delta}{2} - 2 r_s \right) + r_s = 0. \tag{D7}
\]
which is the familiar quadratic equation, whose solution is

$$\alpha = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \text{ where } a = -\frac{3}{2}, b = \frac{5}{2} - 2r_s, c = r_s.$$  \hspace{1cm} (D8)

There is one real root for \(b^2 - 4ac = 0\) and two real roots for \(b^2 - 4ac > 0\). There must be at least one real root, since the minimum point on the jnd curve exists, and there is no maximum point. To obtain this minimum point, we must rewrite in terms of sound level \(x\), where equating the \(\alpha\)'s in Equations D5 and D8 and rearranging terms gives

$$x = \frac{\lambda}{2} \left[ 1 - \ln \left( \frac{2a}{2b \pm \sqrt{b^2 - 4ac}} \right) \right] + \varepsilon \text{ where } \frac{2a}{2b \pm \sqrt{b^2 - 4ac}} = \frac{-3\delta}{2r_s \pm \left( \frac{\delta}{2} + 2r_s \right)^2 + 2\delta r_s}.$$  \hspace{1cm} (D9)

Thus \(x\) has two real roots, only one of which can be the absolute minimum. Before solving for specific values of \(x\) we will turn our attention to the double-logistic, Equation 3:

$$r(x) = y \left( \frac{r_{\text{max}} - r_s}{1 - 2(e^{-1})^\frac{1}{\lambda_1}} + r_s \right) + (1 - y) \left( \frac{r_{\text{max}} - r_s}{1 - 2(e^{-1})^\frac{1}{\lambda_2}} + r_s \right).$$  \hspace{1cm} (D10)

This equation was developed to describe rate-level plots for sloping-saturating fibers. These units typically have spontaneous firing rates \(r_s\) that are small compared to saturation firing rates \(r_{\text{max}}\). Since \(r_s < 1\) spike/sec for these units, and since \(r_s\) appears as an additive but not a multiplicative term, we can safely set \(r_s = 0\). Using the notation

$$\alpha_1 = \frac{1}{1 + 49^{\frac{1}{\lambda_1}}}, \quad \alpha_2 = \frac{1}{1 + 49^{\frac{1}{\lambda_2}}}$$  \hspace{1cm} (D11)

we can solve Equation D2, obtaining

$$\frac{1}{2} = \frac{\left[ \gamma \alpha_1 + (1 - \gamma)\alpha_2 \right] (1 - \gamma) \alpha_1 (1 - 2\alpha_1) + (1 - \gamma) \alpha_2 (1 - 2\alpha_2)}{\lambda_1^2 \alpha_1 (1 - \alpha_1) (1 - 2\alpha_1) + \lambda_2^2 \alpha_2 (1 - 2\alpha_2)}.$$  \hspace{1cm} (D12)

An analytical solution was not apparent and attempts to solve using the Mathematica software failed. However, Equation D12 can be solved numerically. First the sought-after value, \(x\), is cast as a parameter to be established by least-squares regression of the function \(y = f(\zeta, \gamma, \alpha, \lambda_1, \lambda_2, r_{\text{max}})\) on the single data point \(\{\zeta, y(\zeta)\} = \{0, 1/2\}\):
We must be careful to avoid the false minima that can occur with any curvefitting routine, and bear in mind that the jnd curves of Figures 33-35 sometimes display a second minima on the upper branch of the jnd curve. Once the lowest value of \( x \) is found, which can be done to the high degrees of precision allowed by curvefitting routines, the corresponding \( \Delta x \) is calculated, and then a criterion number \( \eta \) of \( dB \) is added to \( \Delta x \). The resulting jnd value \( \eta \cdot \Delta x \) corresponds to 2 \( x \)-values, call them \( x_1 \) and \( x_2 \), each of which can be separately found by numerically solving Equation D1 as done above for Equation D12, that is, by solving

\[
y(\zeta) = \frac{d' \sqrt{\gamma \alpha} \cdot (1-\gamma) \alpha_2}{2 \ln 49 \sqrt{\rho \tau \lambda^2} \left[ \frac{\gamma \alpha_1 (1-\alpha_1) + (1-\gamma) \alpha_2 (1-\alpha_2)}{\lambda_1} \right]^2} \]

where \( \alpha_1 = \zeta + \frac{1}{1 - \frac{\lambda_1}{\lambda_1 - 1} e^{-\frac{\lambda_1}{\lambda_1 - 1}}} \), \( \alpha_2 = \zeta + \frac{1}{1 + 49} e^{-\frac{\lambda_1}{\lambda_1 - 1}} \). (D13)

for the two values of \( x \) satisfying \( \{ \zeta, y(\zeta) \} = \{ 0, \eta + \Delta x \} \). Here \( \alpha_1 \) and \( \alpha_2 \) are defined as in Equation D13. The width of the jnd curve for the criterion \( \eta \) is then \( x_1 - x_2 \).

This width can be found analytically for the single-logistic function (Equation D4) by working backwards from Equation D9. Alternatively, we can follow the steps used in determining the width of the double-logistic. That is, we can numerically solve for the minima on the jnd curve for the single-logistic (sigmoid), by least-squares regression of \( y(\zeta) \) on the single data point \( \{ \zeta, y(\zeta) \} = \{ 0, 1/2 \} \) where

\[
y(\zeta) = \frac{(1-2\alpha) \left( \frac{\alpha(r_{\max} - r_0)}{\alpha(r_{\max} - r_0) + r_0} \right)}{\alpha(1-\alpha)(r_{\max} - r_0)}, \text{ where } \alpha = \zeta + \frac{1}{1 + 49} e^{-\frac{\lambda_1}{\lambda_1 - 1}} . \]

(D15)

Following the steps for the double-logistic, \( x_1 \) and \( x_2 \) can now be found by solving

\[
y(\zeta) = \frac{d' \sqrt{\alpha(r_{\max} - r_0) + r_s}}{\sqrt{\rho \tau \lambda^2} \left[ \frac{2 \ln 49 \sqrt{\rho \tau \lambda^2} \left( \frac{\alpha(r_{\max} - r_0)}{\alpha(r_{\max} - r_0) + r_0} \right)}{\alpha(1-\alpha)(r_{\max} - r_0)} \right]} . \]

(D16)

for the two values of \( x \) satisfying \( \{ \zeta, y(\zeta) \} = \{ 0, \eta + \Delta x \} \).

Note that we have assumed a fixed mean-to-variance ratio, \( \rho \). It may turn out that \( \rho \) varies with intensity \( x \), as modelled in Figure 34c and Figures 64 and 65. If so, \( \rho = \rho(x) \), and the equations above must be altered or supplemented accordingly to include the derivatives of \( \rho(x) \).
**APPENDIX E**

**MATHEMATICAL EXPRESSIONS FOR SOUND INTENSITY AND DISCRIMINABILITY**

Although some papers are careful to refer to sound pressure (usually called \( p \) or \( P \)), others still speak in terms of sound intensity \( I \) where generally

\[
\text{sound intensity } I = P^2. \tag{E1}
\]

In particular, the unitless Weber fractions \( \Delta I/I \) are often reported as terms in algebraic expressions that are in units of decibels. To examine the relationship between the different ways in which Weber fractions are expressed, we let

\[
y = 10 \log_{10} \left( \frac{\Delta f}{I} \right), \quad z = 10 \log_{10} \left( 1 + \frac{\Delta f}{I} \right)
\]

where \( y \) is a common way of reporting the Weber fraction as decibels, and \( z \) is the actual just-noticeable-difference in dB. For example, we employ Equation E2 for \( z \) when converting from Taub’s (1969) Weber fractions (tabulated as \( \Delta I/I \)) to jnds \( z \) in dB (Figure 61). Returning to Equation E2, rearranging terms gives

\[
\frac{y}{10^y} = \frac{\Delta f}{I}, \quad 10^z = 1 + \frac{\Delta f}{I} = 1 + 10^y, \quad \text{so that } z(\text{as } dB) = 10 \log_{10} \left( 1 + 10^y \right). \tag{E3}
\]

Further, to examine the relationship between just-noticeable-differences in decibels as functions of \( \Delta I/I \) and of \( \Delta P/P \), we let

\[
I = \beta P^2, \quad \text{also } \Delta P = P_{\text{upper}} - P,
\]

where \( P_{\text{upper}} \) is sound level at the jnd and \( P \) is the pedestal on which the jnd is imposed, so that

\[
z = 10 \log_{10} \left( \frac{\beta P^2}{P_{\text{upper}}^2} - \frac{\beta P^2}{P^2} \right) = 10 \log_{10} \left( \frac{P_{\text{upper}}^2 - P^2}{P^2} \right) = 20 \log_{10} \left( 1 + \frac{\Delta P}{P} \right) \tag{E5}
\]

i.e. the jnd in terms of pressure \( P \) is equal to the jnd in terms of intensity \( I \). Hence when jnds are expressed as \( y \), the equivalent jnd in terms of pressure is just given as \( z \) by Equation E3.

Now, to compute just-noticeable-differences for the range of post-masker times \( t \) in which masker and probe overlap (Chapter 5), recall that the masker (normalised to an amplitude of 1) is given by

\[
M(t) = 0.84 \cos(2\pi 2000t) \sum_{i=1}^{201} A_i e^{-\frac{(t-2(0.0005)(i-1))^2}{2(0.0005)^2}}, \quad \text{where } A_i = 1 \text{ for } i = 1, 201, \quad A_i = 0.935 \text{ for all other } i. \tag{E6}
\]

To this masker is added the pip, whose amplitude is

\[
P(t) = 10^{\frac{97-2(0.0005)(i-1))^2}{2(0.0005)^2}}, \tag{E7}
\]

where \( dB_{1} = 97 \text{ dB SPL minus the tone-pip detection threshold} \).
Adding Equations E6 and E7 gives the complete waveform which the subject hears in one of the two intervals of the 2IFC task. Thanks to the probe's infinitely long Gaussian tails, the jnd has no unique duration, even though the masker and superimposed probe are always in-phase. Therefore some duration must be chosen over which the masker level is considered to be incremented by the tone-pip. We will chose $\pm n\sigma$, $n=1, 2, 3$, as these durations. The squared amplitude of the waveform over this duration is proportional to energy, so that the jnd can be defined as

$$jnd = 10 \log_{10} \left( \frac{\text{energy of masker plus pip}}{\text{energy of masker}} \right) = 10 \log_{10} \left( \frac{\int_{-M(t) + P(t)}^{M(t) + P(t)} dt}{\int_{-\infty}^{\infty} P(t)^2 dt} \right)$$ (E8)

where the constant of proportionality between squared waveform amplitude and energy has been set to unity, since it drops out of the calculations regardless of its value.
APPENDIX F

FRISHKOPF'S (1956) MODEL OF THE N, POTENTIAL

In modelling $N_1$ in terms of contributions made by individual primary auditory afferents, Frishkopf (1956) made an number of assumptions, namely that:

1. $N_1$ was the sum of action potentials of the primary afferents;
2. certain measurable aspects of the compound action potential vary directly with the number of active neurons, although Frishkopf concerned himself in particular with the peak-to-peak amplitude of $N_1$;
3. responses to successive 0.1 msec clicks were independent of each other (verified experimentally by Frishkopf);
4. a neuron has a threshold level $S_T$ of stimulus level $S$, below which it does not fire and above which it must fire;
5. $S_T$ fluctuates in time in a way only specified by the percentage of time (i.e. probability that) it will be found within a given stimulus interval;
6. the time characteristic of this fluctuation is small compared to the experimental interstimulus presentation time;
7. neurons have a refractory period after firing, during which firing is impossible; this is assumed to be less than 2 msec;
8. all units respond with the same amplitude.

All of these assumptions are still accepted today. Frishkopf dealt with neural populations under the further assumptions that:

9. a given population is a group of neurons with identical behaviour (same probability distribution $D(S)$ for each unit's threshold);
10. all units in a population fire synchronously, so that total amplitude in response to a stimulus is the sum of the amplitudes of all neurons responding.

Now the probability, at a given time, of finding a neuron's threshold $T$ between two stimulus levels $S_1$ and $S_2$ is

$$p(S_1 < T < S_2) = \int_{S_1}^{S_2} D(S) \, dS \text{ where } \int_0^{S_1} D(S) \, dS = 1, \, D(S) \geq 0.$$  \hfill (F1)

Hence the probability that a neuron will fire is the probability that it's threshold lies below $S$, which is

$$p(S) = p(T < S) = \int_0^S D(S) \, dS.$$  \hfill (F2)

$D(S)$ is itself an average over time obtained by repeated measurement of neural response, and hence all properties obtained from it are also averages; the response amplitude on a single trial cannot be predicted. The average number of units firing over repeated presentation of stimulus level $S$, and the respective average response amplitude, are

$$\bar{n}(S) = Np(S), \, \bar{A}(S) = r \bar{n}(S).$$  \hfill (F3)

Having found a general expression for response amplitude $A$, Frishkopf then asked what $A$'s standard deviation would be over repeated stimulus presentations at level $S$. Since a neuron’s probability of firing in response to a level $S$ is $p(S)$, then its probability of not firing is $(1-p(S))$. For a single population of $N$ identical, independent units, the probability
that $n$ fibers will respond is given by the binomial distribution, so that
\[
\bar{n} = Np(S), \sigma_n^2 = Np(S)(1-p(S)).
\] (F4)

At $p(S)=0$, no unit ever fires and hence variability is zero; at $p(S)=1$, all units always fire, and again variability is zero. Now amplitude $A=\bar{m}$, and $r$ does not vary. Hence
\[
\sigma_{A(S)}^2 = (A(S)-\bar{A}(S))^2 = r^2(n-n) = r^2 \sigma_n^2
\] (F5)

and from Equations F3 and F4,
\[
\bar{A}(S) = rNp(S) = A_{\text{max}} p(S)
\] (F6)

which using Equations F4 and F5 leads to the standard deviation in amplitude
\[
\sigma_A(S) = \frac{A_{\text{max}}}{\sqrt{N}} \left( \frac{\bar{A}(S)}{A_{\text{max}}} \right)^{\frac{1}{2}}.
\] (F7)

This reasoning can be extended to multiple populations. For $k$ populations, the respective constant response amplitudes are $\{r_1, \ldots, r_k\}$, $\{n_1, \ldots, n_k\}$ being the number of neurons actually firing out of total available neurons $\{N_1, \ldots, N_k\}$ whose respective units are ruled by distribution functions $\{D_1(S), \ldots, D_k(S)\}$ which are each normalised according to Equation F1. The total amplitude evoked by stimulus level $S$ is now
\[
A = \sum_{i=1}^{k} n_i r_i
\] (F8)

whose expectation, which is non-decreasing, is
\[
\bar{A}(S) = \sum_{i=1}^{k} n_i p_i(S), \text{ where } \bar{A}(S+\Delta S) \geq \bar{A}(S) \forall S, \therefore p_i(S+\Delta S) \geq p_i(S) \forall S.
\] (F9)

The variance of this total is
\[
\sigma_A^2(S) = (A(S)-\bar{A}(S))^2 = \left( \sum_{i=1}^{k} (A_i(S)-\bar{A}_i(S)) \right)^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} (A_i(S)-\bar{A}_i(S))(A_j(S)-\bar{A}_j(S)).
\] (F10)

For independent populations the terms for $i \neq j$ are zero so that
\[
\sigma_A^2(S) = \sum_{i=1}^{k} (A_i(S)-\bar{A}_i(S))^2 = \sum_{i=1}^{k} \sigma_{A_i(S)}^2
\] (F11)

where $\sigma_{A_i(S)}^2$ is given by Equation F7. There may be other sources of variability in the nervous system, assumed to be statistically independent, chiefly biological noise (in contrast to stimulus variability, which was negligible, and equipment noise). Hence total amplitude $A_T = A + A_\theta$ where $A_\theta$ is assumed to take positive and negative values with equal probability, so that
where $\sigma_0$ is assumed stimulus-independent.

Armed with these equations, Frishkopf measured the average $N_1$ potential and its variance, in several cats. Standard deviation $\sigma$ followed the anticipated rise and fall between about 0 dB SL (human thresholds) and 35 dB SL (-60 dB on Frishkopf's scale). This is the shallower portion of the $N_1$ curve, which allows computation of the $N_1$ variance vs. level $S$ in a one-population model:

$$\sigma^2_r(S) = \frac{A_{max}^2}{N} p(S)(1-p(S)) + \sigma_0^2,$$

where $p(S) = \frac{\bar{A}(S)}{A_{max}}$ is the mean $N_1$ and $A_{max} = rN$ is its plateau level.

Noise variance $\sigma_0$ was estimated from the data; $\sigma_0^2 = \sigma_r^2$ at $p(S)=0$. The fitted parameter is population size $N$, chosen for the best fit at the level $S$ where $p(S)=1/2$. Thus computed, $N=360, 320$ or 175 for three cats.

One problem with this model is that the non-decreasing functions for $A$ do not account for the slight dip sometimes seen in the $N_1$ (e.g. Frishkopf, 1956, Figs. 31 & 32) in the medium intensity levels in which a brief plateau, spanning as much as 10 dB, sometimes appears. However, the shape of the $N_1$ is sensitive to recording location and may vary from animal to animal (Frishkopf, 1950), and Frishkopf attempted to fit only the $N_1$ curves showing a mid-level plateau.

There is another problem. Frishkopf's inferred $\sigma(S)$ roughly fits the measured standard deviations over the first increasing portion of the $N_1$, up to the mid-level plateau. But, as Frishkopf noted, at -65 dB, just before or in that plateau region, the standard deviation suddenly peaks again and then drops towards a minimum at the highest levels tested (about -40 dB, corresponding to 55 dB SL for a human listener). In fact the standard deviation can show as many as 3 distinct peaks (ibid., Figs. 30 and 32). Two of Frishkopf's three cats show a general mid-level rise which could be inferred as the single distinct peak in an otherwise jagged pattern. This is exactly opposite to Frishkopf's model of two disjoint populations, in which variability drops at mid-levels where the two populations do not overlap.
APPENDIX G

THE MID-LEVEL HUMP: COMPARISON OF THE FRISHKOPF-RADIONOVA MODEL TO SIGNAL DETECTION THEORY

Below, it is demonstrated that Radionova’s (1963) model (Figure 62) is not compatible with a signal-detection model of the mid-level hump, and also that one of Radionova’s assumptions, that neural threshold density is reflected in the variance of $N_t$, cannot be correct. It is also shown that if signal-detection theory is not to be rejected, plots of $N_t$ that show an outright plateau cannot be used to infer the increase in sensory response that accompanies an increase in stimulus level.

To see these points, let us start by assuming a signal-detection-theory model for the perception of a just-detectable intensity-increment. In this model, the sensory response to the stimulus is normally-distributed over repeated stimulus presentations. The sensory response distributions for the pedestal and that for the pedestal-plus-increment are assumed to be the same except for their mean value. We may assume that detection of an intensity-increment always corresponds to the same detectability index $d^*$, which is equal to the difference in means of the response distributions for the pedestal and that for the pedestal-plus-increment, divided by their assumed common standard deviation $\sigma$. To maintain a given value of $d^*$, a greater separation is required between the response distribution for the pedestal and that for the pedestal-plus-increment when $\sigma$ is greater, and a lesser separation when $\sigma$ is lesser. A greater separation, corresponding for example to the mid-level hump, implies that a greater increase in sensory response is required to maintain a given $d^*$. That is, the mean of the response distribution must be higher. In usual interpretations of signal detection theory, this corresponds to a greater increase in stimulus level and a correspondingly slower increase in the psychometric function, i.e. a shallower slope. Similarly, a steeper slope is predicted for a smaller $\sigma$. Thus the psychometric functions for increment-detection should be shallower at the mid-level hump, which is indeed true here for both observers at all three time conditions (Figures 50 and 51).

To have meaning, such a model requires that we be able to specify what is the internal response. Let us do this by assuming that the signal-detection model is true and that the Radionova (1963) model is also true, then working backwards, bearing in mind that the standard deviation $\sigma$ of internal response must have the same units as the internal response. If $\sigma$ is higher at the peak in the plot of just-detectable intensity-increments (signal-detection model), and if this corresponds to the knee in $N_t$ (Radionova model), then as a first approximation we may assume that $\sigma$ is inversely proportional to the rate of increase of $N_t$. If $N_t$ is proportional to the number of active afferents, then the rate of increase of $N_t$ has units of afferents/db, so that $\sigma$ has units of db/afferent (or in physical units, just dB). Thus the internal response also has units of dB. This does not make sense; the internal response should have units of “number of afferents” which, assuming one spike/afferent, corresponds to “number of spikes” (which corresponds, for synchronised spikes, to a voltage). Thus, the Radionova model and the signal-detection model are not compatible.

To probe this quandary, we return to Radionova’s (1963) model. Radionova (1963) integrated $\sigma_{ni}^2$ over the range of possible click intensities (in dB) to get the total active-population size, assumed proportional to $N_t$:

$$\text{number of active neurons } n = \int \sigma_{ni}^2 dS, \text{ where } S \text{ has units of } \text{dB SL re } N_t \text{ threshold.}$$  \hspace{1cm} (G1)

($N_t$ has units of micro-volts, so to obtain population size there is assumed an invisible proportionality constant [implied in Radionova, Fig. 2] equal to 1 neuron/μV). To perform the above integration, $\sigma_{ni}^2$ must have units of neurons/db; but it’s actual units must be (neurons/db)$^2$ since $\sigma_{ni}$ must have units of neurons/db, the same units as $N_t$. To compensate, a multiplicative constant is needed, having units of dB/neuron. Radionova specified no such constant. Radionova’s integral is not correct; $\sigma_{ni}^2$ cannot reflect population density.

In Radionova’s model, changes in $\sigma_{ni}^2$ represent changes in population density. But if $N_t$ is proportional to the number of active afferents, the quantity $\sigma_{ni}^2$ reflects the variability in the size of the responding population. These two interpretations are incompatible. It has been proposed here that changes in discriminability represent changes in the rate-of-growth of the active neural population, a rate-of-growth which is the slope of the plot of $N_t$. In the Radionova model, the slope of the plot of $N_t$ is the density of neural thresholds. But in signal detection theory, detectability $d^*$ rises (i.e. discriminability improves) as response variance drops. If the sensory response is the potential produced by the massed spikes from the synchronised firing of the active population, and the number of these spikes is
crudely quantified by $N_1$, then the response variance is reflected by $\sigma_{n_1}^2$. Hence the just-detectable intensity-increment should drop as $\sigma_{n_1}^2$ drops. But the Radionova model says otherwise; the just-detectable intensity-increment rises as $\sigma_{n_1}^2$ drops (Figure 62).

The conclusion that the Radionova model is incompatible with signal detection theory can be obtained by another argument. In Section 3.2 the just-detectable intensity-increments were compared to what Zeng (1994) calls loudness variability. Now if loudness variability reflects the underlying variability in the sensory representation, then we would expect the sensory distribution (whatever its nature) to be more variable at the location of Zeng et al.'s (1991) mid-level hump. If $N_1$ is the presumed correlate of loudness, then the response variance $\sigma_{n_1}^2$ is the variance of loudness. Thus if loudness variance peaks when the just-detectable intensity-increment peaks, then $\sigma_{n_1}^2$ must also peak, in opposition to the model of Figure 62.

Let us assume that the sensory response is the count of synchronised spikes, represented crudely by the amplitude of $N_1$. If the mid-level hump corresponds to the knee in $N_1$, as Radionova computed from statistics of $N_1$ (Radionova, 1963, Fig. 4), then a lack of increase in sensory response (an outright plateau) nonetheless corresponds to an increase in stimulus level. But the signal-detection model requires that a greater increase in sensory response correspond to a greater increase in stimulus level. Thus for a signal detection model to work: under an $N_1$ showing an outright plateau, the sensory response cannot be the count of synchronised spikes.

But it is hard to imagine the relevant sensory response not being the spike count. Thus if signal detection theory is not to be rejected, plots of $N_1$ that show an outright plateau cannot be used to infer the increase in sensory response that accompanies an increase in stimulus level. Now in the Radionova formulation, two threshold densities that do not overlap at all must produce a momentary plateau. Hence the Radionova formulation cannot be used to infer the increase in sensory response that accompanies an increase in stimulus level. This can only be done if the two assumed threshold distributions overlap to some unknown degree. The plateau is not ubiquitous; not all plots of $N_1$ show the plateau (see for example Cacace and Smith, 1986) and the shape of the $N_1$ plot varies according to recording site, as noted by Frishkopf (1956, cat) and many others. Other models that infer the neural population size from $N_1$ plots showing a plateau (e.g. Ozdamar and Dallos, 1976) must also be re-evaluated.

It is noteworthy that Radionova's computed intensity-increments peaked at 60 dB SL. This may be 20 dB higher than where the peak is found for measured just-detectable intensity-increments (Raab and Taub, 1969; present data, both in Figure 61). The difference might be artefactual, due to the use of an SL scale for data plotting, which obscures the differences in the amount of energy reaching the basilar membrane for different kinds of stimuli (see caption to Figure 61). Nonetheless, there is uncertainty as the sound pressure level where the two hypothesised threshold distributions intersect, and the possibility of species differences (humans vs. Radionova's cats) cannot be ignored.

Considering these problems, we might be tempted to explain the growth of $N_1$ using a single population distribution. Such an attempt has already been made with unconvincing results. The need for two neural channels was challenged by Ozdamar and Dallos (1976), who computed the $N_1$ produced by a tone burst with a "not too abrupt rise" by first assuming that all nerve fibers contributed (and equally so) to $N_1$ once their thresholds were exceeded. Ozdamar and Dallos took these thresholds from the frequency-tuning curves (FTCs) of individual nerve fibers; as tone intensity rose, the thresholds of more and more remote fibers would be reached. These thresholds were established graphically by plotting the cohort of fiber FTCs on a graph of sound pressure level vs. frequency, drawing a vertical line at the stimulus frequency, and noting the data points formed when the line intersected each known FTC. An across-fiber tuning curve was then drawn from these data points, in the same coordinate system. It was then assumed that fiber innervation density along the basilar membrane was uniform. Combined with the other assumptions, this meant that the width of the across-fiber tuning curve at a given SPL represented the amplitude of $N_1$ at that level. This across-fiber tuning curve appeared highly congruent to the $N_1$ amplitude plot for the gerbil exposed to a 7 kHz tone burst, suggesting a single fiber population as an alternative to two populations of differing mean threshold. However, this success is misleading. The Ozdamar and Dallos model simply ignores the possibility of a distribution of fiber thresholds at a given basilar locus, a distribution which would allow fiber populations to be accessed without significant lateral spread of the excitation pattern across the basilar membrane. Furthermore, closer inspection reveals that the plot of the width of the across-fiber tuning curve vs. level is not congruent to the plot of $N_1$ vs. level. The former shows a sharply-defined slope change at 60 dB SPL, whereas the latter shows a similarly profound change at 75 dB SPL. The width curve also saturates at about 85 dB SPL, whereas the $N_1$ plot has not saturated by 90 dB SPL.
### GLOSSARY OF TERMS

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>adaptive tracking</td>
<td>A way of establishing thresholds, by which sound intensity is changed, according to an algorithm, each time a subject makes a forced choice (see 2IFC). Generally, the process starts out with an easily heard intensity, and incorrect choices cause intensity to be raised and correct choices cause intensity to be lowered. These changes are made smaller and smaller until some minimum step size (typically 1 dB) is reached.</td>
</tr>
<tr>
<td>afferent</td>
<td>(Referring to nerve fibers) Leading away from the periphery and toward the brain.</td>
</tr>
<tr>
<td>backward masking</td>
<td>The effect of a following sound upon the detection and discrimination of a preceding sound (opposite of forward-masking).</td>
</tr>
<tr>
<td>band, bandwidth</td>
<td>Terms (used interchangeably) to indicate a range of contiguous frequencies of which a sound is constructed.</td>
</tr>
<tr>
<td>CAP (compound action potential)</td>
<td>A voltage (electrical potential) created by the initial coordinated firing of the primary afferent auditory neurons that synapse with the auditory receptors, the inner hair cells (IHCs). The CAP lasts a few msec and has several distinct components labeled $N_1$, $P_1$, $N_2$, $P_2$, and so on ($N = \text{negative-going trace}$, $P = \text{positive-going trace}$). CAP is usually recorded (in anesthetized animals) by an electrode placed on the round window, with an indifferent electrode on the neck or elsewhere.</td>
</tr>
</tbody>
</table>
| channel                             | A stream of processing. Specifically, a group of neurons that pool their spikes at some anatomical locus where they are counted over some counting time, which, along with the location, may be unknown. (Delgutte, 1987: “A channel is a processing element that receives inputs from fibers innervating a restricted portion of the
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>characteristic frequency (CF)</td>
<td>Regarding a nerve fiber, that frequency of pure tone which elicits the greatest rate of spikes discharged.</td>
</tr>
<tr>
<td>click</td>
<td>A square sound impulse consisting of an infinitely sharp pressure rise, maintained for a brief duration on the order of 0.1 msec, followed by an identical sharp drop, usually back down to ambient pressure. The infinite rise and fall times give the pressure wave an infinite number of component frequencies, thus perceived as a toneless “click”</td>
</tr>
<tr>
<td>COCB</td>
<td>Crossed OCB (see OCB)</td>
</tr>
<tr>
<td>cochlear nerve</td>
<td>Otherwise called 8th nerve; one per ear. The nerve bundle composed primarily of axons of primary afferent nerve fibers (neurons) ascending from the cochlea. These are the neurons which carry signals about sound up to the next mass of nerve-cell bodies, the cochlear nucleus (CN). The cochlear nerve also includes the axons of the efferent fibers that descend to the cochlea from higher nuclei</td>
</tr>
<tr>
<td>contralateral</td>
<td>On the opposite side of the head from [something]</td>
</tr>
<tr>
<td>dB (decibel)</td>
<td>A unitless (i.e. relative) measure of sound pressure. The number of decibels $z$ separating two sound levels $x$ and $y$, where $x&gt;y$, is given by $z = 20 \log_{10}(x/y)$. $X$ and $y$ are both expressed as RMS pressures (see below for definition of RMS)</td>
</tr>
<tr>
<td>dB SPL</td>
<td>A unitless (i.e. relative) measure of sound pressure. Sound level $x$ in Newtons/meter$^2$ (decibels sound-pressure-level) (abbreviated N/m$^2$) is expressed as $z$ dB SPL where</td>
</tr>
</tbody>
</table>
detection threshold (absolute threshold)

That sound level (dB SPL) at which a probe sound can be distinguished from an absence of sound (an empty interval) with a certain degree of reliability (see Chapter 4 for the actual 2IFC procedure).

dynamic range

- For a nerve fiber, the continuous range of sound intensity over which a noticeable change in sound pressure level (i.e. 1 dB) causes a noticeable change in averaged firing rate, counted over some brief counting interval (usually about 40-50 msec). Usually estimated by the experimenter, for example by visual inspection of the fiber's experimental rate-vs.-level plot. (There is no strict operational definition of neural dynamic range.)

- For the auditory system, the contiguous range of sound intensity over which a trained experimental subject can differentiate relatively small increases or decreases (on the order of 0.5-1 dB) in sound intensity.

afferent

Opposite of afferent

eighth (VIII or 8th) nerve

see Cochlear nerve

forward masking

- In psychophysics: The effect of a preceding sound upon the detection and discrimination of a following sound (Zeng and Turner, 1992)

- In neurophysiology: A shift in the threshold or a reduction in the magnitude of the response evoked by the probe stimulus, caused by the introduction of the preceding (masking) stimulus (Harris and

$$z = 20 \log_{10} \frac{x}{2 \cdot 10^{-5}}$$
Dallo, 1979)

**gating**

The ignoring of one neural input in favour of another. Inhibition of a neural activity by another neural activity, such that one turns on as the other turns off, may or may not be involved

**Gaussian**

Otherwise known as "normal". The probability density

\[
p(\zeta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{\zeta-\mu}{\sigma^2}\right)}
\]

of a randomly-distributed characteristic (call it \(\zeta\)) with mean value \(\mu\) and variance \(\sigma^2\) (a relationship abbreviated as \(N(\mu, \sigma^2)\))

**Hz (Hertz)**

Cycles/second of a sinusoidal waveform, such as a pure tone. 
1 kHz = 1,000 cycles/second

**IHC (inner hair cell)**

The cell that transforms basilar-membrane oscillation into neural spikes in the primary afferent with which it synapses. Embedded in a single row running down the length of the Organ of Corti

**intensity-increment threshold**

That sound level (dB SPL) at which a probe sound can be distinguished from a fixed level (pedestal level) of a less intense sound, with a certain degree of reliability (see Chapter 4 for the actual procedure)

**ipsilateral**

On the same side of the head, with respect to a vertical plane passing from front (nose) to back and top (crown) to bottom

**jnd (just-noticeable-difference)**

Often used in place of "just-detectable intensity-increment" or "discriminability limen". The difference in something (i.e. intensity or frequency), usually an increment, that is just detectable
by an experimental subject (usually, according to some statistical criterion of detectability). Can be extended to refer to nerve fibers that fulfill such a criterion.

**just-detectable intensity-increment**

See jnd. The actual size of a change that is deemed just-detectable. In terms of the intensity-increment threshold, the just-detectable intensity-increment (here, in dB) is the intensity-increment threshold (in dB SPL) minus the pedestal level (in dB SPL) used as the comparison.

**LOC**

Lateral olivo-cochlear system (distinct from MOC)

**loudness**

The perceived intensity of sound. Usually expressed as a growth-of-loudness curve established, for example, by plotting, versus actual sound intensities, the loudnesses of sounds of various intensities (but the same frequency composition) as compared to the loudness of a sound standard of the same frequency composition. This form of psychophysical scaling is called magnitude estimation.

**MOC**

Medial olivo-cochlear system (see Figure 32)

**N1**

The first prominent negative-going deflection in the CAP (compound action potential)

**neural pool**

A group of neurons whose activity (voltage spikes) is gathered together at some collection point (sometimes called a counting center)

**OCB**

Olivo-cochlear bundle (part of the MOC system)

**pedestal**

A sound level upon which another sound is (usually)
superimposed. The two sounds are usually of the same frequency composition, and in order to avoid providing extra discrimination cues (besides intensity difference), the pedestal is just increased in level to provide the \( \text{[pedestal + intensity increment]} \)

**phase-locking**

The synchronisation of the firing of a neurons’ spike, to a particular phase of the waveform of the sinusoidal stimulus. One spike may fire each time the waveform completes one period, or only on every second or third period if the stimulus frequency is high enough

**postsynaptic**

Beyond the synapse between two cells; “upstream” in the direction of neural spike movement

**presynaptic**

On the “lower” (axon) side of the synapse

**PSTH**

Post-stimulus-time histogram. A time histogram (number of event occurrences plotted vs. time) of the number of neural spikes that occur within successive counting bins of equal duration, determined after a stimulus is turned off (here, the forward-masker)

**psychometric function**

A series of data points relating a changing stimulus characteristic (i.e. sound pressure level in dB) to a correspondingly changing response on the part of the subject (i.e. correct discrimination of a target stimulus from a comparison stimulus)

**psychophysics**

The study of the relationship between mental events and physical events

**ramping**

The practice of starting (and ending) a stimulus by steadily increasing (or decreasing) the magnitude of the sound wave’s peak pressure. An infinitely rapid climb from 0 sound pressure to some constant level causes spectral splatter, that is, a momentary spread in the component frequencies of the sound
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition and Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>recovery</td>
<td>In present use, the return to nonmasked threshold of a forward-masked sound. Used in a similar context for vision</td>
</tr>
<tr>
<td>ringing</td>
<td>After the termination of a sound, the continuance of basilar-membrane motion, and/or neural firing above spontaneous rate. Caused by the finite width (in Hz) of the auditory filter (critical band)</td>
</tr>
<tr>
<td>RMS</td>
<td>Root-mean-square, referring to sound pressure level. The RMS level of a time-varying pressure $P(t)$ is $\sqrt{\frac{1}{\Delta t} \int_{t_0}^{t_0+\Delta t} P^2(t)dt}$ where $\Delta t$ covers 1 cycle of the sound's periodic waveform. The energy of the sound over this time period is $k \sqrt{\int_{t_0}^{t_0+\Delta t} P^2(t)dt}$ where $k$ is a constant</td>
</tr>
<tr>
<td>Signal Detection Theory (SDT)</td>
<td>A model of discriminability of stimuli, based on the assumption of a Gaussian distribution of sensory responses over repeated presentations of a stimulus. In audition, this distribution shifts along the auditory-response axis as the sound level changes</td>
</tr>
<tr>
<td>SL (or “dB SL”)</td>
<td>Sensation level, which is sound level in dB with respect to the absolute detection threshold for a sound (rather than to the fixed level of $2 \cdot 10^{-4}$ Newtons/m²; see dB SPL)</td>
</tr>
<tr>
<td>Spectral splatter</td>
<td>See Ramping</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>spike</td>
<td>The rapid voltage change across the neuron’s cell membrane; the neural response to the stimulus’ presence</td>
</tr>
<tr>
<td>SPL</td>
<td>Sound pressure level, usually in dB SPL</td>
</tr>
<tr>
<td>synapse</td>
<td>The juncture between the neural axon carrying an electrical signal, and the subsequent dendrite (non-spike-firing nerve-cell projection) or nerve cell body, from which another axon extends</td>
</tr>
<tr>
<td>TTS (temporary threshold shift)</td>
<td>Elevation in detection threshold caused by exposure to a loud sound presented for a relatively long time, i.e. 110 dB SPL for several minutes. Not to be confused with the shorter-lasting elevations caused by forward masking</td>
</tr>
<tr>
<td>timbre</td>
<td>Quality of a sound, resulting from more than one frequency component; what allows the sound of one musical instrument to be differentiated from another, even with the same pitch and loudness (Gulick, Gescheider, and Frisina, 1989)</td>
</tr>
<tr>
<td>two-interval forced-choice</td>
<td>An experimental scheme for measuring detection, or discrimination, thresholds. The subject listens to two intervals of equal duration, one containing the target sound (i.e. an incremented tone pip), the other containing the comparison sound (i.e. a non-incremented tone pip). The subject then judges which interval contains the target. The target is then randomly re-assigned to either interval (with the other interval containing the comparison), and the next judgement is made</td>
</tr>
</tbody>
</table>