Psychophysical and physiological assessment of the representation of high-frequency spectral notches in the auditory nerve

Ana Isabel Alves Pinto Lopes da Silva

May 2007

Enrique Alejandro Lopez Poveda, Investigador del programa "Ramón y Cajal", adscrito al Departamento de Informática y Automática de la Universidad de Salamanca y al Instituto de Neurociencias de Castilla y León,

CERTIFICA

que la tesis doctoral titulada "Psychophysical and physiological assessment of the representation of high-frequency spectral notches in the auditory nerve" describe el trabajo de investigación realizado por Dª Ana Isabel Alves Pinto Lopes da Silva durante los últimos cuatro años bajo mi dirección.

La memoria de tesis describe un conjunto de estudios psicofísicos, fisiológicos y de simulación diseñados con el objetivo de aclarar de qué forma los valles espectrales de alta frecuencia son codificados en el nervio auditivo. Los resultados de este trabajo cuestionan la noción comúnmente aceptada de que los valles espectrales de alta frecuencia son codificados en el perfil de descarga de las fibras del nervio auditivo, aportando además posibles alternativas a su codificación. Los resultados constituyen una aportación relevante para el campo de la codificación neuronal auditiva y consecuentemente son del interés de la comunidad cientifica especializada.

Considero, por tanto, que esta tesis doctoral reúne la calidad y el rigor científicos necesarios para ser defendida en nuestra Universidad como requisito parcial para que Dª Ana Isabel Alves Pinto Lopes da Silva opte al grado de Doctor.

Salamanca, 22 de Mayo de 2007

Enrique A. Lopez Poveda
Director de la tesis

Manuel Sánchez Malmierca
VºBº Tutor de la tesis

Acknowledgments

My first thanks goes to my supervisor Enrique: his guidance, his teachings, and his example as a researcher are inevitably present in this work. I am grateful to him not only for giving me the opportunity to do this thesis but more, for the time he spent with me teaching me many of the skills I lacked, for his advice, and for his infinite patience. I also thank him for always caring that I learn how to solve the problems, instead of solving them in the easiest and least time-consuming-way for him. He allowed me the fullest opportunities to grow as a researcher. His superior capacity as a scientist and enthusiasm for science, together with his personal qualities, in particular his generosity, make of him the best example a young researcher could ever have. I feel truly honoured for having had his guidance and support in this thesis.

I am indebted to the volunteers that participated in the psychophysical experiments of this thesis: Almudena, Glenda, Luisfer, Nelson, Carmen, Fernando, Flora, Virginia, Nacho and Vanessa. I am deeply grateful to all of them for the time they spent doing these long experiments. Without their participation this thesis would not have been possible. Thank you.

My most sincere thanks goes to Alan Palmer, who collected all the physiological data for this work and who was always available for helping us with the analysis of data and with the publication of the work. His time, help and support were crucial. Thank you.

I am also grateful to Trevor Shackleton for lending us the use of his software, for helping us with our doubts about the software and in the analysis of the physiological data.

My thanks to David Eddins, Ray Meddis and Chris Plack, for their comments and suggestions, fundamental for the development of this work.

My thanks to my lab companions, particularly Almudena, who had always been discretely in the background, supporting me, and for always bringing a touch of humour and joy to the lab.

During this thesis I had the opportunity to live in a different country, to know a different culture, and to work in such enjoyable environments as both the group of Miguel Merchán in Salamanca and the group of Fernando de Castro. It was a very enriching experience and I feel very grateful to all of them for being such nice people.

This work also reflects the never-ending support and patience of my dearest friends, in Salamanca and in Lisbon: Glenda, Joaquín, Catarina and

Sandra. And of course, no need to say, of my family, that, though being far away, has always been by my side and has always supported me in my decisions.

Work supported by the health department of Junta de Comunidades de Castilla-La-Mancha, FIS (PI02/0343 and G03/203), by MEC (CIT-390000-2005-4, BFU2006-07536) and IMSERSO (131/06).

Abbreviations and Acronyms

- AN Auditory Nerve
- BM Basilar Membrane
- CF Characteristic Frequency
- dB decibels
- DRNL Dual-Resonance Non Linear
- FFT Fast Fourier Transform
- HSR High Spontaneous Rate
- HRTF Head Related Transfer Function
- HT Hearing Threshold
- Hz Hertz
- IHC Inner Hair Cell
- LSR Low Spontaneous Rate
- MSR Medium Spontaneous Rate
- RMS Root Mean Square
- SEP Sensitivity based on Excitation Pattern
- SD Standard Deviation
- SMP Sensitivity based on Modulation Pattern
- SPL Sound Pressure Level
- SR Spontaneous Rate
- TH Threshold Level

Contents

Chapter 1

Introduction

1.1 Motivation

High-frequency spectral notches generated by the filtering action of the pinna (Shaw and Teranishi, 1968; Shaw, 1974; Lopez-Poveda and Meddis, 1996) are important cues for sound localization (Hebrank and Wright, 1974; Butler and Belendiuk, 1977; Butler and Humanski, 1992; Carlile *et al.*, 2005). The ability to detect and employ these notches for sound localization requires that they be adequately represented in the auditory nerve (AN) response, since the AN is the only transmission path of acoustic information to the central auditory system. The common view is that these spectral notches are encoded in the distribution of AN activity (average discharge rate) across fibers tuned to different frequencies, the so-called rate profile (Sachs and Young, 1979). This hypothesis is based on physiological studies (Poon and Brugge, 1993; Rice *et al.*, 1995), but there are no psychophysical studies showing that this is the case. The lack of psychophysical evidence corroborating this hypothesis motivated the present work.

Understanding how high-frequency spectral information is encoded in the AN may be crucial to develop more efficient sound processors for auditory prosthesis (Wilson *et al.* 2005, 2006).

1.2 Background

The interaction of sound within the pinna cavities changes the original spectra of sounds. The ratio between the transformed spectrum at the eardrum and the original spectrum reflects the acoustic properties of the head and is

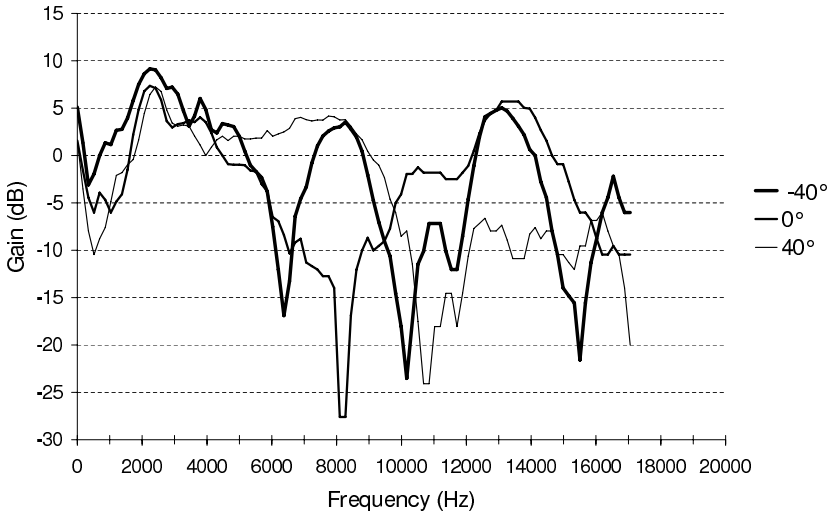

Figure 1.1: HRTFs for a KEMAR manikin (ear DB-065). The sound source was in front of the manikin at different elevations (source: Lopez-Poveda, 1996).

thus referred to as the Head Related Transfer Function (HRTF). Destructive interference at the eardrum resulting of sounds taking different paths through the convolutions of the pinna produce sharp spectral notches in the HRTF (Shaw and Teranishi, 1968; Shaw, 1974; Lopez-Poveda, 1996). It is believed that some of these notches are used by humans as cues to judge the vertical location of sound sources (Butler and Belendiuk, 1977; Butler and Humanski, 1992; Hebrank and Wright, 1974; for a review see Carlile *et al.*, 2005). Particularly important is the notch whose center frequency increases gradually from around 6000 to 11000 Hz as the vertical location of the sound source moves from -40º to +40º relative to the horizontal plane (Fig. 1.1; for a review see Lopez-Poveda, 1996). The bandwidth of this notch at its 5-dB-down points ranges from 1000 Hz at -40º elevation to >4000 Hz at +40º elevation (cf. Shaw and Teranishi, 1968; Chapter 4 in Lopez-Poveda, 1996).

The AN is the only transmission path of acoustic information to the central auditory system. Therefore, the ability to detect and employ high-frequency spectral notches for sound localization requires that they be adequately represented in the AN response. The spectrum of a sound is encoded in the AN activity in at least two ways: in terms of the fibers' average dis-

charge rate as a function of their characteristic frequency (CF; Sachs and Young, 1979), the so called rate profile; or in the timing of the individual spikes in fibers tuned to different frequencies (a temporal profile; Young and Sachs, 1979; Palmer *et al.*, 1986). The common notion is that both rate and temporal profiles may contribute simultaneously to encoding low-frequency spectral features (like those present in speech signals), but fine grained temporal information is unlikely to contribute to the coding of high-frequency spectral features (Delgutte and Kiang, 1984; Rice *et al.*, 1995; Lopez-Poveda, 2005). Indeed, the ability of AN fibers to convey the fine structure of the stimulus waveform in their discharge times (termed phase-locking) falls off rapidly with increasing stimulus frequency above approximately 4000 Hz (Johnson, 1980; Palmer and Russell, 1986). Spectral notches like those generated by the pinna have frequencies ($>$ 6000 Hz; Fig. 1.1) beyond the cut-off of phase-locking and, therefore, are thought to be encoded only in the rate-profile of AN fibers (Poon and Brugge, 1993; Rice *et al.*, 1995).

AN fibers have been classified in either two (Evans and Palmer, 1980; Kiang *et al.* 1965) or three (Liberman, 1978; Winter *et al.*, 1990) types on the basis of their spontaneous rate (Fig.1.2). Fibers with high spontaneous rates (HSR) ($>$ 15 spikes/s) amount approximately 61% of the population. These have low thresholds ($<$ 10 dB SPL) and dynamic ranges of approximately 30 to 40 dB (Fig. 1.2A; Sachs and Abbas, 1974; Evans and Palmer, 1980; Winter *et al.*, 1990). The remaining 40% of the fibers are of a medium- (MSR) or low-spontaneous rate (LSR) type (Figs. 1.2B-C). These have higher thresholds ($>$ 15 dB SPL) and wider dynamic ranges ($\sim$50-60 dB; Evans and Palmer, 1980; Sachs and Abbas, 1974; Winter *et al.*, 1990).

The existence of at least two fiber types with distinct thresholds and dynamic ranges have led several investigators (Delgutte and Kiang, 1984a; Rice *et al.*, 1995; Sachs and Young, 1979) to suggest that the high-frequency spectral characteristics of a stimulus could be conveyed to the central auditory system in the rate profile of HSR and LSR fibers at low and high levels, respectively. However, the apparent quality[1] of rate profiles degrades as the stimulus level is increased even when the rate profile of LSR fibers is considered separately from that of HSR fibers (Fig. 1.3; Delgutte and Kiang, 1984a, b; Rice *et al.*, 1995; Sachs and Young, 1979). Among the factors that may contribute to the deterioration of the quality of the rate profiles at high levels are the broadening of the fibers frequency response with level (Rose *et*

[1] 1 In this context, the word quality must be understood to refer to the degree of similarity between the stimulus spectrum and the AN rate-profile representation.

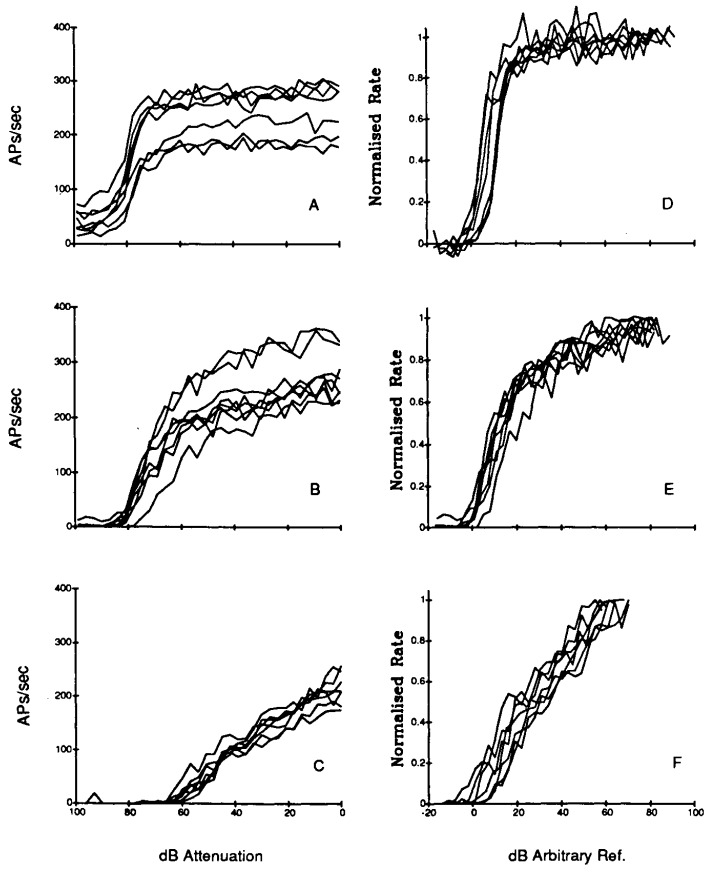

Figure 1.2: Figure 2 of Winter *et al.* (1990) illustrating three different types of AN rate-intensity functions. Top pannels [(a) and (d)] illustrate saturating functions, characteristic of HSR fibers, pannels in the middle [(b) and (e)] illustrate sloping saturating functions, characteristic of MSR fibers, and bottom pannels straight functions [(c) and (f)], characteristic of LSR fibers. Right pannels illustrate the same rate-intensity functions plotted on the left, but the rate is normalised between zero (the SR) and 1 (the maximum rate of the fiber).

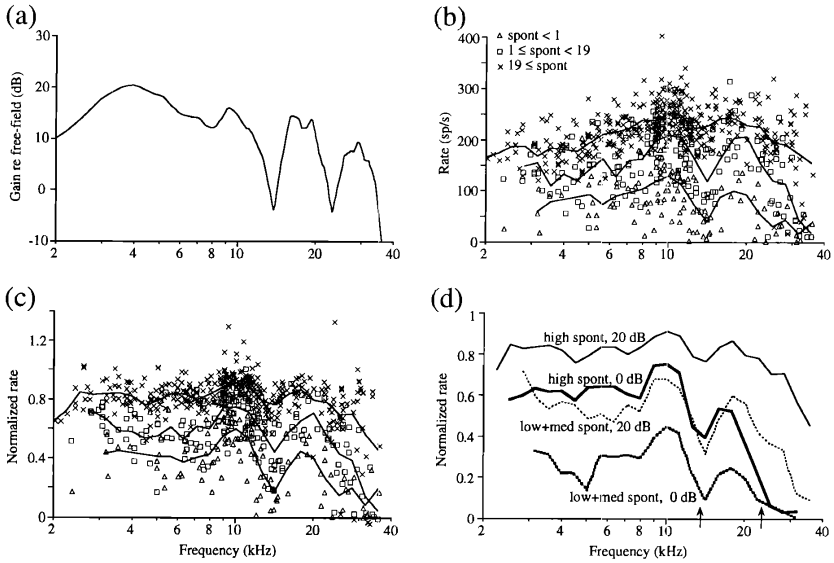

Figure 1.3: Figure 2 of Rice *et al.* (1995) illustrating AN population rate responses for a given HRTF. **(a)** Cat HRTF for a given localization of the sound source. **(b)** Average rate responses of a population of AN fibers in response to the HRTF shown in (a). Different symbols represent fibers with different SR. **(c)** The same as in (b) but for normalized rate profiles. **(d)** Smoothed normalized rate profiles, computed for different goups of fibers and different stimulus levels, classified according to their SR.

al., 1971), the saturation of their discharge rate[2] (Sachs and Young, 1979), and the fiber-to-fiber variation in rate (Rice *et al.*, 1995).

Studies on the encoding of spectral features in the AN have concentrated mainly on the representation of vowels and spectral features relevant to speech perception (Sachs and Young, 1980; Sinex and Geisler, 1983; Miller and Sachs, 1983; Delgutte and Kiang, 1984a, 1984b; Carney and Geisler, 1986; Conley and Keilson, 1995; Miller et al. 1997; Stevens and Wickesberg, 2005; Loebach and Wickesberg, 2006). These spectral features are mainly restricted to frequencies below 4 kHz and therefore can be encoded either through rate-place and/or temporal codes. Relatively less physiological studies have specifically investigated the encoding of other complex stimuli, such

[2]Some studies report non-saturating LSR fibers (e.g. Winter *et al.*, 1990). The existence of this type of fibers seems to be species and frequency specific but the issue is still controversial.

as broadband noises, and of high-frequency spectral features in the AN (Rice *et al.*, 1995; Poon and Brugge, 1993; Louge *et al.* 2004). Additionally computational modelling studies of AN responses to HRTFs have been developed (e.g. May and Huang, 1997; Lopez Poveda, 1996). In general their results support a rate-place representation of spectral features. Several recent studies have focused on studying the representation of spectal features in the AN, as well as in finding correlations between psychophysical performance in a given task and neuronal performance prediced from responses of AN fibers (e.g. Bandyopadhyay and Young, 2004). Most of these concentrate on spectral features below approximatelly 4 kHz.

The encoding of spectral features in the AN has been inferred mainly from the responses of individual AN fibers. Even though the physiological properties of single fibers influence the encoding of spectral features, it is almost certainly the activity of the population of fibers that determines perception of these features. The information conveyed by the population of fibers might be particularly important to encode and perceive complex spectra, such as broadband noises, especially considering that the nonlinear character of the peripheral auditory system makes it difficult to predict the response of the system to the stimuli (Holmes, *et al.*, 2004). In this regard, some physiological and modelling studies have been aimed at finding correlations between psychophysical and neural performance in certain tasks (e.g., Siebert, 1970; Heinz *et al*., 2001; Bandyopadhyay and Young, 2004)

1.3 Objectives

This work aimed at investigating the neural code underlying the perception of high-frequency spectral notches, through searching for behavioral-neural correlations on the representation of these notches at the AN. This implied carrying out the following specific objectives:

1. To investigate the effect of sound level on the discrimination between broadband noises with and without high-frequency spectral notches.

2. To assess the quality of the internal representation of high-frequency notches at increasing stimulus levels.

3. To measure and analyse the responses of guinea-pig AN fibers to noises with and without high-frequency spectral notches in search for neural correlates of the psychophysical results.

4. To investigate different forms of representation of high-frequency spectral notches in the peripheral auditory system by means of a computational auditory model.

1.4 Original Contributions

The main original contributions of this work are:

- An extensive group of psychophysical data relative to the perception of high-frequency spectral notches at increasing sound levels. Data were gathered pertaining to discrimination and detection tasks.
- AN responses to broadband noises with and without high-frequency spectral notches at increasing levels.
- MATLAB^{TM} software for generating broadband noises, in the digital domain.
- Software for running spectral discrimination psychophysical experiments, with a TDT^{TM} System 3 psychoacoustic workstation. The code is based on the software RPvds and PsychRP provided by Tucker Davies Technologies together with the equipment.
- MATLAB^{TM} software for the statistical analysis of psychophysical data.
- MATLAB^{TM} software for analysis of individual and population AN responses.

1.5 Overview of the thesis

This thesis is basically organized according to the objectives listed above.

Chapter 2 describes a psychophysical study aimed at evaluating the ability of listeners to discriminate between broadband noises with and without high-frequency spectral notches. The main result was that discrimination becomes increasingly harder as the level of the noise increases from 32 up to 70-80 dB SPL and then, for levels higher than 80 dB SPL, discrimination improves again. This result was paradoxical because it is not consistent with the notion that high-frequency spectral notches are represented in the rate profile of AN fibers. Finding a physiologically-based explanation for

this result constituted the objective of the studies described in the following chapters.

Chapter 3 describes a psychophysical study aimed at testing the hypothesis that discrimination between flat-spectrum and notch noises is based on the rate-profile representation of the stimuli spectra. The quality of the internal representation of the spectral notch was estimated at increasing noise levels. It was observed that this representation deteriorates gradually at increasing levels, even at levels at which discrimination improves. If discrimination were based on the comparison of the internal representations of stimuli spectra, then the quality of the representation of the spectral notch should vary nonmonotonically with level, as discrimination performance did in the experiments described in Chapter 2.

Chapter 4 describes a physiological study aimed at finding a physiological correlate of the non-monotonic effect of level in the discrimination between flat-spectrum and notch noises (Chapter 2). The analysis of AN responses with different methods suggests that discrimination between flat-spectrum and notch noises is unlikely to be based on rate-profiles and that temporal codes are likely to be involved.

Chapter 5 describes a computational study that explores the likelihood of two possible neural codes (rate vs. temporal) considered in chapters 2-4. The model was used to simulate the response of inner hair cells to stimuli similar to those employed in the discrimination experiment. The analysis of both the average and the instantaneous amplitude of this simulated responses is consistent with the conclusions derived in Chapters 3 and 4.

Chapter 6 discusses the main findings of the present work and its implications. Further, it provides ideas fo future work.

Chapters 2 to 5 are based on several papers that have been already published or submitted for publication in peer-reviewed scientific journals or books. Appendix C provides a list of these papers. Therefore, each chapter is self-contained and may be read separately.

A CD-ROM accompanies this text. It contains the software and the data that resulted from this work. Appendix A provides a description of contents of the CD-ROM.

Chapter 2

Detection of high-frequency spectral notches as a function of level[1]

2.1 Introduction

High-frequency spectral notches are important cues for sound localization. They occur at frequencies (>6000 Hz; see for example Lopez-Poveda, 1996) beyond the cut-off of phase-locking ($\sim$ 4000Hz; Johnson, 1980; Palmer and Russell, 1986) and are therefore believed to be encoded in the profile of average discharge rates of AN fibers only. Our ability to detect them must depend on the quality of their representation in AN rate profiles. Given that the majority of AN fibers have low threshold levels and short dynamic ranges, becoming saturated at very high stimulus levels, the rate-profile representation of high-frequency spectral features is expected to deteriorate as the level inceases. As a consequence, the perception of high-frequency spectral notches should become increasingly harder as the sound level increases. This chapter describes a psychophysical study aimed at testing this hypothesis. Several experiments were carried out. In the main experiment, the threshold notch depth necessary to discriminate between a flat spectrum broadband noise and a similar noise with a spectral notch centered at 8000 Hz was measured at levels from 32 to 100 dB SPL (Fig. 2.1). It will be shown that, contrary to the above hypothesis, the threshold notch depth for discrimination is a *nonmonotonic* function of stimulation level.

[1]This chapter is based on the published paper: Alves-Pinto, A and Lopez-Poveda, E. A. (2005). "Detection of high-frequency spectral notches as a function of level," J. Acoust. Soc. Am. **118**: 2458-2469.

Spectral features can be represented either in the rate profile of AN fibers or in the temporal pattern of fibers' responses. Spectral features beyond the cutoff of phase locking [$>$ 4000 Hz; Johnson (1980); Palmer and Russell (1986)] are most likely to be represented in terms of AN discharge rate alone (Rice *et al.*, 1995). LSR units have higher thresholds and wider dynamic ranges, and thus would be able to encode spectral features at levels at which HSR units are already saturated. This has led several investigators to suggest that high-frequency spectral features could be transmitted in the rate-profiles of HSR and LSR fibers at low and high levels, respectively. LSR units have higher thresholds and wider dynamic ranges, and would, thus, be able to encode spectral features at levels at which HSR units are already saturated. However, rate-profile representation of spectral features deteriorates even when the activity of different goups of fibers is considered separately (Delgutte and Kiang, 1984a, b; Rice *et al.*, 1995; Sachs and Young, 1979; Lopez-Poveda, 1996).

In addition to the role of LSR fibers, other mechanisms have been proposed to explain how high-frequency spectral features may be encoded in the AN response at high levels. For instance, Delgutte and Kiang (1984a, b) suggested that at high levels, when the adapted (steady-state) response of many of the fibers is saturated, the stimulus spectrum may still be represented reasonably well in the onset rate profiles. Their suggestion is based on the fact that AN fibers have a wider dynamic range over the first few (5 to 20) milliseconds of their response (Smith and Brachman, 1980).

Direct evidence is lacking that the dynamic range of AN fibers gets wider the shorter the stimulus rise time. This, however, is likely to be the case because the maximum onset rate of AN fibers depends on the stimulus rise time [see Fig. 3 in Delgutte (1980)] and models of AN adaptation suggest that the level at which the maximum onset rate occurs gets higher the shorter the rise time [e.g. Fig. 2 in Meddis (1988)]. Therefore, following Delgutte and Kiang's suggestion (Delgutte and Kiang 1984a, b), it is reasonable to hypothesize that the AN rate-profile representation of the stimulus spectrum should be clearer for sounds with abrupt onsets, particularly at high stimulus levels. Moreover, if at high levels an important proportion of the spectral information were indeed conveyed in the onset AN rate profile, spectral shape discrimination should hardly depend on the stimulus duration.

These hypotheses were tested by repeating the spectral discrimination experiment described above but using different noise burst durations (20 *vs.*

220 ms) and rise/fall times (2, 10, and 30 ms). It will be shown that contrary to the hypotheses, threshold notch depth is hardly affected by the stimulus rise time and depends strongly on stimulus duration, even at high levels.

2.2 Methods

2.2.1 Procedure

Notch depths at threshold were measured using a three-interval, three alternative forced-choice (3AFC) paradigm. In two of the intervals (standard intervals), the stimulus consisted of a noise burst with a flat spectrum from 20 to 16000 Hz. In the other (target) interval, the stimulus consisted of a burst of noise with a notch in its spectrum (Fig. 2.1). The three intervals were presented in random order to the listener who was instructed to detect the odd one out. The silence period between interval presentations was 500 ms.

The initial notch depth (ΔL in Fig. 2.1) was always fixed at 20 dB below the reference spectrum level of the noise. An adaptive procedure was employed to estimate the 70.7% correct point in the psychometric function (Levitt, 1971). During the adaptive procedure, notch depth decreased after two consecutive correct responses and increased after an incorrect response. Notch depth increased or decreased by 6 dB for the first four turn points and by 1 dB thereafter. Sixteen turn points were recorded for each measurement and the threshold estimate was taken as the mean of the notch depths for the last 12 turn points. When the corresponding standard deviation (SD) exceeded 6 dB, the measurement was discarded (invalid threshold) and a new threshold was obtained. The thresholds reported in this study correspond to the *geometric* mean of at least three valid measurements (see Sec. 2.3.1 for more details). Sometimes detecting the notch was impossible (the adaptive procedure did not converge). Whenever this happened and to prevent endless sessions, the automatic adaptive paradigm was set to stop after 80 stimulus presentations.

Notch depths at threshold were measured as a function of the overall level of the stimulus. Overall levels ranged from 32 to 100 dB SPL (these correspond to spectrum levels of -10 to 58 dB SPL for the flat-spectrum noise). 10-dB steps were used for levels from 32 to 92 dB SPL. Sessions were organized in blocks of eight runs; each run for a single overall level. Within each block, levels were ordered quasi-randomly, although the stimuli with the

higher levels ($>$ 70 dB SPL) were presented at the end of the block. This measure aimed at minimizing any effect that temporary threshold shifts may have on spectral discrimination of low-level sounds.

Listeners were tested individually in a double-wall sound-attenuating chamber. Interaction was provided by means of a response box. Lights were used to mark the presentation of the stimuli and to give listeners a trial-by-trial feedback on their responses. Stimuli were delivered monaurally to the listener via Etymotic Research^{TM} ER2 insert earphones. This earphone is designed to have a flat frequency response at the eardrum up to approximately 16000 Hz. The sound pressure levels (SPLs) reported below are calibrated values [for a flat-spectrum broadband noise (20 to 16000 Hz), the calibrated SPL output at a microphone coupled to a Zwislocki occluded ear simulator was 100 dB SPL for 2.3 V_{RMS}].

2.2.2 Stimuli

The center frequency (f_C) of the spectral notch was fixed at 8000 Hz. The reference condition was with a notch bandwidth (BW) of 2000 Hz, and noise bursts duration of 220 ms, including 10-ms cosine-squared rise/fall times. The effect of notch bandwidth was investigated by testing notch BWs of 1000, 2000 and 4000 Hz. The effect of the stimulus duration was investigated by testing stimulus durations of 20 and 220 ms, both including 10-ms rise/fall times. The effect of the stimulus rise time was investigated by setting the stimulus duration at 220 ms and using rise/fall times of 2, 10 and 30 ms.

a. *Noise generation*

The standard and the target noise bursts were made by adding three separately generated noises with different narrowband flat spectra (N1, N2, and N3 in Fig. 2.1): N1 with frequency components from 20 to f_C-0.5BW Hz; N2 with a spectrum centered at f_C= 8000 Hz and a bandwidth equal to the notch bandwidth; and N3 with components ranging from f_C +0.5BW to 16000 Hz.

Each of the three noises, N1, N2, and N3, were generated digitally in the time domain (sampling frequency = 48828 Hz) by adding sinusoids of equal amplitude but random phases uniformly distributed between 0 and 2π radians. The frequencies of the sinusoids spanned the spectral bandwidth of each of the noises in steps of 1 Hz. The waveforms of N1, N2, and N3 were then added together on a sample-by-sample basis in a TDT^{TM} System 3 psychoacoustics workstation. The noise burst corresponding to a standard

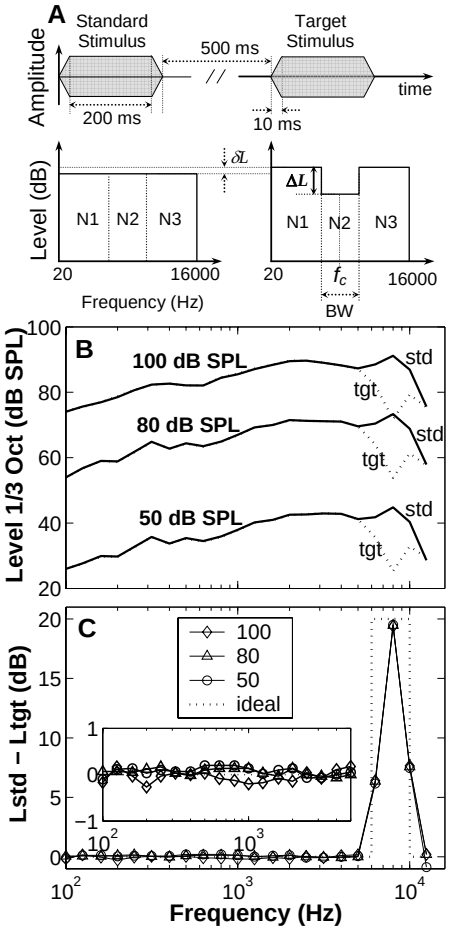

Figure 2.1: **(a)** Waveforms and spectra of the standard and target noises. **(b)** Actual SPLs for the flat-spectrum (continuous lines) and the notched (dotted lines) noises delivered by the Etymotic^{TM} ER2 transducer. The SPLs were measured with a Brüel&Kjaer^{TM} sound level meter (mod. 2238A) coupled to a Zwislocki occluded ear simulator over a frequency range from 100 to 12500 Hz in 1/3 octave bands. The notched noise contained a 4000-Hz wide, 20-dB deep spectral notch. The reference spectrum level was the same for the flat-spectrum and the notched noises, and such that the overall SPL for the flat-spectrum noise was equal to 50, 80, and 100 dB SPL (indicated by the number next to each line). **(c)** SPL differences between the flat-spectrum and the notched noises for stimuli with overall levels of 50, 80, 100 dB SPL. The inset provides greater detail on the differences. The thick dotted line illustrates an ideal difference spectrum. Note that the actual difference profiles are almost identical at the three overall levels tested.

interval was generated by adding N1, N2 and N3, all with identical spectrum levels. The notched-noise burst (target interval), however, was generated by attenuating N2 by ΔL dB (notch depth) as required (Fig. 2.1), and adding the resulting waveform to those of N1 and N3. Notch depths (ΔL in Fig. 2.1) are, therefore, expressed in dB relative to the spectrum level of N1 or N3.

Before delivering the stimuli, the overall level of the combined noise (N1+N2+N3) was set digitally by attenuating the signal as required. Signal clipping at high levels was avoided by ensuring that no sample of the combined digital noise reached the maximum output voltage of the TDT™ system (±10 V). This is an important point because the distortion associated with clipping could smear the spectrum of the notched noise in the target interval, hence making notch detection and the interpretation of the results more difficult. Furthermore, transducer distortion did not affect the results at high levels, as shown in Fig. 2.1(b) and (c), because the voltage required (2.3 V_{RMS}) to produce the maximum level considered in this study (100 dB SPL) was within the nominal operational limits (2.5 V_{RMS}) of the Etymotic ER2 transducer.

b. *Strategies for making level differences unreliable discrimination cues*

The presence of the notch in the target stimulus makes its overall level lower than that of the standard stimuli. For example, a level difference of 1.23 dB occurs for the broadest (4000-Hz wide) notch considered and with the maximum notch depth measured (18.4 dB as shown in Fig. 2.4). Such a level difference could be used as a cue in a task aimed at measuring the listeners ability to discriminate between spectral shapes. Hence, it could complicate the interpretation of the results.

Two different strategies were employed to reduce this unwanted effect. The first strategy consisted of presenting the standard and the target stimuli with equal overall levels. This was achieved by reducing the spectrum level of the standard noise as necessary (δL in Fig. 1) to make its overall level identical to that of the notched-noise (target) stimulus. It can be shown that the necessary reduction (in dB) is equal to:

$$\delta L = 10 \times log_{10}\left[1 + \frac{BW_{N2}}{BW_T}\left(10^{-\frac{\Delta L}{10}} - 1\right)\right] \tag{2.1}$$

where BW_{N2} is the bandwidth of noise N2 (Hz); BW_T is the total bandwidth (Hz) of the noise (15980 Hz); and ΔL is the notch depth (dB). The mathematical derivation of this equation is shown in Appendix B. To get an idea of the necessary reduction in spectrum level, δL equals -0.58 dB for a 2000-Hz wide, 27-dB deep notch. Notice that δL (Eq. 2.1) is also equal to the overall level difference between the standard, flat-spectrum noise of bandwidth BW_T and the target noise with a spectral notch of bandwidth BW_{N2} and depth ΔL.

Whenever this equalized-level strategy was employed, the same noise tokens N1, N2 and N3 were used for the standard and the target intervals and for every block of every condition. This condition will be hereon referred to as *frozen-noised/equalized-level* (FN-EL).

The second strategy to prevent the use of level differences as discrimination cues consisted of presenting each of the intervals in a given trial at a different overall level (level roving). The idea is to make differences in overall level *and also* level differences within any given frequency band unreliable cues for detecting the spectral notch, forcing stimulus discrimination to be based solely on differences in overall spectral shape. This method of level randomization has been used in other notch detection experiments (Moore *et al.*, 1989) and is a common practice in spectral-shape discrimination tasks [for a review see Green (1988)]. Here, for any given trial, the overall level of the stimuli was randomly attenuated by an amount between 0 and 10 dB with respect to the corresponding reference level (uniform distribution, 0.25-dB steps). As a result of this, for any given reference spectrum level, the actual spectrum level for this condition was on average 5 dB lower. This will be taken into account when illustrating the results (see caption of Fig. 2.3).

Even with a roving level, a notch that is deep enough could in principle be detected on the basis of comparing the overall level across intervals (cf., Green, 1988). According to Eq. 2.1, the overall level difference between the target and the standard stimuli increases asymptotically with increasing notch depth (ΔL in dB). Such a difference approaches 1.25 dB for the broadest (4000-Hz wide) notch considered in the present study. Given that level was roved between 0 and 10 dB with a uniform distribution, this implies that spectral discrimination based on overall level comparisons can still occur despite the level roving, but only in 12.5% of the occasions and in the least favorable case; that is, for the broadest and deepest notch. In summary,

the present level roving strategy guarantees that the listeners responses be based on spectral shape discrimination in more than 87.5% of the occasions.

Whenever this level-roving strategy was employed, different (*random*) N1, N2 and N3 noise tokens were employed for different intervals and for different conditions. This was intended to investigate to what extent the use of stimuli with different temporal structures makes notch detection more difficult relative to the FN-EL condition. This condition will be hereafter referred to as random-noise/roving-level (RN-RL).

No significant differences were observed in the results obtained with either strategy (see Fig. 2.3 and Sec. 2.3.2). For this reason and for convenience, the frozen-noise/equalized-level strategy was more generally used in the experiments reported below.

2.2.3 Listeners

Data were collected for eight listeners with ages ranging from 20 to 40 years, although some of the listeners were not tested in all conditions. Their absolute thresholds were within 20 dB re. ANSI 3.6-1996 (Specifications for Audiometers) at the audiometric frequencies (250-8000 Hz). All listeners were given at least one training session in the task. Only S1 (the author) had previous experience in psychoacoustic tasks.

2.3 Results

2.3.1 The statistical distribution of notch depth

Figure 2.2 illustrates that there is a strong correlation ($R = 0.84$, $p < 0.0001$) between the *arithmetic* mean and the corresponding SD of threshold notch depth. Furthermore, threshold notch depths are limited to values ≥ 0 dB. These properties demonstrate that threshold notch depth does not conform to an equal-variance distribution (cf., Bland and Altman, 1996a). Indeed, the data in Fig. 2.2 conform to a *lognormal* distribution with a probability of $p = 0.94$ (two-tailed Kolmogorov-Smirnoff test). For this reason, the results are hereafter illustrated as the *geometric* mean plotted on a *logarithmic* scale (Bland and Altman, 1996a, 1996b, 1996c). Furthermore, the variability of the results is illustrated as the 68% confidence interval of the geometric mean (equivalent to the more conventional arithmetic mean plus and minus one SD) (Bland and Altman 1996b, 1996c, 1996d).

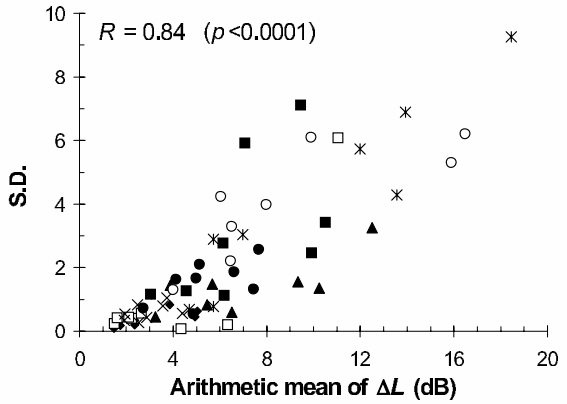

Figure 2.2: Standard deviation of threshold notch depths (SDs) as a function of their associated arithmetic means. Data are pooled from Fig. 2.4 and thus are for eight listeners and the whole range of levels. Different symbols illustrate data for different listeners. The strong significant correlation (Pearsons R = 0.84, $p < 0.0001$) between the arithmetic mean and SD suggests that threshold notch depth does not conform to an equal-variance distribution.

This procedure is similar to and inspired by a previous level discrimination study (Buus and Florentine, 1991) where it is suggested that level-discrimination thresholds are appropriately represented by plotting the difference limen for a change in sound level on a logarithmic scale. In that case, the argument in favor of this procedure was that this type of representation provides a straightforward relation between measurements of discrimination thresholds and the sensitivity of the auditory system as estimated in terms of d-prime (Buus and Florentine, 1991, p. 1379 and their Fig. 7).

2.3.2 Discrimination is based on detecting differences in overall spectral shape

Figure 2.3 illustrates threshold notch depth *vs.* level functions for four representative listeners (S1 to S4) for a notch bandwidth of 2000 Hz. Each panel illustrates data for a different listener (as indicated in the upper left corner of the panels). Shaded triangles on the abscissa illustrate the listeners absolute threshold for the flat-spectrum (standard) noise. Crosses indicate conditions for which the notch was undetected consistently in three or more trials. Note that the y-axis scale differs across panels.

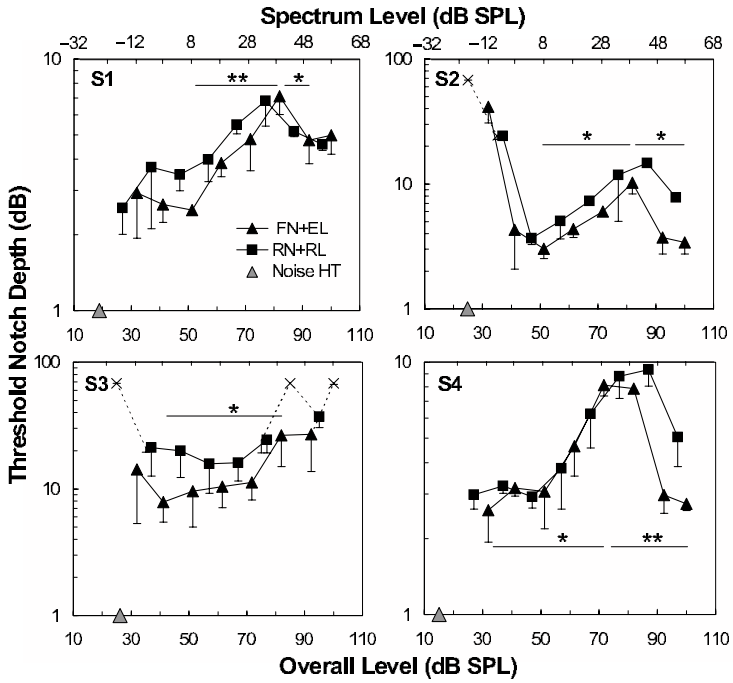

Figure 2.3: Threshold notch depths for discriminating between a flat-spectrum noise and a notched noise plotted as a function of the stimulus overall level (the corresponding spectrum level scale is also shown at the top of the upper panels). The notches were centered at 8000 Hz and had a bandwidth of 2000 Hz. Each panel illustrates data for a single listener as indicated in the upper-left corner of each panel (note the different scales on the y-axis of each panel). Each data point corresponds to the geometric mean of at least three measurements. Error bars illustrate the 68% confidence interval (see text for details). Crosses indicate conditions for which no discrimination between the target and the standard stimuli was possible. Shaded triangles on the abscissa indicate the listeners absolute hearing threshold (HT) for the standard noise. Thresholds were measured in two experimental conditions: in the frozen noise/equalized level condition (FN-EL) condition and in the random noise/roving level (RN-RL) condition. Horizontal bars and asterisks apply to the FN-EL condition only, and denote statistically significant differences between the maximum threshold notch depth that occurs around 70-80 dB SPL and the lowest values that occur on each side of the maximum (single-tailed paired Students t-test; one and two asterisk denote $p < 0.1$ and $p < 0.01$, respectively).

Threshold notch depth values for the random-noise/roving-level condition (squares; RN-RL) are comparable or slightly larger than those for the frozen noise/equalized-level condition (triangles, FN-EL). Except for listener S3, the poorest performer, the main differences between the results for the two conditions occur at the highest overall levels. For the most part, these differences may be attributed to roving the level rather than to using random noise. If the use of frozen noise facilitated notch detection as a result of the listeners memorizing its temporal pattern, depths would be considerably lower at all levels for the FN-EL condition (triangles in Fig. 2.3) and this is not the case. Furthermore, all participants reported that level randomization was an important distracter, particularly at high levels. This agrees with the observations of Moore *et al.* (1989), who showed a deterioration in the detection of notches centered at 8000 Hz with level randomization. Moore *et al.* attributed this deterioration to a reduction of the listeners attention towards the fainter spectral features when more prominent perceptual changes are introduced.

Given that the observed differences are small and that randomizing the level for each stimulus presentation makes energy within or outside the notch band an unreliable cue for stimulus discrimination, it can be reasonably concluded that even in the frozen-noise/equalized-level condition listeners discriminate between the target and the standard stimuli by detecting a notch in a flat reference spectrum rather than by monitoring the level over certain frequency regions (i.e., in the notch band or elsewhere).

2.3.3 Notch depth increases *non*monotonically with level

Figure 2.3 illustrates that for three of the four listeners (S1, S2 and S4), threshold notch depth increases nonmonotonically with level. Consider, for example, the frozen-noise/equalized-level condition (triangles). Depth values increase from approximately 3 dB at 40 dB SPL to approximately 9 dB at 70-80 dB SPL and then decrease again to 3 dB (S2 and S4) or 5 dB (S1). For these three listeners, the differences between the largest and the smallest threshold notch depth values are statistically significant, as indicated by the horizontal bars and their associated asterisks in Fig. 2.3 [single-tailed paired Students t-test on the ln-transformed data (Bland and Altman, 1996d); one asterisk: $p < 0.1$; two asterisks: $p < 0.01$]. The other listener (S3) was the poorest performer and reported that the task was very difficult. Her threshold depth values are larger across levels than those for the other three listeners. They increase from 8 dB at 42 dB SPL to 10 dB at 72 dB SPL, and

then increase more rapidly with level until the notch becomes undetectable at 100 dB SPL (indicated by a cross). Also, the variability of her results is overall larger than those for the other three listeners. This may occur as a result of her having worse frequency resolution and/or her finding it more difficult to follow an appropriate spectral discrimination cue.

It is noteworthy that for listeners S2 and S3, threshold notch depth values are larger at the lowest overall level (32 dB SPL) and decrease rapidly for the next level tested (42 dB SPL). It is likely that this result reflects simply that 32 dB SPL is only 4 and 5 dB above the noise absolute threshold of listeners S2 and S3, while it is 11 and 15 dB above the absolute threshold of listeners S1 and S4, respectively.

To confirm that nonmonotonic threshold notch depth *vs.* level functions are the norm, the experiment was carried out for eight listeners (S1 to S8) for the frozen-noise/equalized-level condition. This time, however, the notch bandwidth was set to 4000 Hz because it is an easier condition. The results are illustrated in Fig. 2.4. Although the actual values vary considerably across listeners, notch depth increases nonmonotonically with level for *all* of them except S3, at least qualitatively. Quantitatively, however, the difference between the largest threshold notch depth value and the lowest one that occurs at a higher level is statistically significant for four (S1, S2, S4, and S6) out of the eight listeners only (single-tailed paired Students t-test on the *ln*-transformed data; one asterisk: $p < 0.1$; two asterisks: $p < 0.01$). A maximum in the nonmonotonic functions occurs at approximately the same level (70-80 dB SPL) for *all* listeners except S3.

2.3.4 Notch depths are comparable at overall levels of 42 and 100 dB SPL

The quality of the rate-profile representation of the spectral notch is expected to deteriorate at high levels as a result of the broadening of the fibers frequency response with level (Rose *et al.*, 1971), the saturation of their discharge rate (Sachs and Young, 1979), and/or the fiber-to-fiber variation in rate (Rice *et al.*, 1995). However, the nonmonotonic character of the threshold notch depth *vs.* level functions suggests that the spectral notch is equally well represented in the AN response at 42 and 100 dB SPL. To investigate whether this is actually the case, threshold notch depths at 100 dB SPL were statistically compared with those at 42 dB SPL for each listener (single-tailed, paired Students t-test on the *ln*-transformed data). Results

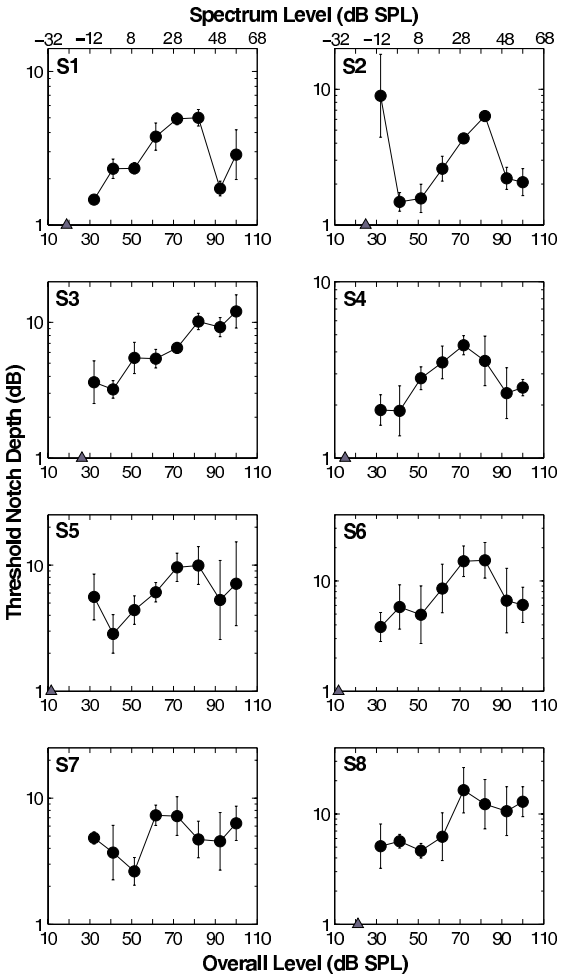

Figure 2.4: Threshold notch depth *vs.* level functions for a spectral notch centered at 8000 Hz and a bandwidth of 4000 Hz. Except for the notch bandwidth, the experimental conditions were the same as for the FN+EL condition of Fig. 2.3. Each panel illustrates data for a single listener as indicated in the upper-left corner of each panel (note the different scales on the y-axis of each panel). Error bars illustrate the 68% confidence interval of the geometric mean. Shaded triangles illustrate the hearing thresholds for the standard noise.

are illustrated in Fig. 2.5 for the two notch bandwidths tested (2000 and 4000 Hz) and for the frozen-noise/equalized-level condition. Threshold notch depth values are larger at 100 dB SPL for approximately 50% of the listeners. However, statistically significant ($p < 0.05$) differences are rare (denoted by asterisks in Fig. 2.5). The significance of this result is discussed below (Sec. 2.4.1)

2.3.5 The effect of notch bandwidth

Figure 2.6 illustrates threshold notch depth *vs.* level functions for the frozen-noise/equalized-level condition for three notch bandwidths: 1000, 2000, and 4000 Hz. Crosses illustrate conditions for which notch detection became erratic and three valid measurements could not be obtained.

Results vary widely across listeners. Generally, however, threshold notch depth increases as the notch bandwidth decreases. Interestingly, a significant increase in threshold notch depth occurs for the narrowest, 1000-Hz wide notch. This is true particularly for listeners S3 and S5 and for overall levels higher than 70 dB SPL. For these listeners and conditions, obtaining a threshold notch depth was impossible.

Similar results have been described elsewhere. Moore *et al.* (1989) reported that none of their listeners could detect 1000-Hz wide notches centered at 8000 Hz whilst they had no problem detecting wider notches. Heinz and Formby (1999) also reported that detecting energy decrements in the spectrogram of a random-level noise was possible only if the spectral bandwidth of the decrement was ≥ 500 Hz.

It should also be noted in Fig. 2.6 that the nonmonotonic effect, when present, is comparable for different notch bandwidths.

2.3.6 The effect of stimulus rise time and stimulus duration

Figures 2.7 and 2.8 illustrate the effects of stimulus rise time (in four listeners) and of stimulus duration (in five listeners), respectively. The data are for a notch bandwidth of 2000 Hz and for the frozen-noise/equalized-level condition. Clearly, the stimulus rise time has no significant or systematic effect on notch depth at any of the levels tested (Fig. 2.7). On the other hand, stimulus duration has a clear effect (Fig. 2.8). Threshold notch depth values are significantly larger for the short (20-ms) than for the long (220-ms) noise bursts at *all* levels with very few exceptions. The bottom right

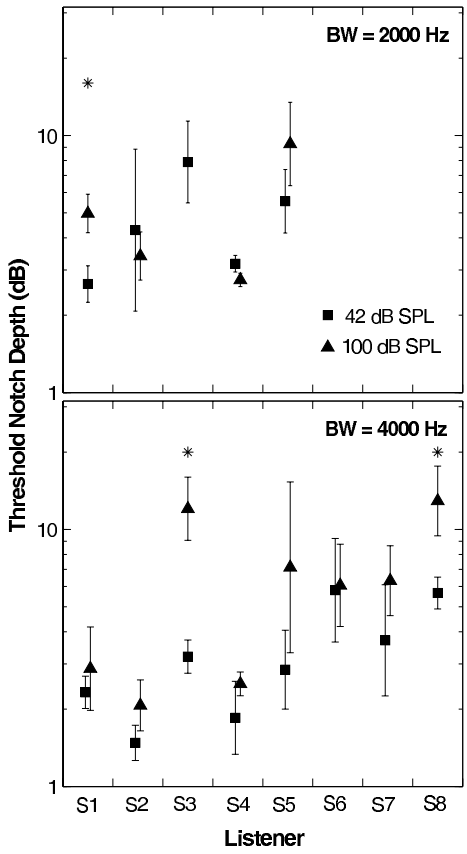

Figure 2.5: Comparison of individual threshold notch depths at overall levels of 42 and 100 dB SPL, for notch bandwidths of 2000 (top panel) and 4000 Hz (bottom panel). To facilitate the comparison between the results at the two sound levels, data points for any given listener appear slightly displaced horizontally as necessary. Error bars indicate the 68% confidence interval of the geometric mean. Asterisks indicate significant differences between the thresholds at 42 and 100 dB SPL (single-tailed, paired Students t-test; $p < 0.05$).

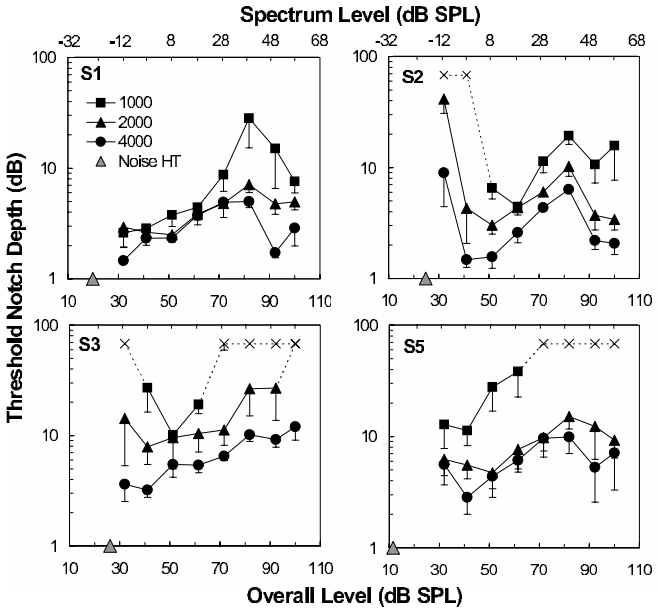

Figure 2.6: Threshold notch depth *vs.* level functions for notch bandwidths of 1000, 2000 and 4000 Hz (as indicated by the inset in the top left panel). Error bars indicate the 68% confidence interval of the geometric mean. Data were obtained using the FN-EL strategy. Stimuli had a total duration of 220 ms, including rise/fall ramps of 10 ms. Each panel illustrates the results for a single listener. Shaded triangles on the abscissa illustrate the hearing thresholds (HTs) for the standard noise.

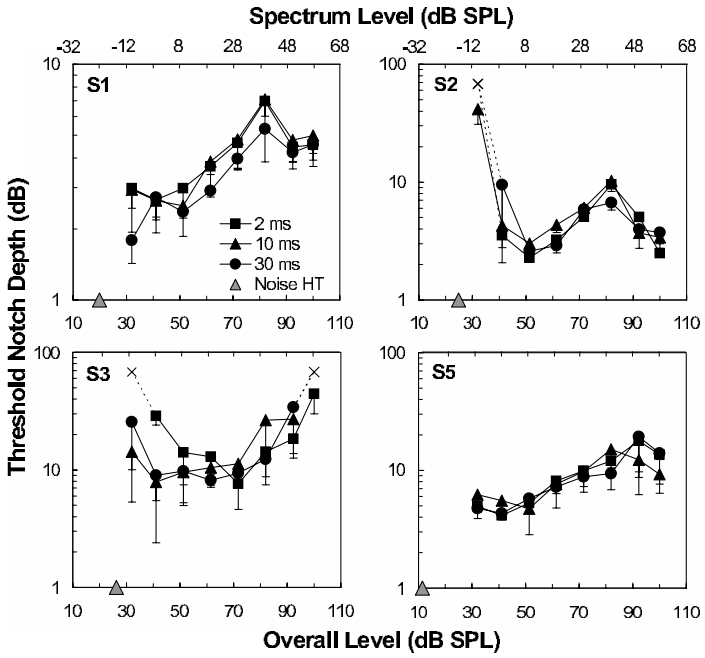

Figure 2.7: Threshold notch depth *vs.* level functions for stimuli with different rise times (as indicated by the inset in the top left panel). Error bars indicate the 68% confidence interval of the geometric mean. In all conditions, the notch had a bandwidth of 2000 Hz and stimuli had a total duration of 220 ms including the rise/fall times. Data were obtained using the FN-EL strategy. Each panel illustrates the results for a single listener. Shaded triangles on the abscissa illustrate the hearing thresholds (HTs) for the standard noise.

panel of Fig. 2.8 illustrates the ratios[2] of threshold notch depths for the long and the short stimuli for every listener and for the average. Although the ratio varies widely across listeners, its average value equals 2.5 and is approximately independent of level. This shows that the effect of level is comparable for the long and the short stimuli.

[2]Since threshold notch depth conforms to a lognormal distribution (Sec. 2.3.1), the ratio is more appropriate than the difference to quantify the effect of duration or of any other stimulus parameter.

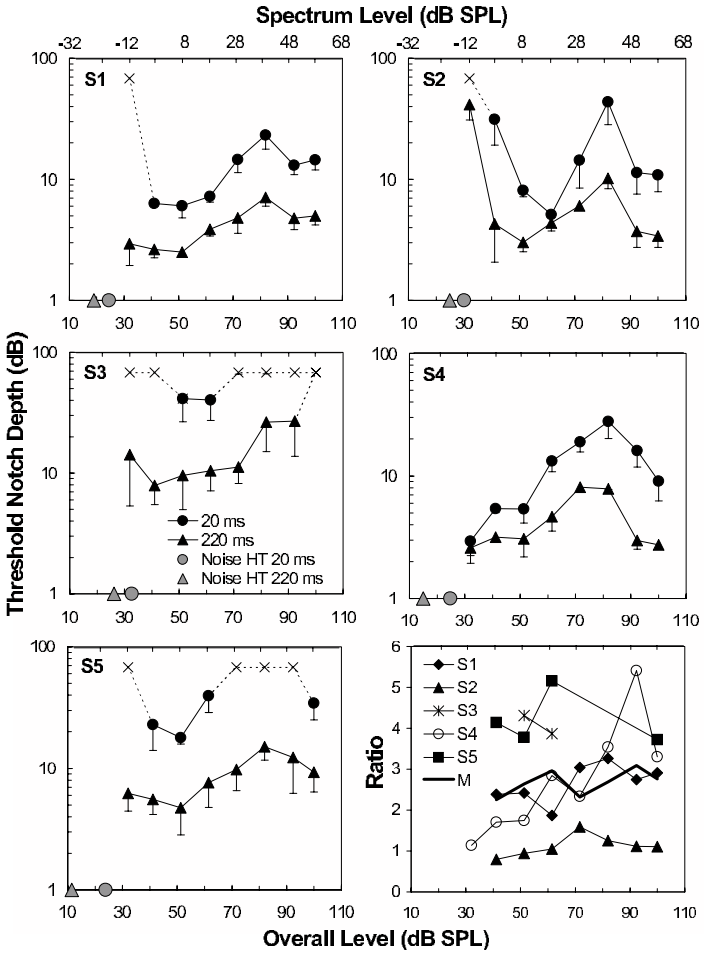

Figure 2.8: Threshold notch depth *vs.* level functions for stimuli with different durations (as indicated by the inset in the mid left panel). Error bars indicate the 68% confidence interval of the geometric mean. Stimuli rise/fall times were equal to 10 ms in both conditions. Notches had a bandwidth of 2000 Hz. Data were obtained using the FN-EL strategy (see text for details). Shaded triangles and circles on the abscissa indicate the listeners absolute hearing thresholds (HTs) for the standard noises of 220 and 20 ms of duration, respectively. Each panel illustrates the results for a single listener. The bottom right panel illustrates the ratio of threshold notch depths for the short and long (short/long) stimuli. Each symbol corresponds to a different listener. Note that on average (thick continuous line) the ratio is independent of stimulus level.

2.4 Dicussion

The aim of the study was to investigate how the detection of spectral notches with center frequencies well above the cutoff of phase locking depends on stimulus level. The effects of other factors such as the notch bandwidth, the stimulus duration and the stimulus rise time have also been investigated. Two hypotheses were tested. First, since the AN rate-profile representation of the stimulus spectrum deteriorates at high levels, threshold notch depths for discrimination should increase monotonically with increasing stimulus level. Second, if at high levels the stimulus spectrum were encoded mostly in the onset rate-profile, threshold notch depths should be smaller for stimuli with abrupt onsets (short rise times) and comparable for long and short stimuli.

Results show that none of these hypotheses hold. For notches 2000 and 4000-Hz wide, although the actual values vary widely across listeners, threshold notch depth is a *non*monotonic function of level (Figs. 2.3 and 2.4). A clear maximum in the function occurs for overall levels around 70-80 dB SPL for most listeners. Furthermore, threshold notch depths are hardly affected by the stimulus rise time (Fig. 2.7) and are clearly larger for noise bursts of 20 ms of duration than for bursts 220-ms in duration (Fig. 2.8).

2.4.1 Physiological interpretation

a. *The nonmonotonic character of the threshold notch depth vs. level functions*

Nonmonotonic threshold notch depth *vs.* level functions have been observed for both the equalized and roving level conditions (Fig. 2.3). For the reasons explained above (Sec. 2.2.2.b), this suggests that discriminating between the flat-spectrum and the notch noise bursts is based on detecting differences in overall spectral shape, rather than on detecting energy differences in the frequency region of the notch or elsewhere. The notches tested span a region of frequencies higher than the cutoff of phase locking (Johnson, 1980). Hence, they are unlikely to be encoded in the temporal aspects of the AN response. Instead, they must be encoded in the rate of the response, possibly as a rate profile. Therefore, the nonmonotonic character of the threshold notch depth *vs.* level functions suggest that the quality of the AN rate profile representation of the stimulus spectrum must be a nonmonotonic function of level. In other words, the stimulus spectrum must be reasonably well

encoded in the AN rate profile at low and high levels, but more poorly represented at 70-80 dB SPL.

Considering the classification of AN fibers suggested by Winter *et al.* (1990), this result may be interpreted as an indication that humans have HSR and LSR fibers but no MSR fibers, at least in the 8000-Hz region. In that case, it is possible that the notch be encoded in the rate profile of HSR fibers at low levels and in that of LSR fibers at high levels. The maximum that occurs for overall levels around 70-80 dB SPL (spectrum levels of 28-38 dB SPL) would reflect the transition between the dynamic ranges of the two. That is, it would occur at levels for which HSR fibers are almost saturated and LSR fibers are just below threshold. MSR fibers have thresholds that are intermediate between those of HSR and LSR fibers (Winter *et al.*, 1990, p. 199). Furthermore, over the range of levels (100 dB) used by Winter *et al.*, their dynamic range appears to be wider than that of HSR and LSR fibers as a result of their showing sloping saturation (cf. Fig. 2 of Winter *et al.*, 1990). Therefore, if MSR fibers were present in the human AN, they could probably encode for the spectral notch at moderate-to-high levels. As a result, the threshold notch depth *vs.* level functions would be unlikely to show a maximum and instead threshold notch depth would probably appear independent of level.

The exact way that the two fiber populations translate into nonmonotonic threshold notch depth *vs.* level functions is unclear. However, the above interpretation imposes the restriction that the saturation threshold of HSR fibers and the rate threshold of LSR fibers should both occur around the level for which the maximum threshold notch depth occurs. That is, both thresholds would be around an overall level of 70-80 dB SPL or its equivalent spectrum level of 28-38 dB SPL (Figs. 2.3 and 2.4). For this range of spectrum levels, the effective SPL seen by an AN fiber with a CF = 8000 Hz and an effective bandwidth of 1000 Hz would be approximately equal to 38-48 dB SPL. These values agree well with those estimated from pure-tone AN rate-level functions for other mammalian species (e.g., Fig. 1 of Winter and Palmer, 1991).

The above interpretation also agrees with the conclusions of Zeng *et al.* (1991). They studied the effects of prior stimulation upon intensity discrimination and found unusually large just-noticeable differences for midlevel sounds. Zeng *et al.* (1991) argued that their result is consistent with humans having only two types of auditory nerve fibers, HSR and LSR, each

with a different time of recovery from adaptation and a different threshold [see Zeng *et al.* (1991) for a full discussion].

It has been shown (Fig. 2.5) that threshold notch depths are, with some exceptions, comparable for stimulation levels of 42 and 100 dB SPL so long as the notches are wide enough ($\geq$ 2000 Hz). This is somewhat surprising considering that the frequency response of AN fibers broadens with increasing sound level (Rose *et al.*, 1971), hence the spectral notch should be less clearly represented in the AN rate profile at the highest level tested. One possible explanation for the result is that peripheral suppression may enhance the spectral notch (as represented in the AN rate profile) at high levels but less so (or nothing at all) at low levels, somehow compensating for the broadening of the fibers frequency response at high levels. This is a likely explanation because, as discussed by Poon and Brugge (1993), suppression would be expected to cause a decrease in the discharge rate of AN fibers with CFs in the notch region, hence enhancing the contrast between the energy level at the spectral notch frequency and that at the surrounding frequencies, and the amount of suppression is more prominent at high levels (e.g. Schalk and Sachs, 1980).

Other explanations exist for the nonmonotonic character of threshold notch depth *vs.* level functions. For example, discrimination based on AN rate responses is likely to be directly proportional to the rate difference between the two stimuli but inversely proportional to their corresponding rate variances (e.g., Young and Barta, 1986; Winter and Palmer, 1991). Consequently, small rate variances may lead to discriminable percepts even when the associated rate difference is small. The variance-to-mean ratio of AN rate responses decreases with increasing mean rate, that is, with increasing stimulus level (e.g., Winter and Palmer, 1991; Young and Barta, 1986). Therefore, it is possible that the improvement in spectral discrimination observed at high levels results from a reduction in the fibers variance-to-mean ratio in combination with (or perhaps instead of) the above mentioned spectral-enhancing effects of suppression.

Basilar membrane (BM) compression may also contribute to the nonmonotonic character of the threshold notch depth *vs.* level functions. Some studies (though not all; see Robles and Ruggero, 2001, pp. 1308-1309) suggest that the basal region of the BM responds linearly at low ($<$ 40 dB SPL) and high ($>$ 90 dB SPL) levels, but compressively at intermediate levels. If this were the case, the spectral notch would be more clearly represented in the BM excitation pattern at low and high levels, where linear

responses occur, than at moderate levels, for which the notch would appear shallower in the BM excitation pattern as a result of compression. Additionally, cochlear distortion may negatively affect the BM representation of the spectral notch as a result of distortion-product energy traveling from remote BM sites, where they are generated, to those with CFs within the notch band (cf. Robles *et al.*, 1997). This effect is likely to be maximal at moderate levels for which compression is greatest. On the other hand, the effects of BM compression are unlikely to explain by themselves the nonmonotonic character of the function. BM compression is approximately constant over a wide range of input levels (∼40-90 dB SPL; e.g., Ruggero *et al.*, 1997); hence, should the previous explanation be correct, it would lead to a plateau-shaped nonmonotonic function rather than to the peak-shaped curve that is more common in the present results.

It must be acknowledged that all the above interpretations implicitly assume that listeners use the same cues at all levels and that the internal representation of the spectral notch degrades at levels around 70-80 dB SPL. However, an alternative explanation could be that the cues used for spectral discrimination are different at low and high levels. It could well happen that the cues that are salient at low levels degrade as the level increases and that new cues improve with increasing level to become most salient at high levels. The nature of these cues is uncertain at this time. It is unlikely that spectral discrimination at high levels be facilitated by distortion in the stimuli because, as shown in Fig. 2.1(c), the actual difference spectrum between the target and the standard noises are comparable for overall levels of 50, 80 and 100 dB SPL.

It should be noted that the explanations discussed above are not mutually exclusive. They can all contribute simultaneously to explain the data. On the other hand, it is not the purpose of this report to provide a convincing explanation of the nonmonotonic effect.

b. *The effect of notch bandwidth and its interaction with level*

Detection of spectral notches with bandwidths equal or greater than 2000 Hz was generally possible at all levels. On the other hand, detection of notches 1000-Hz wide was clearly much more difficult, particularly at high levels.

Notch detection must be based on spectral discrimination. Given the nature of the stimuli, the most salient information for spectral discrimination is likely to be provided by AN fibers with CFs within and near the notch

band (cf. Poon and Brugge, 1993). Those fibers will be driven with less energy by the target, notched stimulus than by the reference, flat-spectrum one. The effective bandwidth of those fibers is around 1000 Hz at low SPLs (e.g. Fig. 10 of Evans, 1975) and increases as the SPL increases (Rose *et al.*, 1971). Therefore, at high levels, the fibers in question will also respond to energy in frequency regions adjacent to the 1000-Hz wide notch (see Chapter 8 in Lopez-Poveda, 1996). Hence, their response to the notched stimulus will be only slightly smaller than that to the flat-spectrum stimulus, making discrimination between the two more difficult even when the notch is very deep. In psychophysical terms, this relates to the fact that the critical band at 8000 Hz is comparable to the bandwidth of the 1000-Hz wide notch (Glasberg and Moore, 1990) and increases with level (e.g. Baker *et al.*, 1998). Furthermore, the inter-listener variability observed in the present results, particularly for the narrowest notch (Fig. 2.6), is likely to relate to the wide inter-listener variability in auditory filter bandwidth (Patterson and Moore, 1986). As noted above (section 2.3.2), this may account for the poorer performance of listener S3.

Spectral discrimination is also likely to be affected, though to a lesser extent, by the number of fibers with CFs within the notch band. This number increases as the notch bandwidth increases and this might explain, at least in part, that threshold notch depths be lowest for stimuli with the widest notch. In any case, the shape of the threshold notch depth *vs.* level function is comparable for different notch bandwidths (see Fig. 2.6), which suggests that the nonmonotonic effect is independent of the number of fibers signaling for the notch.

c. *The effect of stimulus duration*

AN fibers show a wider dynamic range at the onset of the stimulus (Smith and Brachman, 1980). For this reason, Delgutte and Kiang (1984a, b) suggested that at high levels, when the rate response of the majority of AN fibers is saturated, the stimulus spectrum may still be conveyed in the onset rate profile. Above, it has been shown that maximizing the effects of the stimulus onset by reducing the stimulus rise time hardly affected threshold notch depths (Fig. 2.7). It has also been shown that threshold notch depths were considerably lower for a stimulus duration of 220 ms than for a duration of 20 ms, even at the highest levels tested (Fig. 2.8). If at high levels notch detection were based solely on the onset representation of the stimulus spectrum, no differences would be observed between the long and the short

stimuli. Therefore, although the present results do not rule out a possible contribution of the onset rate profile to the encoding of the stimulus spectrum, they suggest that the contribution in question is less important than previously suggested (Delgutte and Kiang, 1984a, b; Lopez-Poveda, 1996).

Since both the stimulus and the nature of the AN response are stochastic, the rate-profile representation of the stimulus spectrum surely varies over time. The fact that threshold notch depths are smaller for the longer stimuli across *all* levels tested (Fig. 2.8) suggests that the longer stimulus provides a clearer representation of the stimulus spectrum. It is possible that the brain integrates (or averages) the rate profile over time with a relatively long time window (Green, 1960). Alternatively, it is possible that the decision be based on taking multiple looks at rate profiles computed over a much shorter time window, as suggested by Viemeister and Wakefield (1991). Either of the two mechanisms serves to explain (at least qualitatively) the smaller values of notch depth for the longer stimulus.

Furthermore, it is noteworthy that the effect of level is comparable for the short and the long stimuli when threshold notch depths are plotted on a logarithmic scale. This indicates that the nonmonotonic character of the threshold notch depth *vs.* level function is independent of stimulus duration and, consequently, that it is not determined by temporal integration or multiple looks.

2.4.2 Relation with level discrimination results

The dynamic-range problem has been widely investigated by previous psychophysical studies of level discrimination. Indeed, the present results resemble those of level discrimination experiments in a number of respects. First, here it has been shown (Figs. 2.1 and 2.4) that threshold notch depth varies nonmonotonically as a function of level. Similarly, the threshold for detecting a level change in a high-frequency stimulus is a nonmonotonic function of level, both for tones (e.g. Carlyon and Moore,1984; Florentine, 1986; Florentine *et al.*, 1987) and noise (e.g. Buus, 1990; Nizami *et al.*, 2001). Second, the nonmonotonic aspect of the threshold notch depth *vs.* level function appears more obvious for the short (20 ms) than for the long (220 ms) stimuli when the ordinate is on a linear scale, rather than on the logarithmic scale used in Fig. 2.8. The same is true for the level-change *vs.* level function (e.g., Carlyon and Moore, 1984; Nizami *et al.*, 2001). Finally, here it has been shown that dynamic responses to the signal onset play little role in

spectral discrimination (Fig. 2.7) and this seems to be the case also for level discrimination (Carlyon and Moore, 1984).

The results of level-discrimination experiments led Carlyon and Moore (1984) to suggest a bimodal distribution of thresholds in human primary auditory neurons. Such a suggestion is in line with the two types of fibers interpretation proposed above for the nonmonotonic aspect of the threshold notch depth *vs.* level function. Of course, detecting a level change is not the same as detecting a spectral notch. Furthermore, as discussed above, the detection of the spectral notch is unlikely to be based on comparing the levels of the standard and the target noises (neither the overall level or the level within or outside the notch band). Instead, it is likely to depend on the quality with which the spectral notch is represented in the overall auditory nerve rate profile. Nevertheless, the operation range of auditory nerve fibers will determine, in a similar way, the level-dependence of the quality of the internal representation of the notch (i.e., its depth in the rate profile) and the fibers difference response to a change in the stimulus level.

2.4.3 Relation with sound localization results

Sound localization relies partly on accurately detecting the spectral content of the HRTF. Specifically, it has been reported that spectral notches with sound source elevation-dependent center frequencies constitute prominent cues for judging the vertical location of sound sources (e.g., Butler and Belendiuk, 1977; Butler and Humanski, 1992). On the other hand, the bandwidth and the depth of HRTF notches vary widely across listeners [see, for instance, Fig. 3 in Shaw (1982) or Chap. 3 in Lopez-Poveda (1996)], possibly reflecting differences in the shapes and sizes of the listeners ears (Lopez-Poveda and Meddis, 1996). The present study shows that the ability to detect high-frequency spectral notches varies widely across listeners (Fig. 2.4), and depends on the notch bandwidth (Fig. 2.6) as well as on the stimulus level and duration (Figs. 2.3 and 2.8). As a result, the ability of listeners to actually use the notches in their individual HRTFs as cues for sound localization must depend on a complex combination of their level of performance in notch detection tasks, the shape of their ears, and the characteristics of the stimulus (duration and level). In any case, since vertical localization relies on detecting HRTF notches, the present results suggest that, in general, localization judgments should be more precise for long stimuli than for short ones and for levels below 60-70 dB SPL than for levels around 70-80 dB SPL and this is indeed the case (Hartmann and Rakerd, 1993; Macpherson

and Middlebrooks, 2000; Vliegen and Van Opstal, 2004). Furthermore, in the light of the present results an improvement in precision should occur for levels higher than 80 dB SPL, although this remains to be tested.

Despite the emphasis given here to spectral notches, it must be acknowledged that other authors (e.g., Blauert, 1969/70; Humanski and Butler, 1988) have suggested that spectral peaks may be as important for sound localization as spectral notches, if not more. The auditory nerve rate-profile representation of spectral peaks at high levels would be also negatively influenced by the limited dynamic range of most nerve fibers. On the other hand, the broadening of the fibers frequency response with increasing level would not deteriorate the rate-profile representation of peaks. If anything it would enhance it by spreading the energy of the peak to fibers with a wider range of CFs. Consequently, comparing the present results with those of experiments aimed at discriminating between flat-spectrum and peaked noise bursts may help elucidating the relative contribution of both mechanisms (i.e., limited dynamic range and filter broadening) to the auditory nerve representation of spectral features.

2.5 Conclusions

The extent that the detection of high-frequency spectral notches is affected by stimulus level was investigated by measuring the threshold notch depth necessary to discriminate between a flat-spectrum wideband noise and a similar noise with a rectangular spectral notch centered at 8000 Hz. The main conclusions are:

1. High-frequency spectral notches are detected by detecting differences in the overall spectral shape of the stimulus rather than by detecting level differences over certain frequency regions.

2. For a large proportion of listeners, threshold notch depth is clearly a *non*monotonic function of stimulation level. It increases for levels up to 70-80 dB SPL and decreases for higher levels; thus a maximum in the function occurs at levels around 70-80 dB SPL. Interpretations of this result have been discussed based on the physiological response properties of the mammalian auditory nerve.

3. The nonmonotonic character of the threshold notch depth *vs.* level function is independent of notch bandwidth and stimulus duration.

4. Threshold notch depths are hardly affected by the stimulus rise time and depend strongly on the stimulus duration. Notch depth values are larger for shorter stimuli. These results suggest that the onset rate-profile contributes less than previously thought to encoding for the stimulus spectrum at high levels. It is also consistent with the idea that the detection of the stimulus spectrum involves either integrating the AN rate profile over a relatively long time window or multiple looks at the rate profiles computed over relatively narrow time windows.

Chapter 3

Psychophysical assessment of the level-dependent representation of high-frequency spectral notches in the auditory nerve[1]

3.1 Introduction

In the previous chapter, it was observed that discriminating between broadband noises with and without high-frequency spectral notches is, for most listeners, more difficult around 70-80 dB SPL than at lower or higher levels (Alves-Pinto and Lopez-Poveda, 2005). This challenges the common notion (reviewed by Lopez-Poveda, 2005) that spectral features beyond the cut-off frequency of phase locking (>4 kHz; Palmer and Russel, 1986) must be represented in the AN by means of a rate-place code (Sachs and Young, 1979, Rice *et al.*, 1995; Lopez-Poveda, 1996, 2005) and that the quality of this representation deteriorates gradually with increasing sound level (Rice *et al.*, 1995; Lopez-Poveda, 1996). As an explanation to this challenging observation, it was proposed that the result could be consistent with the existence of two types of AN fibers with different thresholds and dynamic ranges (the so-called high- and low-spontaneous rate fibers; Liberman, 1978) and that discrimination might be most difficult at 70-80 dB SPL because this could be the transition level between the dynamic ranges of the two fiber types.

[1]This chapter is based on the paper Alves-Pinto, A. and Lopez-Poveda, E.A. (**submitted**). "Psychophysical assessment of the level-dependent representation of high-frequency spectral notches in the auditory-nerve," J. Acoust. Soc. Am.

This explanation would be consistent with the AN rate-place representation of high-frequency spectral features and would imply that the AN representations of two broadband stimuli with different high-frequency spectra should be most similar at 70-80 dB SPL than at lower or higher levels. The study described here aimed at testing this hypothesis behaviorally by comparing the masking patterns of the noise stimuli used in the spectral discrimination task of Chapter 2.

A masking pattern (or masked audiogram) is a graphical representation of the detection thresholds of masked probe tones of different frequencies as a function of the probe frequency. Psychophysical forward masking is thought to reflect (to a large extent) the recovery of auditory nerve fibers from previous stimulation and/or the persistence of neural (post AN) activity (Meddis and O'Mard, 2005; Oxenham, 2001). Whatever the case, the detection of a low-level tonal probe is assumed to depend on the discharge rate evoked by the probe in AN fibers with CFs approximatelly equal to the probe frequency. When preceded by a masker sound, such rate depends almost certainly on the discharge rate evoked by the masker on those fibers (Harris and Dallos, 1979; Meddis and O'Mard, 2005). Temporal interactions between the masker and the probe (e.g. suppression) are minimized in forward masking. Therefore, the masking pattern provides possibly the best psychoacoustical correlate of AN rate-place (or rate-profile) representation of the masker spectrum (Moore, 1997, 2005). If the abovementioned explanation for the result of chapter 2 were correct, then the difference between the masking patterns of broadband noises with and without high-frequency spectral notches should be negatively correlated with the ability to discriminate between the two spectra. In other words, the more similar the masking patterns, the harder the discrimination task should be.

Two experiments were carried for the same group of listeners. Experiment I was identical to that of the discrimination experiment of Chapter 2. Experiment II consisted in measuring the masking patterns of two broadband-noise maskers, one with a flat spectrum and one with a spectral notch at 8 kHz, for several masker levels. Correlations were sought between the results of the two experiments as a way to test the above mentioned explanation for the results of chapter 2. The results indicate that the deterioration in spectral discrimination with increasing level up to 70-80 dB SPL (Alves-Pinto and Lopez-Poveda, 2005) is consistent with a deterioration of the spectral-notch representation in the masking patterns. By contrast, the improvement in spectral discrimination above 80 dB SPL does *not* corre-

spond to an improvement in the quality of the representation of the spectral notch in the masking pattern. It will be discussed that the discrimination of high-frequency spectral features is unlikely to rely solely on comparisons of the rate-place representation of the spectra, at least above 80 dB SPL, and that some form of temporal representation may contribute.

3.2 Experiment I - Discrimination between flat spectrum and notch noises

The experiment was identical to the main experiment of chapter 2. It consisted in measuring the level-dependence of the threshold notch depth for discriminating between a flat-spectrum broadband noise and a similar noise with a high-frequency rectangular spectral notch (Fig. 1).

3.2.1 Methods

a. Stimuli

Stimuli consisted of bursts (total duration = 220-ms, including 10 ms rise/fall ramps) of random broadband (20-16000 Hz) noise with either a flat spectrum or with a rectangular spectral notch (bandwidth = 2 kHz) centred at 8 kHz (Fig. 3.1). The noises were generated as described in Section 2.2.2. The reduced energy in the notch frequency band would have made the overall level of the notch noise slightly lower than that of the flat-spectrum noise. To prevent that this level difference be used as a cue for discrimination, the overall levels of the two noises were made equal by reducing the spectrum level of the flat-spectrum noise with respect to that of the notch noise. The reduction required for any given notch depth was determined as described in Eq. 2.1.

b. Procedure

The procedure was identical to that described in chapter 2 (see also Alves-Pinto and Lopez-Poveda, 2005). Two flat-spectrum and one notch noise bursts were played in random order to the listener (with an inter-stimulus period of 500 ms), who was instructed to identify the odd-one out. A two-down, one-up adaptive procedure with feedback was employed to estimate the notch depth that produced 70.7% correct responses (Levitt, 1971). The initial notch depth was fixed at 20 dB below the reference spectrum level of the noise. The notch depth decreased/increased by 6 dB for the first

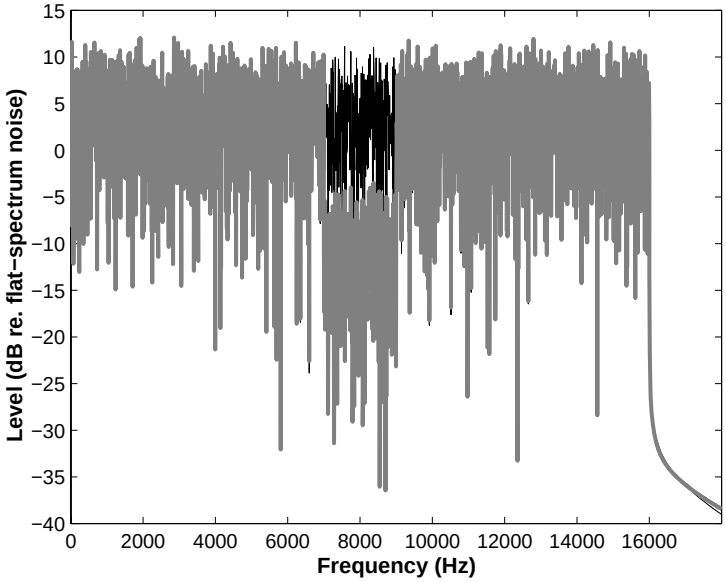

Figure 3.1: Spectra of the broadband flat-spectrum (dark trace) and notch noises (grey trace). The spectral notch was centred at 8000 Hz and its bandwidth was 2000 Hz. The notch illustrated had a depth of 15 dB relative to the spectrum level outside the notch band. Note that the two traces overlap outside the notch frequency band. The plotted spectra correspond to the FFT of the time-domain digital signal.

six turn points and by 1 dB thereafter. Sixteen turn points were recorded for each measurement and the threshold estimate was taken as the mean of the notch depths for the last 10 turn points. When the corresponding standard deviation exceeded 6 dB, the measurement was discarded and a new threshold estimate was measured. The thresholds reported correspond to the *geometric* mean of at least three valid measurements (see Chapter 2 and Alves-Pinto and Lopez-Poveda, 2005). Notch depths at threshold were measured for stimulus overall levels from 40 to 90 dB SPL, in 10-dB steps, and for a last level of 95 dB SPL.

Listeners were tested individually in a double-wall sound attenuating chamber. Stimuli were generated digitally (24-bit, sampling rate of 48.8 kHz) with a TDT^{TM} Psychoacoustics workstation (System 3) and delivered via Etymotic^{TM} ER2 earphones. The sound pressure levels (SPLs) reported below correspond to calibrated values (see Chapter 2 and Alves-Pinto and Lopez-Poveda, 2005 for details on the calibration procedure).

c. Listeners

Four volunteers (aged 26 to 33) participated in this experiment. They all had hearing thresholds within 20 dB re. ANSI 3.6-1996 (Specifications for Audiometers) at the audiometric frequencies (250-8000 Hz). All listeners were given at least one training session in the task. Listener S1 was the author.

3.2.2 Results and discussion

Figure 3.2 illustrates the threshold notch depths as a function of noise overall level. Each panel illustrates the results for a different listener. Error bars illustrate one standard error of the geometric mean (see Section 2.3.1). Triangles illustrate absolute thresholds for the flat-spectrum noise.

For the four listeners, threshold notch depths vary nonmonotonically with level: they increase gradually with increasing level up to 70-80 dB SPL and then decrease with further increases in level. In other words, the results suggest that discrimination is most difficult at 70-80 dB SPL than at lower or higher levels. The non-monotonic trend is less clear for S4, but still the threshold notch depth is highest at 70 dB SPL. In conclusion, these results are in line with those of Chapter 2.

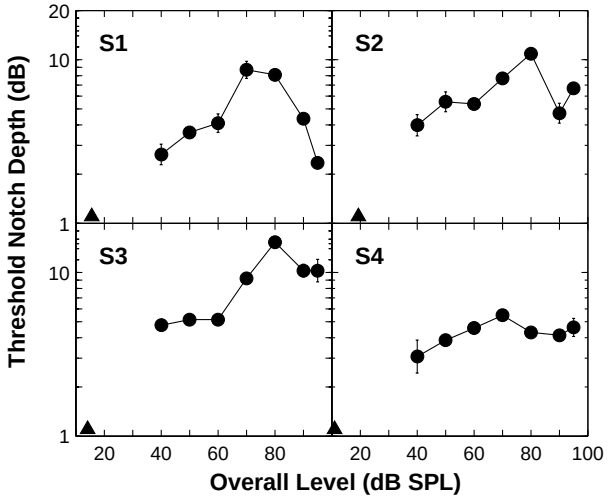

Figure 3.2: Threshold notch depths for discriminating between a flat-spectrum noise and a notch noise plotted as a function of the stimulus overall level. Notches were centred at 8000 Hz and had a bandwidth of 2000 Hz. Each panel illustrates data for a single listener as indicated in the upper-left corner of each panel. Each data point corresponds to the geometric mean of at least three measurements. Error bars illustrate one standard error of the geometric mean. Shaded triangles on the abcissa indicate the listeners absolute hearing threshold for the flat-spectrum noise.

3.3 Experiment II - Masking patterns of flat-spectrum and notch noises

3.3.1 Rationale

The aim was to measure the masking patterns for the flat-spectrum and the notch noises used in Experiment I and for the same group of listeners. The nonlinear processing characteristic of the basilar membrane (BM) the inner hair cell (IHC), and the IHC-AN synapse may all contribute to deteriorating the quality of the internal (rate-place) representation of the spectral notch. Consider, for instance BM compression with a degree of 4 to 1 dB/dB (Ruggero *et al.*, 1997). For stimulus spectrum levels within the range of BM compression, the spectral notch would appear approximately 4 times shallower in the BM excitation pattern than in the original spectrum. In other words, compression, among other factors, may alter the quality of the internal representation of the spectral notch.

On the other hand, in forward masking, it is commonly assumed that neural activity recovers exponentially after masker excitation and that probe detection occurs when the response to probe reaches a certain threshold value. An alternative interpretation would be that the masker excitation decays exponentially with time and that the probe detection threshold occurs when the internal probe excitation just exceeds the residual masker excitation at the time of the probe (Meddis and O'Mard, 2005; Oxenham and Moore, 1994; Oxenham, 2001). Whatever the interpretation, however, evidence exists that the time constants of recovery (or decay) are level independent over a wide range of levels. If the masker-probe time interval remained constant for all masker levels, then the probe detection thresholds would increase with increasing masker levels, and the increase would be approximately comparable for the masker and the probe, regardeless of the amount of compression. Thus, varying the masker level while keeping the masker-probe interval fixed would not reveal the negative effect of the compression on the quality of the internal representation of the spectral notch.

A better approach would be to vary the masker-probe time interval so that the probe detection thresholds remain approximately constant for all masker levels. This is the approach considered here. Masking patterns were measured using forward masking for masker overall-levels of 50, 70, 80 and 90 dB SPL. Different masker-probe time intervals were used for different masker levels and listeners with the aim to maintain the probe thresholds approxi-

matelly constant accross masker levels and equal to the masked thresholds for the 50-dB SPL masker.

The probe absolute thresholds were also measured as a control to verify that there was a masking effect.

3.3.2 Methods

a. Procedure

The thresholds for detecting probe tones of different frequencies in the presence of a forward noise masker were measured using a two-interval two-alternative forced-choice adaptive paradigm. In one of the intervals the masker was presented alone; in the other interval, the masker was followed by a brief probe tone. The two intervals were presented in random order (with an inter-stimulus period of 500 ms) and the listener was instructed to identify the interval containing the probe. Feedback was immediately given to listener after her response. The initial probe level was set so that the probe was perfectly audible at the beginning of the trial. A two-down, one-up adaptive procedure was used to estimate the probe level corresponding to 70.9% correct responses in the psychometric function (Levitt, 1971). The probe-level increment was 6 dB for the first two reversals and 2 dB thereafter. The measurement ended after twelve reversals were recorded and the threshold was estimated as the arithmetic mean of the probe level for the last 10 reversals. When the corresponding standard deviation exceeded 6 dB, the measurement was discarded and a new estimate was obtained. Three threshold estimates were obtained in this way, the mean of which was taken as the true threshold. When the standard deviation of those three estimates exceeded 3 dB, an additional estimate was measured and included in the mean.

Measuring a masking pattern consisted in measuring the detection thresholds of fifteen tonal probes of different frequencies (see the following text). Each experimental block consisted in measuring one group of fifteen probe tones, but listeners were given a short (∼5-10 min) resting period in the middle of each block to avoid fatigue.

The absolute thresholds for the probes (i.e., without the masker) were also measured using the same procedure.

Listeners were always tested in the same conditions and with the same equipment as described in Experiment I. The sound pressure levels reported

below correspond to calibrated values. The details of the calibration procedure can be found elsewhere (Alves-Pinto and Lopez-Poveda, 2005).

b. Stimuli

To measure a masking pattern consisted in measuring masked detection thresholds for tonal probes of the following frequencies: 5, 6, 6.5, 6.75, 7, 7.25, 7.5, 8, 8.5, 9, 9.25, 9.5, 9.75, 10, 11 kHz. The probes had a total duration of 10 ms, including 5-ms onset/offset ramps and no steady-state portion.

Masking patterns were measured for two broadband random-noise maskers (20-16000 Hz) with different spectra: one was flat and one was similar except for it contained a rectangular notch centred at 8 kHz, with a bandwidth of 2 kHz and a fixed depth of 15 dB (Fig. 3.1). These noise signals were similar to those used in Experiment I and were generated as described in Chapter 2. The overall level of the notch-noise masker was 0.56 dB lower than that of the flat-spectrum noise because of its having less energy in the notch frequency band. Unlike in Experiment I, this level difference was not compensated for and so the two maskers had identical spectral levels outside the notch band. Maskers had a total duration of 110 ms, including 5-ms onset/offset ramps. Masking patterns were measured for masker overall levels of 50, 70, 80 and 90 dB SPL.

A fixed 2-ms masker-probe time interval (defined from masker offset to probe onset) was used for the 50 dB SPL masker and for all listeners. For higher masker levels, longer masker-probe intervals were used as necessary (see above) so that the masked threshold of the 8-kHz probe was approximately constant across masker levels. An 8-kHz tone was used because it is at the centre of the notch frequency band. The actual intervals were determined as follows. The masked threshold for the 8-kHz probe tone was measured (as explained above) for the flat-spectrum noise masker for several masker-probe intervals; typically 2, 10, 20 and 30 ms. The flat-spectrum noise masker was used because it was the standard stimulus in Experiment I. The resulting data points (Fig. 3.3) were then fitted with either a 1st or 2nd order polynomial (continuous lines in Fig. 3.3), depending on the number of points measured and the fit error. The fitted polynomial was then used to predict the masker-probe intervals that for each masker level would give a masked threshold for the 8-kHz probe approximately equal to the one observed for the 50-dB SPL masker. The predicted value was then confirmed

Listener	8 kHz Probe Threshold (dB SPL)	Masker Level (dB SPL)			
		50	70	80	90
S1	22.5	2	16	19	23
S2	27.6	2	19	27	34
S3	28.5	2	23	37	100
S4	19.5	2	11	12	24

Table 3.1: Individual masker-probe time intervals (ms) used to measure the masking patterns for each of the four masker levels tested. The individual threshold for the 8-kHz probe in the absence of a masker are also shown.

experimentally (by measuring the masked probe threshold for the predicted time interval) and, if necessary, readjusted. This occurred for example for listener S3 at 80 and 90 dB SPL, for whom masked-probe thresholds varied by ∼6 dB across measures. In other words, the actual masker-probe intervals used (given in Table I) were approximations to the values predicted by the fit.

c. Listeners

The same four listeners who participated as volunteers in Experiment I took part in this experiment. They were given one or two training sessions on the task.

3.3.3 Results

Figure 3.4 illustrates the masking patterns for the flat-spectrum (open circles) and notch noises (filled circles) for different listeners at different masker levels. Each row corresponds to a different listener. Each column corresponds to a different masker level, as indicated at the top. The squares in the first column illustrate the absolute thresholds for the probes. Error bars represent one standard error of the mean. Vertical dotted lines depict the notch frequency band.

Note, first, that masked thresholds for any given listener are similar across masker levels, as was intended (see the preceding text). Further, they are always higher than the absolute thresholds, which confirms that there was a masking effect.

The masking patterns are not flat even for the flat-spectrum noise and

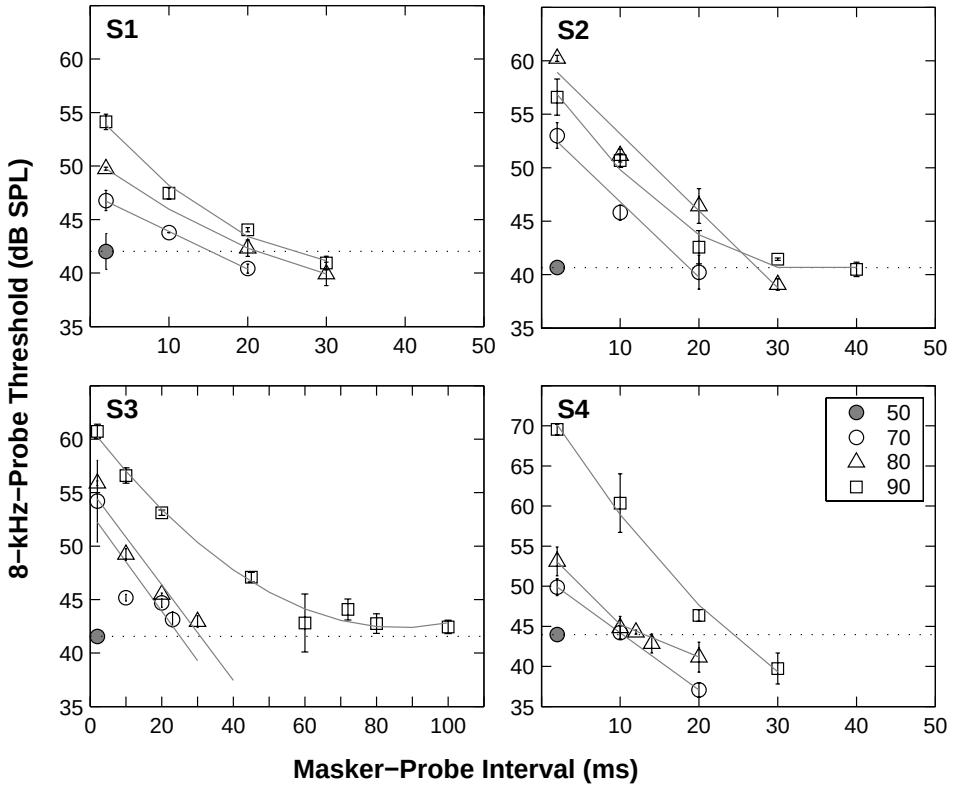

Figure 3.3: Detection thresholds of an 8-kHz probe tone masked by a forward notch noise masker as a function of the masker-probe time interval. Each panel illustrates the results of a different listener. Different symbols illustrate the results for different masker levels, according to the inset of the lower right panel (values in dB SPL). Error bars represent one standard error of the mean. The horizontal dotted line indicates the detection threshold for a masker level of 50 dB SPL and a masker-probe time interval of 2 ms. Continuous lines illustrate 1^{st}or 2^{nd} order polynomial fits to the experimental functions.

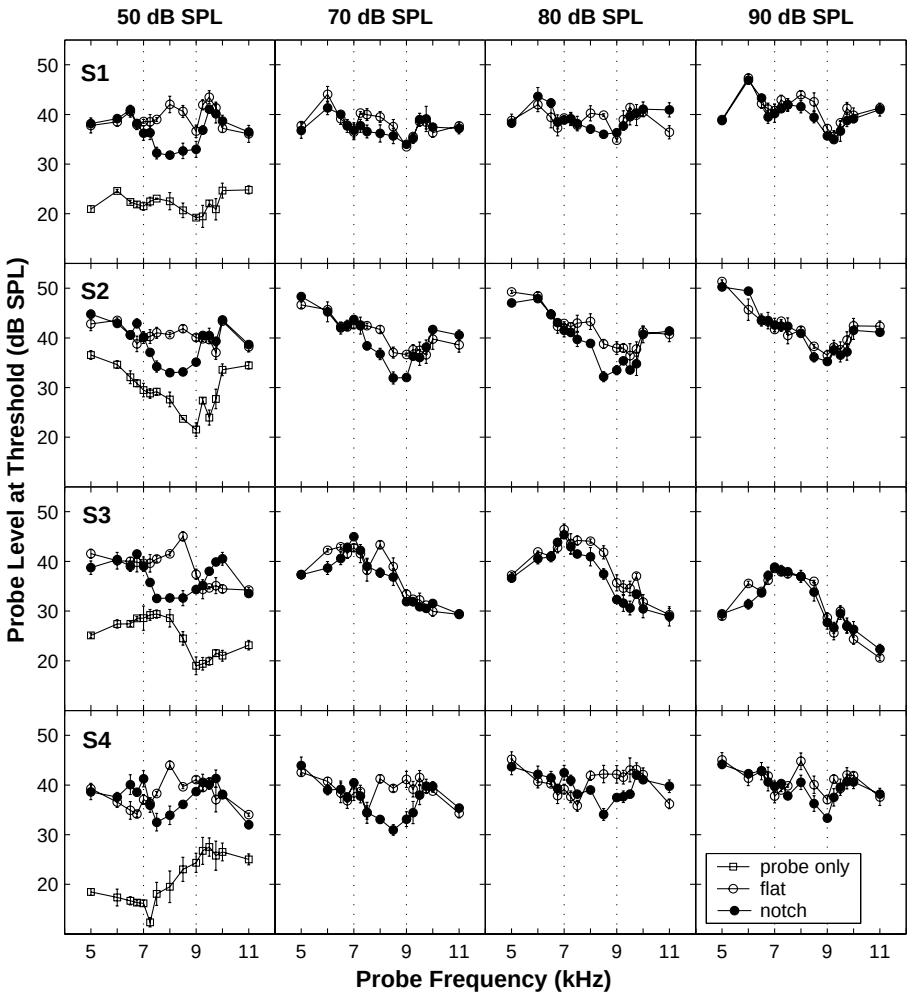

Figure 3.4: Masking patterns for the flat-spectrum (open circles) and the notch (filled circles) noise forward maskers. Each panel illustrates the results for a different listener and level. Different rows correspond to different listeners. Each column illustrates the masking patterns for the masker level indicated at the top. The squares in the first column represent the absolute detection thresholds for probes, i.e. the thresholds in the absence of the masker. Error bars represent one standard error of the mean. Dashed lines illustrate the boundaries of the spectral notch in the notch-noise masker.

their shapes vary significantly across listeners. Still, masked thresholds for probes around the notch frequency band are clearly lower in the notch-noise masking pattern than in the flat-spectrum one, while those for probes outside the notch band are very similar for the two maskers. The notch is particularly evident in the masking pattern of the 50-dB SPL masker. It is less obvious but still visible at 70 dB SPL, especially for listeners S2 and S4, and at 80 dB SPL for listener S4, but faints at 90 dB SPL for all listeners.

The effect of level on the relative shape of the masking patterns can be better seen by analysing the difference masking pattern (Fig. 3.5). This form of analysis eliminates the dependence of the masked thresholds (hence of the masking patterns) on the individual frequency sensitivity of the listeners. Each row of Fig. 3.5 illustrates the difference masking patterns for a different listener and the bottom row illustrates the mean across the four listeners. Each column corresponds to a different masker level. Error bars in the bottom row represent the corresponding standard error of the mean across listeners. Error bars in any other row represent the standard error of the mean difference. Dotted lines indicate the difference between the spectra of the two maskers.

For all listeners there is a clear trough in the difference masking pattern at 50 dB SPL which becomes gradually less obvious as the masker level increases. Furthermore, the trough seems to become wider and its tip shifts slightly towards the higher frequencies. The latter is consistent with the upward spread of masking described elsewhere (Moore, 2005).

For listeners S2, S3 and S4 peaks occur at the notch edges, which can be as high as 5 dB. These peaks almost certainly occur as a result of suppression effects (Poon and Brugge, 1993).

3.3.4 Discussion

The masking patterns of flat-spectrum and notch noise maskers were measured at increasing masker levels to evaluate the effect of level on the quality of the internal representation of the spectral notch. The main result was that the quality deteriorates gradually as the overall level of the noise increases from 50 up to 90 dB SPL (Figs. 3.4 and 3.5).

Probe detection thresholds were similar across masker levels (Fig. 3.4). This suggests that any given probe must have produced approximately the same excitation for all masker levels. Therefore, the effects of masker level on

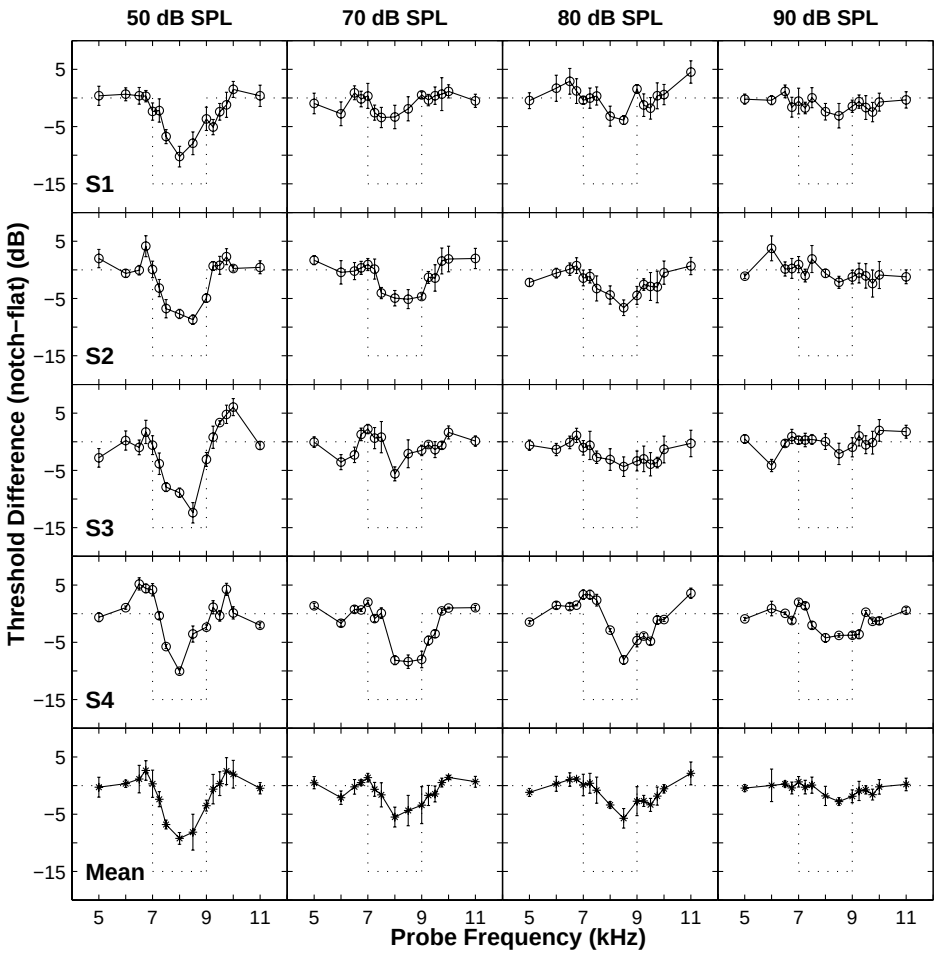

Figure 3.5: . Difference masking patterns. The difference between the masking patterns for the flat-spectrum and notch noise maskers (notch - flat). Each column illustrates the results for a different masker level (as indicated at the top). Each row contains the results for a different listener. Error bars represent one standard error of the mean difference. The bottom row illustrates the mean difference across listeners and the error bars illustrate one standard error of the mean. The dotted lines illustrate the difference between the spectra of the noise maskers.

the masking patterns are likely to reflect differences in the internal excitation of the masker only. The thresholds of listener S3 were slightly lower for the 90-dB SPL masker than for the other masker levels. Still, the notch was already less clearly represented at 80 dB SPL and therefore the conclusion stated above remains valid even for this case.

As explained in the preceding text, the shapes of the masking patterns may have also been influenced by off-frequency listening (Moore, 2005). Off-frequency listening, however, does not compromise the conclusion that the internal representation of high-frequency spectral notches deteriorates gradually as the stimulus level increases. One reason is that the amount (and effects) of off-frequency listening would have been approximately similar for the two noise maskers used, and hence would have cancelled out in the difference masking patterns. Second, masked thresholds were similar across masker levels and hence any off-frequency listening effects would have been comparable across masker levels.

A similar earlier study (Moore and Glasberg, 1983) investigated the effect of level on the quality of the internal representation of vowel formants by measuring the masking patterns of synthetic vowels in forward masking for overall levels of 50, 70 and 90 dB SPL. The masker was one of two synthetic vowels (/æ/ and /i/), both with a fundamental frequency of 100 Hz. Probe tones were presented immediately after the maskers and their detection thresholds were measured as a function of probe frequency. The authors concluded that the representation of the formant structure was impaired *only slightly* at high masker levels. The apparent contradiction between this conclusion and the present results may reflect that the spectral features considered in the two studies extended to very different frequency ranges: vowel formants were below 3.5 kHz while the present notch extended from 7 to 9 kHz. For broadband stimuli, the effective level driving an AN fibre is proportional to the fiber's bandwidth. The near-threshold bandwidth of AN fibres depends on the characteristic frequency of the fibre; the higher the CF, the larger the bandwidth (Evans, 1975). As a result, a broadband stimulus that drives low-CF AN fibres below saturation might drive high-CF AN fibres at saturation. At high levels, this effect would be much stronger for high- than for low-CF units, as their bandwidth increases substantially more with increasing level than that of low-CF units (Rose *et al.*, 1971). The probe frequencies used by Moore and Glasberg (1983) were much lower than those used here because they were used for measuring masking patterns over a lower-frequency range. Those probes would have been detected primarily

by monitoring the activity of AN fibers with lower CFs than the CFs of the fibers signaling the present probes. This might explain why the negative effect of increasing masker level was less obvious in the vowel masking patterns of Moore and Glasberg (1983) than in the present ones.

3.4 Correlation between the results of Experiments I and II

The masking patterns measured in Experiment II inform about the quality of the internal excitation patterns of the flat-spectrum and notch noises used in the discrimination task of Experiment I. These, in turn, are assumed to be closely related to the AN rate-place representation of the spectra of the two noises (see above). Therefore, to test the rate-place representation of high-frequency spectral features, qualitative correlations were sought between the threshold notch depths of Experiment I with the total difference between the two masking patterns of Experiment II. The latter was calculated as the square root of the sum (across frequencies) of the squared difference masking patterns (Fig. 3.5). This measure is akin to the Euclidean distance between the masked probe thresholds for the two maskers.

The lower panels of Fig. 3.6 illustrate the total difference between the masking patterns as a function of masker level. Different symbols illustrate the results for different listeners. For comparison, the upper panels of Fig. 3.6 show the discrimination threshold notch depths at the same four levels (adapted from Fig. 3.2). The results of both experiments are plotted separately for two level ranges: 50-80 (left panels) and 80-90 dB SPL (right panels). This facilitates the visual comparison of the results below and above 80 dB SPL, the level at which the trend in discrimination performance changes from increasing to decreasing with increases in level. Below 80 dB SPL, discrimination threshold notch depth increases with increasing level while the total masking-pattern difference decreases. In other words, discriminating between flat-spectrum and notch noise stimuli becomes gradually more difficult as the level increases from 50 to 80 dB SPL and this corresponds to a gradual increase in the similarity between the masking patterns of the two stimuli. This would be consistent with the notion that high-frequency spectral notches are represented internally with a rate-place code and that the quality of this representation deteriorates gradually with increasing level due to the broadening of the frequency response and to the saturation of the response of AN fibres (Rice *et al.*, 1995; Lopez-Poveda, 1996).

This notion, however, does not account for the results above 80 dB SPL (right panels of Fig. 3.6). The total difference between the masking patterns continues to decrease with increasing level from 80 to 90 dB SPL and yet discrimination improves. This occurs for all listeners.

3.5 General Discussion

The aim was to investigate behaviorally to what extent the discrimination between broadband noise stimuli with and without high-frequency spectral notches relies on AN rate-place representations of the spectral notch. The quality of the rate-place representation of a stimulus spectrum depends on level (Sachs and Young, 1979, 1980; Rice *et al.*, 1995; Lopez-Poveda, 1996) and forward-masking patterns are regarded as reasonable approximations of the AN representation of a stimulus spectrum (see above and Moore, 2005). Accordingly, the approach consisted in running the spectral discrimination task in question at different stimulus levels (see Alves-Pinto and Lopez-Poveda, 2005) and comparing the results with the corresponding shapes of the forward masking patterns of the two noises. The hypothesis was that if discrimination relied on the difference between the excitation patterns of the two stimuli, varying the stimulus level should have a consistent effect on the results of the two experiments: the more similar the masking patterns, the more difficult the discrimination task. Data have shown that this hypothesis holds true below but not above 80 dB SPL (Fig. 3.6). This suggests that a form of representation other than rate-place mediates high-frequency spectral discrimination above 80 dB SPL.

The approach stands on the assumption that a forward-masking pattern is a reasonable approximation of the AN rate-profile representation of the stimulus spectrum, a plot of the average discharge rate of the population of AN fibres as a function of their CFs. Arguments in support of this assumption are reviewed above. Rate-profiles and masking patterns are, however, different measures and despite precautions have been taken to maximize the correspondence between the two (e.g., by using forward masking and by ensuring that the probe level was reasonably low and approximately constant for all masker levels), the correspondence may not be straight forward.

Indeed, the quality of the notch representation in the present masking patterns deteriorates gradually with level in a way consistent with most studies on the AN rate-profile representation of high-frequency spectral notches (Rice *et al.*, 1995; Lopez-Poveda, 1996). It is important, though, to bear

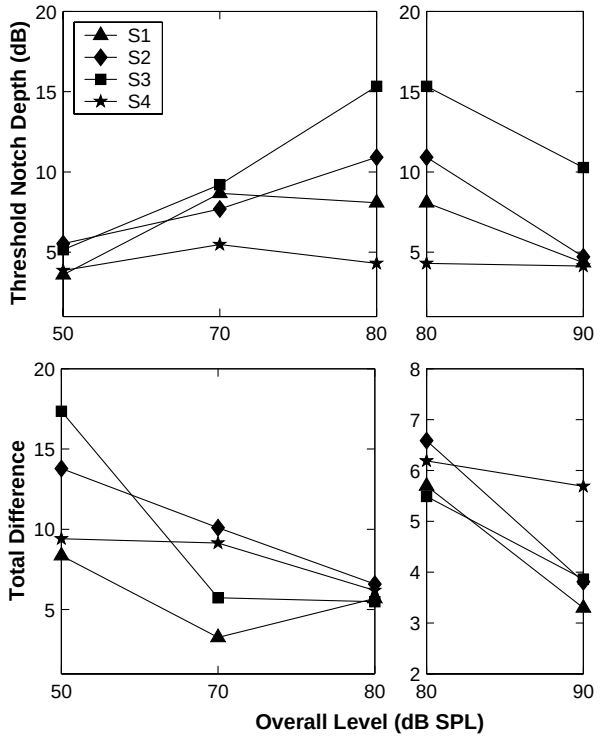

Figure 3.6: The effect of increasing level on the threshold notch depths for discriminating between flat-spectrum and notch noise stimuli (upper panels; adapted from Fig. 3.2) and on the total difference between the masking patterns of flat-spectrum and notch noise maskers (lower panels). Different symbols illustrate the results for different listeners. The left and right panels illustrate the results over two different level ranges: 50-80 (left) and 80-90 dB SPL (right). This is done to facilitate comparing the trends of the two variables (total difference and discrimination thresholds) below and above (80 dB SPL) the level above which the trend in discrimination performance changes.

in mind that forward masking is not only the result of a post-stimulatory reduction in AN activity, but also of efferent inhibitory processes and of the persistence of neural activity (see Meddis and O'Mard, 2005 for a brief review on the topic).

The present results suggest that discriminating between two stimuli with different high-frequency spectral features is unlikely to rely on comparing the excitation pattern of the two spectra, at least above 80 dB SPL. Assuming that the excitation pattern relates closely to the AN rate-profile, this undermines the conjecture put forward on Chapter 2 that the nonmonotonic aspect of the threshold-notch-*vs.*-level function (Fig. 3.2) reflects the existence of only two fibre types (HSR and LSR fibers) with different thresholds and dynamic ranges (see Chapter 1). While two fibre types (with low and high-spontaneous rates) probably exist in the human AN, as they do in other mammals (Liberman, 1978; Yates, 1991), the present results suggests that the statistical distributions of their thresholds and dynamic ranges are likely be continuous rather than bimodal, as it also occurs for other mammals (Yates, 1991).

If the quality of the excitation pattern representation of high-frequency spectral notches deteriorates gradually with increasing level, what explains the improvement in spectral discrimination above 80 dB SPL? It is possible that above 80 dB SPL, when the rate-place representation of the different spectra becomes indistinct, discrimination relies on comparing the discharge times of AN fibres evoked by the different spectra. Recent studies have shown that although the number of AN spikes phase locked to the stimulus waveform rolls-off rapidly with increasing frequency, statistically-significant phase locking may still occur for stimulus frequencies as high as 14 kHz (Recio-Spinoso *et al.*, 2005; Temchin *et al.*, 2005). Furthermore, the cut-off frequency of phase locking varies largely across species (reviewed by Palmer, 1986) and so the cut-off frequency for human may be higher than is conventionally accepted.

The conjecture that some form of temporal code may mediate the discrimination of high-frequencies has already been put forward based on a computational model of the limits of human auditory perception for single tones (Heinz *et al.*, 2001). The present study, however, suggests that the principle might apply to the spectral discrimination of broadband aperiodic stimuli, which is a more natural type of auditory task than pure tone discrimination.

One final remark: several acoustic reflexes activate at high levels (reviewed by Möller, 2000). Among these is the medial olivo-cochlear reflex, whose effect is to shift the dynamic range of AN fibres toward higher levels (Guinan, 1996). Therefore, one might think that the improvement in spectral discrimination above 80 dB SPL [Figs. 2.3 and 3.2] relates to the activation of this reflex (Winslow and Sachs, 1987; Pickles, 1988). This, however, is unlikely. If it had been the case, it would have led to a clearer representation of the spectral notch in the masking patterns above 80 dB SPL and this was not the case (Figs. 3.4 and 3.5).

3.6 Conclusions

The shape of the masking patterns of broadband noise forward maskers with and without a high-frequency spectral notch become gradually more similar with increasing the overall level of the masker. This suggests that the auditory-nerve rate-place representation of high-frequency spectral notches deteriorates gradually with increasing level. This is consistent with the increased difficulty in discriminating between flat-spectrum and notch noise stimuli for increasing stimulus levels between 50 and 80 dB SPL but not with the improvement observed for further level increases above 80 dB SPL. In conclusion, a form of spectral representation other than rate-place (possibly temporal) is likely to mediate high-frequency spectral discrimination, at least above 80 dB SPL.

Chapter 4

Physiological assessment of the representation of high-frequency spectral notches in the auditory nerve[1]

4.1 Introduction

The following chapter describes our efforts in finding physiological correlates of behavioral human performance in discriminating between flat-spectrum and notch noises. The study is based on analysing guinea-pig responses to the two stimuli. Two analysis of AN responses are described: 1) the rate-profile representation of stimulus spectrum and 2) an analysis of instantaneous discharge rates of AN fibers based on a statistical "optimum observer" model. There are almost certainly differences in sound processing between humans and guinea pigs that must be taken into account when establishing any correlations between human behavioral performance and predictions from guinea-pig responses. Still, the exercise is worthwhile, not only because animal physiology is the only way of searching for physiological correlates of behavioral performance, but also because it may suggest more adequate human psychophysical experiments to investigate the AN encoding of high-

[1]Parts of this chapter were published in Lopez-Poveda, E. A., Alves-Pinto, A. and Palmer, A. R. (**in press**). "Psychophysical and physiological assessment of the representation of high-frequency spectral notches in the auditory nerve," in *Hearing - from basic research to applications*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp and J. Verhey (Springer Verlag, Heidelberg).

frequency spectral notches. The validity of our approach is supported by the fact that guinea-pig HRTFs, estimated through both acoustic (Dickson *et al.*, 2002, 2003; Sinyor and Laszlo, 1973) and physiological studies (Carlile and Pettigrew, 1987, King and Palmer, 1985), are similar to human HRTFs, though displaced towards higher frequencies. This is likely to reflect similarity in the structures of the peripheral systems of the two species but with different dimensions (Carlile, 2005).

The spectrum of a sound can be encoded in the AN activity in at least two ways: 1) in the distribution of activity (average discharge rate) across fibers tuned to different frequencies (a rate profile); or 2) in the timing of the individual spikes in fibers tuned to different frequencies (a temporal profile). Both rate and temporal profiles may contribute simultaneously to encoding low-frequency spectral features like those present in speech signals, but fine grained temporal information does not contribute to the coding of high-frequency spectral features (Delgutte and Kiang, 1984; Rice *et al.*, 1995; Lopez-Poveda, 2005). The ability of AN fibers to signal the details of the stimulating waveform in the fine-time structure of their discharges (phase-locking) falls off with stimulus frequency, and is therefore limited to frequencies below approximately 4000 Hz (Johnson, 1980; Palmer and Russell, 1986). Spectral notches like those generated by the pinna have frequencies ($>$ 6000 Hz) beyond the cut-off of phase-locking and are, therefore, thought to be encoded only in the rate-profile of auditory nerve fibers (Poon and Brugge, 1993; Rice *et al.*, 1995).

Most auditory nerve fibers have low response thresholds (10 dB SPL) and narrow dynamic ranges over which they can signal sound level changes ($\sim$20-40 dB). As a result, the quality of the rate-profile representation of high-frequency spectral features deteriorates at high sound levels (Rice *et al.*, 1995). Paradoxically, psychophysical evidence reveals that our ability to discriminate noise stimuli with and without a high-frequency (8 kHz) rectangular spectral notch is worse at levels around 70-80 dB SPL than at lower or higher levels (see Figs. 2.3 and 2.4; Alves-Pinto and Lopez-Poveda, 2005). If behavioral discrimination were based on comparing the quality of the auditory nerve rate-profile representation of the spectral notch, as is commonly believed, then discrimination should have been increasingly harder with increasing stimulus level.

So, what underlies the improvement in high-frequency spectral-notch discrimination at high sound levels? One possibility is that the notch be represented in the rate profile of LSR fibers. This mechanism as been traditionally

considered as one way that the AN handles information over a much wider range of sound levels than the dynamic range of individual fibers. That is, as a solution for the dynamic range problem of hearing (Delgutte, 1996; Viemeister, 1988). We address this question here by recording the activity of guinea-pig AN fibers in response to stimuli like those used in the related psychophysical study (Chapter 2). The results are not consistent with the hypothesis that the discrimination of high-frequency spectral notches is based on the information present in the simple AN average rate profiles.

4.2 Methods

4.2.1 Physiological Recordings

The method used for recording from AN fibres of the aneasthetised guinea pig has been described in detail previously (Palmer *et al.*, 1986). Data were collected from fibers with CFs between 2 and 16 kHz; a CF range sufficient to cover the relevant spectral content of the stimulus. The response of any given fiber was recorded for a minimum of 5 and a maximum of 100 repetitions per stimulus condition. A different stimulus condition, defined by the notch depth and the overall sound level of the stimulus, was presented every 880 ms. Conditions were presented in random order. Not all stimulus conditions were used to stimulate all the fibers (see below). All experiments were conducted in accordance with the 1986 UK Animals (Scientific Procedures) Act.

All the physiological recordings were made by Prof. Alan R. Palmer at the MRC Institute of Hearing Research, Nottingham (UK).

4.2.2 Stimuli

AN fibers were stimulated with bursts of broadband (0.02 to 16 kHz) noise. The spectrum of the noise was flat except for a frequency region centered at 7 kHz where it had a rectangular spectral notch. The spectrum level in the notch band was 0 (i.e., flat spectrum), 3, 6, 9, 15, 21 or 27 dB below the spectrum level outside the notch band. Notch bandwidths of 2 and 4 kHz were used. Noise bursts had a total duration of 110 ms, including a 10-ms rise time (no fall ramp was applied). Stimuli were presented for overall levels ranging from 40 to 100 dB SPL in 10-dB steps. A single noise token was generated in the digital domain for each notch depth and used for repeated measures of AN responses at all levels.

The noise bursts were generated as described in the related behavioral study (Section 2.2.2.a; Alves-Pinto and Lopez-Poveda, 2005). The noise bursts used in the present study were shorter (110 ms *vs.* 220 ms) and the notch center frequency was lower (7 kHz *vs.* 8 kHz) than those used in the behavioral study. Despite these differences, the fundamental characteristics of the stimuli remained the same: 1) in both cases the notch frequency band was beyond the cut-off of phase-locking [∼4 kHz according to Palmer and Russell (1986)], and 2) the stimulus duration was longer than the fast-adaptation period of AN fibers [∼30 ms according to Westerman and Smith (1984)].

4.2.3 Rate Profiles

A rate profile is a graphical representation of the mean discharge rates of a sample of AN fibers as a function of the fibers CFs; the mean discharge rate calculated over the stimulus duration. Raw rate profiles are uninformative of the spectral content of the stimulus due to the large across-fiber variability in spontaneous and saturated rates (Rice *et al.*, 1995). To account for such variability, normalized rate profiles (varying from 0 to 1) were used instead. The normalization was done as follows (Rice *et al.*, 1995):

$$R_{norm} = \frac{R - SR}{R_{max} - SR} \tag{4.1}$$

where R is the average discharge rate of the fiber, SR its spontaneous rate, and R_{max} its maximum discharge rate. Here, SR and R_{max} were estimated as the average discharge rates for a flat-spectrum noise stimulus of 40 and 100 dB SPL, respectively. In the behavioral task (Chapter 2; Alves-Pinto and Lopez-Poveda, 2005), subjects were asked to discriminate between a flat-spectrum noise and a noise with a spectral notch. Therefore, difference rate profiles for the two stimuli were also calculated as they provide a more relevant neural correlate of behavioral performance than do normalized rate profiles. All rate profiles were smoothed by applying a running average calculated over $1/3^{rd}$-octave-band intervals.

4.2.4 "Ideal observer" analysis of auditory nerve responses

The behavioral threshold notch depth for discriminating between notch and no-notch stimuli, $\triangle\alpha$, was predicted from the *instantaneous* discharge rate (see below for details) of the population of AN fibers as follows (Siebert, 1970; Heinz *et al.*, 2001):

$$\Delta\alpha = \left\{ \sum_i \int_0^T \frac{1}{r_i\left(t,\alpha\right)} \cdot \left[\frac{\partial r_i\left(t,\alpha\right)}{\partial \alpha} \right]^2 \cdot dt \right\}^{-0.5} \tag{4.2}$$

where t denotes time, T denotes stimulus duration, and $r_i(t,\alpha)$ the instantaneous discharge rate of the *i-th* fiber in response to the stimulus with notch depth α.

Equation 4.2 was derived on the assumption that the times at which AN spikes occur follow a Poisson distribution (i.e. that spikes occur at times that are independent of each other). Furthermore, it assumes that behavioral discrimination thresholds reflect optimal use of every bit of information available in the activity of the population of fibers; i.e. on the performance of an "ideal observer". Neither of these two conditions apply here (Siebert, 1965; Siebert, 1968; Siebert, 1970), thus one should not expect the resulting $\Delta\alpha$ values to match the absolute values of the behavioral thresholds directly. However, we assume that the error in using Eq. 4.2 for predicting the behavioral thresholds is similar for all sound levels so Eq. 4.2 predicts the *shape* of the threshold notch depth *vs.* level function, as reported in the related behavioral study (Chapter 2; Alves-Pinto and Lopez-Poveda, 2005).

It is noteworthy that Eq. 4.2 predicts the threshold notch depth for spectral discrimination using the *instantaneous* discharge rate of the population of AN fibers. This contrasts with the rate-profile principle described above that only considers the information conveyed in the overall discharge rate of the fibers.

For obvious reasons, in applying Eq. 4.2, we had to consider a discrete version of the instantaneous discharge rate, $r_i(\Delta t,\alpha)$, rather than the continuous-time $r_i(t,\alpha)$. Note that $r_i(\Delta t,\alpha)$ may be interpreted as a mean-rate post-stimulus time histogram for a bin width duration of Δt. $\Delta\alpha$ was computed for different sampling periods (or time windows), Δt, from 2 to 110 ms. Notice that in the extreme case that Δt equals the stimulus duration, the resulting $\Delta\alpha$ corresponds to performance based on a rate-profile code only. The discrete version of Eq. 4.2 used to predict the threshold notch depth $\Delta\alpha$ is as follows:

$$\Delta\alpha = \left\{ \sum_i \sum_{k=1}^{nbins} \frac{1}{r_i\left(\triangle t_k, 0\right)} \cdot \left[\frac{r_i\left(\triangle t_k, 0\right) - r_i\left(\triangle t_k, 3\right)}{3-0} \right]^2 \cdot \triangle\, t_k \right\}^{-0.5} \tag{4.3}$$

In Eq. 4.2, the term between square brackets denotes the change in instantaneous discharge rate for an incremental change in parameter α. As shown in Eq. 4.3 it was calculated as the difference in discharge rate for the flat-spectrum noise ($\alpha = 0$ dB) and the noise with the smallest notch depth used in experimental recordings ($\alpha = 3$ dB). $\Delta\alpha$ would become unrealistically equal to zero when the average discharge rate of any fiber equals zero. To prevent this artifactual result, a small, arbitrary constant of 0.1 Hz was added to the measured discharge rate in all bins of all fibers. The actual value of this constant does not alter results significantly.

It should be noted that predicted thresholds $\Delta\alpha$ are inversely proportional to the square root of the *sensitivity* of the population of fibers to the two stimuli that are being discriminated. In Eq. 4.2 and 4.3, the sensitivity of the population of fibers is given by the term in curly brackets.

4.3 Results

The results shown below are based on responses for a sample of 164 AN fibers from 18 animals. Figure 4.1 shows the distribution of the sample of fibers in terms of CFs, SR and threshold levels. The sample is representative of the population of guinea-pig AN fibers over the CF range considered (2 to 16 kHz) in terms of spontaneous rates and threshold levels (Yates, 1991). Results will be illustrated for a notch bandwidth of 2 kHz only, but the conclusions also apply to the 4-kHz notch bandwidth.

4.3.1 *Notch representation in auditory-nerve rate profiles is not consistent with behavioral discrimination performance*

We first tested whether behavioral spectral discrimination could be accounted for using only the AN rate-profile representation of the stimulus spectrum. Rate profiles were analyzed and plotted for a sub-sample of 105 fibers from 16 animals for which at least 5 (and typically 10) repetitions were measured per stimulus condition[2]. Figure 4.2 illustrates the rate profiles calculated from AN responses to the noise with a notch bandwidth of 2 kHz only.

A simple visual analysis of both normalized and difference rate profiles [Figs. 4.2(a) and (b)] revealed a lower discharge rate for those fibers with CFs

[2]Rate profiles are not based on the responses of the full sample of fibers because for 60 of them responses were measured for a notch depth of 3 dB only.

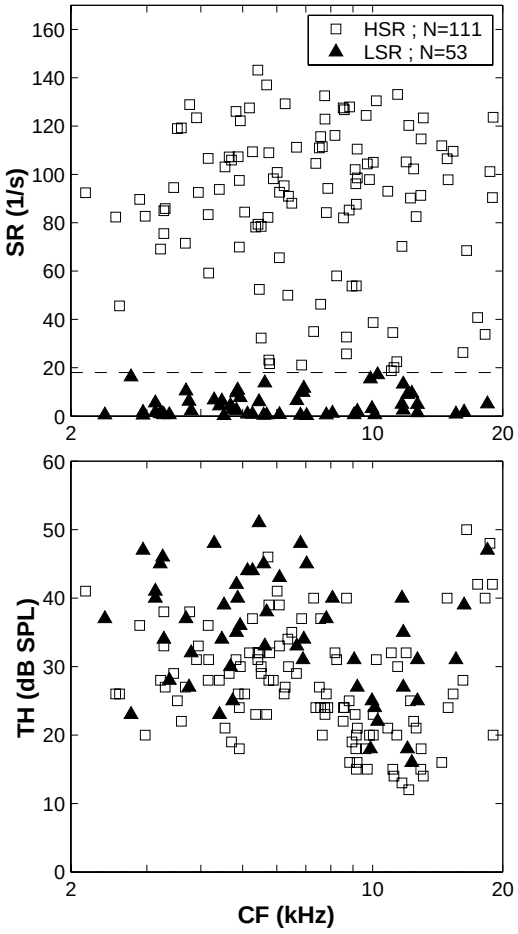

Figure 4.1: Characteristics of the sample fibers. **(a)** Spontaneous rate *vs.* characteristic frequency. **(b)** Rate threshold *vs.* characteristic frequency. Different symbols illustrate two fiber groups with spontaneous rate higher (HSR) or lower (LSR) than 18 spikes/s.

around the frequency band of the notch. The discharge rate was proportional to the notch depth at mid levels. This would suggest that AN rate-profile comparisons constitute a reasonable physiological basis for behavioral discrimination of high-frequency spectra. However, a closer look disproves this suggestion. The absolute difference rate was largest for overall levels around 60-80 dB SPL. This implies that discrimination should be easiest around these levels, in clear contrast with the actual behavioral results showing that discrimination is hardest at these intermediate sound levels (Figs. 2.3 and 2.4).

4.3.2 *Predicted performance based on an "ideal observer" analysis over the stimulus duration of auditory nerve responses*

As referred above, Eq. 4.3 can be used to test the rate-profile code hypothesis. Setting a time window equal to the duration of the stimulus implies that any possible information conveyed in the timing of spikes is disregarded and only changes in the average discharge rate are considered. Fig. 4.3 illustrates the threshold notch depth predicted from the average discharge rate of the sample of 164 fibers calculated from their responses to between 5 and 100 stimulus repetitions. The threshold notch depths for discrimination were, in absolute terms, about an order of magnitude lower than the behavioral ones [notice the different scales on the left and right axes of Fig. 4.3(a)]. This may reflect differences in cochlear processing between humans and guinea pigs, and/or that humans do not operate as optimal spectral discriminators, as others have suggested (Siebert, 1965; Siebert, 1968; Siebert, 1970; Delgutte, 1996; Heinz *et al.*, 2001). Otherwise the absolute values should be approximately similar.

Remarkably, the predicted and the behavioral functions [Fig. 4.3(a)] clearly have different *shapes*. The predicted threshold notch depth is lowest at 60 dB SPL and its value increases gradually with increasing sound level. No improvement in spectral discrimination is observed for levels above 80 dB SPL, which is inconsistent with the behavioral function. In summary, this analysis, which is based on the idea that behavioral discrimination uses whatever rate-place information may be available, confirms that the rate-profile is unlikely to provide the basis for high-frequency spectral discrimination.

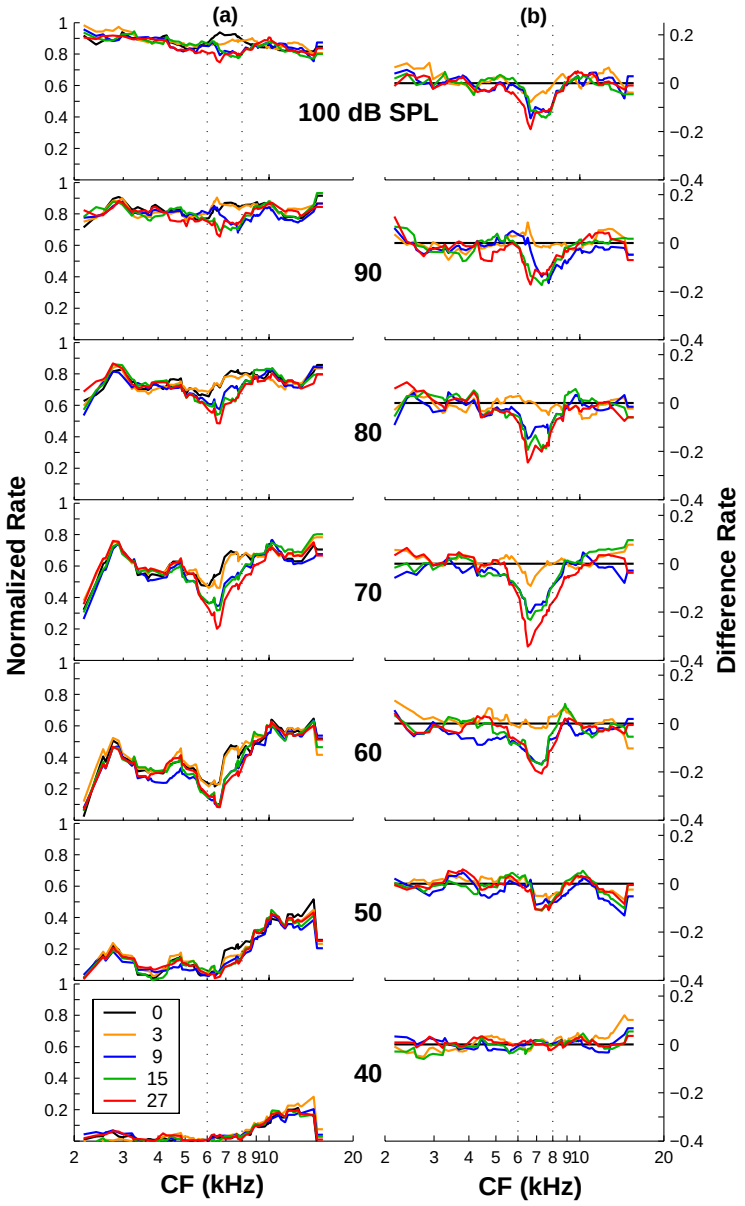

Figure 4.2: **(a)** Normalized and **(b)** difference rate profiles at different overall sound levels. Each curve in (a) is for a different notch depth, as indicated by the inset. The curves in (b) illustrate the difference between the rate profiles for the flat-spectrum and the notched noises. The numbers in the inset denote notch depths in dB re spectrum level of the flat-spectrum noise. Vertical dashed lines illustrate the frequency band of the spectral notch.

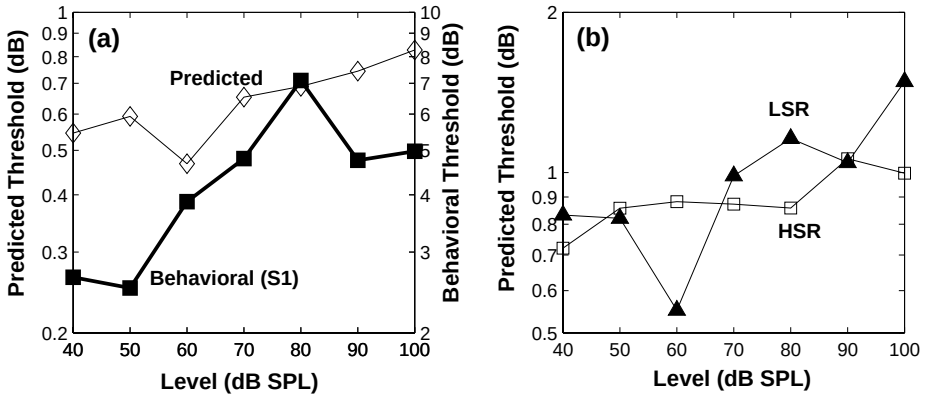

Figure 4.3: Predicted threshold notch depth *vs.* level functions based on an "ideal-observer" type of analysis of physiological average-rate auditory nerve responses (see Methods). **(a)** Predicted thresholds considering all the fibers (open symbols, left ordinate axis) *vs.* behavioral thresholds (filled squares, right ordinate axis) for an example listener [S1 of Fig. 2.3]. **(b)** Predicted thresholds considering LSR and HSR fibers separately.

4.3.3 *Selective use of different fiber types does not account for the behavioral results*

We have previously conjectured (Chapter 2; Alves-Pinto and Lopez-Poveda, 2005; Alves-Pinto et al., 2005a) that the nonmonotonic shape of the behavioral threshold notch depth *vs.* level functions indicates the existence of two fiber types with different thresholds and dynamic ranges. The idea was that the spectral notch would be encoded in the rate profile of fibers with low thresholds and narrow dynamic ranges (HSR fibers) at low sound levels, and in that of fibers with higher thresholds and wider dynamic ranges (LSR fibers) at high levels. The peak in the behavioral function would occur at the transition sound level from the dynamic ranges of the two fiber types. Behavioral thresholds would thus be determined by the activity of HSR and of LSR fibers at low and high levels, respectively.

To test this conjecture we applied the "ideal observer" analysis to the two groups of nerve fibers. Units were classified in one of the two groups depending on their spontaneous rate being higher or lower than 18 spikes/s (Liberman, 1978). The resulting HSR and LSR groups contained 111 and 53 fibers, respectively (Fig. 4.1).

Predicted threshold notch depth *vs.* level functions differed for the two groups [Fig. 4.3(b)]. Nevertheless, the shape of the predicted functions for neither fiber group or for the combined sample [Fig. 4.3(a)] matched that of the behavioral functions [filled squares in Fig. 4.3(a)]. This result is inconsistent with our conjecture and suggests that the nonmonotonic shape of the behavioral discrimination functions is unlikely to reflect a transition between the dynamic ranges of the two fiber types.

4.3.4 *Predicted performance based on the analysis of instantaneous rate by an "ideal observer"*

An alternative hypothesis, that spectral discrimination is based on the *instantaneous* discharge rates of the population of AN fibers, was also tested. Threshold notch depths were predicted by applying Eq. 4.3 to AN responses using time windows shorter than the stimulus duration, that is, using the *instantaneous* rather than the *average* discharge rate. This analysis implies that any temporal information available also contributes to discrimination. Of course, different degrees of temporal information may be gained by sampling the instantaneous discharge rate in nonoverlapping time windows of different durations; the shorter the time window, the more precise the timing information, the greater the discrimination capability of the system, and the lower the discrimination thresholds. This was indeed found to be the case. Figure 4.4 illustrates the results of this analysis for a few time windows: 2, 5, 7, 8.5, 9, 22, 55 and 110 ms. The numbers adjacent to each curve indicate the time window used to sample the discharge of the fiber. The behavioral thresholds measured for listener S1 (see Fig. 2.3) are also plotted (filled squares) for comparison with the predicted thresholds. The "ideal observer" analysis is based on the responses of 164 AN fibers. For any given sound level, the predicted threshold notch depths decreased with shortening the sampling time window. In absolute terms, however, the predicted thresholds were about two orders of magnitude lower than the behavioral ones.

The shape of the predicted threshold notch depth *vs.* level functions varied greatly depending on the time window. Only for time windows within the range from 4 to 9 ms were the predicted functions nonmonotonic with a peak at or around 80 dB SPL, thus resembling the shape of the behavioral functions. This would suggest that high-frequency spectral discrimination may be based on sampling the instantaneous discharge rate of AN fibers in nonoverlapping time windows of between 4 and 9 ms.

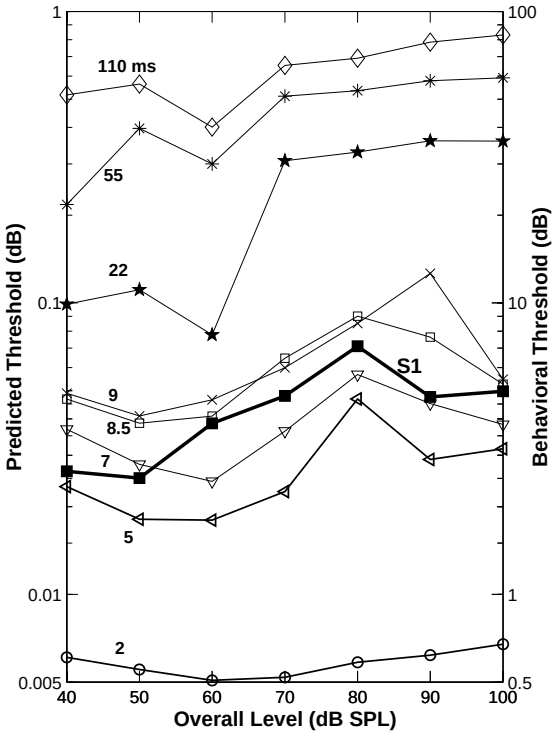

Figure 4.4: Predicted and behavioral thresholds as a function of noise overall level. Behavioral (filled squares, right ordinate axis) data are for listeners S1 of Fig. 2.3. Predicted thresholds (open symbols, left ordinate axis) were obtained based on an "ideal observer" type of analysis of physiological audiotry nerve responses (see Section 4.2.4). Different curves illustrate predicted thresholds when auditory nerve activity is sampled in time windows of different durations, as indicated by the numbers next to each trace (in ms). The analysis is based on responses of N=164 units with the number of repeats ranging from 5 to 100.

"Noise" in discrimination performance predicted with the "ideal observer" analysis

Predicted neural threshold notch depths depends strongly on the sensitivity of individual fibers. For example, in Eq. 4.3, an instantaneous discharge rate equal to zero in a given bin yields an infinite threshold. That is, sensitivity of single fibers in specific bins may influence strongly predicted thresholds. It is therefore important to clarify whether the shapes of the curves in Fig. 4.4 truly reflect differences in the responses of fibers to the flat-spectrum (α = 0 dB) and notch noises (α = 3 dB) or they result artifactually from the method. This was done by comparing the curves of Fig. 4.4 with equivalent curves obtained from responses to the reference flat-spectrum noise only. Theoretically, the sensitivity of the population of fibers to the same stimulus should be zero and should therefore be independent of stimulus level. That is, predicted threshold *vs.* level functions obtained from the responses to the reference stimulus only should be flat and close to zero. A similar approach was employed by Conley and Keilson (1995) to estimate the 'noise' discrimination performance calculated from experimental responses.

Thresholds were predicted in this case by comparing instantaneous discharge rates calculated from different subgroups of responses evoked by the flat-spectrum noise. Therefore Eq. 4.3 was applied as follows:

$$\Delta\alpha = \left\{ \sum_{i} \sum_{k=1}^{nbins} \frac{1}{r_i\left(\triangle t_k, 0\right)} \cdot \left[\frac{r_i\left(\triangle t_k, 0\right) - r_i\left(\triangle t_k, 3\right)}{3-0} \right]^2 \cdot \triangle\, t_k \right\}^{-0.5} \tag{4.4}$$

where $r_i\left(\Delta t_k, 0\right)$and $r'_i\left(\Delta t_k, 0\right)$ denote instantaneous discharge rates calculated from different subgroups of responses to the reference stimulus (α= 0 dB). The denominator $\Delta\alpha$ was kept equal to 3 dB to avoid dividing by zero and to allow the comparison with thresholds obtained from responses to reference and target (α= 3 dB) stimuli.

Figures 4.5 (a) and (b) show the predicted thresholds *vs.* level functions obtained with the responses to the reference flat-spectrum noise only [Eq. 4.4; Fig. 4.5(a)] and to the flat-spectrum *vs.* notch noise [Eq. 4.3; Fig. 4.5(b)]. Evaluations are based on instantaneous rates of 98 [Fig. 4.5(a)] and 106 [Fig. 4.5(b)] fibers respectively, estimated from between 5 and 50 stimulus repeats. The same time windows as those shown in Fig. 4.4 were also used in this case.

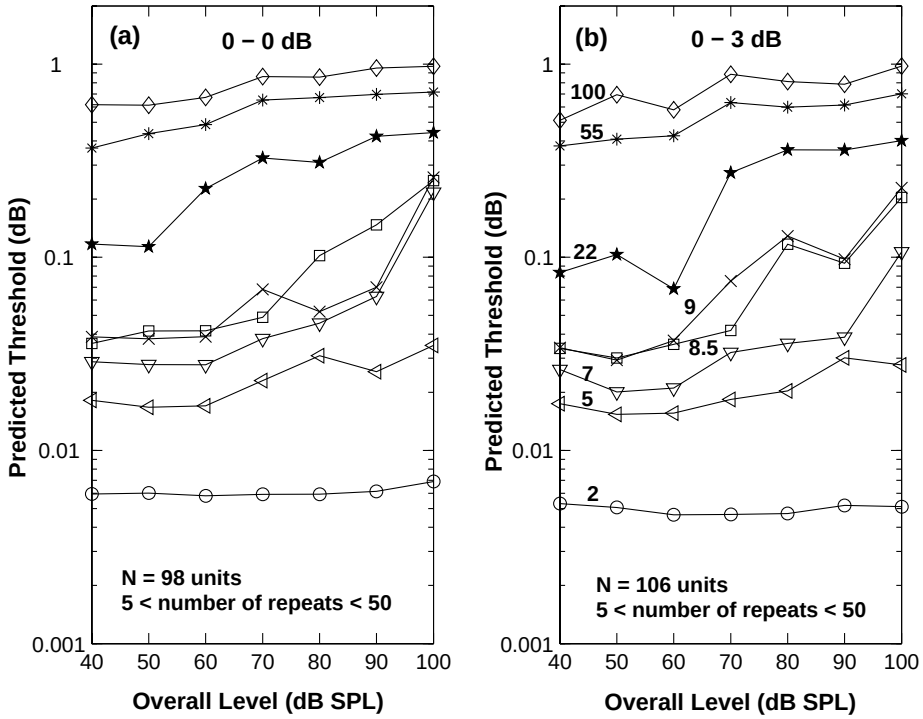

Figure 4.5: Threshold notch depths predicted from responses to **(a)** the reference stimulus (the flat-spectrum noise;$\alpha = 0$ dB) only and **(b)** the reference ($\alpha = 0$ dB) and target ($\alpha = 3$ dB) stimuli. Predicted thresholds were calculated according to the "ideal observer" analysis described in section 4.2.4. The numbers next to each trace indicate the time window used to sample fibers' activity.

In both cases (0-0 dB and 0-3 dB) the predicted threshold *vs.* level functions are, in general, monotonic. The functions obtained for the case 0-3 dB [Fig. 4.5(b)] are quite different from the corresponding functions illustrated on Fig. 4.4, which is likely to result of being based on the responses of a smaller group of units. Moreover, the predicted threshold functions for the reference stimulus only [Fig. 4.5(a)] are not flat as one would expect.

The lack of clear differences between the two plots together with the high variability in the shape of the functions for different time windows, undermines the above reported correlation between behavioral and neural performance when the instantaneous activity of the population of fibers is considered (Fig. 4.4).

4.3.5 *Limitations of the "ideal observer" analysis of AN instantaneous discharge rates*

Threshold notch depths predicted by the analysis of AN responses according to Eq. 4.3 depend not only on the number of fibers analysed and contributing to the sensitivity of the population of fibers. They are also determined by the *instantaneous* discharge rates, as estimated by averaging discharge rates in each time window across the several stimulus repeats. These averages depend on the number of repeats measured for each condition (more repeats yield more accurate estimates of the instantaneous rate). The effect of the number of repeats on the predicted neural thresholds becomes increasingly more important as time windows get shorter because the shorter the time window the more likely it is that $r_i\left(t_k, \alpha\right) = 0$ and if this occurred it would yield $\Delta\alpha$ being artificially equal to zero (Eq. 4.3).

The effect of the number of repetitions on the predicted threshold *vs.* level functions can be more clearly illustrated by applying the "ideal observer" analysis to *simulated* AN responses. The advantage of this approach is that it is easier to manipulate the number of repeats involved in the rate calculation. Spike trains with statistical properties similar to those of AN responses were simulated with a Poisson generator and analysed in the same way as the physiological responses. Briefly, a Poisson generator is a mathematical function that retrieves a series of events, in this case a series of spike times, in such a way that the number of events ocurring in a given time interval is a random variable with a Poisson distribution (see for example http://en.wikipedia.org/wiki/Poisson_process). The probability of a spike

occurring imediately after another spike is close to zero, due to refractoriness in AN activity. Trains of spikes were simulated such that the mean rate of the simulated responses over the whole stimulus duration was identical to the corresponding mean rate obtained experimentally. The MATLAB function used to simulate trains of spikes was downloaded from the web site: *http://www.mit.edu/~linuo/Matlab%20Scripts.html* and adapted for our purpose. It was confirmed that this function produces realistic AN spike trains: the statistics of the spike trains generated with the function were consistent with those reported from experimental studies (e.g. Fig. 5 of Winter and Palmer, 1991) when a refractory period of 1.5-2 ms was used.

Figures 4.6 (a), (b) and (c) illustrate predicted neural thresholds as a function of level based on simulations of AN responses for N=164 fibers. This was the number of experimental fibers for which spike trains were recorded for at least 5 stimulus repetitions. Each pannel illustrates the results for different numbers of simulated repeats. In (a) the sensitivity of each fiber was calculated for a number of simulated repeats equal to the number of measured repeats. This number varied between 5 and 100. In (b) and (c) the sensitivity of each fiber was calculated for 30 and 75 repeats respectively.

Clearly, the shapes of threshold *vs.* level functions are more variable when they are based on a number of repeats equal to the number of repeats cosidered experimentally. When the number of repeats is increased, the functions for different time windows become more similar and less variable. It is also noteworthy that predicted thresholds are higher (i.e. the sensitivity of the population of fibers decreases) the larger the number of repeats. This may reflect larger rate differences or that divisions by a small denominator are more likely to occur when, in Eq. 4.3, they are calculated from fewer repeats.

Based on this insight, the "ideal obsever" analysis was applied separately to 39 units for which responses were measured for between 50 and 100 repeats in each experimental condition. The results of this analysis are plotted in Fig. 4.7. The shape of the curves hardly changes depending on the time interval. All of them are nonmonotonic and suggest that thresholds should be lower for mid-level sounds (60-80 dB SPL) than for sounds with lower or higher levels. This is almost the opposite to behavioral performance. Nevertheless the analysis is based on 39 fibers only.

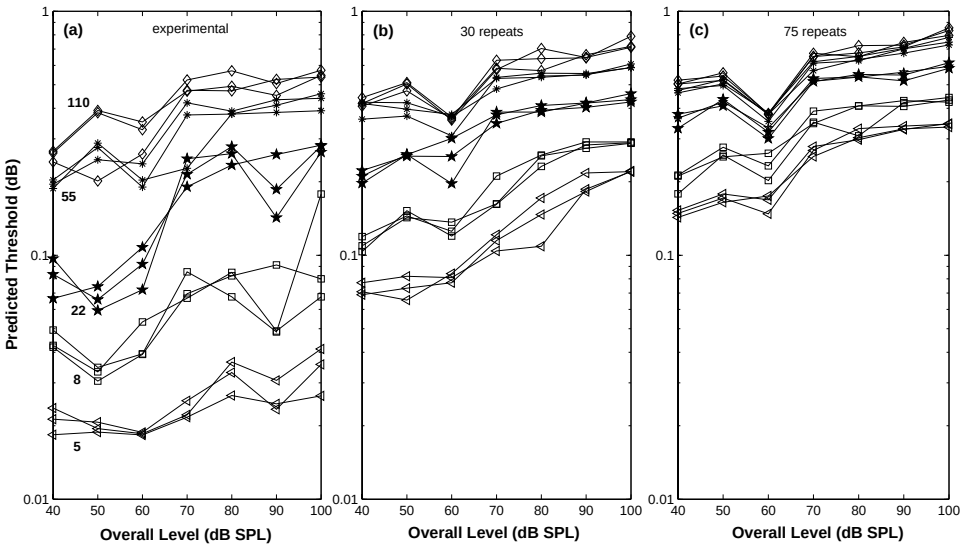

Figure 4.6: Predicted thresholds as a function of level based on simulated Poisson spike trains. Each panel represents the results based on different numbers of stimulus repeats. Results illustrated in **(a)** were obtained for a number of repeats equal to that measured experimentally. Panels **(b)** and **(c)** illustrate the results for 30 and 75 repeats respectively. For each panel, different curves illustrate different evaluations for the same condition and different symbols identify the different time windows used (as indicated by the numbers next to each curve). Three independent evaluations were made for each time window to test whether the number of repeats is the same for all time windows.

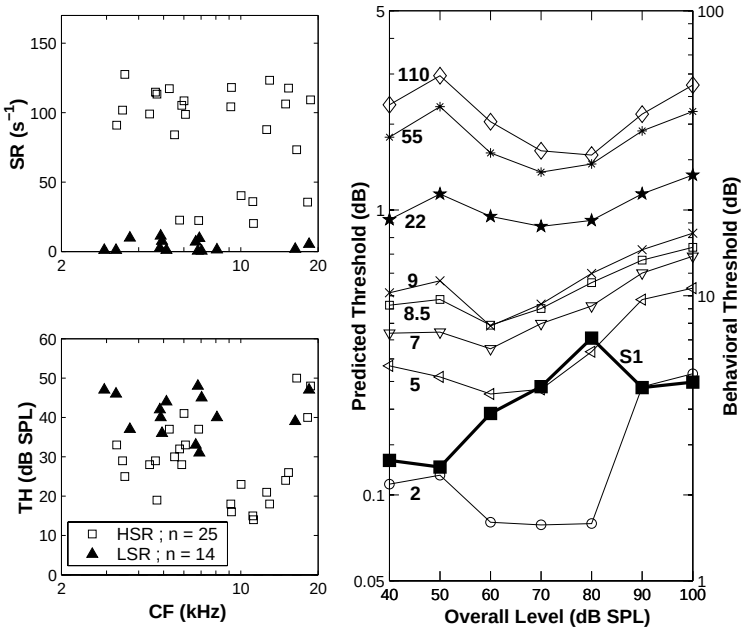

Figure 4.7: Threshold notch depths as a function of level predicted from responses of AN fibers with 50 and 100 repeats measured in each experimental condition (right pannel). Each curve corresponds to a different time window as indicated by the numbers next to each trace. Panels on the left illustrate the distribution of fibers by CF, SR and TH.

4.4 Discussion

The physiological rate profiles show that the notch is better represented at midlevels than at low and high levels. This is not consistent with with either the results of discrimination experiments of Chapter 2 or with the masking patterns of Chapter 3. At levels at which behavioral discrimination is deteriorated the rate-profile representation is optimal. On possibility that cannot be excluded is that this results may be indicative of a mismatch in the transmission of energy by the middle ear between humans and guinea pigs.

4.4.1 *Auditory spectral discrimination is not based on comparing auditory-nerve rate profiles*

We have shown that behavioral discrimination between auditory broadband stimuli with and without high-frequency spectral notches (Fig. 3.1) is uncorrelated with the differences in the overall AN rate-profile representations of their spectra [Figs. 4.2 and 4.3(a)]. Further, the ideal observer analysis reveals that discrimination at different sound levels does not rely on the average-rate information conveyed by subpopulations of nerve fibers with different thresholds and dynamic ranges [Fig. 4.3(b)]. Therefore, the results are inconsistent with the notion that high-frequency spectral features are encoded in AN average-rate profiles.

The ideal observer analysis implies that even one single highly-sensitive fiber at high sound levels might be sufficient to improve behavioral discrimination substantially in any particular stimulus condition. Such an explanation has been put forward to resolve (at least theoretically) the so-called dynamic range problem of hearing; the apparent mismatch between the wide range of sound levels over which good intensity discrimination can be shown and the dynamic range of most AN fibers (Delgutte, 1996). Therefore, a possible explanation for the inconsistency between the shapes of the predicted and observed threshold notch depth *vs.* level functions [Fig. 4.3(a)] would be that our AN fiber sample excludes key fibers which could improve spectral discrimination at high levels. This, however, seems unlikely as the average discharge rate of AN fibers is thought to be proportional to the average receptor potential of the inner hair cells from which they innervate and, as will be shown in Chapter 5, the average-receptor potential representation of the spectral notch deteriorates gradually with increasing level as a result of the saturation of the receptor potential.

4.4.2 *Discrimination could be based on monitoring the cadence of auditory-nerve spikes*

Since average rate-place or temporal codes are the only available ways to represent the sound spectrum in the AN, our results suggest, by exclusion, that the discrimination of high-frequency spectral notches is likely to be based on comparing the times of the discharges evoked by the stimuli to be discriminated. The idea that some form of temporal code may be used for high-frequency spectral discrimination has already been put forward based on a computational model of the limits of human auditory perception for single tones (Heinz et al., 2001). Our study suggests that the principle could also apply to spectral discrimination of broadband aperiodic stimuli, which is a more natural type of auditory task than pure tone discrimination. In support of this conjecture, recent physiological studies have shown that, contrary to common belief, significant phase-locking can occur for frequencies as high as 14 kHz (Recio-Spinoso *et al.*, 2005; Temchin *et al.*, 2005). In other words, the spectral notch could be encoded in the discharge times of AN fibers.

We tested this hypothesis (that discrimination is based on the *instantaneous* discharge rates of the population of AN fibers) by applying Eq. 4.3 to the AN data available. The results suggest that the discharge rate over narrow time windows conveys useful information for discriminating between flat-spectrum and notch noise stimuli. The results also suggest that humans somehow sample the discharge rate of AN fibers in non-overlapping time windows between 4 and 9 ms (Fig. 4.4).

Three questions arise: 1) what is the temporal code in question; 2) how is it generated; and 3) how does it relate to the predicted time window. The answer to these questions requires further analysis of the AN responses and we can only speculate at present.

The effective drive to any AN fiber is a half-wave rectified, low-pass filtered version of the BM response waveform at its corresponding place in the cochlea. With broadband noise stimulation, this can be described as a randomly amplitude-modulated carrier with a carrier frequency near the fibers CF, where the range of modulation frequencies is limited by the bandwidth of the cochlear filter (Louage *et al*,. 2004) or the cut-off of phase locking. The bandwidth of BM filters, and thus the range of modulation frequencies, increases with increasing the SPL. Similarly, the phase of the BM response waveform depends on the filter bandwidth and thus on the stimulus SPL.

AN fibers can phase-lock to the envelope of BM excitation even at high levels, when their discharge rate is at saturation (Cooper *et al*, 1993). Fibers with CFs near the notch frequency certainly see a different level than those with CFs well away from it. It is, therefore, possible that spectral discrimination be based on detecting either the range of modulation frequencies or the phase differences implicit in AN spike trains (or both). On the basis of this conjecture, the psychophysical threshold notch depth *vs.* level functions reported in Chapter 2 (Alves-Pinto and Lopez-Poveda, 2005) would reflect the dynamic range of envelope-following rather than of discharge rate of AN fibers.

Unfortunately, the results of this analysis must be taken with strong reservations given that they do not differ significantly from those obtained when the analysis is based on different spike trains measured in response to the same flat-spectrum stimulus. In this latter case, different responses to the same stimulus should be similar and the threshold *vs.* level functions should be flat, which is clearly not the case. The fact that predicted threshld *vs.* level functions are not flat [Fig. 4.5(a)] is likely to be a result not having enough repetitions for each fiber in each experimental condition. Indeed, the mean rate to the flat-spectrum noise (i.e., the denominator in Eq. 4.3) as well as the difference between the mean rates to the flat-spectrum and notch noises (i.e. the numerator in Eq. 4.3) increases with increasing sound level. However, if the increase in mean rate were not directly proportional to the increase in rate difference, the net effect would be that the overall sensitivity changes as the level increases, even when its calculation is based on responses that do not differ substantially between each other. Furthermore, the probablility of rate differences being nonzero is higher the fewer the stimulus repeats. Threshold *vs.* level functions for different time windows do not change significantly when mean rates are based on a higher number of simulated spike trains (Fig. 4.6), which suggests that the reliability of the "ideal observer" analysis of depends on the number of repeats, especially for short time windows. The number of stimulus repeats or the stimulus duration considered here were chosen *a priori* to obtain good estimates of the average rate of the fibers, but almost certainly were insufficient to apply an "ideal observer" type of analysis.

4.5 Conclusions

1. The representation of high-frequency spectral notches in the difference profile of average discharge rates of AN fibers, measured in response to

broadband noises with and without a high-frequency spectral notch, is clearer at midlevels, around 70-80 dB SPL, than at lower or higher levels. This result is inconsistent with behavioral results, which undermines the notion that discrimination between flat-spectrum and notch noises is based on the rate-profile representation of the spectra of the two stimuli.

2. Applying an "ideal observer" type of analysis reliably requires measuring AN responses for many more stimulus repeats (or longer stimuli) that we have considered here.

3. Separate analysis of the average activity of HSR and LSR fibers does not support the hypothesis that the notch is encoded in the activity of HSR fibers at low levels and in that of LSR fibers at high levels.

Chapter 5

Rate versus time representation of high-frequency spectral notches in the peripheral auditory system: a computational modelling study[1]

5.1 Introduction

Behavioral discrimination between two noises with and without a high-frequency spectral notch was shown to be inconsitent with a mechanism based on the discrimination between the AN rate profiles evoked by the two stimuli (Chapter 4). Indeed, the representation of the high-frequency spectral notch in the profile of average discharge rates of AN fibers was observed to be optimal around 70-80 dB SPL, while behavioural discrimination betwen the two noises was shown to be harder at midlevels. This inconsistency suggests that other types of physiological mechanisms underlie behavioural discrimination. Possibly, some form of temporal information is likely to be involved. The present chapter describes a modeling study undertaken to explain the paradoxical behavioural result and, hence, to clarify how high-frequency spectral features may be represented in the AN response.

AN fibers transmit acoustic information from the cochlea to the central

[1]This chapter is based on the paper: Lopez-Poveda, E. A., Alves-Pinto, A., Palmer, A. R. and Eustaquio-Martín, A. (**accepted for publication**). "Rate *versus* time representation of high-frequency spectral notches in the peripheral auditory system: A computational modelling study," Neurocomputing.

auditory system in the form of spike trains. Every AN fiber is most sensitivite to a particular sound frequency, its CF, and responds to a restricted range of frequencies around it. Therefore, every AN fiber can be viewed as a spectral filter with a certain CF and the population of AN fibers as a bank of filters with CFs spanning approximately the frequency range of hearing (Evans, 1972; reviewed by Lopez-Poveda, 2005).

There are at least two ways in which the spectrum of a signal may be obtained from the output of a filter-bank system. First, it may be obtained by plotting the root-mean-square (RMS) amplitude of the output signals from the filters as a function of the filters center frequencies. This representation will be hereinafter referred to as the *excitation pattern* [Fig. 5.1(c)]. An alternative way would be to estimate the spectra of the output signals from all the filters and add them together to obtain the total filter-bank output spectrum. For reasons that will become clear in the following text, this representation will be hereinafter referred to as the *modulation-rate pattern* [Fig. 5.1(d)]. For a linear filter-bank, both the excitation and the modulation-rate patterns resemble the coarse stimulus spectrum, assuming a large number of perfectly overlapping filters [compare Fig. 5.1(a) with Figs. 5.1(c) and 5.1(d)].

The central auditory system may obtain the stimulus spectrum from the response of AN fibers in two corresponding ways: from the rate profile and from the cadence of spikes for the population of AN fibers. The rate profile is directly related to the abovementioned excitation pattern. AN fibers fire in synchrony with the positive cycles of the driving stimulus waveform (but see the following text), the so-called called phase locking (Rose *et al.*, 1967). Therefore, the cadence of spikes for the AN fiber population informs of the stimulus waveform and thus of its spectrum. This representation of the stimulus spectrum relates closely to the abovementioned modulation-rate pattern.

In Chapter 2 (Alves-Pinto and Lopez-Poveda, 2005) several explanations were proposed for the paradoxical nonmonotonic character of the threshold notch depth *vs.* level function. It was conjectured that the result was consistent with the existence of two types of AN fibers with different thresholds and dynamic ranges, as explained above. Discrimination might be worst at 70-80 dB SPL because this would be the intensity at which the transition between the dynamic ranges of the two fiber types occurs. Direct evidence in support of this explanation (see Chapter 4) is inconclusive. Alternative explanations

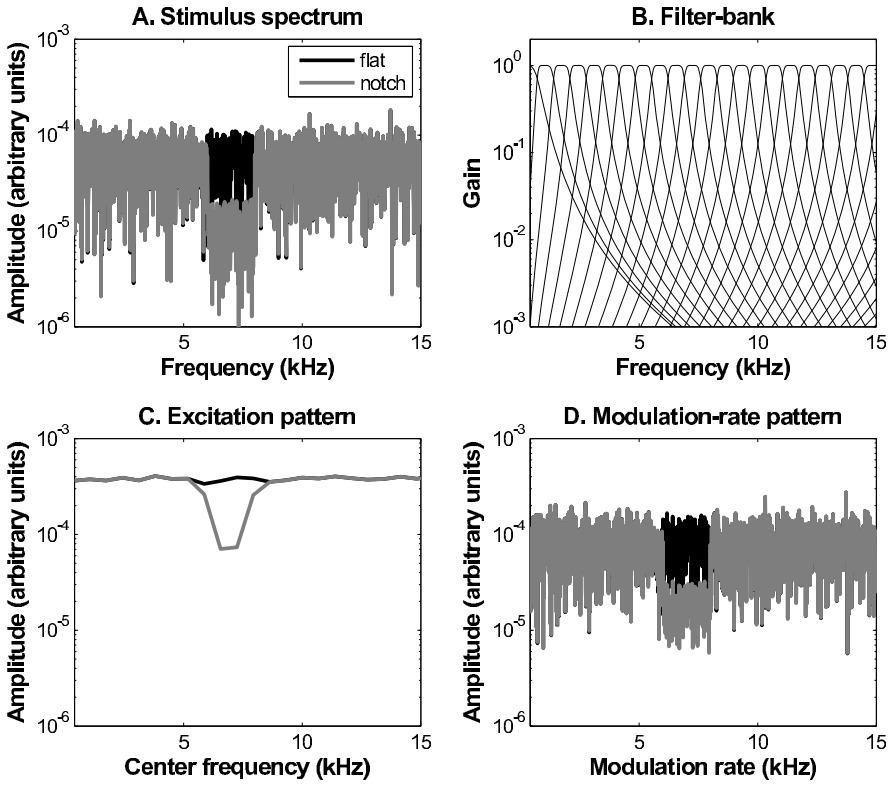

Figure 5.1: Two possible representations of the stimulus spectrum at the output of a linear filter-bank. **A**. Example spectra of flat-spectrum (dark trace) and notch (light trace) frozen-noise stimuli. Both noises have a broad-band spectrum (20 Hz to 16 kHz). The notch is centered at 7 kHz and has a 2-kHz bandwidth. Its depth is 15 dB re spectrum level in notch side bands. The two spectra are identical except for the amplitude in the notch band. **B**. An example linear filter-bank. **C**. and **D**. The excitation and modulation-rate patterns of the two stimuli at the filter-bank output, respectively.

were also discussed in Chapter 2 (Alves-Pinto and Lopez-Poveda, 2005), all of which were based on the assumption that high-frequency spectral notches must be represented in the rate profile of the AN.

A recent computational modeling study (Heinz *et al.*, 2001) has suggested that the just-noticeable difference in frequency between two tones can be explained by considering that even high-frequencies (up to 10 kHz) are encoded in the discharge times of AN fibers. In support of this result, recent physiological studies have shown that, contrary to common belief, significant phase-locking can occur for frequencies as high as 12 kHz (Recio-Spinoso *et al.*, 2005). Based on these ideas, an alternative explanation of the result of Chapter 2 (Alves-Pinto and Lopez-Poveda, 2005) could be that high-frequency spectral notches may still be encoded in the temporal pattern of auditory nerve discharge and that discrimination may rely on comparing the AN modulation-rate pattern representations of the spectra (Chapter 4).

The aim of this study is therefore to investigate which of the two representations of high-frequency spectral features, the excitation (rate) or the modulation-rate (time) patterns, is more consistent with the observation of Chapter 2 (Alves-Pinto and Lopez-Poveda, 2005). A computational model of the peripheral auditory system was used to simulate receptor potential waveforms (the signals driving AN firing) for a bank of inner hair cells with different CFs in response to the stimuli used in the behavioural and physiological studies. The simulations suggest that only the modulation-rate pattern representation is consistent with the behavioral results of Chapter 2 (Alves-Pinto and Lopez-Poveda, 2005). It will be discussed that this supports recent studies that suggest that high-frequency features may still be encoded in the temporal pattern of AN firing. Additionally, it will be shown that the improvement in spectral discrimination for intensities above 70-80 dB SPL is associated to the IHC nonlinearities.

5.2 Methods

The IHC instead of the AN representation of the stimulus spectrum was considered for two reasons. First, the receptor potential drives auditory nerve spiking. Therefore, the quality of the spectrum representation in the IHCs imposes a limit on the quality of its associated auditory-nerve representation. In other words, the AN representation of the spectrum cannot be better than that observed at the level of the IHCs. For example, it is likely that the excitation pattern representation of the stimulus spectrum degrades at high

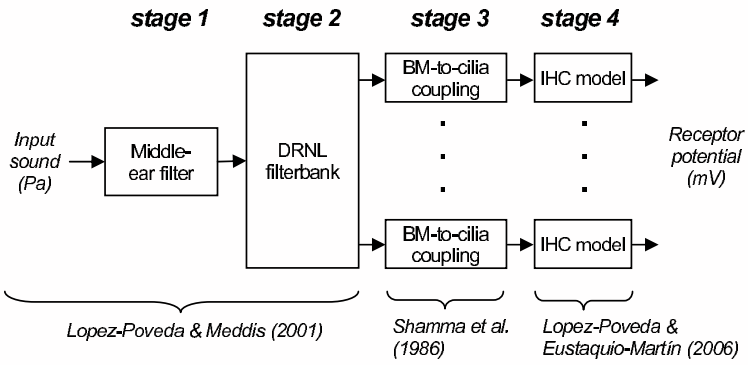

Figure 5.2: Model structure.

sound levels because of the saturation of the receptor potential (e.g. Russell and Sellick, 1978). If this were the case, it would undermine the idea of Alves-Pinto and Lopez-Poveda (2005) that the peak in the threshold notch depth *vs.* level function (Figs. 2.3 and 2.4) relates to the transition between the dynamic ranges of HSR and LSR auditory-nerve fibers. A second reason for using an IHC model is that it outputs a continuous, deterministic signal (the IHC receptor potential), which is easier to analyze than trains of stochastic spikes.

5.2.1 The model

The model consisted of four signal-processing stages, each of which represented a physiological stage of the peripheral auditory system (Fig. 5.2). The input to the model was a digital sound waveform (in units of Pascal) assumed to represent an auditory stimulus at the eardrum[2]. The model output was a collection of waveforms representing the receptor potential for a bank of IHCs with different CFs.

The first model stage simulated the middle-ear transfer function. Its output was the stapes velocity (in units of m/s). The second model stage was

[2]The effect of the outer-ear was not included in the model because in the behavioural study (Chapter 2; see also Alves-Pinto and Lopez-Poveda, 2005) the stimuli were delivered through Etymotic ER2 earphones, which are designed to give a flat-frequency response at the eardrum.

a bank of dual-resonance (DRNL) nonlinear filters (Meddis *et al.*, 2001) that simulated human cochlear frequency selectivity (Lopez-Poveda and Meddis, 2001). The output from each filter was a waveform representing the velocity of vibration (in units of m/s) of a given point along the human BM. A first-order low-pass filter (cut-off frequency of 1 Hz) was applied to the filter-bank output signals to transform their units from velocity to displacement (in units of m), as required for the next model stage. The third model stage was a high-pass filter that coupled BM displacement to the displacement of IHC stereocilia (in units of m) (Shamma *et al.*, 1986). The last model stage was a biophysical model of the IHC that outputs the IHC receptor potential (in volts) as a function of the cells stereocilia displacement (Lopez-Poveda and Eustaquio-Martin, 2006). Details on these models can be found in the original publications and will not be reproduced here.

It is noteworthy that the DRNL filter-bank and the IHC models were specifically and carefully designed to simulate the frequency-response and the nonlinear (compressive) characteristics of their physiological counterparts.

5.2.2 Implementation and parameters

The model was implemented in Matlab R14 as a dynamic signal-processing chain. The middle-ear and filter-bank stages were implemented as described in Lopez-Poveda and Meddis (2001). A bank of 100 filters was considered, with a lowest and a highest CF of 100 Hz and 20 kHz, respectively. The parameter values provided by the same authors for an average human listener [see Fig. 2B and Table II in Lopez-Poveda and Meddis (2001)] were used here. These were obtained by linear regression fits to the values optimized at six CFs, from 250 Hz to 8 kHz in octave steps.

The basilar membrane-to-cilia coupling stage was implemented as described in Shamma *et al.* (1986, Eq. 1) (coupling gain C = 2; cut-off frequency of 530 Hz). The *in vivo* IHC model of Lopez-Poveda and Eustaquio-Martin (2006) was used with the parameter values provided by these authors (see their Table 1). It was assumed that the parameter values for these two model stages were identical across frequency channels.

5.2.3 Evaluations

The model was evaluated in the time domain (sampling frequency = 44100 Hz) in response to broadband (20 Hz to 16 kHz) noise bursts with flat-spectrum (dark trace in Fig. 5.1A) and with a rectangular spectral notch

centered at 7 kHz (light trace in Fig. 5.1A). These were generated as described in Chapter 2. The bandwidth of the notch was 2 kHz, and its depth 15 dB re the spectrum level on the notch side bands (Fig. 5.1). The spectra of the two noises were identical except for the amplitude in the notch band (Fig. 5.1). The noise bursts had a total duration of 110 ms including raised cosine rise/fall ramps of 10 ms. The model was evaluated for overall sound intensities from 38 to 98 dB SPL in 10-dB steps, which correspond to spectrum levels from -2 to 58 dB SPL. These were identical to those considered in the behavioural study (Chapter 2).

5.2.4 Output analysis and estimates of the spectral discrimination sensitivity

For any given stimulus, the model output was analyzed to obtain the excitation pattern and the modulation-rate pattern representations of the stimulus spectrum. The average discharge rate of AN fibers is thought to be proportional to the average receptor potential of IHCs (Cheatham and Dallos, 2001). Therefore, the excitation pattern was obtained by plotting the average receptor potential of each IHC as a function of the cells CF (e.g. Fig. 5.3A). For each IHC and each stimulus condition, the average receptor potential was calculated by summing the samples in the receptor potential waveform with amplitudes greater than zero and dividing the result by the total number of samples in the waveform (Cheatham and Dallos, 2001). The modulation-rate pattern representation was obtained by applying a fast Fourier transform (FFT) to the receptor potential waveform of each IHC and adding the resulting modulation-rate spectra of all IHCs in the frequency domain (e.g. Fig. 5.3C).

For either representation, the *ad hoc* assumption was made that behavioral spectral discrimination would be based on the Euclidean distance between the logarithms of the IHC representations of the flat-spectrum and notch noise spectra; that is, on the squared difference between the dark and light traces in Figs. 5.3A and 5.3C. Consequently, at any given intensity, L, the discrimination sensitivity based on a comparison of the excitation patterns of the two stimuli, $S_{EP}(L)$, was estimated as:

$$S_{EP}(L) = \left\{ \sum_{i=1}^{N} \left[log_{10} \left(\frac{V_{0_i}(L)}{V_i(L)} \right) \right]^2 \right\}^{0.5}$$

where $V_{0_i}$and V_i are the average values of the receptor potential for the i-th

IHC in response to the flat-spectrum and notch noise bursts, respectively, and N is the number of frequency channels.

Likewise, the discrimination sensitivity based on the modulation-rate patterns of the two stimuli, $S_{MP}(L)$, was estimated as:

$$S_{MP}(L) = \left\{ \int_0^\infty \left[log_{10} \left(\frac{S_0(f,L)}{S(f,L)} \right) \right]^2 df \right\}^{0.5}$$

where $S_0(f, L)$ and $S(f, L)$ are the modulation-rate patterns in response to the flat-spectrum and notch noise bursts, respectively.

Given that the auditory peripheral system is highly compressive, S_{EP} and S_{MP} were expected to have different values for different sound intensities and thus were evaluated for every sound intensity considered (cf. Sec. 5.2.3). It was pointless to compare the numerical values of S_{EP} and S_{MP} because they depend on unrelated arbitrary parameters; S_{EP} depends on the number of CFs and the CF range (Eq. 5.1), while S_{MP} depends on the sampling frequency, df [Eq. (5.2)]. It was more important, however, to compare how they varied with stimulus intensity. To facilitate such comparisons, their values were normalized relative to their respective maximum values across intensities.

5.3 Results

5.3.1 Excitation patterns

Figure 5.3A illustrates the IHC excitation pattern representation of the spectra of the flat-spectrum (dark traces) and the notch (light traces) stimuli for the seven stimulus intensities considered. For the lowest intensity, the excitation pattern peaks around 1 kHz due to the band-pass characteristics of the middle-ear filter [cf. Fig. 2 in Lopez-Poveda and Meddis (2001)]. The pattern becomes all-pass at high stimulus intensities. This is due to a combination of BM compression [cf. Fig. 4 in Lopez-Poveda and Meddis (2001)] and the saturation of the IHC receptor potential at high intensities [cf. Figs. 9 and 10 in Lopez-Poveda and Eustaquio-Martin (2006)]. Figure 5.3A also shows that the quality of the excitation-pattern representation of the spectral notch degrades gradually with increasing stimulus intensity. This is more clearly seen in the zoomed inset and in Fig. 5.3B. The latter illustrates the difference excitation patterns [i.e., $\log_{10}(V_0/V)$] at each intensity normalized to the maximum difference found across CFs and intensities. Note that

differences occur for IHCs with CFs within or around the notch band only and that the largest difference occurs for the lowest intensity (40 dB SPL).

5.3.2 Modulation-rate patterns

Figure 5.3C illustrates the simulated modulation-rate patterns for the flat-spectrum (dark traces) and the notch (light traces) noise stimuli. Each pair of traces is for a different stimulus intensity. The noisiness of the patterns reflects the noisiness of the stimulus spectrum (Fig. 5.1A). To facilitate the visual distinction between the patterns for the flat-spectrum and the notch noises, Fig. 5.3C illustrates 5-point running averages of the original patterns.

Unlike the excitation patterns of Fig. 5.3A, the modulation-rate patterns appear as *low*-pass at all levels. This reflects the low-pass transfer characteristics of IHCs (Palmer and Russell, 1986), which are approximately intensity independent (Russell and Sellick, 1978). Note that despite its high frequency, the spectral notch is still observed in the patterns, particularly at the lowest and highest intensities (zoomed inset). Figure 5.3D shows the normalized difference modulation-rate patterns (i.e. $\log_{10}[S_0(f)/S(f)]$) for the seven intensities considered. Clearly, the difference is smaller at mid levels than at lower or higher intensities. Furthermore, significant differences occur for modulation rates outside those corresponding to the notch frequency band (6-8 kHz). These are larger the higher the intensity.

5.3.3 Spectral-discrimination sensitivity

The sensitivity for discriminating between the flat-spectrum and the notch noises was assumed to be directly proportional to the areas under the squared difference excitation and modulation-rate patterns of Figs. 5.3B and 5.3D. Figure 5.4 shows how these two sensitivity measures depend on sound intensity. The sensitivity based on the difference excitation pattern (S_{EP}) decreases monotonically with increasing sound intensity (triangles in Fig. 5.4). By contrast, the sensitivity based on the difference modulation-rate pattern (S_{MP}) is a non-monotonic function of intensity (circles in Fig. 5.4). S_{MP} is smallest at mid-levels and higher at the lowest/highest intensities tested. Only S_{MP} is consistent with the behavioral results (Chapter 2).

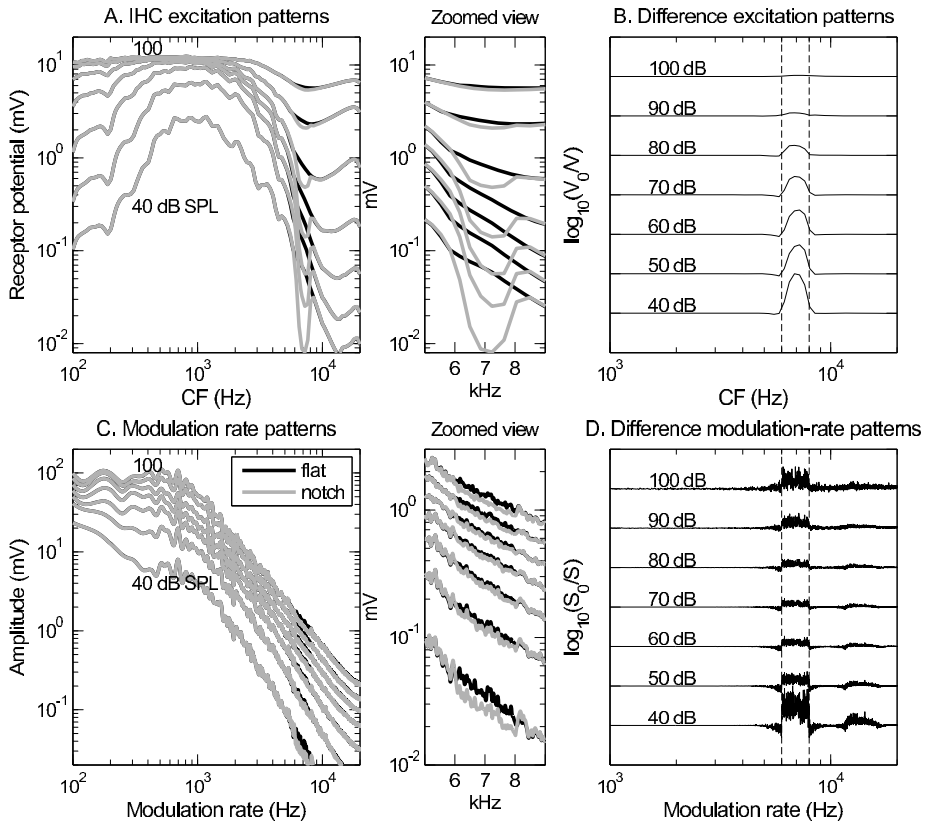

Figure 5.3: **A**. IHC excitation pattern representation of the flat-spectrum (dark traces) and notch noises (light traces). The numbers next to each trace indicate stimulus intensity in decibels SPL. **B**. The difference excitation patterns $[\log_{10}(V_0/V)]$ normalized to the maximum value across CFs and intensities. Traces are vertically displaced to ease the visual comparison of the amplitudes. The numbers next to each trace indicate stimulus intensity in decibels SPL. **C**. IHC modulation-rate pattern representation of the flat-spectrum (dark traces) and notch noises (light traces). **D**. The difference modulation rate patterns $[\log_{10}(S_0/S)]$ normalized to the maximum value across modulation-rates and intensities. The vertical dashed lines in B and D indicate the notch frequency band.

5.3.4 Why does S_{EP} decrease monotonically with increasing sound intensity?

S_{EP} decreases gradually with increasing sound intensity due a combination of three effects: (1) the gradual broadening of the DRNL filters with increasing level (cf. Fig. 7 in Lopez-Poveda and Meddis, 2001); (2) the compression applied by the DRNL filters at moderate to large intensities (cf. Fig. 4 in Lopez-Poveda and Meddis, 2001); (3) the saturation (compression) of the IHC receptor potential at large intensities (cf. Fig. 12 in Lopez-Poveda and Eustaquio-Martin, 2006).

5.3.5 Why does S_{MP} vary non-monotonically with sound intensity?

This question is particularly important, because it may clarify why spectral discrimination improves at high levels in the behavioral results (Chapter 2; Alves-Pinto and Lopez-Poveda, 2005). To answer it, S_{MP} was evaluated for four different model configurations with different nonlinear characteristics. The one considered so far (Fig. 5.2) was denoted DRNL/IHC. In this configuration, both the DRNL and the IHC models were nonlinear (see Lopez-Poveda and Meddis, 2001; Lopez-Poveda and Eustaquio-Martin, 2006). In another configuration (denoted LIN/IHC), the DRNL filter-bank was replaced by a bank of *linear* asymmetric gammatone filters (de Boer, 1975). This was obtained from the DRNL filter-bank simply by setting its parameter "a" to zero (see Lopez-Poveda and Meddis, 2001). The resulting filter-bank may be thought of as the cochlear filter-bank for a listener with total outer-hair-cell dysfunction, hence without the cochlear "active mechanism" (Robles and Ruggero, 2001). Only the IHC was nonlinear in this configuration. In a third configuration (denoted DRNL/HWR), the IHC model was replaced by a half-wave rectification stage followed by a 1^{st}-order Butterworth low-pass filter (cut-off frequency of 1400 Hz). The idea was to preserve the low-pass, half-wave rectification characteristics of the IHC without the nonlinearities associated to the IHC transduction; specifically, without the expansive and compressive growth of the dc receptor potential with increasing intensity that occurs at low and high intensitites, respectively (Patuzzi, and Sellick, 1983; Lopez-Poveda and Eustaquio-Martín, 2006). There were two sources of nonlinearity in this configuration: the DRNL filter-bank and the half-wave rectification. The last model configuration (denoted LIN/HWR) consisted of a linear filter-bank followed by a half-wave rectification/low-pass filter stage, thus, the half-wave rectification was the only nonlinearity in the model.

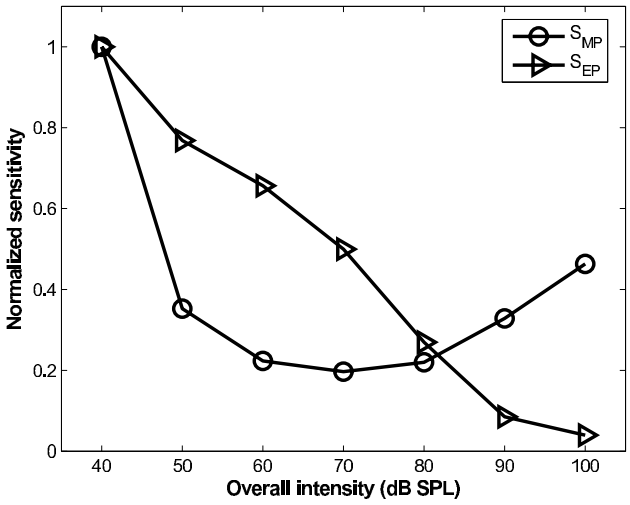

Figure 5.4: Normalized sensitivity to spectral notches as a function of overall sound intensity for the two models tested, S_{EP} and S_{MP}.

Figure 5.5 illustrates the resulting normalized S_{MP}-*vs*-level functions for the four model configurations as indicated in the legend. The figure shows some striking results. First, the sensitivity is intensity independent for the linear-like[3] model (LIN/HWR, triangles in Fig. 5.5). This suggests that it is the nonlinearities (other than the associated to the half-wave rectification) what causes spectral discrimination sensitivity to be intensity-dependent. Second, when the DRNL filter-bank is the only nonlinear stage (DRNL/HWR), sensitivity decreases with increasing intensity up to 80 dB SPL and then increases slightly. In other words, the filter-bank compressive nonlinearity causes sensitivity to decrease for the most part of the intensity range. Third, when the IHC is the only source of nonlinearity, the sensitivity increases rapidly with increasing intensity above 60 dB SPL. Indeed, maximum sensitivity occurs at the highest intensity considered, at which the saturation of the IHC receptor potential is maximum.

Combined, these results indicate (1) that BM compression is responsible for the decrease in sensitivity with increasing intensity up to around 60-70 dB SPL; and (2) that the improvement in spectral discrimination at high inten-

[3]The resulting model is not fully linear because the half-wave rectification is a nonlinear operation.

sities observed psychophysically is strongly associated to IHC nonlinearities. How this happens is discussed in the following section.

5.4 Discussion

We aimed to clarify how high-frequency spectral features are represented in the AN. Our approach consisted in modeling the paradoxical result that discriminating between broadband noises with and without high-frequency spectral notches is most difficult at 70-80 dB SPL than at lower or higher intensities (Chapter 2). Two possibilities were tested: (a) that discrimination is based on the difference between the IHCs excitation patterns for the two stimuli, a representation related to the AN difference rate profile; and (b) that it is based on the difference IHC modulation-rate pattern, a representation related to the difference in the temporal pattern of AN discharges. The results support the latter. This suggests that spectral features as high in frequency as 8 kHz may be encoded in the temporal pattern of AN responses despite the rapid decay of phase locking above 2 kHz. The results also indicate that the improvement at high intensities in the behavioral discrimination task is associated to the saturation of the IHC receptor potential.

5.4.1 Model limitations

The highest CF in our filter-bank (20 kHz) was above the highest CF (8 kHz) for which the filter-bank of Lopez-Poveda and Meddis (2001) was designed. Therefore, the model representations of spectral features above around 8-9 kHz may be inaccurate to some (uncertain) extent. It is unlikely, however, that this affected the representation of the notch, as it extended from 6 to 8 kHz.

We have assumed that the IHC model of Lopez-Poveda and Eustaquio-Martin (2006) is adequate to simulate *human* IHCs. Additionally, we have assumed that its parameters have identical values across CFs, even though this is not the case. At least the endocochlear potential and the relative size of two of the IHC ionic currents are known to vary along the length of the cochlea (Conlee and Bennett, 1993; Kimitsuki *et al.*, 2003). The impact of these assumptions on the IHC representations of the stimulus spectrum is uncertain.

The modulation-rate pattern model (Fig. 5.4, circles) suggests that

spectral-discrimination sensitivity is lowest at 60-70 dB SPL, while according to the psychophysical data (Fig. 2.3), it should be lowest a 70-80 dB SPL. This 10-dB discrepancy may be due to a mismatch between the model and experimental SPL calibration. Alternatively, the gain of one or several model stages may be inadequate. We made no attempt to vary the original model parameters to optimize the fit of the model predictions to the experimental data. Indeed, the 10-dB discrepancy could be solved by reducing the middle-ear gain by 10 dB (results not shown).

5.4.2 Implications of the results

The quality of the IHC excitation pattern representation of the spectral notch, and hence the associated discrimination sensitivity, decreases gradually with increasing sound intensity (Fig. 5.3A and Fig. 5.4). This implies that the quality of the AN rate profile must necessarily decrease with increasing intensity, regardless of the threshold and dynamic range of AN fibers. This undermines the conjecture put forward (Chapter 2; Alves-Pinto and Lopez-Poveda, 2005; Alves-Pinto *et al.*, 2005) that the peaks in the behavioral threshold notch depth *vs.* level functions (Figs. 2.3 and 2.4) reflect the transition between the dynamic ranges of AN fibers with low and high thresholds. This is not to negate the existence of two types of AN fibers in the human auditory nerve or to challenge that two fiber types with different properties can account for other phsychoacoustical phenomena (e.g. Viemeister, 1988; Zeng *et al.*, 1991).

The difference between the IHC modulation-rate pattern representations for the flat-spectrum and notch noises decreases with increasing intensity up to around 60-70 dB SPL and increases again at higher intensities (Fig. 5.3D). Therefore, the associated spectral-discrimination sensitivity is a non-monotonic function of intensity (circles in Fig. 5.4), which is consistent with the observations of the psychophysical study (Chapter 2; Alves-Pinto and Lopez-Poveda, 2005). Since AN discharge follows the IHC receptor potential, this supports two important conjectures. First, spectral discrimination may be based on comparisons of internal representations of the spectra obtained by precise analysis of the timing of AN spikes. The actual mechanism that allows the central auditory system to extract such a representation is unknown but may be similar in effect to a Fourier transform. Indeed Young and Sachs (1979) showed that a method of this kind allows recovering vowel spectra from AN responses even at very high intensities at which most fibers are saturated. One could argue that the results of Young and Sachs are not

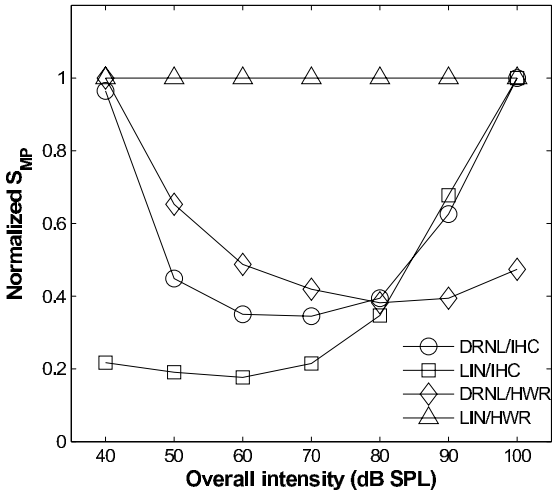

Figure 5.5: Intensity dependence of the discrimination sensitivity based on the modulation rate pattern representation of the spectrum for different model configurations with different sources of compression. See main text for details.

surprising because vowel formants are in the low-frequency region, where phase locking is strong (Palmer and Russell, 1986), and significant phase locking occurs even when the fibers discharge at saturation (Rose *et al.*, 1967; Cooper *et al.*, 1993). What is most interesting, perhaps, is that the present results support that a similar mechanism may apply to *nonperiodic*, high-frequency stimuli. This leads to the second conjecture.

The present results also indicate that the high-frequency spectrum (at least up to 8 kHz) is likely to be encoded in the timing of AN-spike occurrences. Heinz *et al.* (2001) reached a similar conclusion based on a computational model of the just-noticeable difference in frequency between two pure tones. The present study extends the conclusion of Heinz *et al.* (2001) to broad-band, *nonperiodic* stimuli. Whatever the stimuli, however, this would be possible only if AN fibers fired synchronously with modulation rates of their driving receptor potential waveforms higher than 4 kHz. While this seems impossible in view of early reports (e.g. Johnson, 1980; Russell and Palmer, 1986; Joris and Yin, 1992), recent studies demonstrate that significant phase locking occurs for frequencies as high as 14 kHz (Recio-Spinoso *et al.*, 2005; Temchin *et al.*, 2005). Therefore, the present conclusions as well as those of Heinz *et al.* (2001) are physiologically plausible.

5.4.3 What is the reason for the non-monotonic aspect of the threshold notch depth *vs.* level function?

It has been shown (Fig. 5.5) that the improvement in spectral discrimination sensitivity at high intensities almost certainly relates to inherent IHC nonlinearities different from mere half-wave rectification. The actual mechanism is uncertain but may be related to the saturation of the receptor potential at high intensities. The saturation of the IHC transducer current would alter the receptor potential waveform at high intensities relative to the corresponding waveform at low intensities. For example, a perfectly sinusoidal receptor-potential waveform at low intensities would become akin to a square waveform at high intensities. The receptor potential waveform drives neurotransmitter release from the IHC to the synapses between the cell and its neighbor AN fibers. As a result, the latter would show different patterns of phase-locked AN spikes at low and high intensities. Indeed, assuming perfect phase-locking, at low intensities spikes would occur with a periodicity equal to that of the driving sinusoidal receptor potential. At high intensities, by contrast, spikes would occur not with the frequency of the sinusoidal waveform, but also with those of its harmonics. Of course, perfect phase locking is unlikely to occur, but the principle would still apply.

The nonlinearity associated to the IHC transducer current occurs instantaneously (at least in the present model). Therefore, its effects would also alter the shape of complex receptor potential waveforms like the ones produced by the flat-spectrum and notch noises considered here. Consider now an IHC with a CF equal to the notch center frequency. At low intensities, both the flat-spectrum and the notch noises would drive the IHC into a linear region. Discriminating between the two would be purely based on the amplitude of the receptor produced (or equivalently on the associated number of spikes) evoked by the two noise stimuli. At high intensities, however, the flat-spectrum noise would drive the IHC in question more into saturation than the less-energetic notch noise stimulus. As a result, discriminating between the two stimuli could be based not only on the amplitude of the receptor potential (or the number of spikes), but also on the timing of the spikes. The spike trains in response to the flat-spectrum noise would contain distortion frequencies not present (or present with less amplitude) in the spike trains evoked by the notch noise. In other words, harmonic distortion associated to the saturation of the receptor potential would explain why the differences between the modulation-rate patters of Fig. 5.3D extend to frequencies outside the notch band.

The effect just described would affect primarily IHCs (or AN fibers) with CFs within the notch frequency band, but would not be restricted to them. IHCs (or AN fibers) with CFs above the notch band would have a broad-enough frequency response at high intensities to feel energy difference between the two stimuli and translate into spike trains with different high-frequency timing.

This proposed mechanism is physiologically plausible because isolated IHCs introduce distortion (Jaramillo *et al.*, 1993).

5.5 Conclusions

A computational model has been used to simulate two different representations of the spectrum of broadband sounds at the level of the inner hair cell: the excitation pattern and the modulation-rate pattern. Only the latter appears consistent with the behavioral result that is more difficult to discriminate between broadband noise stimuli with and without a high-frequency spectral notch at 70-80 dB SPL than at lower or higher intensities. This suggests that discrimination between two sounds with different high-frequency spectra relies on differences in the timing of AN spikes, rather than on spike rates evoked by the two stimuli. It further suggests that AN spikes occur in phase with the receptor potential for frequencies as high as 8 kHz. Additionally, the simulations suggest that the improvement in spectral discrimination above 70 dB SPL is associated to inherent inner hair cell nonlinearities, possibly to the saturation of the receptor potential at high intensities.

Chapter 6

Overall Discussion

The present work was aimed at elucidating the nature of the code underlying the representation of high-frequency spectral notches in the AN. As explained in the Introduction, the common notion is that these notches are encoded in the rate profile of AN fibers only, and that the quality of this code deteriorates gradually with increasing sound level. The present work tested this notion by investigating our ability to perceive high-frequency spectral notches at increasing stimulus levels. Additionally, it analyzed responses of AN fibers and simulated the output of the peripheral auditory system in response to stimuli with and without high-frequency spectral notches. The main result was that, contrary to the common notion, behavioral discrimination varies non-monotonically with level, being harder at 70-80 dB SPL than at lower and higher levels. Furthermore, several of the results point to a code based on the temporal characteristics of AN activity.

6.1 High-frequency spectral notches are unlikely to be represented in the rate profile of discharge rates of AN fibers

Strong evidence against a rate-place code for high-frequency spectral notches arises from a correlation analysis between the results of a noise discrimination experiment (Chapter 2) and the masking patterns of flat-spectrum and notch noises (Chapter 3). As explained in Chapter 3, the difference between the masking patterns of the two noises suggests that the internal representation of the notch deteriorates gradually with increasing level, even over a range of levels for which discrimination improves with increasing level.

Further evidence against a rate-place code comes from the analysis of guinea-pig AN responses to stimuli similar to those used in the psychophysical experiments. The level dependence of the rate-profile representation of high-frequency spectral notches (Figs. 4.2 and 4.3) is inconsistent with the effect of level on the discrimination between the flat-spectrum and notch noises (Figs. 2.3-2.4, 2.6-2.8, 3.2).

However, the conclusions drawn from the AN rate profiles also differ from those drawn from the masking patterns. Indeed the quality of the rate-profile representation of high-frequecy spectral notches is better around 70-80 dB SPL than at lower and higher levels (Fig. 4.2), while the notch is less clearly represented in the masking patterns at 70-80 dB SPL than at 50 dB SPL (Fig. 3.4). This discrepancy might reflect differences in signal processing between the peripheral auditory systems of guinea-pigs and humans. Such differences might also account for the lack of correlation between the rate-profile representation of spectral notches (Fig. 4.2) and the results of the noise discrimination task.

The results of the computational modeling study described in chapter 5 also dispute the notion that high-frequency spectral features are encoded by means of a rate-place code. Indeed the analysis of simulated responses of a population of IHCs indicates that the excitation patterns of flat-spectrum and notch noises become gradually more similar as the stimulus level increases (Fig. 5.4), in agreement with the masking patterns (Figs. 3.4-3.5).

6.2 On the use of a temporal code for discrimination of high-frequency spectral features

If high-frequency spectral notches are not represented in the rate profile of AN fibres, how are they encoded? Acoustic information may be represented in only two possible ways (see Introduction): in a rate-profile or in the discharge times of the population of AN fibers. Therefore, by exclusion, the discrimination of broadband stimuli with and without high-frequency spectral notches is likely to rely on comparisons of the temporal pattern of AN discharges, at least above 80 dB SPL.

This conjecture is supported by several results reported in this thesis. First, a reasonable degree of correlation has been shown to exist between the psychophysical and physiological results pertaining to the effect of sound level on a spectral discrimination task, but only when the instantaneous

discharge rates of the AN fibres were analyzed in time windows between 4 and 9 ms according to an 'optimal observer' analysis. Unfortunately, this result is inconclusive because the number of stimulus repeats considered, though adequate for a rate-profile analysis, was insufficient to draw reliable conclusions from an 'ideal observer' analysis of AN responses, especially for short time windows (shorter than about 20 ms).

It is worth pointing out that spectral detection and spectral discrimination are in fact two different tasks. Spectral discrimination does not actually require that the listener identifies any spectral features of the stimuli. Therefore, performing such task correctly could have been based on detecting any difference (however small) between the AN spike trains evoked by the two noises. In other words, the amount of information necessary for performing the task correctly would be much less than that required for an identification task. This implies that any available temporal information, even if "noisy" or insignificant relative to other types of information, could still be critical and therefore could be used to discriminate between the two noises.

This conjecture is consistent with the explanation posed in Chapter 5 for the origin of the differences between the internal signals produced by the two noises. According to this explanation, IHC nonlinearity would be related to the improvement in spectral discrimination above 80 dB SPL.

The use of temporal information for high-frequency discrimination has been put forward previously by studies looking for physiological correlations of psychophysical frequency discrimination (Heinz *et. al.*, 2001). This study concluded that performance in frequency discrimination predicted from an 'ideal observer' analysis of simulated AN responses correlates better with psychophysical performance when the temporal information is included in the analysis. The authors further suggested that "there is significant temporal information in the AN for frequency discrimination up to at least 10 kHz, and thus, temporal schemes cannot be rejected at high frequencies, based on the decrease in phase locking in the AN" (Heinz *et. al.*, 2001, pp. 2309). The present study arrives at a similar conclusion but for the discrimination of broadband stimuli. As further support to this conclusion, recent physiological studies (Recio-Spinoso *et al.*, 2005; Temchin *et al.*, 2005) have shown that there may exist significant phase locking at frequencies as high as 14000 Hz.

To confirm the contribution of temporal information to the discrimination between flat-spectrum and notch noise with an "optimal processor" analy-

sis would require a larger fiber sample, more stimulus repeats, and longer stimuli. This may also occur in the auditory system: small amounts of information, gathered accross hundreds of fibers distributed along a given range of frequencies, may sum in a significant and useful amount of information not possible to achieve with a small number of fibers and with short stimuli.

What kind of temporal information could be involved? Two possibilities exist: information given 1) in the fine-time structure or 2) in the envelope of the trains of spikes. The first possibility was essentially explored in the modelling study described in chapter 5. The 'ideal observer' analysis could have been used to test this possibility. Unfortenately, we did not have enough data. This analysis would require using a very short time window (FFT analysis for frequencies up to 7000 Hz would require post-stimulus time histograms with binwidths of 0.14 ms), which requires in turn measuring spike trains for many more stimulus repeats than we considered to minimize analysis artifacts. By contrast, testing that temporal information is conveyed in the envelope of AN discharges would require using longer time windows, typically longer than of 2 ms. The results show a close match between behavioral and neural performance for time windows between 4 and 9 ms (Fig. 4.4). Interpretations for this result were discussed in Chapter 4. Unfortenately, it is possible that the results of this analysis may be artifacts of the method.

A neural code based on the instantaneous acytivity of AN fibers would be compatible also with the other behavioral results described in chapter 2. It was observed that threshold notch depths are higher for short than for long stimuli. Assuming that there is relevant information in the cadence with which spikes occur, then more 'bits' of useful information would be acumulated for longer stimuli this way improving discrimination performance.

In chapter 2 it was observed that discrimination is harder for narrower notches (Fig. 2.6) but the nonmonotonic effect of level is still observed for some listeners. A temporal code would not change the fact that narrower notches are signaled by fewer fibers, and thus give higher thresholds, and it would still determine the non-monotonic effect of level.

The observation that the nonmonotonic effect of level occurs when the noise level is roved was interpreted as indicating that discrimination is based in comparing the overall spectral shapes of the flat-spectrum and notch noises (see section 2.4.1 a). This implies that somehow the information is integrated across CFs. This interpretation matches with the existence of a temporal

code based on the fine-time structure of the trains of spikes accross CFs, as described above (section 5.4.3). In this case it would be the comparison of modulation rates across CFs that determines the threshold of discrimination.

Several studies investigted the temporal aspects of spectral processing in sound localization (for a review see Carlile, 2005). One of them (Jin, 2001), relates the time window the localisation system needs to process spectral information with the response and decay times of cochlear filters. It was noted that reliable information on the spectrum level in a given filter can only be provided if the input signal to that filter is integrated over a given time interval (Jin, 2001). The study suggests that the lower limit for this time window is about 5ms. This time window is within the range of time windows (∼4-9 ms) for which the 'ideal observer' analysis of instantaneous discharge rates of AN fibers provided a good match with behavioral performance on noise discrimination.

6.3 Ideas for future work

It has been shown that the amount of perceptually-relevant information for high-frequency spectral discrimination is less for sound levels around 80 dB SPL than for lower or higher levels. This still needs an explanation. The results described suggest that it is not consistent with notches being encoded in the rate profile of AN fibers (Figs. 4.3 and 4.4(a)) nor to having two fiber populations with different thresholds and dynamic ranges (Fig. 4.4(b)). They suggest, alternativelly, that the cadence of spikes may provide useful information for noise discrimination. This conjecture lacks experimental confirmation. Therefore, proposed future work includes:

1. Carrying out psychophysical studies aimed at testing whether temporal information is useful in the encoding of high-frequency spectral notches. If there is then, the next step would be to investigate the nature of this temporal information: whether it is in the fine-time structure or in the envelope of AN discharge rates.

2. Recording AN responses with the objective of investigating the temporal aspects of AN spike trains. This would imply to gather data for longer stimuli and for as many repeats for each fiber as possible. Besides applying the "ideal observer" analysis to those responses, another possible way exploring the existence of temporal information in the fine-time structure of AN fibers would be through an FFT analysis of

AN responses. In this case, the number of repeats would not be be so critical, but it would require to gather data for longer stimuli.

3. Comparing animal behavioral data with physiological data on AN activity, as was done in the present thesis for humans and guinea pigs. This would avoid having to deal with differences between species and therefore would allow to establish reliable correlations between behavioral and predicted neural performance.

4. Investigate whether the present conclusions apply to different (lower or higher) notch frequency bands and to the discrimination of high-frequency features in realistic HRTFs.

5. Analyzing AN spike trains with a procedure different from the optimal processor. For instance, the analysis of Chase and Young (2000) could be applied. Basically it analyzes the spike times and estimates the mutual information between each one of the spike trains and the different stimuli. This analysis does not depend so critically on the number of repeats as the "ideal observer" type of analysis described in Chapter 4.

6. Develop the computational modeling study of Chapter 5 to simulate AN responses and apply the optimal processor analysis (as well as other analysis) to the simulated spike trains. This has the advantages that it allows considering a large number of fibers and stimulus repeats, as well as the use of longer stimuli.

7. Search for correlations between human performance in sound localization in the median vertical plane with their ability to discriminate between noise stimuli with high-frequency spectra.

Conclusions

For most listeners, discriminating between two broadband noises, with and without a rectangular spectral notch centered at a high frequency, varies *non*monotically with level: it becomes increasingly more difficult as the stimulus level increases from 30 up to 70-80 dB SPL and improves again above 80 dB SPL.

1. The nonmonotonic character of the threshold notch depth *vs.* level function is independent of notch bandwidth and stimulus duration. Furthermore, threshold notch depths are hardly affected by the stimulus rise time and depend strongly on the stimulus duration. Notch depth values are larger for shorter stimuli.

2. The effect of level on the discrimination between the two noises does not match with an encoding mechanism at the AN in which spectral notches are represented in the profile of discharge rates of AN fibers. Evidence is diverse: 1) the effect of level on the masking patterns of the two noises does not match with that on the discrimination between them; 2) the representation of high-frequency spectral notches in the profile of discharge rates of guinea pig AN fibers is better around 70-80 dB SPL than at low and high levels. This would suggest that discrimination is easier around 70-80 dB SPL, which clearly does not occurr in the discrimination task; and 3) the representation of the spectral notch in the excitation pattern difference at the output of the inner hair cell, calculated with a model of the peripheral auditory system, decreases monotonically with level rather than nonmonotonically as the discrimination task suggests.

3. The mismatch in the level dependence between noise discrimination and the internal representation of high-frequency spectral notches in the rate-profile of AN fibers is also inconsistent with the notch being represented in the profile of discharge rates of HSR fibers at low levels

and on that of LSR fibers at high levels, with the peak in the function reflecting the transition between the two types of fibers.

4. Discrimination between flat-spectrum and notch noises is probably based on cues different from the rate-profile representation of the spectal notch. Analysis of the output signal of inner hair cells, simulated with a computational model the peripheral auditory system, suggest that information carried in the temporal variations of the signal might provide information useful to discriminating between the two noises.

References

Alves-Pinto A, Lopez-Poveda E.A. (**2005**) "Detection of high-frequency spectral notches as a function of level," J. Acoust. Soc. Am. **118**: 2458-2469.

Alves-Pinto, A., Lopez-Poveda, E. A., and Palmer, A. R. (**2005a**). "Auditory nerve encoding of high-frequency spectral information," in *Mechanisms, Symbols, and Models Underlying Cognition*, edited by J. Mira, and J. R. Álvarez, Lecture Notes in Computer Science **3561**, 223-232.

Alves-Pinto, A., Lopez-Poveda, E. A., and Palmer, A. R. (**2005b**). "Auditory nerve response to broadband noise with high-frequency spectral notches," Abstract 4pPP16, *149th Meeting of the Acoustical Society of America*, Vancouver (Canada), 16-20 May; J. Acoust. Soc. Am. **117**: 2654.

Baker, R. J., Rosen, S., and Darling, A. M. (**1998**). "An efficient characterization of human auditory filtering across level and frequency that is also physiologically reasonable," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London), pp. 81-88.

Bland, J.M., and Altman, D.G. (**1996a**). "Measurement error and correlation coefficients," Brit. Med. J. **313**: 106.

Bland, J.M., and Altman, D.G. (**1996b**). "Transformations, means and confidence intervals," Brit. Med. J. **312**: 1079.

Bland, J.M., and Altman, D.G. (**1996c**). "Transforming data," Brit. Med. J. **312**: 770.

Bland, J.M., and Altman, D.G. (**1996d**). "The use of transformation when comparing two means," Brit. Med. J. **312**: 1153.

Blauert, J. (**1969/70**). "Sound localization in the median plane," Acustica **22**: 205-213.

Butler, R. A., and Belendiuk, K. (**1977**). "Spectral cues utilized in the localization of sound in the median sagittal plane," J. Acoust. Soc. Am. **61**: 1264-1269.

Butler, R. A., and Humanski, R. A. (**1992**). "Localization of sound in the vertical plane with and without high-frequency spectral cues," Percept. Psychophys. **51**: 182-186.

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**: 2643-2654.

Buus, S., and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am. **90**: 1371-1380.

Carlile, S. and Pettigrew, A.G. (**1987a**). "Directional propeties of the auditory periphery in the guinea pig," Hear. Res. **31**: 111-122.

Carlile, S. and Pettigrew, A.G. (**1987b**). "Distribution of frequency sensitivity in the superior colliculus of the guinea pig," Hear. Res. **31**: 123-136.

Carlile, S., Martin, R., McAnally, K. (**2005**). "Spectral information in sound localization," in *International Review in Neurobiology*, edited by M. Malmierca and D.R.F. Irvine (Elsevier Academic Press), **70**: 399-434.

Carlyon, R.P. and Moore, B.C.J. (**1984**). "Intensity discrimination: A severe departure from Webers law," J. Acoust. Soc. Am. **76**: 1369-1376.

Chase, S.M. and Young, E.D. (**2006**). "Spike-timing codes enhance the representation of multiple simultaneous sound-localization cues in the Inferior Colliculus," J. Neurosc. **26**: 3889-98.

Cheatham, M.A. and Dallos, P. (**2001**). "Inner hair cell response patterns: implications for low-frequency hearing," J. Acoust. Soc. Am. **110**: 2034-2044.

Conley, J.W. and Bennet, M.L. (**1993**). "Turn-specific differences in the endocochlear potential between albino and pigmented guinea pigs," Hear. Res. **65**: 141-150.

Conley, R. A. and Keilson, S. E. (**1995**). "Rate representation and discrimination of second formant frequencies for $/\varepsilon/ - like$ steady-state vowels in cat auditory-nerve," J. Acoust. Soc. Am, **98**: 3223-3234.

Cooper, N. P., Robertson, D., and Yates, G. K. (**1993**). "Cochlear nerve fiber responses to amplitude modulated stimuli: Variations with spontaneous rate and other response characteristics," J. Neurophysiol. **70**: 370-386.

de Boer, E., (**1975**). "Synthetic whole-nerve action potentials for the cat," J. Acoust. Soc. Am. **58**: 1030-1045.

Delgutte, B. (**1980**). "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers," J. Acoust. Soc. Am. **68**: 843-857.

Delgutte, B., and Kiang, N. Y. S. (**1984a**). "Speech coding in the auditory nerve: III. Voiceless fricative consonants," J. Acoust. Soc. Am. **75**: 887-896.

Delgutte, B., and Kiang, N. Y. S. (**1984b**). "Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics," J. Acoust. Soc. Am. **75**: 897-907.

Delgutte B (**1996**) Physiological Models for Basic Auditory Percepts, in *Auditory Computation*, edited by H.L. Hawkins, T.A. McMullen, A.N. Popper, R.R. Fay (Springer-Verlag, New York), pp. 157-220.

Dickson, B., Berhend, O., Anandan, A.and Carlile, S. (**2002**). "Changes in the spectral filtering due to variation in the auditory periphery of the guinea pig," Proceedings of the Australian Neuroscience Society, **13**: 209.

Dickson, B., Berhend, O., and Carlile, S. (**200**3). "Acoustic properties of the outer of the guinea pig," Proc. Aust. Neurosc. Soc., **1**4: 267.

Evans, E. F. (**1972**). "The frequency response and other properties of single fibers in the guinea-pig cochlear nerve," J. Physiol. **226**: 263-287.

Evans, E. F. (**1975**). Cochlear nerve and cochlear nucleus, in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer, Berlin), Vol. 5/2, pp. 1-108.

Evans, E. F., and Palmer, A. R. (**1980**). "Relationship between the dynamic range of cochlear nerve fibers and their spontaneous activity," Exp. Brain Res. **40**: 115-118.

Florentine, M. (**1986**). "Level discrimination of tones as a function of duration," J. Acoust. Soc. Am. **79**: 792-798.

Florentine, M., Buus, S., and Mason, C.R. (**1987**). "Level discrimination as a function of level for tones from 0.25 to 16 kHz," J. Acoust. Soc. Am. **81**: 1528-1541.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched noise data," Hear. Res. **47**: 103-138.

Grant, K. W., and Walden, B. E. (**1996**). "Spectral distribution of prosodic information," J. Speech Hear. Res. **39**: 228-238.

Green, D. M. (**1960**). "Auditory detection of a noise signal," J. Acoust. Soc. Am. **32**: 121-131.

Green, D. M. (**1988**). *Profile Analysis* (Oxford University Press, New York).

Greenwood, D.D. (**1971**). "Aural combination tones and auditory masking," J. Acoust. Soc. Am. **50**: 502-543.

Guinan, J.J. Jr (**1996**) Physiology of olivocochlear efferents, in *The Cochlea*, edited by P. Dallos, A.N. Popper, and R.R. Fay, (Springer), **8**: 435-502.

Harris, D. M. and Dallos, P. (**1979**). "Forward masking on auditory nerve fiber responses," J. Neurophysiol. **42**: 1083-1107.

Hartmann, W. M. (**1998**). *Signals, Sound and Sensation* (AIP Press, New York).

Hartmann, W. M., and Rakerd, B. (**1993**). "Auditory spectral discrimination and the localization of clicks in the sagittal plane," J. Acoust. Soc. Am. **94**: 2083-2092.

Hebrank, J., and Wright, D. (**1974**). "Spectral cues used in the localization of sound sources on the median plane," J. Acoust. Soc. Am. **56**: 1829-1834.

Heinz, M. G., and Formby, C. (**1999**). "Detection of time- and band-limited increments and decrements in a random-level noise," J. Acoust. Soc. Am. **106**: 313-326.

Heinz MG, Colburn HS, Carney LH (**2001**). "Evaluating Auditory Performance Limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Computation **13**: 2273-2316.

Houtgast, T. (**1972**). "Psychophysical evidence for lateral inhibition in hearing," J. Acoust. Soc. Am. **51**: 1885-1894.

Houtgast, T. (**1973**). "Psychophysical experiments on tuning curves and two-tone inhibition," Acustica. **29**: 168-179.

Houtgast, T. (**1974**). "Lateral suppression in hearing," Ph.D. thesis. Free University of Amsterdam.

Humanski, R. A., and Butler, R. A. (**1988**). "The contribution of the near and far ear toward localization of sound in the sagittal plane," J. Acoust. Soc. Am. **83**: 2300-2310.

Jaramillo, F., Markin, V.W. and Hudspeth, A.J. (**1993**)."Auditory illusions and the single hair cell," Nature **364**: 527-529.

Jin, C. T. (2001). "Spectral analysis and resolving spatial ambiguities in human sound localization, " Ph.D. thesis. University of Sydney (Sydney , Australia).

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**: 1115-1122.

Joris, P.X. and Yin, T.C. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**: 215-232.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge).

Kimitsuki, T., Kawano, K., Matsuda, K., Haruta, A., Nakajima, T. and Komune, S. (**2003**). "Potassium current properties in apical and basal inner hair cells from guinea-pig cochlea," Hear. Res. **180**: 85-90.

Levitt, H. (**1971**). "Transformed up down methods in psychoacoustics," J. Acoust. Soc. Am. **49**: 467-477.

Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**: 442-455.

Louage, D. H., van der Heijden, M., and Joris, P. X. (**2004**). "Temporal properties of responses to broadband noise in the auditory nerve," J. Neurophysiol. **91**: 2051-2065.

Lopez-Poveda, E. A. (**1996**). "The physical origin and physiological coding of pinna-based spectral cues," Ph.D. thesis. Loughborough University (Loughborough, U.K.).

Lopez-Poveda, E. A., and Meddis, R. (**1996**). "A physical model of sound diffraction and reflections in the human concha," J. Acoust. Soc. Am. **100**: 3248-3259.

Lopez-Poveda, E. A., and Meddis, R. (**2001**). "A human nonlinear cochlear filterbank," J. Acoust. Soc. Am. **110**: 3107-3118.

Lopez-Poveda EA (**2005**). Spectral processing by the peripheral auditory system: Facts and models, in *International Review in Neurobiology*, edited by M. Malmierca and D.R.F. Irvine (Elsevier Academic Press), **70**: 7-48.

Lopez-Poveda, EA, Alves-Pinto, A, Palmer, AR and Eustaquio-Martín, A (**accepted for publication in Neurocomputing**) "Rate versus time representation of high-frequency spectal notches in the peripheral auditory system: a computational modeling study,".

Lopez-Poveda, E.A., Alves-Pinto, A. and Palmer, A.R. (in press) "Psychophysical and physiological assessment of the representation of high-frecuency spectral notches in the auditory nerve," in *Hearing - from basic research to applications*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, J. Verhey (Springer Verlag, Heidelberg), scheduled for publication in 2007.

Lopez-Poveda, E. A., Eustaquio-Martín, A. (**2006**). "A biophysical model of the inner hair cell: The contribution of potassium current to peripheral compression," J. Assoc. Res. Otolaryngol. **7**: 218-235.

Macpherson E. A., and Middlebrooks, J. C. (**2000**). "Localization of brief sounds: effects of level and background noise," J. Acoust. Soc. Am. **108**: 1834-1848.

May, B. J. (**2003**). "Physiological and psychophysical assessments of the dynamic range of vowel representations in the auditory periphery," Speech Communication **41**: 49-57.

Meddis, R. (**1988**). "Simulation of auditory-neural transduction: Further studies," J. Acoust. Soc. Am. **83**: 1056-1063.

Meddis, R., O'Mard, L.P.O., and Lopez-Poveda, E.A. (**2001**). "A computational algorithm for computing non-linear auditory frequency selectivity," J. Acoust. Soc. Am. **109**: 2852-2861.

Meddis, R. and O'Mard, L.P.O. (**200**5). "A computer model of the auditory-nerve response to forward-masking stimuli," J. Acoust. Soc. Am. **117**: 3787-3798.

Moller, A.R. (**2000**). *Hearing - Its Physiology and Pathophysiology* (Academic Press).

Moore, B. C. J., Glasberg, B. R. (**1983**). "Masking patterns for synthetic vowels in simultaneous and forward masking," J. Acoust. Soc. Am. **73**: 906-917.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**: 820-836.

Moore, B. C. J., Alcántara, J.I. and Dau, T. (**1998**). "Masking patterns of sinusoidal and narrow-band noise maskers," J. Acoust. Soc. Am. **104**: 1023-1038.

Moore, B. C. J. (**2005**). "Basic psychophysics of human spectral processing," in *International Review in Neurobiology*, edited by Malmierca, M and Irvine, D.R.F. (Elsevier Academic Press), **70**: 49-86.

Nizami, L., Reimer, J.F., and Jesteadt, W. (**2001**). "The intensity-difference limen for Gaussian-enveloped stimuli as a function of level: Tones and broadband noise," J. Acoust. Soc. Am. **110**: 2505-2515.

Oxenham, A.J. (**2001**). "Forward masking: adaptation or integration?," J. Acoust. Soc. Am. **109**: 732-741.

Oxenham, A.J. and Moore, B.C.J. (**1994**). "Modeling the additivity of nonsimultaneous masking," J. Acoust. Soc. Am. **80**: 105-18.

Palmer, A. R. and Evans, E. F. (**1980**). "Cochlear fibre rate-intensity functions: no evidence for basilar membrane nonlinearities," Hear. Res. **2**: 319-326.

Palmer, A. R. and King, A. J. (**1983**). "A monaural space map in the guine pig superior colliculus," Hear. Res. **17**: 267-280.

Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hear. Res.**24**: 1-15.

Palmer AR, Winter IM, Darwin CJ (**1986**) "The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea-pig cochlear nerve and primary-like cochlear nucleus neurons," J. Acoust. Soc. Am. **79**: 100-113.

Patterson, R. D., and Moore, B. C. J. (**1986**). Auditory filters and excitation patterns as representations of frequency resolution, in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 123-177.

Poon, P. W. F., and Brugge, J. F. (**1993**). "Sensitivity of auditory nerve fibers to spectral notches," J. Neurophysiol. **70**: 655-666.

Pralong, D., and Carlile, S. (**1996**). "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," J. Acoust. Soc. Am. **100**: 3785-3793.

Recio-Spinoso A, Temchin AN, van Dijk P, Fan YH, Ruggero MA (**2005**). "Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers," J. Neurophys. **93**: 3615-3634.

Rice, J. J., Young, E. D., and Spirou, G. A. (**1995**). "Auditory-nerve encoding of pinna-based spectral cues: Rate representation of high-frequency stimuli," J. Acoust. Soc. Am. **97**: 1764-1776.

Robles, L., and Ruggero, M.A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**: 1305-1352.

Robles, L., Ruggero, M.A., and Rich, N.C. (**1997**). "Two-tone distortion on the basilar membrane of the chinchilla cochlea," J. Neurophysiol. **77**: 2385-2399.

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked response to low-frequency tones in single auditory nerve fibers of squirrel monkey," J. Neurophysiol. **30**: 769-793.

Rose, J. E., Hind J.E., Anderson, D. J., and Brugge, J. F. (**1971**). "Some effects of stimulus intensity on response of auditory-nerve fibers in the squirrel monkey," J. Neurophysiol. **34**: 685-699.

Ruggero, M.A., Rich, N.C., Recio, A., Narayan, S.S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**: 2151-2163.

Russell, I. J., and Sellick, P. M. (**1978**). "Intracellular studies of hair cells in the mammalian cochlea," J. Physiol. **284**: 261-290.

Sachs, M. B., and Abbas, P. J. (**1974**). "Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli," J. Acoust. Soc. Am. **56**: 1835-1847.

Sachs, M. B., and Young, E. D. (**1979**). "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," J. Acoust. Soc. Am. **66**: 470-479.

Sachs, M. B., and Young, E. D. (**1980**). "Effects of nonlinearities of speech encoding in the auditory nerve," J. Acoust. Soc. Am. **68**: 903-913.

Schalk T.B., and Sachs, M. B. (**1980**). "Nonlinearities in auditory-nerve fiber responses to bandlimited noise," J. Acoust. Soc. Am. **67**: 903-913.

Shamma, S. A., Chadwick, R. S., Wilbur, W.J., Morrish, K. A., Rinzel, J. (**1986**). "A biophysical model of cochlear processing: intensity dependence of pure tone responses," J. Acoust. Soc. Am. **80**: 133-145.

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**: 1848-1861.

Shaw, E. A. G., and Teranishi, R. (**1968**). "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**: 240-249.

Siebert WM (**1970**). "Frequency Discrimination in Auditory System - Place Or Periodicity Mechanisms," Proceedings of the Institute of Electrical and Electronics Engineers **58**: 723-&.

Siebert WM (**1965**). "Some Implications of Stochastic Behavior of Primary Auditory Neurons," Kybernetik **2**: 206-&.

Siebert WM (**1968**) "Stimulus transformation in the peripheral auditory system," in: *Recognizing Patterns*, edited by P.A. Kolers and M. Eden (MA: MIT Press, Cambridge), pp. 104-133.

Simpson, A., McDemott, H.J. and Dowell, R.C. (2005). "Benefits of audibility for listeners with severe high-frequency hearing loss," Hear. Res. **210**: 42-52.

Sinex, D. G. and Geisler, C. D. (**1993**). "Response of auditory-nerve fibers to consonant-vowel syllables," J. Acout. Soc. Am. **73**: 602-615.

Smith, R. L., and Brachman, M. L. (**1980**). "Operating range and maximum response of single auditory nerve fibers," Brain Res. **184**: 499-505.

Sinyor, A. nas Lazlo, C.A. (**1973**). "Acoustic behavior of the outer ear of the guinea pig and the influences of the middle ear," J. Acoust. Soc. Am. **54**: 916-921.

Temchin AN, Recio-Spinoso A, van Dijk P, Ruggero MA (**2005**). "Wiener kernels of chinchilla auditory-nerve fibers: Verification using responses to tones, clicks, and noise and comparison with basilar-membrane vibrations," J. Neurophys. **93**: 3635-3648.

Viemeister, N.F. (**1983**). "Intensity discrimination at high frequencies in the presence of noise," Science **221**: 1206-1208.

Viemeister, N. F. (**1988**). "Intensity coding and the dynamic range problem," Hear. Res. **34**: 267-274.

Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**: 858-865.

Vliegen, J., and Van Opstal, A. J. (**2004**). "The influence of duration and level on human sound localization," J. Acoust. Soc. Am. **115**: 1705-1713.

Westerman LA, Smith RL (**1984**) "Rapid and short-term adaptation in auditory nerve responses," Hear Res **15**: 249-260.

Wilson BS, Schatzer R, Lopez-Poveda EA, Sun X, Lawson DT, Wolford RD. (**2005**). "Two new directions in speech processor design for cochlear implants," Ear and Hearing, **26**: 73S-81S.

Wilson BS, Schatzer R, Lopez-Poveda EA. (**2006**). "Possibilities for a closer mimicking of normal auditory functions with cochlear implants," in *Cochlear Implants*, edited by S.B. Waltzman and J.T. Roland (Thieme Medical Publishers, New York), 2nd Edition, pp. 48-56.

Winter, I. M., and Palmer, A.R. (**1991**). "Intensity coding in low-frequency auditory-nerve fibers of the guinea pig," **90**: 1958-1967.

Winter, I. M., Robertson, D., and Yates, G. K. (**1990**). "Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres," Hear. Res. 45: 191-202.

Yates GK (**1991**) "Auditory-Nerve Spontaneous Rates Vary Predictably with Threshold," Hear. Res. **57**: 57-62.

Young, E.D., and Sachs, M.B., (**1979**). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers," J. Acoust. Soc. Am. **66**: 1381-1403.

Young, E. D., and Barta, P.E. (**1986**). "Rate responses of auditory nerve fibers to tones in noise near masked threshold," J. Acoust. Soc. Am. **79**: 426-442.

Zeng, F.-G., Turner, C. W., and Relkin, E. M. (**1991**). "Recovery from prior stimulation II; effects upon intensity discrimination," Hear. Res. **55**: 223-230.

A. CD-ROM with data and software

The CD-ROM appended to this thesis contains the following files and folders:

- a pdf copy of this thesis: **Ana Alves Pinto thesis.pdf**
- a summary of this thesis, including conclusions, in spanish: **resumen en español.pdf**
- a group of folders, one for each study, containing the data and the software resulting from this work. Please read the file **README.doc**, in the CD-ROM, for more details.
- a pdf copy of the papers resulting from this thesis already published.

B. Derivation of Equation 2.1

This appendix describes the derivation of Eq. 2.1, used to equalize the overall level of the flat-spectrum noise with that of the notch noise, in the psychophysical experiment of noise discrimination. The presence of the notch in the target stimulus introduces a level difference between the two stimuli that is compensated by attenuating the flat-spectrum noise by the amount L_{att}.

The overall level of a broadband noise with bandwidth BW and spectrum level SL is given by (Hartmann, 1998):

$$L\,(dB\,SPL) = SL\,(dB\,SPL) + 10 \cdot log_{10}\,(BW) \tag{6.1}$$

Let $L_R$and $L_{R'}$be the overall level of the reference flat-spectrum noise before and after being attenuated by the amount L_{att} respectivelly. Then:

$$L_{att} = L_R - L_{R'} \tag{6.2}$$

$L_R$can be expressed as a function of the intensity of the reference stimulus, I_R, as follows:

$$L_R\,(dB\,SPL) = 10 \cdot log_{10}\left(\frac{I_R}{I_0}\right) \tag{6.3}$$

In the same way, $L_{R'}$, can be expressed as a function of the intensity of the reference stimulus attenuated $I_{R'}$ as follows:

$$L_{R'}\,(dB\,SPL) = 10 \cdot log_{10}\left(\frac{I_{R'}}{I_0}\right) \tag{6.4}$$

Combining the last three equations, L_{att}can be expressed as a function of $I_R$and $I_{R'}$as follows:

$$L_{att}\,(dB) = 10 \cdot log_{10}\left(\frac{I_{R'}}{I_R}\right) \tag{6.5}$$

From the latter formula, $I_{R'}$ can be expressed as a function of I_R and L_{att} as:

$$I_{R'} = I_R \cdot 10^{\frac{L_{att}}{10}} \tag{6.6}$$

The difference in intensisty ΔI between the reference flat-spectrum noise before (I_R) and after ($I_{R'}$) being attenuated can be expressed, then, as:

$$\Delta I = I_{R'} - I_R = I_R \cdot \left(10^{\frac{L_{att}}{10}} - 1\right) \tag{6.7}$$

This diference ΔI equals the difference in intensity between the target notch noise (I_T) and the reference flat-spectrum noise (I_R), that is introduced by the presence of the notch in the target stimulus. Therefore, it must equal the difference in intensity between the noise N2 in the target (I_{R2T}) and in the reference (I_{R2R}) noises (see section 2.2.2 for details on the way notch and flat-spectrum noises are generated):

$$\Delta I = I_T - I_R = I_{R2T} - I_{R2R} = \Delta I_{R2} \tag{6.8}$$

The same procedure used to derive expression B.7 can be used to express ΔI_{R2} as follows:

$$\Delta I_{R2} = I_{R2R} \cdot \left[10^{\frac{\Delta L_{R2}}{10}} - 1\right] \tag{6.9}$$

where ΔL_{R2} is the difference in overall noise level of noise N2 between the target (L_{R2T}) and the reference (L_{R2R}) noises:

$$\Delta L_{R2} = L_{R2T} - L_{R2R} \tag{6.10}$$

ΔL_{R2} actually is the depth of the notch, in dB. This result follows immediately from Eq. B.1. The overall level of N2 in the reference noise and target noises is given respectivelly by:

$$L_{R2R} = SL_{R2R} + 10 \cdot log_{10}\left(BW_{R2}\right) \tag{6.11}$$

and

$$L_{R2T} = SL_{R2T} + 10 \cdot log_{10}\left(BW_{R2}\right) \tag{6.12}$$

According to Eq. B.10:

$$\Delta L_{R2} = SL_{R2T} - SL_{R2R} \tag{6.13}$$

which is equal to notch depth (in dB).

The amount of attenuation that must be applied to the reference flat-spectrum noise, L_{att}, as a function of the notch depth, ΔL, which is a known parameter of the experiment (it is actually the dependent variable), can be derived from equations B.7, B.8 and B.9:

$$L_{att} = 10 \cdot \left[\frac{I_{R2R}}{I_R} \cdot \left(10^{\frac{\Delta L}{10}} - 1 \right) + 1 \right] \tag{6.14}$$

Therefore, the gain that must be applied to the reference flat-spectrum noise to make it having the same overall level of the target notch noise is given by:

$$G_{att} = 10^{\frac{L_{att}}{20}} = \left[\frac{I_{R2R}}{I_R} \cdot \left(10^{\frac{\Delta L}{10}} - 1 \right) + 1 \right]^{0.5} \tag{6.15}$$

Finally, the relation between the intensities $I_{R2}$and I can be expressed as a function of the bandwidths of the noise N2, BW_{N2}, and of the reference stimulus, BW_R, that are also known:

$$\frac{I_{R2R}}{I_R} = \frac{BW_{R2}}{BW_R} = \frac{BW_{R2}}{BW_{R1} + BW_{R2} + BW_{R3}} \tag{6.16}$$

C. Publications and conference communications resulting from this thesis

Publications

- Alves-Pinto, A. and Lopez-Poveda, E. A. (**submitted**). "Psychophysical assessment of the internal representation of high-frequency spectral notches," Journal of the Acoustical Society of America.

- Lopez-Poveda, E. A., Alves-Pinto, A., Palmer, A. R. and Eustaquio-Martín, A. (**accepted for publication**). "Rate *versus* time representation of high-frequency spectral notches in the peripheral auditory system: A computational modelling study," Neurocomputing.

- Lopez-Poveda, E. A., Alves-Pinto, A. and Palmer, A. R. (**in press**). "Psychophysical and physiological assessment of the representation of high-frequency spectral notches in the auditory nerve," in *Hearing - from basic reseach to applications*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M.Mauermann, S. Uppenkamp and J. Verhey (Springer Verlag, Heidelberg), scheduled for publication in 2007.

- Alves-Pinto, A. and Lopez-Poveda, E. A. (**2005**). "Detection of high-frequency spectral notches as a function of level," Journal of the Acoustical Society of America **118**, 2458-2469.

- Alves-Pinto, A., Lopez-Poveda, E. A. and Palmer, A. R. (**2005**). "Auditory-nerve encoding of high-frequency spectral information," Lecture Notes in Computer Science **3561**, 223-232.

Published conference abstracts

- Alves-Pinto, A., Lopez-Poveda, E. A. and Palmer, A. R. (**2006**). "Psychophysical and physiological assessment of high-frequency spectral notches in the auditory nerve," 151$^{\text{st}}$ Meeting of the Acoustical Society of America, Journal of the Acoustical Society of America **119**, 3235.

- Alves-Pinto, A., Lopez-Poveda, E. A. and Palmer, A. R. (**2005**). "Auditory-nerve response to broadband noise with high-frequency spectral notches," 149$^{\text{th}}$ Meeting of the Acoustical Society of America, J. Acoust. Soc. Am. **117**, 2564.

- Alves-Pinto, A. and Lopez-Poveda, E. A. (**2004**). "Detection of high frequency spectral notches as a function of level," 4$^{\text{th}}$ Forum of European Neurosciences.

- Alves-Pinto, A. and Lopez-Poveda, E. A. (**2004**). "Detection of high frequency spectral notches as a function of level," Midwinter Meeting of the Association for Research in Otolaringology.

Conference communications

- Alves-Pinto, A., Lopez-Poveda, E. A. (**2007**). "The internal representation of high-frequency spectral notches," poster, 153$^{\text{rd}}$ Meeting of the Acoustical Society of America, Salt Lake City, USA.

- Alves-Pinto, A., Lopez-Poveda, E. A. and Palmer, A. R. (**2006**). "Psychophysical and physiological assessment of high-frequency spectral notches in the auditory nerve," poster, 151$^{\text{st}}$ Meeting of the Acoustical Society of America, Providence, Rhode Island, USA.

- Alves-Pinto, A., Lopez-Poveda, E. A. and Palmer, A. R. (**2005**). "Auditory-nerve response to broadband noise with high-frequency spectral notches," poster, 149$^{\text{th}}$ Meeting of the Acoustical Society of America, Vancouver, Canada.

- Alves-Pinto, A. and Lopez-Poveda, E. A. (**2004**). "Codificación de la información espectral auditiva de alta frecuencia," poster, III Congreso de Neurociencias de Castilla y León, Valladolid, Spain.

- Alves-Pinto, A. and Lopez-Poveda, E. A. (**2004**). "Detection of high frequency spectral notches as a function of level," poster, 4^{th} Forum of European Neurosciences, Lisbon, Portugal.

- Alves-Pinto, A. and Lopez-Poveda, E. A. (**2004**). "Detection of high frequency spectral notches as a function of level," poster, Midwinter Meeting of the Association for Research in Otolaringology, Daytona Beach, Florida, USA.