Selectivity and Adaptation in the Human Auditory System

Andrew J.R. Simpson

Submitted in partial fulfilment of the requirements of

the Degree of Doctor of Philosophy

Declaration

I, Andrew J.R. Simpson, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature: Andrew J.R. Simpson
Date: 31/03/2014

Some of the work presented in this thesis was previously presented in the following peer-reviewed publications:

Chapter 2: Simpson AJR, Reiss JD (2013) The Dynamic Range Paradox: A Central Auditory Model of Intensity Change Detection, *PLoS ONE* 8(2): e57497

Chapter 3: Simpson AJR, Reiss JD, McAlpine D (2013) Human Modulation Filter Tuning is Carrier-Frequency Dependent, *PLoS ONE* 8(8): e73590

Chapter 4: Simpson AJR, Harper NS, Reiss JD, McAlpine D (2014) Selective Adaptation to "Oddball" Sounds by the Human Auditory System, *J Neurosci* 34:1963-1969

Appendix: Simpson AJR, Terrell MT, Reiss JD (2013) A Practical Step-by-Step Guide to the Time-Varying Loudness Model of Moore, Glasberg and Baer (1997; 2002), *in Proc. 134th AES Conv., Rome*.

I would like to thank my supervisor Joshua Reiss and collaborators David McAlpine, Nicol Harper and Michael Terrell. I would also like to thank the respective anonymous reviewers of the four papers described in this thesis. I was financially supported by an EPSRC studentship.

Abstract

Two fundamental principles dominate the signal processing of the auditory system: *selectivity* and *adaptation*. The response of the auditory system is selective for various acoustic features and the representation of these acoustic features adapts over time. This thesis is concerned with the characterisation of selectivity and adaptation in the human auditory system. Initially, selectivity for modulation rate and adaptation to intensity are characterised in a central auditory model. Next, selectivity for temporal modulation rate and *selective adaptation* to both intensity and temporal modulation rate are characterised in psychophysical data.

Table of Contents

…Table of Contents

Figures and Tables

Chapter 1: **General Introduction**

Recent developments in auditory neuroscience have challenged the idea that the auditory brain is a static processor of sound and have shifted the spotlight away from the ear to the brain. In particular, two signal processing strategies have captured the imagination of auditory neuroscientists: *selectivity* and *adaptation*. Selectivity may be defined as enhanced neural response to a given acoustic feature (e.g., frequency). Adaptation may be defined as a change in neural representation for a given acoustic feature that occurs over time. This chapter gives an introduction to the literature on auditory selectivity and adaptation and relates the two through an overview of the role they play in the general signal processing of the auditory system. In this context, we develop the motivation and rationale for the work presented in this thesis and we outline the main research questions and objectives.

1.1. Background

1.1.1. Feature based representation

In the periphery, sound pressure variations at the ear drum are mechanically transmitted as vibrations through the middle ear to the cochlea (Pickles, 2008). However, between the cochlea and the central nervous system the representation of sound is transmitted by auditory neurons in the form of electrical discharges known as spikes (Dayan and Abbot, 2001). As the neural representation of sound ascends the auditory pathway it is first decomposed by frequency in the cochlea and then further decomposed by periodicity between the midbrain and cortex. This decomposition yields a feature-based representation and represents a systematic transformation of the various acoustic features of the sound into a topographic neural map, where the location of a given neuron encodes the feature(s) for which the cell selects. This is known as a 'place code'.

1.1.2. Selectivity

Selectivity of auditory neurons for sound frequency is instigated in the cochlea (Pickles, 2008). The basilar membrane is lined with inner hair cells which shear in response to local resonance on the basilar membrane and so act as place-selective transducers. The inner hair cells are innervated by afferent (ascending) auditory nerve fibers, whose neurons fire in proportion to the degree of shearing. Mass-stiffness variation along the length of the basilar membrane cause it to act as an array of resonant filters which decompose the frequency components of the input signal into a tonotopic (arranged in order of frequency) place-code that is maintained by the systematic arrangement of afferent auditory nerve fibers. This tonotopic representation is retained throughout the ascending auditory pathway until at least the primary auditory cortex (Humphries *et al.*, 2010).

Selectivity for modulation rate (i.e., periodicity) emerges at the level of midbrain (Joris *et al.*, 2004; Baumann *et al.*, 2011) and is further refined in auditory cortex (Sadagopan and Wang, 2008; Barbour,

2011; Pasley *et al.*, 2012; Ding and Simon, 2012; 2013; Xiang *et al.*, 2013; Garcia-Lazaro *et al.*, 2011; Wang *et al.*, 2012; Lakatos *et al.*, 2013). Selectivity for modulation rate is also systematically arranged in the form of a periodotopic (arranged in order of period) map in midbrain (Baumann *et al.*, 2011) and cortex (Barton *et al.*, 2012) and there is evidence that the tonotopic and periodotopic dimensions are orthogonal (Baumann *et al.*, 2011; Barton *et al.*, 2012). This central selectivity for modulation rate has been suggested to play a key role in speech perception (Drullmann *et al.*, 1994; Shannon *et al.*, 1995; Ding and Simon, 2013; Zion Golumbic *et al.*, 2013; Lakatos *et al.*, 2013) and music perception (Zarate and Zatorre, 2012).

1.1.3. Adaptation

The coding of auditory neurons is not static but evolves over time to reflect the recent history of neuronal activity. Adaptation by auditory neurons to sound statistics has been reported in several neurophysiological studies involving small mammals (Dean *et al.*, 2005, 2008; Watkins and Barbour, 2008; Wen *et al.*, 2009; Rabinowitz *et al.*, 2011; Sadagopan and Wang, 2008; Barbour, 2011; Jaramillo and Zador, 2011; Walker and King, 2011; Ulanovsky *et al.*, 2003, 2004; Nelken, 2004; Perez-Gonzalez *et al.*, 2005; Malmierca *et al.*, 2009; Yaron *et al.*, 2012). Adaptation is typically characterised as changes in the spiking rate-level function (e.g., Dean *et al.*, 2005; Rabinowitz *et al.*, 2011) and has been argued to enhance coding accuracy (Dean *et al.*, 2008). Furthermore, auditory neurons have been shown to adapt over various timescales, from milliseconds to minutes (Dean *et al.*, 2005; 2008; Ulanovsky *et al.*, 2004; Yaron *et al.*, 2012; Jaramillo and Zador, 2011), suggesting adaptation to both long- and short-term sound statistics.

This *statistical selectivity* is further refined by tuning for the timescale over which the statistics are computed (Dean *et al.*, 2008; Ulanovsky *et al.*, 2004; Yaron *et al.*, 2012; Jaramillo and Zador, 2011); some neurons are tuned to adapt to short term statistics and others to long term statistics. Furthermore, sounds occurring in the natural world are known to exhibit low-order statistical regularities (Voss and Clarke,

1975; 1978) and auditory selectivity for 'natural' acoustic statistics has been demonstrated (Garcia-Lazaro *et al.*, 2006, 2011; Lesica and Grothe, 2008). Therefore, statistical selectivity might play a key role in how the brain represents sound in the natural world.

1.2. Motivation and rationale

The main motivation for studying selectivity and adaptation in human auditory perception is to generalise the above findings and principles from in-vivo electrophysiology in small mammals. Equivalent adaptation has not yet been demonstrated to exist in human auditory perception, nor has it been shown to confer any enhancement of perception. The application of psychophysics to this problem has two specific advantages. The first advantage is that it is a non-invasive method, and hence is convenient for use on human subjects. The second advantage is that, arguably, human auditory perception remains more sensitive than the currently available neuroimaging methods. Hence, psychophysics provides a nuanced window into the human auditory system that cannot be attained in any other way.

While the feature-based representation has obvious advantages for signal processing, perception is typically more object oriented. For example, speech or music signals contain multiple components but are typically perceived as a whole (Bregman, 1990). This is useful in communication because the perceptual object is used to attribute sound to its likely source. It would appear that a primary function of feature decomposition in the auditory system is to provide the basis for arbitrary recombination into object-based representations. Object-based representations emerge in auditory cortex (Mesgarani and Chang, 2012; Pasley *et al.*, 2012; Ding and Simon, 2012, 2013; Shamma *et al.*, 2011; Teki *et al.*, 2013), where sound features sharing a common temporal envelope are fused (Shamma *et al.*, 2011; Teki *et al.*, 2013; see Bregman, 1990). These auditory objects are then subject to top-down influences such as voluntary attention (Mesgarani and Chang, 2012; Pasley *et al.*, 2012; Ding and Simon, 2012, 2013). Therefore,

understanding the nature of cortical and subcortical feature-based representation is critical to understanding how auditory objects are ultimately maintained.

1.3. Thesis overview

This thesis is structured as self-contained chapters, including their local motivations and contexts. In the next chapter, a model is presented which provides evidence of adaptation and selectivity in human auditory perception. At this stage, we remain agnostic as to whether adaptation actually provides any enhancement of perception. In chapter 3, data is presented which characterises the human auditory brain as selective for modulations which are similar to those of speech. This chapter sets the stage for the fourth chapter, in which this selectivity is important. In chapter 4, the findings of the two previous chapters are generalised and combined to provide an argument that human auditory perception is enhanced by adaptation. Data is presented which demonstrates an interaction between selectivity and adaptation, suggestive of a sophisticated and general processing strategy for enhanced representation of novel and unusual sound events.

1.4. Aims and contributions

The main aim of this thesis is to advance the state of knowledge of selectivity and adaptation in the human auditory system. In particular, this thesis is focused on providing evidence and characterisation of selectivity and adaptation in auditory perception. This thesis contributes new perceptual data on auditory selectivity for modulation rate (Chapters 2, 3, 4), new perceptual data on adaptation (Chapter 4), new psychophysical methods (Chapters 3 and 4) and a new computational model (Chapter 2) of central auditory processing of intensity.

Chapter 2: A Central Auditory Model

In this chapter we use empirical loudness modelling to explore a perceptual sub-category of ***the dynamic range problem*** of auditory neuroscience. Humans are able to reliably report perceived intensity (loudness), and discriminate fine intensity differences, over a very large dynamic range. It is usually assumed that loudness and intensity change detection operate upon the same neural signal, and that intensity change detection may be predicted from loudness data and vice versa. However, while loudness grows as intensity is increased, improvement in intensity discrimination performance does not follow the same trend, and thus dynamic range estimations of the underlying neural signal from loudness data contradict estimations based on intensity just-noticeable difference (JND) data. In order to account for this apparent paradox we draw on recent advances in auditory neuroscience. We test the hypothesis that a central model, featuring central adaptation to the mean loudness level and operating on the detection of maximum central-loudness rate of change, can account for the paradoxical data. We use numerical optimization to find adaptation parameters that fit data for continuous-pedestal intensity change detection over a wide dynamic range.

2.1. Loudness and the Intensity Just-Noticeable Difference

Human hearing is known to function over an extremely wide dynamic range. In contrast, at a neural level the auditory system is known to have a very limited dynamic range. In auditory neuroscience, this is known as *the dynamic range problem* (e.g., see Dean *et al.*, 2005). In this chapter we address a somewhat paradoxical sub-category of the dynamic range problem which has arisen in psychoacoustics.

Loudness (L) is the perceived intensity (I) of a sound and the just-noticeable change in intensity is called the intensity just-noticeable difference (JND). Both loudness and intensity change detection are typically assumed to operate upon the same neural signal, generated in the cochlea and transmitted on the auditory nerve. This assumption gives rise to the intuitive anticipation of a relationship between loudness and the intensity JND, such that one may be predicted from the other and vice versa. However, previous researchers (Hellman and Hellman, 1990; 2001; Allen and Neely, 1997) were not able to provide a unified model due to the apparently paradoxical observation that loudness growth, beyond a certain level, is not reflected in improvement in intensity discrimination performance (Allen and Neely, 1997; Miler, 1947). From a neural coding point of view, the problem can be stated as follows; Spike rate is known to be intensity dependent, and loudness is assumed to scale with spike rate, and since information scales with spike rate (*Fisher information scales with spike rate under reasonable assumptions*, Dayan and Abbot, 2001), then why do more spikes not provide a better encoding of intensity change?

The work of Hellman and Hellman (1990, 2001) and Allen and Neely (1997) resulted in the theoretical construct of the loudness JND, which represents the just-noticeable change in loudness that corresponds to the intensity JND, and the assumption that a reciprocal relationship between loudness and loudness change detection should exist. Focusing on the intensity *discrimination* paradigm, Hellman and Hellman (1990) predicted loudness functions for pure tones from intensity JND data, following the suggestion of McGill and Goldberg (1968a, 1968b) that the loudness JND is the square root of loudness

($\Delta L_{jnd} = L^{0.5}$). Allen and Neely (1997) tested this for tones and noise using equivalent loudness and intensity JND (ΔI_{jnd}) data as follows:

$$\Delta L_{jnd} = L(I + \Delta I_{jnd}) - L(I) \tag{2.1}$$

Using the loudness function of Fletcher and Munson (1933) and the equivalent intensity discrimination data of Riesz (1928), Allen and Neely showed (via Eq. 2.1) that the square root exponent of Hellman and Hellman (1990) required modification above 20 dB sensation level (SL) and introduced a 'saturation of internal noise' to account for the modification. This showed that loudness and loudness change detection may not be modelled reciprocally and thus, their paradox was defined.

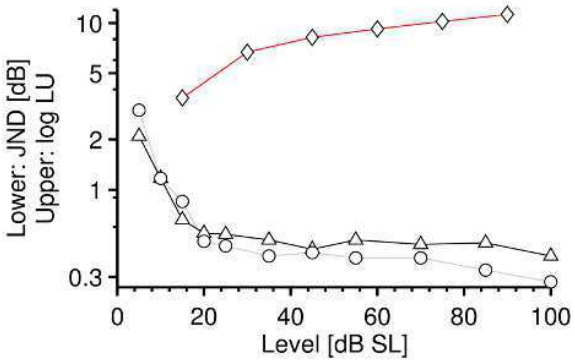

Figure 2.1. Loudness versus intensity JND. Miller's averaged data for loudness (diamonds) and the intensity JND (circles/triangles) for broadband noise for two individual listeners, as a function of sound level (SL). Loudness data (diamonds), presented in log loudness units (LU), are taken from Neely and Allen (1998) who converted them from loudness level data of Miller using the loudness function of Fletcher and Munson. Above about 20 dB SL, the JND is approximately constant (i.e., Weber's Law) but loudness increases.

To illustrate the paradox, Fig. 2.1 shows a comparison of Miller's (1947) wide-band noise data for the intensity JND and for loudness levels as a function of intensity. Miller's (1947) loudness level data are converted into loudness units (LU), taken from Neely and Allen (1998) according to the loudness function of Fletcher and Munson (1933), and plotted in *log*(LU) for comparison to the intensity JND. At medium levels and above, loudness rises while the intensity JND remains almost constant.

Recent auditory neuroscience literature appears to provide a promising solution; Dean *et al.* (2005, 2008), Wen *et al.* (2009) and Rabinowitz *et al.* (2011) have addressed the dynamic range problem in terms of adaptive neural coding. It has been demonstrated (in animals) that central neural adaptation to mean sound level acts to improve coding of sound at the most likely (mean) sound level, mitigating neural dynamic range limitations. Dean *et al.* (2005) showed that input/output functions of neuronal populations in the inferior colliculus of the guinea pig are able to shift their operating points to suit the prevailing (most likely) stimulus sound pressure level. Dean *et al.* also showed that the result of such neural adaptation may be characterized as an imperfect dynamic range normalization of the neural signal. The general parameters that define the adaptation process are the time constant (how fast the adaptation occurs), threshold (central neural dynamic range) and amount (how much adaptation occurs).

In order to resolve the paradox, in this chapter we assume that central adaptation to mean sound level occurs in humans during psychoacoustic experiments (Pienkowski and Hagerman, 2009). We also assume that the small change that constitutes a typical intensity JND falls at the lower limit of the fixed central neural dynamic range, and that adaptation to high mean levels necessarily raises the lower limit accordingly. This adaptive raising of the lower limit effectively degrades intensity discrimination performance relative to the performance limitations imposed by the peripheral processor.

There are no physiological data available to characterize central adaptation in human listeners. Therefore, in a numerical optimization sense, the time constant, threshold and amount are effectively free parameters within an empirical model of central adaptation. The main objective of this chapter is to

establish, by a process of optimization, working central adaptation parameter values from the empirical data available in the psychoacoustic literature.

Although there are data available over a wide enough dynamic range to establish the free parameters of adaptation threshold and amount, the majority of psychoacoustic experiments on intensity discrimination do not control or report the mean sound level over the entire course of the experiment. Hence, there are no data available to establish the time constant.

To overcome this problem, we look to the continuous-pedestal (carrier) paradigm, where the reported pedestal level provides a good approximation to the long-term average level. Two such studies exist with data over a very wide dynamic range; one for tones (Viemeister and Bacon, 1988) and the other for noise (Miller, 1947). Both studies remain definitive, in terms of data and in terms of phenomena characterized by the data, and are ideal for our optimization problem.

The theoretical foundation for our modelling is the excitation pattern model (Florentine and Buus, 1981). The excitation pattern model is an empirical model of the cochlea and auditory nerve representation of a sound – hence we may classify it as a *peripheral model*. The output of this model may then be integrated in order to calculate loudness (Moore *et al.*, 1997). This is known as the integrated auditory nerve formulation of loudness (Fletcher and Munson, 1933; Allen and Neely, 1997).

The excitation pattern loudness model (Moore *et al.*, 1997; see Appendix) incorporates functionality, based on peripheral auditory physiology, which approximates the major phenomena of psychoacoustic theory (i.e., cochlear compression, spread of excitation, the auditory filter, etc). A full account of this model is given in the Appendix. The parameters of the model are set to fit a broad range of empirical data. We take this model as input to our central model, much as the auditory nerve is peripheral to the (central) auditory cortex. We extend the peripheral excitation pattern model to include a central adaptive representation which we call a *central excitation pattern model*. This approach is similar to that of Parra and Pearlmutter (2007), who proposed a central adaptive model of tinnitus and the 'Zwicker tone'.

Since the excitation pattern model of loudness is well established, we optimize the central adaptation parameters of our central excitation model to relate the fixed parameters of the loudness model to intensity change detection. In keeping with the paradoxical data, we make the implicit assumption that loudness, and loudness change, are coded independently at a central neural level, based on common input from the auditory nerve.

In the first stage of this chapter we briefly review the related literature and describe an analysis of the empirical data based on simulation of the experiments that produced the data. This analysis is used to assess the scope of the problem. Next we propose a central excitation pattern model with a maximum rate-of-change detector. The free parameters of the model are optimized to fit the tone and noise intensity JND data over a wide dynamic range. The resulting optimized model is shown to perform well at predicting independent pseudo-continuous intensity JND data from the literature. We report an experiment based on the detection of linearly ramped up-down increments in pseudo-continuous noise pedestals. This experiment shows that slowly-ramped increments are hard to detect and validates our use of a rate-of-change model. In this chapter we provide empirical evidence to support an argument that loudness reflects peripheral coding, and the intensity JND reflects central coding.

2.2. Modelling Background and Methods

We base our analysis, and subsequent modelling, on the time-varying excitation pattern loudness model of Moore *et al.* (1997; Glasberg and Moore, 2002) – which we term *peripheral*. The model has been adequately described by the authors and we do not repeat the description here except to summarize the temporal integration of the model. Glasberg and Moore's time-varying loudness model produces a time-varying excitation pattern which is integrated over short time intervals to produce 'instantaneous loudness'. Two successive exponential temporal windows are then used to estimate short-term loudness (STL) with respect to instantaneous loudness, and long-term loudness (LTL) with respect to STL. STL is

used to account for loudness of brief duration sounds of fixed intensity, and LTL is used to account for overall loudness impression of continuous amplitude modulated sounds.

Each temporal window is defined by a pair of exponential functions and time constants for 'attack' and 'release' respectively. The STL integration times are not symmetrical, the attack time is 25 ms and the release time is 50 ms, in order to account for greater forward masking than backward masking. The attack and release times for LTL are similarly asymmetrical. The attack time is 100 ms and the release time is 2 s, allowing for the persistence of loudness impression after the stimuli has ceased.

Because the present chapter is concerned with amplitude modulation for continuous pedestals, we apply the loudness model using the LTL integration window. While the LTL attack time was deliberately set (see Glasberg and Moore, 2002) to fit data for loudness of amplitude modulated sounds, the 2 s LTL release time is merely intended to produce a lasting impression of loudness after the stimulus has ceased. Since this release time is not justified in terms of any specific asymmetry in the temporal integration of loudness, in our modelling the LTL release time was set to 100 ms (the same as the attack time), which produced a symmetrical temporal window for LTL with respect to STL. The combination of the two temporal windows remains asymmetrical due to the asymmetry in the short-term temporal window.

2.2.1. Magnitude or Envelope?

When the intensity of a signal changes over some time frame, the temporal shape (or profile) of the intensity function is known as the envelope. An important question is whether it is the size or envelope of the intensity increment that determines the detection threshold. Hellman and Hellman (1990, 2001) and Allen and Neely (1997) have defined the loudness JND in terms of magnitude of loudness change caused by the intensity increment (Eq. 2.1). This means that for envelope ramps which are long (slow) compared to temporal integration of loudness the intensity JND is assumed to be constant.

A single study exists which does not support this assumption. Riesz's (1928) study of the intensity JND is rarely considered, by today's standards, to be strictly intensity discrimination. However, this study

was the first to introduce evidence to suggest a rate-of-change detector process. It involved the detection of amplitude (or envelope) modulation produced when two sine waves, closely spaced in frequency, are

summed to produce a modulating envelope and is known as the method of beats. Riesz used continuous 1 kHz signals to test the amplitude modulation (beat) detection thresholds, as a function of beat rate, and found the smallest thresholds at a rate of around 3 to 4 Hz. He also found that at lower and higher rates of modulation, the threshold of detection increased almost symmetrically (on a logarithmic scale) about the 3 to 4 Hz point. This result is not predicted by Eq. 2.1. In section 2.4 we describe an experiment designed to confirm the generality of Riesz's results as a function of beat rate.

Eq. 2.1 provides a loudness domain subtraction between loudness values at two intensity levels, which relates the difference in intensity to the difference in loudness that is just noticeable by *discrimination*. However, for the rate-of-change detector necessary to explain the data of Riesz (1928), this equation must be transformed into the time domain (Wojczak and Viemeister, 1999). This transformation between the JND domains, for change over a given time frame (Δt), relates change in intensity $\Delta I/\Delta t$ to change in loudness $\Delta L/\Delta t$. Eq. 2.1 becomes:

$$\left(\frac{\Delta L}{\Delta t}\right)_{jnd} = L\left(I + \left(\frac{\Delta I}{\Delta t}\right)_{jnd}\right) - L(I) \qquad (2.2)$$

2.2.2. Choice of Continuous Data

Candidate continuous-pedestal data for increment detection in noise (Miller, 1947) and in pure tones (Viemeister and Bacon, 1988) were selected because of the large dynamic range covered in both studies (>90 dB), and because both studies remain definitive. In Miller's (1947) experiment, the increment envelope for the noise signals was instantaneous (square) and duration was 1.5 seconds. For the

experiment of Viemeister and Bacon (1988), tones contained 10 ms cosine-ramped increments of 200 ms duration. A full description of the stimuli of the respective studies is given in section 2.5.3.

Weber's Law states that the ratio of the intensity JND to intensity should be constant (Weber, 1846). Miller's data showed that this was approximately true for noise signals. However, Weber's Law does not generally hold for pure tones, as is demonstrated by the data of Viemeister and Bacon. The appearance of an 'almost' constant ratio for pure tones has been termed the 'near-miss' to Weber's Law (McGill and Goldberg, 1968a, 1968b). Therefore, the two studies chosen provide a contrast, both in terms of stimuli properties (tones/noise, envelope shape, increment duration) and in terms of qualitative characterization of the data (Weber's Law/'near-miss'). This provides a compelling challenge to the intended unified model.

2.2.3. Transformation of Continuous Data

Here we investigate the question of whether temporal integration of the loudness model is able to unify the two paradigms sufficiently such that we can proceed to optimization of the central adaptation stage. Using the loudness model of Glasberg and Moore (2002), we transform I into L, ΔI_{jnd} into ΔL_{jnd}, and finally $(\Delta I/\Delta t)_{jnd}$ into $(\Delta L/\Delta t)_{jnd}$ for the simulated pedestals-with-increments of Miller and of Viemeister and Bacon. This analysis tells us how much need there is for central adaptation and the range in which it is necessary.

Fig. 2.2(a) shows the re-plotted intensity JND data for Miller and Viemeister and Bacon, illustrating the disparity in function shape that must be overcome within our model. Fig. 2.2(b) shows the loudness functions of intensity for the pedestals of the respective studies, as estimated using the loudness model. In Fig. 2.2(b), for comparison with the loudness model results, we also show the loudness level data of Miller (1947), as converted by Neely and Allen (1998) using the loudness function of Fletcher and Munson (1933) [$I = SL$ + 10 dB (Miller, 1947); 1 *sone* = 975 *LU*]. The shape of the loudness function

estimated by the loudness model is in good agreement with the loudness level data of Miller, but the loudness model predicts lower absolute thresholds than the data of Miller suggests (see section 2.5.3).

Fig. 2.2(c) shows the respective estimated transformed data for $\Delta L_{jnd}(L)$, using Eq. 2.1. Fig. 2.2(d) shows $(\Delta L/\Delta t)_{jnd}(L)$, estimated using Eq. 2.2 for Δt = 1 ms. In Fig. 2.2(d) the two functions are much closer than the two functions of Fig. 2.2(c). This shows that, within the loudness model, the temporal parameters of the stimuli (envelope and duration) allow us to better unify the ΔI_{jnd} data between the tone and noise studies in terms of $(\Delta L/\Delta t)_{jnd}(L)$. In other words, Eq. 2.1 does not take into account the envelopes of the stimuli but, using Eq. 2.2, the 10 ms cosine-ramped increments in tones (Viemeister and Bacon) and the instantaneous changes in noise (Miller) produce similar maximum loudness slopes for a given overall pedestal loudness.

In Fig. 2.2(c), we see a disagreement between the transformed data sets with regards to the smallest ΔL_{jnd} that is detectable, by a factor of around two. This disparity would make it difficult to model using a magnitude of change model. Moore *et al.* (1997) suggest an absolute threshold of 0.003 sones. Assuming that absolute threshold and masked threshold are equivalent, this is not compatible with the minimum loudness JND of approximately 0.01 sones shown in the function of Fig. 2.2(c). Therefore, it is clear that a magnitude-of-change model, with a threshold of 0.003 sones, would not explain the data.

After transforming the data further into $(\Delta L/\Delta t)_{jnd}$, in Fig. 2.2(d) we see that the smallest $(\Delta L/\Delta t)_{jnd}$ is much more in agreement between the two stimuli (~5×10^{-5} sones/ms). Thus, we confirm that our choice of decision variable $[(\Delta L/\Delta t)_{jnd}]$ is useful. Below about 0.25 sones, the slopes of these functions are relatively flat. Between 0.005-0.25 sones there is a slope of around 0.00005 sones/ms but between 0.05 and 2.5 sones there is a far greater slope. These two observations conform to the two necessary conditions of constructing a central, adaptive rate-of-change model; i) that the $(\Delta L/\Delta t)_{jnd}$ functions must be close together (equivalent) and ii) that both functions must be approximately constant in the range below an equivalent loudness threshold (i.e., the two functions represent the same central dynamic range). The point where the two functions take on a marked increase in slope (~0.25 sones) is the starting point in our search

for a common threshold parameter value. During the subsequent optimization, we take the value $5.5x10^{-5}$ sones/ms of $(\Delta L/\Delta t)_{jnd}$ as a constant for our modelling. This might be taken to represent internal noise level.

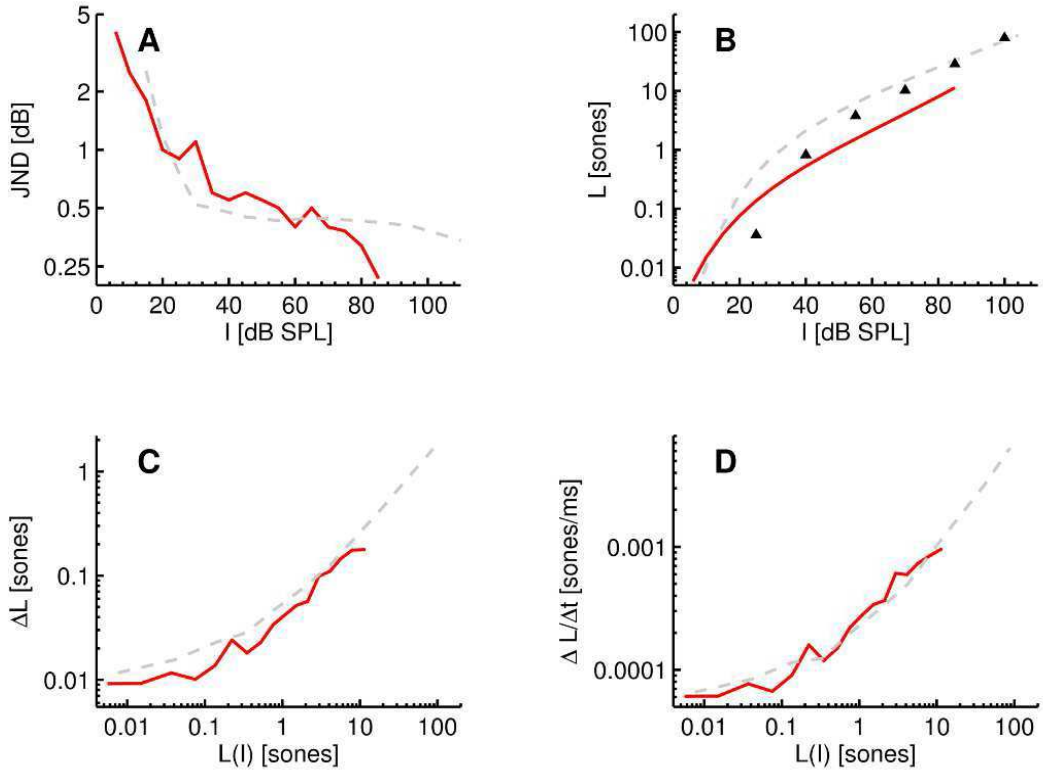

Figure 2.2. Transformation results for the noise data of Miller [dashed grey line] and the pure tone data of Viemeister and Bacon [solid red line]. **A** Average intensity JND data. **B** Estimated loudness functions [$L(I)$] for the stimuli (pedestals). Triangles represent Millers loudness data (I = SL + 10 dB), converted to sones (1 sone = 975 LU) from the calculated values of Neely and Allen. **C** Eq. 2.1: Estimated transformation of ΔI_{jnd} [pane **A**] to ΔL_{jnd}. **D** Eq. 2.2: Estimated transformation of ΔI_{jnd} [pane **A**] to $(\Delta L/\Delta t)_{jnd}$. The two magnitude-of-loudness-change functions in **C** are not consistent at low levels – there is an offset, but the rate-of-loudness-change functions in **D** are closer, indicating that the temporal parameters (duration, envelope) of the stimuli represented in **D** allow the stimuli to be unified. In **D**, below ~0.25 sones the functions are approximately zero slope [i.e., $(\Delta L/\Delta t)_{jnd}$ is constant].

2.3. Central Excitation Pattern Model

A general block diagram of the proposed central excitation pattern model and rate-of-change detector is given in Fig. 2.3. Glasberg and Moore (2002) provided a loudness model that operates on the temporal waveform of a given sound to produce a time-dependent loudness function. We extend this model to produce a time-dependent central loudness contrast function which can be used to predict those changes in the intensity of a sound that may be detectable. It should be noted that our definition of central loudness (change) is purely functional/notational, in order to maintain some consistency with the previous literature regarding the loudness JND.

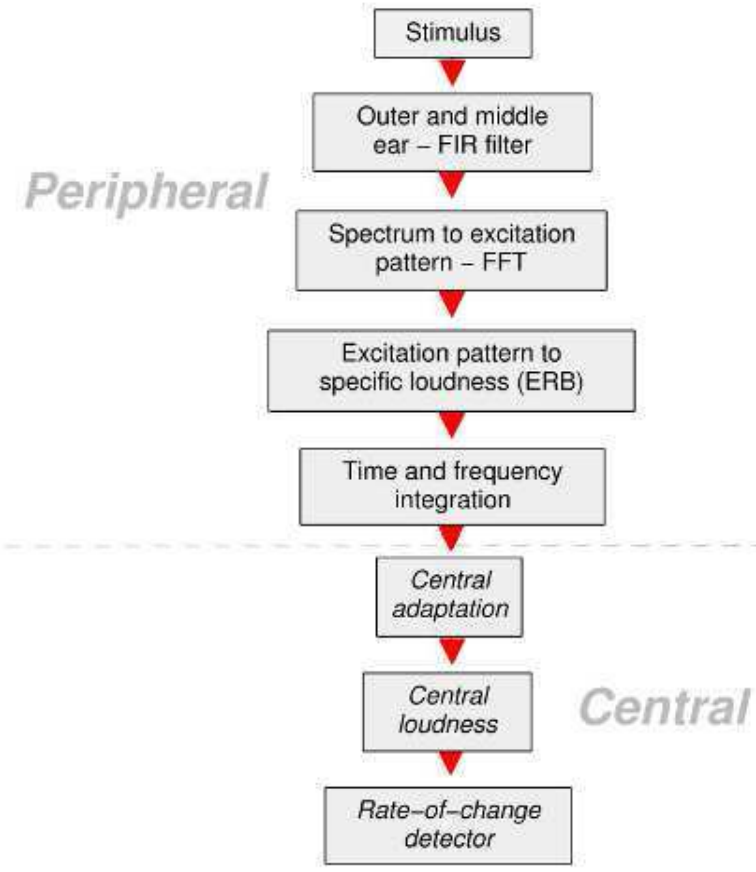

Figure 2.3. Block diagram of the central excitation pattern model and rate-of-change detector process. The area indicated as peripheral contains the loudness model of Glasberg and Moore (2002) and the area indicated as central contains the proposed additions of the present chapter.

2.3.1. Central Loudness Adaptation

Due to our confinement to the continuous pedestal paradigm, we are able to assume that mean level is approximately the same as the reported level of the pedestal. Therefore, only two free parameters are needed to define central adaptation (CA) in our model; threshold (T_{CA}) and normalization amount (α). The value of α determines central threshold shift that results from mean peripheral loudness exceeding the central adaptation threshold (i.e., exceeds the central dynamic range). Consistent with long-term central adaptation to the prevailing sound level (Dean *et al.*, 2005, 2008; Wen *et al.*, 2009; Rabinowitz *et al.*, 2011), central adaptation is implemented in the form of a partial normalization of any time-varying loudness function (L) which has a mean loudness ($\bar{L}$) above the central adaptation threshold, T_{CA}. Since we are concerned with continuous pedestals, mean loudness refers to a single value for tonal pedestals and an average over an arbitrarily long time frame for noise pedestals. The use of the mean loudness for adaptation threshold in continuous pedestals also provides for smoothing of instantaneous loudness changes in noise pedestals. The conditional normalization used to produce the central loudness function, L_{Cen}, is

$$L_{Cen} = \begin{cases} L & for \quad \bar{L} < T_{CA} \\ (1-\alpha)L + \alpha \frac{T_{CA}}{\bar{L}} L & for \quad \bar{L} > T_{CA} \end{cases} \tag{2.3}$$

2.3.2. Central Loudness Just-Noticeable Difference

Unlike tonal pedestals, noise pedestals include inherent loudness changes which must be taken into account (Dau *et al.*, 1997a, 1997b; Glasberg *et al.*, 2001). In our model we treat each noise signal as deterministic (and repeatable), or frozen (Buus, 1990; Agus *et al.*, 2010) and we base detection on the difference between the maximum value of ΔL_{Cen} for the pedestal and the maximum value of ΔL_{Cen} during an increment/decrement applied to that pedestal. Consistent with Eq. 2.2, the threshold constant is defined in sones per ms and the proposed threshold expression is

$$\left(\frac{\Delta L}{\Delta t}\right)_{jnd} = \max\left(\frac{|\Delta L_{Cen}|}{\Delta t}\right)_{inc} - \max\left(\frac{|\Delta L_{Cen}|}{\Delta t}\right)_{ped} \quad (2.4)$$

where the pedestal signal is denoted $(\Delta L_{Cen}/\Delta t)_{ped}$, and the pedestal-plus-change signal is denoted $(\Delta L_{Cen}/\Delta t)_{inc}$. Thus, given a fixed (constant) value for $(\Delta L/\Delta t)_{jnd}$, Eq. 2.4 may be solved by adjusting the increment size so as to affect $(\Delta L_{Cen}/\Delta t)_{inc}$.

Using a fixed value of $(\Delta L/\Delta t)_{jnd}$ extracted from Fig. 2.2d (5.5×10^{-5} sones/ms) a manual, iterative optimization process was conducted by using the central model to predict the value of ΔI_{jnd} for each data point of the two studies using given parameter values of threshold T_{CA} and α. Within each iteration the entire range of stimuli for both studies was simulated. For each simulation within a given iteration, Eq. 2.4 was evaluated numerically using the model to find ΔI_{jnd}. The predicted value of ΔI_{jnd} was compared to the respective data point and an error term calculated. For each iteration the average error term was calculated over the two datasets. This process was repeated, with adjustments made to the free parameters (T_{CA} and α) in order to minimize the error terms until both slopes of the respective minima for each free parameter were located – i.e., until the values of T_{CA} and α were optimal. The JND for the change in intensity (ΔI_{jnd}) is expressed as

$$JND = 10\log_{10}\left(1 + \frac{\Delta I_{jnd}}{I}\right) \quad (2.5)$$

2.4. Experiment 2.1: *Generalising Riesz's Beat Detection Paradigm*

The following experiment was designed to replicate the rate-of-change-detection paradigm of Riesz (1928), within the more controlled conditions of linearly ramped increments in noise pedestals, and to confirm the generality of his rate findings. In a two-interval, forced-choice procedure, listeners were asked to detect linear up-down ramps in wideband noise. The use of linear ramps in broadband noise removes possible confounds, relating to unwanted detection cues of the beat-detection paradigm employed by Riesz.

2.4.1. Experiment 2.1: Stimuli and task

All stimuli were generated digitally at 24 bit resolution. A pair of Beyerdynamic DT100 isolating headphones were used to present the stimulus to the subjects, which was played back directly from a computer at a sampling rate of 44,100 Hz. Presentation was diotic (same in both ears). The pedestal was a broadband (0-20 kHz) Gaussian noise, presented at an overall level of 33 dB SPL (rms). In the target interval, symmetrical, linearly-ramped envelopes with half-ramp durations of between 5 and 50,000 ms were added to the noise pedestals. Half-ramp durations of [5, 10, 100, 1000, 10000, 50000] ms were used. The increment consisted of a linear increment ramp immediately followed by a linear decrement ramp of equal duration. The increments were located in the temporal centre of the target pedestal. For half-ramp durations of 1 second or below, pedestals were of 4 seconds. For half-ramp durations of 10 seconds, the pedestal was of 24 seconds. For half-ramp durations of 50 seconds the pedestal was of 104 seconds. Both target and reference intervals were gated with 10 ms raised-cosine ramps. After hearing each pair of noise signals the listener was asked which contained the ramp.

2.4.2. Experiment 2.1: Procedure

An adaptive three-down one-up, two-interval forced-choice (2IFC) procedure was employed which estimates the 79.4% correct identification (Levitt, 1971). See Figure 2.4. Each trial consisted of two observation intervals, one of which was selected at random to contain the target increment. The inter-stimulus interval was 3 seconds. The level of the increment was defined as the maximum difference (in dB) between the pedestal and the target. The starting value was 20 dB. The initial step size was 5 dB for the first 4 reversals and was subsequently halved. A reversal was defined as an increase in increment size following a decrease, or vice-versa. Three consecutive correct identifications of a ramp resulted in a reduction in size of the increment and one incorrect answer resulted in an increase. After 12 reversals, threshold was taken as the arithmetic mean of the last 10 reversals.

After each trial, subjects were provided with correct/incorrect feedback on their responses. Trials were undertaken in blocks lasting no longer than 20 minutes. Due to the large number of relatively long duration trials necessary, blocks were often interrupted with a break period of 15 minutes, after which the block continued until either the next rest period or completion. For the longest half-ramp duration (50 s) such breaks were occasionally taken in the course of a single threshold determination. On two occasions, within a block, the break was extended overnight and the block was continued on the following day. Prior to the test, each subject was given a brief demonstration to familiarize themselves with the interface and procedure and was allowed a single practice run.

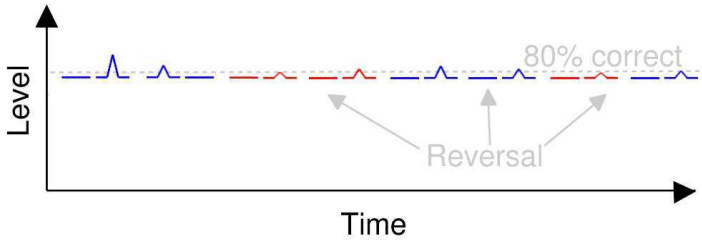

Figure 2.4. Illustration of the adaptive method. Pairs of noise signals are presented, one of which contains a ramped increment. The listening level and ramp duration are fixed throughout, whilst ramp size is adaptively changed until the procedure converges on ~80% correct performance. When the listener correctly identifies the location of the ramped intensity increment the size of the increment is reduced, otherwise the size is increased, depending on a rule (simplified here to a 1-up, 1-down rule). Correct responses (blue) result in decreased ramp size and incorrect responses (red) result in increased ramp size. The ~80% correct threshold level is estimated by averaging the ramp size measured at several points where the adaptive procedure changes direction ('reversals'). The step size of the ramp size change is reduced after a reversal and the procedure eventually converges on the ~80% correct point.

2.4.3. Experiment 2.1: Listeners

Ten unpaid volunteer subjects served as listeners in the experiments. Seven male subjects and three female subjects took part. The mean age of the subjects was 29 (min: 20, max: 36, standard deviation: 5.9). All reported normal hearing and some reported limited previous experience of participating in listening tests. All participants were naïve about the purpose of the test.

2.5. Results and Discussion

In this section we describe the results of the optimization process and of the proposed central excitation pattern model applied to a further set of pseudo-continuous intensity JND data from the literature (see section 2.5.3). For each separate simulation, within the optimization and within the simulation of the pseudo-continuous data, stimulus waveforms were produced to exactly replicate the documented conditions of the respective study. This explicitly included level and envelope.

For comparison, empirical data for intensity JND values are also presented in terms of intensity in the form of Eq. 2.5. Data are plotted on a logarithmic scale to allow easier determination of Weber's Law characteristics, whilst retaining the familiar numerical scale of classical literature for the intensity JND. Goodness-of-fit measures are given, for each dataset, in the form of two-tailed Pearson correlation coefficients (r, P) and root-mean-square error (e, dB). A description of the experimental conditions for each study is given in the 2.5.3.

2.5.1. Central Adaptation Parameters; Optimization Results

From the optimization, the following values were found: T_{CA} = 0.215 sones, and α = 0.95 (i.e., resulting in 95% normalization using Eq. 2.3). The T_{CA} value of 0.215 sones (approximately 25 dB SPL in the 1 kHz pure tone case) corresponds relatively well to the known dynamic range (approximately 35 dB) of primary auditory nerve fibers (Evans and Palmer, 1980; Sachs and Abbas, 1974). The 95% normalization of the central loudness function is approximately consistent with the known sub-optimal adaptation behaviour of auditory neurons (Dean *et al.*, 2005, 2008; Wen *et al.*, 2009; Rabinowitz *et al.*, 2011). In summary, the parameter values found appear reasonable.

Fig. 2.5(a) shows the resulting central loudness (red line) as a function of peripheral loudness (grey, dashed line), illustrating the result of the optimization and the effects of central adaptation. In order to show the effect of central adaptation on the estimated intensity JND functions, Figs. 2.5(b, c) show the rate-of-change predictions of the unaltered peripheral model (grey, dashed line) compared to the optimized central excitation pattern model (red line) for the data of Viemeister and Bacon (Fig. 2.5b) and Miller (Fig. 2.5c). The fit of the optimized central excitation pattern model to the data of Viemeister and Bacon is good (r=0.99, P=1.8x10^{-13}, e=0.04 dB), as is the fit to the data of Miller (r=0.94, P=1.4x10^{-5}, e=0.19 dB). The growth of loudness for both cases (tones/noise) gives a good prediction below central adaptation threshold. However, in both cases, the unaltered peripheral model results diverge strongly from

those of the optimized central model above approximately 0.2 sones and the peripheral model fails to hold to the data at higher levels. As can be expected from looking at Fig. 2.5(b/c), the value of T_{CA} is relatively tightly controlled since a larger value would increase the error for the data of Viemeister and Bacon (Fig. 2.5b) and a smaller value would increase the error for the data of Miller (Fig. 2.5c). The value of alpha is also relatively tightly constrained because smaller values would cause the functions to tend towards the under-estimation of the peripheral model output, and because larger values the model would tend towards Weber's Law for the tonal data.

This modelling result is interesting because the 'near-miss' is often attributed to a combination of cochlear compression and spread of excitation (Florentine and Buus, 1981; Viemiester, 1983), where high-pass noise or high-frequency tones are used to eliminate the near-miss, and hence it is anticipated that the spread of excitation featured in the excitation pattern model should lead to a near-miss. The modelling result for the unaltered peripheral model does not produce a compelling near-miss and so it appears that the addition of central adaptation is necessary to fit the data. To repeat the statement made by Allen and Neely (1997), this account of the near-miss seems different to the spread-of-excitation hypothesis. Furthermore, it should be noted that in this model, adaptation is equivalent to an instantaneous nonlinearity.

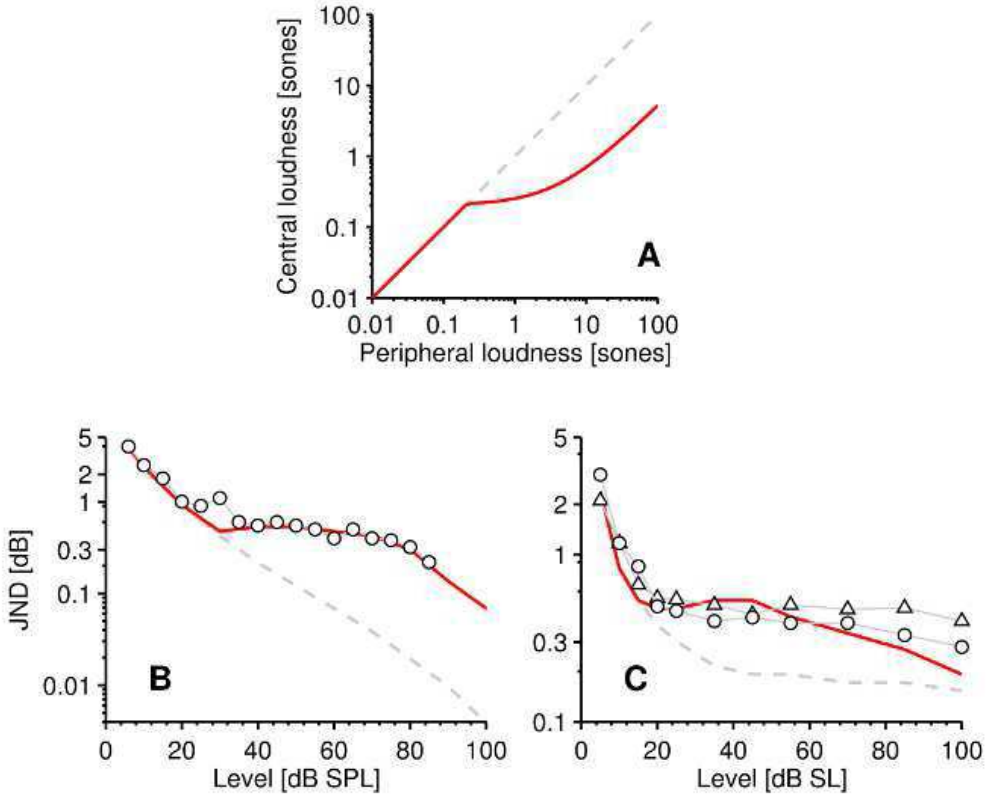

Figure 2.5. Optimization results; *peripheral versus central model*. A Central loudness (solid red line) for continuous pedestals, as a function of peripheral loudness (dashed grey line), illustrating the saturating effect of central adaptation (Eq. 2.3). **B**, **C** Comparison of estimated intensity JNDs from the peripheral and central excitation pattern rate models respectively. **B** circles: the averaged 1-kHz continuous pure tone increment-detection data of Viemeister and Bacon and **C** is the individual (circles and triangles) continuous-noise increment-detection data of Miller.

2.5.2. Results of Experiment 2.1

Fig. 2.6a shows the results of the ramped-noise experiment 2.1. Group mean thresholds for the 10 listeners are given, including error bars representing 95% confidence intervals. The trends shown in the data are significant (P=9.55x10^{-8}, *Friedman Rank Sum Test* – see Hollander and Wolfe, 1973). The results, plotted on a logarithmic (time) scale, show symmetry about the half-ramp of 100 ms 'best detection point' which appears equivalent to that shown around 3-4 Hz by Riesz (Fig. 2.6b). Furthermore, the results confirm Riesz's general finding that slow ramps are hard to detect. It should be noted that short-term memory

(Durlach and Braida, 1969) may play a role in the results at very long ramps (i.e., >4 seconds), in that the listener is forced to assess the intensity change within the short-term memory window.

2.5.3. Simulation of Pseudo-Continuous Experiments

A selection of contemporary intensity JND studies were chosen to test the generality of the model in conditions where the continuity constraint held only loosely but where other parameters important to temporal integration theory were varied. We call these studies pseudo-continuous because the pedestals used would be considered continuous if they were not gated on and off. We also include our ramped-noise experiment (see section 2.5.2). None of these studies varied (roved) the listening level within experimental runs, so the long-term average level should be reasonably close to the reported pedestal levels.

For direct comparison with the results of Viemeister and Bacon (1988), the model was used to obtain detection thresholds for increments of 200 ms in continuous 1 kHz tones over the intensity range from threshold to 85 dB SPL. The increments were gated with 10 ms raised-cosine ramps.

Miller (1947) measured increment detection thresholds for two subjects using continuous, wide-band noise signals. The noise signals were specified as having power spectrum of ±5 dB from 150 to 7,000 Hz and were incremented for 1.5 sec. duration at intervals of 4.5 sec. Since Miller did not specify the spectrum outside of this range, in our modelling a band pass filter was used to reduce the energy outside of this range by 12 dB per octave. We assume that the increment envelope is square (instantaneous). Best fit to the data was found where SL was converted to SPL to be consistent with the threshold predicted by the (peripheral) loudness model (SPL = SL + 4 dB).

For comparison to the results of Oxenham (1997), we used the model to obtain intensity JND thresholds as a function of increment and decrement duration at 55 dB SPL at durations between 4 and 200 ms. Thresholds were obtained both in quiet and in wide-band noise of 0 and 20 dB spectrum level.

Increments and decrements in 4 kHz pure tone pedestals of 500 ms were gated using raised-cosine ramps of 2 ms.

For comparison to the results of Plack *et al.* (2006), the model was used to obtain thresholds for detection of brief symmetrically-ramped increments in a 20 dB spectrum-level (i.e., dB per 1 Hz band) broadband (0 - 20 kHz) noise pedestal. The ramps were linear and of durations between 2.5 and 20 ms. Increments were centrally located within the pedestal.

To test the model against the results of Gallun and Hafter (2006), we employed 477 Hz pure tone pedestals and obtained thresholds for detection of brief symmetrical increments of durations between 10 and 85 ms, gated with 10 ms cosine ramps. Pedestals of 1000 ms were used and the increments were centrally located within the pedestal.

Fig. 2.6a shows the predictions of the model (dashed grey line) compared to the results of the ramped-noise experiment. The model predictions are reasonably close (r=0.94, P=4.8 x10^{-3}, e=0.9) to the data. The model predicts an approximately symmetrical curve about the 'best-detection' rate. The large intensity JNDs at high and low rates of change and best-detection half-ramp duration of 100 ms are in good quantitative agreement. Within the model, Riesz's paradigm and that of the ramped-noise experiment are shown to be equivalent.

Fig. 2.6b shows a comparison of the predictions of the model (dashed grey line) with the data of Riesz's first experiment which determined beat-modulation intensity JND as a function of beat frequency for continuous ~1 kHz pedestals. The shape of these data are similar to the experimental data of the ramped-noise experiment, in that it shows a log-time symmetrical non-monotonic JND as a function of beat rate, where low beat rates are as hard to detect as high beat rates. The shape and form of the function produced by the model is similar (r=0.93, P=1.4 x10^{-5}, e=0.19 dB) to that of Riesz's data, particularly in terms of a minimum JND point and symmetrical shape about the minimum. We note that Riesz's data as a function of level, which almost hold to Weber's Law above about 60 dB SL, do not appear consistent with

other more recent data (Wojtczak and Viemeister, 1999; Allen and Neely, 1997) and so we do not attempt to model them here.

Fig. 2.6c shows the predictions of the model (dashed grey line) compared to the mean data of Plack *et al*. (2006). These data show the effect of duration on brief, linearly ramped increments in noise pedestals. The model shows good agreement with the data (r=0.92, P=7.5 x10^{-2}, e=0.84 dB) in terms of shape, but a small over estimation is evident.

Fig. 2.6d shows the predictions of the model (dashed grey line) compared to the mean data of Gallun and Hafter (2006). These data describe the effect of brief linearly-ramped increments on 477 Hz pure tone pedestals and so represents the pure tone equivalent of the data of Plack *et al.* (2006). The model shows good agreement with the data (r=0.99, P=7.7 x10^{-2}, e=0.1 dB).

Fig. 2.6(e, f) shows selected data points from Oxenham's (1997) data for brief increments and decrements (respectively) in pure tones compared to the predictions of the model (dashed grey line). These data characterize the effect of duration and background (masking) noise on the pure tone intensity JND. The data show a monotonic decrease of JND with increase in duration and a parallel shift upwards in the JND for the addition of masking noise. In our central excitation pattern modelling of these data, we treat the sum of masking noise and tonal pedestal as a single signal and we look for a threshold increase in the maximum loudness slope caused by the increment in the tonal pedestal component. Generally, the model provides reasonable, if not ideal, qualitative and quantitative account of the data (r=0.89, P=2.6 x10^{-8}, e=0.19). For the signals presented in noise, central adaptation provides for an increase of the JND consistent with the data.

Table 2.1. Goodness of fit measures for the central model. Pearson correlation coefficients (*r*, *P*) and *rms* error (*e*) for central excitation pattern rate modelling results compared with the data.

	r	*P*	*e*
Viemeister & Bacon, 1988	0.99	1.8×10^{-13}	0.04
Miller, 1947	0.94	1.4×10^{-5}	0.19
Oxenham, 1997	0.89	2.6×10^{-8}	0.5
Riesz, 1928	0.93	4.8×10^{-4}	0.15
Present study	0.94	4.8×10^{-3}	0.9
Plack *et al.*, 2006	0.99	1.1×10^{-2}	0.84
Gallun & Hafter, 2006	0.99	7.7×10^{-2}	0.1
Overall	**0.91**	$\mathbf{<1} \times 10^{-16}$	**0.09**

Table 2.1 provides a summary of the goodness of fit measures described above and for the overall fit to the whole data set (r=0.91, $P<1 \times 10^{-16}$, e=0.09 dB). Outside of the error margins discussed in the *Error Margins* section, some error in the modelling of the pseudo-continuous data may be explained in terms of assumption of the continuous-levels approximation. It may be that the central adaptation contribution is excessive in these cases.

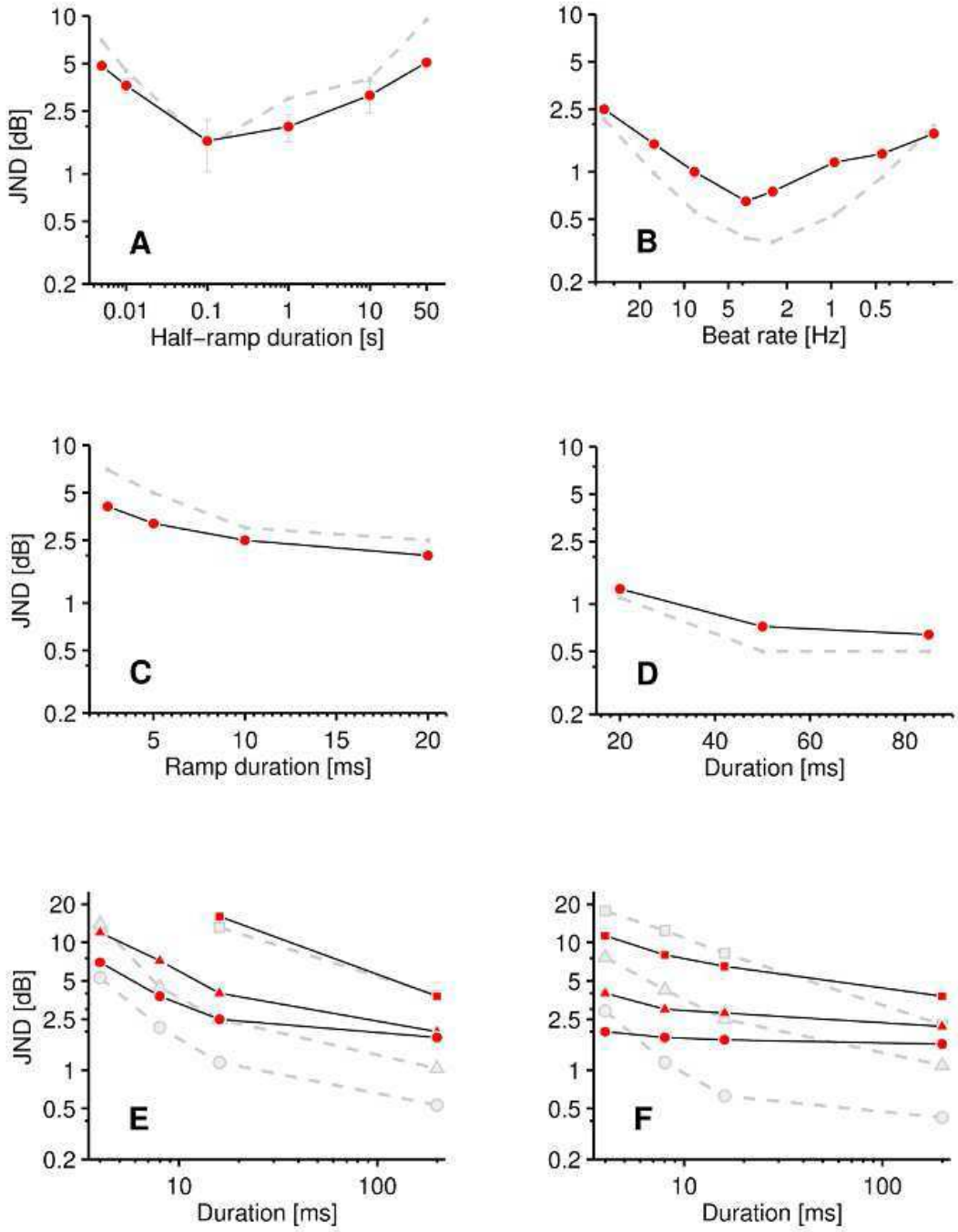

Figure 2.6. Simulation of pseudo-continuous data. Predictions of the central excitation pattern model (dashed grey line); **A** group mean thresholds of the ramped-noise experiment (circles); noise pedestals with up-down ramps, at half-ramp durations of 5, 10, 100, 1000, 10000 and 50000 ms and at an overall listening level of 33 dB SPL (rms). Error bars represent 95% confidence intervals. The trends shown in the data are significant (P=9.55x10^{-8}, *Friedman Rank Sum Test*). **B** Just-noticeable difference for envelope modulation of a 1 kHz tone, as a function of beat frequency, produced with the method of beats

by Riesz for a listening level of 50 dB SL. **C** Just-noticeable difference for detection of symmetrical, linearly-ramped increments in 20-dB spectrum-level noise pedestals, as a function of half-ramp duration (one-sided) - averaged data of Plack *et al*. (circles). **D** Just-noticeable difference for increment detection in 477 Hz pedestals, as a function of increment duration at a peak level of 60 dB SPL - averaged data of Gallun and Hafter (circles). **E, F** JND for increment and decrement detection in 4 kHz pedestals respectively, as a function of duration at a listening level of 55 dB SPL - averaged data of Oxenham for 500 ms pedestals presented in quiet (circles), 0 dB (triangles) and 20 dB (squares) spectrum level noise.

2.5.4. Error Margins

There are several potential sources of error or confusion in the recreation and modelling of the experimental conditions of the studies reviewed in this chapter. First, since much of the data were originally presented in terms of SL, the question of thresholds is important. Riesz (1928), for example, did not obtain absolute thresholds for his subjects but took them from an earlier work by Fletcher and Wegel (1922). Fletcher and Wegel did not describe the method or statistical calculation by which they obtained their thresholds. In any case, the thresholds are sufficiently different to those obtained with modern experimental methods and equipment that some margin must be allowed to account for this. Furthermore, Miller (1947) obtained absolute thresholds for his noise stimulus but did not specify the procedure by which he obtained the absolute thresholds.

Second, there is significant variation in statistical level used to define intensity JND threshold. Miller, for example, defined the threshold according to a 50% correct location on the psychometric function, whereas Viemeister and Bacon defined the threshold at the 70.7% correct point. For our ramped-noise experiment we define threshold at the 79.4% correct point. The model, which is based on loudness data from modern studies (Moore *et al.*, 1997) is likely to provide error in the estimation of intensity JND values for earlier studies.

Third, the data of Miller (1947) were taken with noise stimulus that is only defined as having a spectrum of +-5dB in the range of 150 Hz to 7,000 Hz. Although the +-5 dB appears reasonable for

Gaussian noise, this description does not allow any reasonable assumption to be made about the spectrum of noise outside of the bandwidth specified. Further, Miller did not specify the spectrum of the noise after it had been passed through the filter of the headphone receiver. Generally, the data of the studies reviewed here were obtained with various headphone receivers and other apparatus whose influence is not known.

Fourth, the experimental population size involved in the studies reviewed is highly limited; 2 subjects for Miller, 3 subjects for Viemeister and Bacon, 4 subjects for Oxenham and 3 subjects for Riesz.

Since the intensity JND as a function of listening level is known to be a steep function at low levels, the question of absolute thresholds for a given listener or for a population is critical. Where modelling error is shown in offset but not in slope (i.e., there is an offset in the SPL axis) it is possible that variance in individual thresholds is the source of the error. This is particularly likely in light of the small population sizes described above.

2.5.5. Limitations

The loudness model used here features relatively complex functionality; the transfer function of the outer and middle ear filter is relatively discontinuous, the auditory filters change shape (asymmetrically) with level and many aspects of the nonlinear input/output function are frequency dependent. Our results are therefore somewhat dependent on this model. However, alternate peripheral models should, in principle, produce similar results as far as they show an equivalent (or better) fit to loudness data.

2.6. Chapter Summary

The main objective of this chapter was to establish parameters of a central adaptive model able to relate loudness to the intensity JND. The fit of the model is good, even in the case of pseudo-continuous data, and the adaptation parameters obtained are plausible with regards to the neuroscience literature. The ramped-noise experiment has shown that large intensity JNDs are

obtained at very low rates of intensity change, confirming the generality of Riesz's findings. In the context of the modelling, we have shown that the spread of excitation explanation alone is not sufficient to produce a near-miss. Central adaptation has been used to simultaneously explain data featuring approximate examples of Weber's Law and the near-miss, and to explain the effects of masking noise on increment and decrement detection.

In 1997 Allen and Neely anticipated a role of central adaptation in human auditory perception. We have made explicit the argument that loudness reflects peripheral neural coding, that intensity JND reflects central neural coding and that adaptation has a pronounced effect on human auditory perception. In the next chapter, the selectivity for modulation rate outlined in this chapter is further characterised and related more directly to what is known of the central auditory pathway.

Chapter 3: **Modulation Filters**

Recent studies employing speech stimuli to investigate 'cocktail-party' listening have focused on entrainment of cortical activity to modulations at syllabic (5 Hz) and phonemic (20 Hz) rates. The data suggest that cortical modulation filters (CMFs) are dependent on the sound-frequency channel in which modulations are conveyed. In this chapter, we characterize modulation filters in human listeners using a novel behavioural method. Within an 'inverted' adaptive forced-choice increment detection task, listening level was varied whilst increment size was held constant for ramped increments with effective modulation rates between 0.5 and 33 Hz. The data show frequency dependent trends which suggest that modulation filters are tonotopically organized (i.e., vary systematically along the primary, frequency-organized, dimension). This suggests that the human auditory system is optimized to track rapid (phonemic) modulations at high sound-frequencies and slow (prosodic/syllabic) modulations at low frequencies.

3.1. Central Auditory Contrast Processing

The primary feature represented by the peripheral auditory system is sound frequency. The basilar membrane of the cochlea is arrayed, from base to apex, according to a tonotopic representation, with high frequencies resolved at the basal end and low frequencies at the apical (Pickles, 2008). Tonotopic organization is apparent up to at least primary auditory cortex (Humphries *et al.*, 2010), which has been characterized as showing an intensity-independent representation of sound (Sadagopan and Wang, 2008; Barbour, 2011) responding primarily to stimulus contrast. Numerous studies have revealed a preference for "natural" 1/*f* modulation statistics (Voss and Clarke, 1975, 1978) in the auditory system (Garcia-Lazaro *et al.*, 2006, 2011; Wang *et al.*, 2012) and this selectivity has been localized to auditory cortex (Garcia-Lazaro *et al.*, 2011; Wang *et al.*, 2012). Models comprising central modulation filter-banks have been proposed (Dau *et al.*, 1997a, 1997b; Jepsen *et al.*, 2008), including the existence of independent modulation filters in the human auditory cortex (Xiang *et al.*, 2013). Presumably, these cortical modulation filters (CMF) represent separate neuronal populations, each with different tuning to modulation rate (Ding and Simon, 2013). Xiang *et al.* (2013) have suggested that, much like the 'beating' that occurs within the auditory filters of the cochlea itself, CMFs are nonlinear and produce sum and difference products when two modulations (at different rates) exist within the same filter.

Speech intelligibility has been shown to be dependent on sensitivity to slow temporal amplitude modulations (Drullmann *et al.*, 1994; Shannon *et al.*, 1995). Assuming CMFs play a key role in coding speech, particularly in background noise, i.e. 'cocktail-party' listening (see Ding and Simon, 2013; Zion Golumbic *et al.*, 2013; Lakatos *et al.*, 2013), a potential strategy for separating speech from background noise, and one recently suggested by Ding and Simon (2013), is that CMFs are carrier-frequency dependent. That is, the modulation rate to which CMFs are tuned increases systematically along the tonotopic gradient. This strategy also makes sense from the perspective of the limits imposed by peripheral auditory filters, the bandwidths of which increase (in Hertz terms) with increasing centre

frequency, making it theoretically possible to convey increasingly higher modulation rates. In support of this, Lakatos *et al.* (2013) demonstrated tonotopically-arranged entrainment of neural activity in the cortex of non-human primates, suggestive of a tonotopic arrangement of CMFs. Further evidence in support of a tonotopic arrangement of CMFs comes from neuroimaging studies (as reviewed by Zarate and Zatorre, 2012), where a 'dual stream model' of the cortex has been proposed to account for hemispheric spectro-temporal processing differences (for musical stimuli) equivalent to those observed by Lakatos *et al.* (2013). It follows from this that if CMF tuning is carrier-frequency dependent, it might be the product of tonotopic variation in underlying neuronal physiology.

Since human cortex (like that of the monkey) is tonotopically mapped (Humphries *et al.*, 2010), if CMFs are carrier-frequency dependent, then subcortical spread of excitation across the tonotopic gradient (likely initiated at the level of the basilar membrane) may have an equivalent 'cortical spread of modulation' effect, where the peripheral spread of excitation along the tonotopic gradient spreads modulation across nearby CMFs. This spread of modulation might then result in similar level-dependent, nonlinear interactions to those observed by Xiang *et al.* (2013), such that CMF tuning would broaden with increasing sound level to cause 'simultaneous modulation masking', much as the peripheral auditory filters cause simultaneous energetic masking (Brungart *et al.*, 2006).

Previous psychoacoustic studies have suggested that intensity discrimination is carrier frequency dependent; intensity discrimination varies as a function of stimulus duration (Watson and Gengel, 1969) and as a function of sound level (e.g., Jesteadt *et al.*, 1977; Long and Cullen, 1985; Ozimek and Zwislocki, 1996). However, these findings have not been systematically verified or related to cortical processing of stimulus contrast. In keeping with the approach in Chapter 2, more recent studies have suggested a key role of contrast in detecting changes in sound intensity (Oxenham, 1997; Plack *et al.*, 2006; Gallun and Hafter, 2006; Simpson and Reiss, 2013). Here, we investigated modulation filters using a novel behavioural method derived from psychoacoustics. Listeners were asked to detect linearly-ramped increments (i.e., the just noticeable difference [JND]), in pure tone carriers, at effective modulation rates

between 0.5 and 33 Hz. These rates span the range of prosodic (<5 Hz), syllabic (5 Hz) and phonemic (20 Hz) rates commonly found in speech (Xiang *et al.*, 2013; Drullman *et al.*, 1994; Shannon *et al.*, 1995). By varying the level and frequency of the carrier signal, we characterized the tuning of the modulation filters as a function of carrier frequency and level (in terms of *modulation rate sensitivity* and *modulation depth sensitivity*). Our data support the view, as suggested by Ding and Simon (2013) and implied by Lakatos *et al.* (2013), that modulation filters are systematically dependent on carrier frequency. Given that the cortex is known to be tonotopically organized (Humphries *et al.*, 2010), this suggests that CMFs are similarly organized, in agreement with the well-established tonotopic map, and in support of the 'dual stream' model (Zatorre and Zarate, 2012). We also observe that modulation sensitivity changes as a function of sound level in a manner that may be attributable to spread of excitation across modulation filters as sound level increases. In summary, our data suggest that the human auditory system is optimized to track rapid modulations at high sound-frequencies and slow modulations at low frequencies, and supports a model of cortical function based on tonotopically-organized modulation filters.

3.2. Experiment 3.1

As in Chapter 2, the prevailing experimental paradigm for assessing the intensity JND specifies a fixed listening level and an adaptively-varied increment size (Oxenham, 1997; Plack *et al.*, 2006; Gallun and Hafter, 2006; Simpson and Reiss, 2013). However, due to individual differences in auditory physiology, small changes in listening level produce large changes in the size of the intensity JND (e.g., Viemeister and Bacon, 1988) and, near threshold, the mapping is both extremely nonlinear and highly individualized. When this method is applied to a medium sample size, even if individual listeners are extremely reliable, the mean results for such a sample constitute a gross averaging (blurring) of subtle trends in the data that potentially characterize modulation filter tuning. In previous studies (e.g., Jesteadt *et al.*, 1977; Long and Cullen, 1985; Ozimek and Zwislocki, 1996), listening levels were fixed relative to the absolute threshold

(i.e., sensation level – SL) for each listener in order to provide comparison between intensity JNDs at different carrier frequencies. This resulted in the observation of carrier-frequency dependence in the JND as a function of SL but the findings were not related to temporal integration (see below), a major topic of more recent investigations (Oxenham, 1997; Plack *et al.*, 2006; Gallun and Hafter, 2006; Simpson and Reiss, 2013). Here, we invert the traditional experimental paradigm such that listening level is adaptively varied and the size of the increment is held constant (see Fig. 3.1). This normalizes between-subject variance caused by individual differences in absolute thresholds.

As in Chapter 2, by assessing JNDs at different ramp durations, a modulation rate sensitivity function is produced (Oxenham, 1997; Plack *et al.*, 2006; Gallun and Hafter, 2006; Simpson and Reiss, 2013), characterizing the relative sensitivity of the modulation filter to different ramp (i.e., modulation) rates. From this function, tuning for the modulation filter at each carrier frequency can be estimated. For modulation filters tuned to low modulation rates (e.g., prosodic or syllabic; 5 Hz or less), the modulation rate sensitivity function will show greatest sensitivity to the slowest ramps (1000 ms). For modulation filters tuned to higher modulation rates (e.g., near phonemic; 20 Hz or more), the modulation rate sensitivity function will show greatest sensitivity at the higher modulation rates. By testing at different heights of ramp (with a fixed ramp duration of 5 Hz effective modulation rate), modulation depth sensitivity functions can be produced and level dependence in the modulation filters can be probed. If the tuning of modulation filters varies as a function of carrier frequency, level-dependent trends with carrier-frequency should be observed. This is because, for a fixed modulation rate, as carrier frequency is varied some CMFs will be operating in the tuned peak and other CMFs will be operating in the skirts. Therefore, this also allows us a window into possible spread-of-modulation effects.

3.2.1. Inverted method

Detection threshold levels were obtained for up-down ramped increment envelopes added to the centre of 4 s long pure tone carrier-signals, for nine listeners. Listeners were presented with pairs of matched 4 s long tones, one of which (at random) contained a linear up-down increment. The listening level was started high, so that the increment was clearly audible, and then varied adaptively until threshold level was determined. If the subject correctly selected the tone with the increment the listening level was reduced, and, if incorrectly, the listening level was increased. Thresholds were estimated by averaging the listening level at several such decision rule points.

By separately varying the frequency of the carrier and the size and duration of the increment envelopes, corresponding equal-JND-level contours were produced and, from these contours, threshold-level functions of ramp duration and of ramp size, i.e., *modulation rate sensitivity* and *modulation depth sensitivity* functions obtained. Parametric analysis of the data was employed to reveal systematic trends with carrier frequency.

Two experiments were conducted. The first experiment was designed to illustrate the modulation rate sensitivity tuning of modulation filters as a function of carrier frequency. The second experiment was designed to illustrate the associated modulation depth sensitivity tuning within the modulation filters for modulations at approximately 5 Hz (i.e., syllabic rate). In experiment 3.1 (the *temporal* experiment), the size of the intensity increment was fixed at 3 dB. Half-ramp duration of the increment was set to either [15, 50, 100 or 1000] ms for each block (equivalent to a modulation rates of [33, 10, 5 or 0.5] Hz respectively). This produced a set of four contours, from which modulation rate sensitivity functions of increment ramp duration could be extracted. In experiment 3.2 (the *magnitude* experiment), the increment size was set to either [1,2 or 3] dB for each block, and half-ramp durations of 100 ms (corresponding to a modulation rate of approximately 5 Hz) were used for each respective block. This produced a set of three contours, from which modulation depth sensitivity functions could be extracted. From here onwards, we refer to the ramp durations of [15, 50, 100 or 1000] ms in terms of the equivalent modulation rates of [33, 10, 5 or 0.5] Hz respectively.

A *post-hoc* analysis was performed quantifying systematic trends in the shapes of the modulation rate sensitivity and modulation depth sensitivity functions, and a correlation analysis was employed to assess correlations in the two measures that may be attributable to the properties of the modulation filters.

3.2.2. Near Miss

A prerequisite of this method is that, for a given increment, detection improves with increases in listening level. Weber's Law states that the ratio of intensity to the intensity JND should be constant (Weber, 1846) and has been shown to be approximately true for wideband signals (Miller, 1947). However, in the case of pure tones, Weber's Law has been shown not to hold (e.g., Viemeister and Bacon, 1988) and the characteristic steady (monotonic) decrease in the JND with increasing sound level is referred to as the 'near-miss to Weber's Law' (McGill and Goldberg, 1968). The near miss necessary for the method has been shown to hold for continuous 1-kHz carriers up to 85 dB SPL, corresponding to around 80 dB above threshold (Viemeister and Bacon, 1988). In this study, by using relatively large increments, we limit our investigation to the range between threshold and around 40 dB above threshold. However, it should be noted that non-monotonicity was observed for gated 1-kHz signals above 90 dB SPL in the above-mentioned study (Viemeister and Bacon, 1988), and that the near-miss is less well defined in (or even absent from) studies employing noise maskers (e.g., Peters *et al.*, 1995).

3.2.3. Experiment 3.1: Stimuli

Stimuli were generated digitally at 24 bit resolution. A pair of Beyerdynamic DT100 isolating headphones was used to present the stimulus to listeners directly from a computer, at a sampling rate of 44,100 Hz. Presentation was diotic (identical in both ears). The carriers were gated on and off using 10 ms raised-cosine ramps. In both experiments, detection threshold levels were obtained at carrier frequencies of [62, 125, 250, 500, 1000, 2000, 4000, 5650, 8000, 11300, 16000] Hz. Pure tone (sinusoidal) carriers were

presented in blocks of JND = 1, 2 and 3 dB, where JND is defined as $10log_{10}(1+\Delta I/I)$, I=intensity. Carrier frequency was varied inside blocks. Symmetrical ramped envelopes were added to the tone carriers. A ramped envelope for a given duration consisted of a linear increment ramp of that duration, immediately followed by a linear decrement ramp of the same duration. The ramp envelopes were located in the temporal centre of the 4 s long carrier. The increment was set to a fixed value within any given block. In the *temporal* experiment, linear up-down ramped increments with effective modulation rates of [0.5, 5, 10 or 33] Hz were imposed upon 4-s pure tone carriers. Threshold levels were obtained for JND = 3 dB. In the *magnitude* experiment, 5 Hz modulations were used and threshold levels were obtained for JND = [1,2,3] dB.

The range of JNDs was chosen to lie within the known monotonic range. The range was also limited to relatively large values of JND (>=1 dB) for the reason that very small values of JND at low and high carrier frequencies would have required listening levels beyond those possible with the available apparatus.

3.2.4. Experiment 3.1: Procedure

For each carrier frequency within a block, an adaptive three-down one-up, two-interval forced-choice (2IFC) procedure was employed which estimates the 79.4% correct identification (Levitt, 1971). Each pair of signals that constituted a trial, presented in random order, consisted of one carrier that contained a ramp envelope and a second carrier that contained no ramp. The signal pairs were presented with silent inter-signal intervals of 0.5 s. At the start of the adaptive sequence, the initial listening level was set to be below the threshold of audibility. This was increased in steps of 10 dB until the subject indicated that the carriers (and increment) were clearly audible, at which point the adaptive procedure began. Three consecutive correct identifications of a ramp resulted in a reduction in the listening level and one incorrect answer resulted in an increase. After each trial, subjects were provided correct/incorrect feedback on their

responses. Following a reversal, the step size (starting value of 10 dB) was divided by two. A reversal was defined as an increase in listening level following a decrease, or *vice versa*. After 12 reversals, threshold level was taken as the arithmetic mean of the last 10 reversals. Trials were undertaken in blocks lasting no longer than 20 minutes. Blocks were occasionally interrupted with a break period of 15 minutes, after which the block continued until either the next rest period or completion. Blocks and carrier-frequency orders within blocks were chosen at random. Prior to the test, each subject was provided with a brief demonstration to familiarize themselves with the interface and procedure. A training period was then undertaken which was terminated when the performance of the subject was judged to have stabilized. The data from the training period were not included in subsequent analyses.

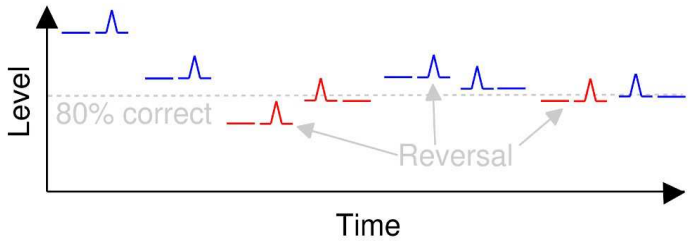

Figure 3.1. Illustration of the inverted method. Pairs of pure tones are presented, one of which contains a ramped increment. The ramp size and duration is fixed throughout, whilst listening level is adaptively changed until the procedure converges on ~80% correct performance. When the listener correctly identifies the location of the ramped intensity increment the listening level is reduced, otherwise the listening level is increased, depending on a rule (simplified here to a 1-up, 1-down rule). Correct responses (blue) result in decreased listening level and incorrect responses (red) result in increased listening level. The ~80% correct threshold level is estimated by averaging the listening level measured at several points where the adaptive procedure changes direction ('reversals'). The step size of the level change reduced after a reversal and the procedure eventually converges on the ~80% correct point.

3.2.5. Experiment 3.1: Listeners

Nine unpaid volunteer subjects served as listeners in the experiments. Six male subjects and three female subjects took part. The mean age of the subjects was 27 (min: 21, max: 33). All reported normal hearing and previous experience of participating in listening tests. All participants were naïve concerning the purpose of the test.

3.3. Results of Experiment 3.1

3.3.1. Modulation filter tuning is carrier-frequency dependent

In order to characterize the tuning of the modulation filter, sensitivity measures must be obtained for two main properties; modulation rate (i.e., rate of change) and modulation depth (i.e., contrast). In the first experiment, we assessed the ability of listeners to detect a change in sound intensity (from a reference intensity), where the change constituted an increment of a defined duration, quantified by the 'half-ramp' duration, i.e. the duration from the start of the ramp to its peak. As all ramps were symmetric in time around their peaks, changing the duration of the ramp provides for a proxy of different modulation rates, i.e. faster ramps represent faster modulation rates and slower ramps represent slower rates. Effective modulation depth was held constant at 3dB, so that threshold levels were obtained by assessing the ability of listeners to detect a 3 dB increment for effective modulation rates of [0.5, 5, 10 or 33] Hz for pure tones spanning the range 62 Hz to 16 kHz, i.e., encompassing much of the frequency range of normal-hearing listeners. Absolute sound level was adaptively varied according to the criteria described in the section 3.2 until ~80% performance was reached.

Figure 3.2a plots group mean threshold levels as a function of carrier frequency for the nine subjects, for increments of 3 dB at effective modulation rates of [0.5, 5, 10 or 33] Hz. Each data point corresponds to the mean absolute sound-level at which 80% performance was reached for 3 dB ramps of the respective modulation rate. The overall shape of these curves (equivalent to equal loudness-level contours e.g., see

Moore *et al.*, 1997) is not greatly affected by ramp duration. However, the overall distance between the functions is smallest at the extremes of the carrier frequency range.

Fig. 3.2b plots the same data as in Fig. 3.2a, but here as modulation rate sensitivity functions, where the data are normalized to remove the effect of absolute threshold. The main effect of half-ramp duration was verified to be significant in all modulation rate sensitivity functions ($P<0.05$, *Friedman Rank Sum test*), with the exception of those for 62 Hz and 16 kHz. This is likely explained by the combined inter and intra-subject variability associated with extremes of carrier frequency and of half-ramp duration.

The modulation rate sensitivity function is monotonic for low carrier frequencies, and non-monotonic (U-shaped) for high carrier frequencies. The monotonic nature of the functions at low carrier frequencies is consistent with data from several contemporary studies (e.g., Oxenham, 1997; Plack *et al.*, 2006; Gallun and Hafter, 2006) suggesting that increments (or decrements) in sound intensity are detectable in terms of a change in energy (with no reference to the rate of change). And the non-monotonic modulation rate sensitivity functions at high carrier-frequencies are consistent with data reported in Chapter 2 for similar ramps conveyed in noise (Simpson and Reiss, 2013), which suggest that increment detection might be determined, at least in part, in terms of a change in stimulus contrast. A gradual transition from monotonic functions at low carrier-frequencies to non-monotonic functions at higher carrier-frequencies is evident in the data, with a transition point around 4 kHz. This is in agreement with the findings of Watson and Gengel (1968), who demonstrated a faster integration time constant with increasing carrier frequency. However, in both cases, it seems likely that non-monotonic functions would be observed given longer durations on the order of minutes such as those employed in Chapter 2.

A critical feature of this method is that the different durations of intensity ramp act as a proxy for modulation rate; short ramps correspond to fast rates and long ramps to slow. By measuring the listening level at which the 80% performance was achieved for the various effective modulation rates, we obtained a measure of the sensitivity of the modulation filter to modulation at each effective rate, i.e. a modulation rate sensitivity function. A monotonic function implies increasing sensitivity to decreasing rates. A non-

monotonic function implies that peak sensitivity is within the range of rates tested. The centre frequency of the modulation filter corresponds to the modulation rate at which it is most sensitive. By measuring the regression slope (*G*) of each modulation rate sensitivity function, we obtained a measure of how well our range of modulation rates captured the centre frequency tuning of the modulation filter at a given carrier frequency. This provides a crude proxy to centre frequency tuning. It should be noted that *G* does not quantify a curve fit to the modulation rate sensitivity function, but rather is a means of quantifying how well the peak of the modulation filter is centrally captured by the function, i.e., *G* is informative as to how well the modulation rates represented by each filter are arrayed around the tuned peak. Thus, $G = 0$ indicates a modulation filter tuned to a carrier frequency at the centre of the function, $G < 0$ indicates a filter tuned to the right of the function's centre and $G > 0$ a filter tuned to the left of the function's centre.

Fig. 3.2c shows an interpretation of the data in terms of illustrative modulation-filters, corresponding to the modulation rate sensitivity functions, which illustrate variation in modulation filter centre-frequency for two example modulation rate sensitivity functions; at low carrier-frequencies there is a large, positive value of *G*, meaning that modulation filters are most sensitive to slow (i.e., near-prosodic) modulations. At high carrier-frequencies there is a smaller (even negative) value of *G*, meaning that the modulation filters are most sensitive to faster (i.e., near-phonemic) modulations. Fig. 3.2d plots *G* as a function of carrier frequency. Although it is not a clear trend, the decrease of *G* with increase in carrier frequency confirms the trend for increasingly high-rate tuned modulation filters along the tonotopic gradient. The narrower dynamic range over which 80% performance was achieved at the extremes of the tonotopic gradient indicates these modulation filters to be relatively broadly tuned, whilst the wider dynamic range at the mid-to-high carrier frequency end indicates these modulation filters to be more selective for modulation rate.

The data plotted in Fig. 3.2 can be summarized as follows. At low carrier frequencies, modulation filters appear to be most sensitive to modulation rates that are near-prosodic (i.e. ~1-5 Hz), but towards higher carrier frequencies the filters appear to be more sensitive to near-phonemic (~20 Hz) modulation

rates. Within the range of modulation rates represented in our data, at low carrier frequencies the modulation filters appear to be low pass and at higher carrier frequencies the filters appear to be pass band. However, our data do not preclude the possibility that, if slower modulation rates were represented in the function, pass band tuning might be observed for low carrier frequencies.

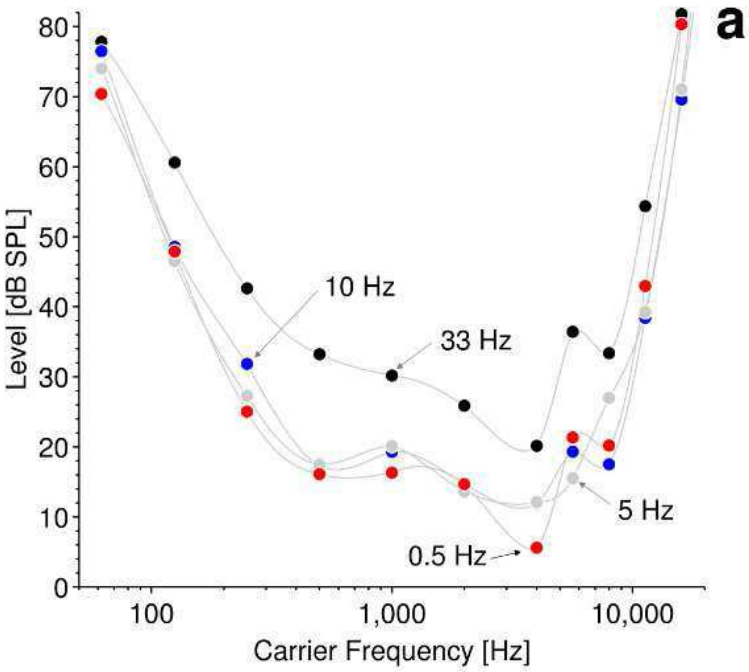

Figure 3.2a. Modulation rate sensitivity contours. Group mean threshold-levels as a function of carrier frequency for the nine subjects, for increments of 3 dB at effective modulation rates of [0.5, 5, 10 or 33] Hz. Each data point corresponds to the mean absolute sound-level at which 80% performance was reached for 3-dB ramps of the respective durations.

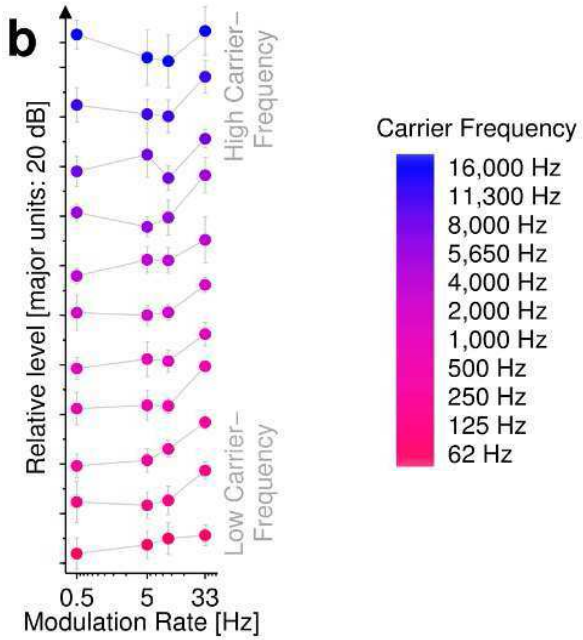

Figure 3.2b. Modulation rate sensitivity functions. b plots the same data as in Fig. 3.2**a** in the form of modulation rate sensitivity functions, where the data are normalized to remove the effect of absolute threshold. Colour scale from red to blue indicates low-to-high carrier frequency. Error bars indicate 95% confidence intervals. Modulation rate sensitivity functions become increasingly non-monotonic with increase in carrier frequency, indicating a smooth transition in modulation tuning from near-prosodic to near-phonemic rates along the tonotopic gradient.

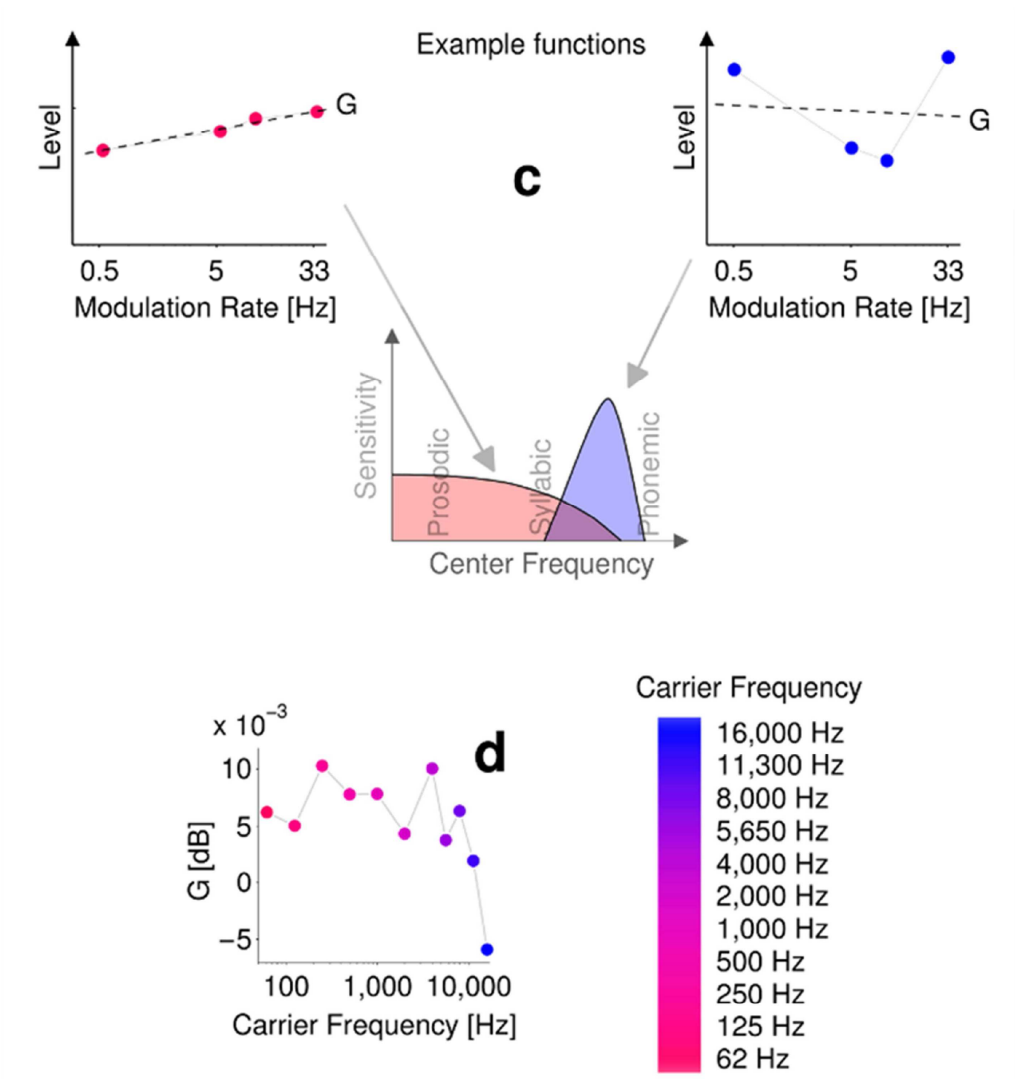

Figure 3.2c/d. *G* as a function of carrier frequency. c illustrates our interpretation of *G* for two example modulation rate sensitivity functions; large, positive values of *G* at low carrier-frequencies, indicating modulation filters to be most sensitive to slow (i.e., near-prosodic) modulations. At high carrier frequencies there is a smaller (even negative) value of *G*, meaning that the modulation filters are most sensitive to faster (i.e., near-phonemic) modulations. **d** plots *G* as a function of carrier frequency.

3.3.2. Modulation filter tuning is listening-level dependent

In the second experiment, modulation rate (half-ramp duration) was held constant and the effective modulation depth varied by varying the height of the ramp, to produce a measure of modulation depth

sensitivity. These data were then assessed with respect to sensitivity to modulation rate from the first experiment.

Figure 3.3a plots contours showing group mean threshold levels at each carrier frequency for the nine subjects, for increments of [1,2,3] dB at effective modulation rates of 5 Hz. Each data point corresponds to the mean absolute sound-level at which 80% performance was reached for 1, 2 or 3 dB ramps, respectively. In general, the contours of both experiments (Fig. 3.2a & 3.3a) resemble equal loudness contours, and hence it is reasonable to assume that a major factor in their shape is the outer- and middle-ear transfer function. This is supported by a correlation between Glasberg and Moore's (2002) combined outer-and-middle ear filter and the average contour from all the data of the temporal and magnitude experiments (r=0.96, p=3.6x10^{-6}, *Pearson two-tailed*).

The contours of the data in Fig. 3.3a are not parallel, but are most widely spaced in the middle of the carrier-frequency range, and the overall dynamic range of the functions is again smallest at the extremes of the carrier-frequency range. This indicates that modulation depth sensitivity varies with level most steeply in the middle of the carrier-frequency range. Fig. 3.3b removes (by normalization) the effects of the absolute threshold, allowing the form of the functions to be compared directly. The curved functions at low carrier-frequencies are comparable to the equivalent functions previously reported (e.g., Viemeister and Bacon, 1988). Thus the results of previous studies most likely reflect the tuning of the relevant modulation filter at a particular carrier frequency and level. The error bars in Fig. 3.3b represent 95% confidence intervals. Main effect of JND size was verified to be significant in all functions (P<0.05, *Friedman Rank Sum test*), with the exception of the modulation depth sensitivity function at 62 Hz. As previously, this is likely a result of the combined inter and intra-subject variability associated with extremes of carrier frequency and of increment size. At high carrier-frequencies, the functions are almost perfectly linear (log-log axes) and so could be predicted with a power law. There is a general trend towards power-law type functions as carrier frequency increases, with a transition after 4 kHz. Furthermore, by comparing the data for 62 Hz and 16 kHz with nearly identical absolute threshold levels

at 1 dB (Fig. 3.3a), differences in the shapes of the functions between low and high carrier-frequencies are most apparent. The same comparison is also evident for 125 Hz and 11.3 kHz.

At high carrier-frequencies, as the JND is increased (Fig. 3.3b) the listening level (at threshold) is reduced proportionally. This suggests that tuning is relatively invariant to sound level. However, at low carrier-frequencies, as the JND is increased the listening level (at threshold) is not reduced proportionally, suggesting that tuning changes with sound level. In order to assess the relative changes in tuning of modulation filters at different listening levels, gradients for the level functions of Fig. 3.3b were calculated. For each function, ΔG was calculated as the change in slope between threshold levels for increments of [1,2] dB and [2,3] dB (where a zero value of ΔG indicates power-law type functions). Fig. 3.3c plots ΔG as a function of carrier frequency and shows a steady rise of ΔG with increase in carrier frequency. Fig. 3.3d plots ΔG as a function of G (a proxy to modulation filter centre frequency), including a quadratic fit to the data (dashed line). It can be seen that G and ΔG are highly correlated (r=-0.945, $p<5\text{x}10^{-7}$, *Spearman two-tailed*).

One way of explaining the trends shown in Fig. 3.3c and 3.3d might be the spread-of-modulation that would result from tonotopically organized CMFs. Fig 3.3e shows a cartoon illustration of this interpretation of ΔG for two example modulation depth sensitivity functions. Near absolute threshold (i.e., for JNDs of 3dB) peripheral spread of modulation plays little role, meaning that coding of the syllabic (5 Hz) modulation at a given carrier frequency is dependent only on the modulation filter located on the tonotopic gradient according to carrier frequency. However, for smaller JNDs level is increased and peripheral spread of the carrier causes spread of modulation. Spread of modulation causes the recruitment of modulation filters that are more or less sensitive to syllabic (5 Hz) modulation. For high frequency carriers (blue), the basal modulation filter is most sensitive to the syllabic (5 Hz) modulation, and so recruitment of less sensitive filters (by peripheral spread of modulation) has little influence on performance. However, for low-frequency carriers (red), the apical modulation filter is insensitive to the

syllabic (5 Hz) modulation and so at high levels (i.e., 1 dB JND) performance is enhanced by more sensitive modulation filters recruited towards the basal end of the tonotopic gradient. This enhancement falls away as level is reduced and hence produces the curved functions seen towards the apical end of the tonotopic gradient. Therefore, small values of ΔG (i.e., at high carrier-frequencies) indicate little effect of spread-of-modulation and large values of ΔG (i.e., at low carrier-frequencies) indicate spread of modulation effects. Following this interpretation, the steady rise of ΔG with increase in carrier frequency indicates a trend describing steady decrease in spread-of-modulation effects across the tonotopic gradient. The correlation shown in Fig. 3.3d provides both a cross validation for both proxy measures of modulation filter tuning, and support for our interpretation of an interaction between modulation-filter tuning and peripheral spread-of-modulation effects. However, it should be noted that spread of modulation is not the only mechanism that may be invoked to explain ΔG. Rather spread of modulation is a mechanism we would expect to see evidence of, based on the suggested cortical tonotopy, and hence is the most plausible interpretation given the correlation with G. Alternative explanations for ΔG might include input/output nonlinearities which are carrier frequency dependent or CMF bandwidths which change with sound level in a carrier frequency dependent way.

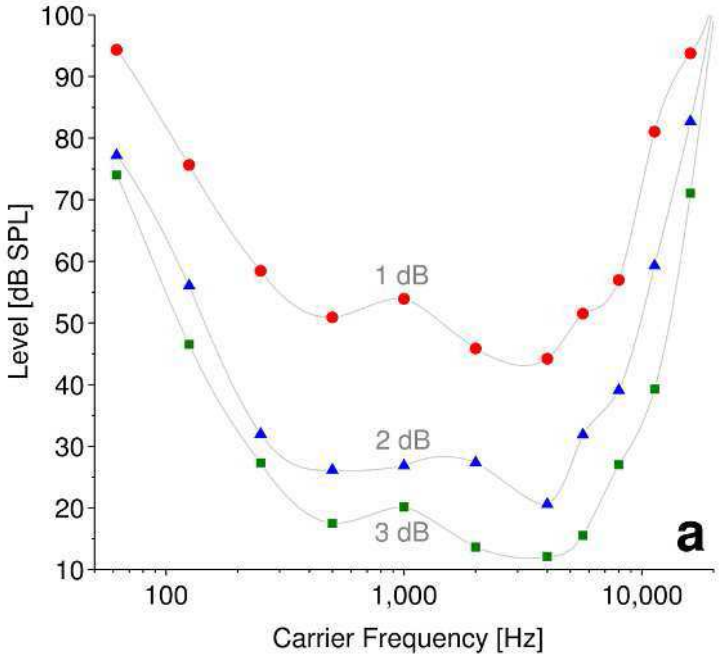

Figure 3.3a. Modulation depth sensitivity contours. a plots contours showing group mean threshold levels at each carrier frequency for the nine subjects, for increments of 1 (red circles), 2 (blue triangles) or 3 (green squares) dB at an effective modulation rate of 5 Hz (i.e., syllabic). Each data point corresponds to the group mean absolute sound-level at which 80% performance was reached for 1, 2 or 3 dB ramps respectively.

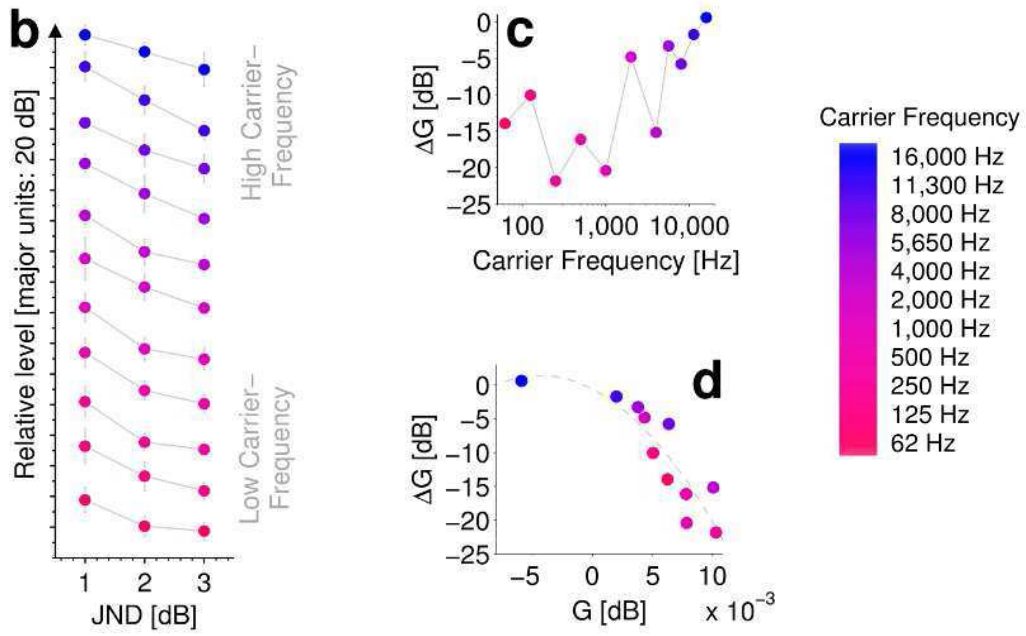

Figure 3.3b/c/d. Modulation depth sensitivity functions. **b** plots the same data as in Fig. 3.3**a**, normalized to produce modulation depth sensitivity functions. The error bars represent 95% confidence intervals. Colour scale (right) from red to blue indicates low-to-high (apical to basal) carrier frequency. **c** plots Δ*G* as a function of carrier frequency. **d** plots Δ*G* as a function of *G* (a proxy to modulation filter centre frequency), including a quadratic fit to the data (dashed line). **e** shows an interpretation of the data in terms of a cartoon illustration of the interpretation of Δ*G* for two example modulation depth sensitivity functions.

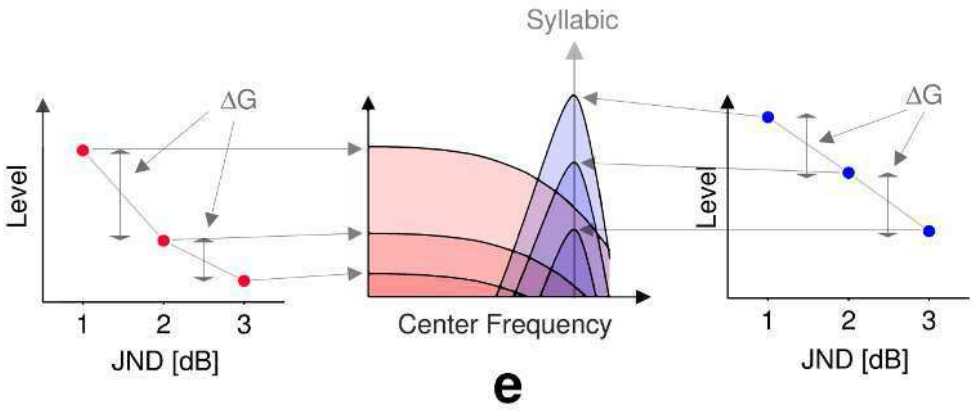

Figure 3.3e. Interpretation of Δ*G*. **e** shows an interpretation of the data in terms of a cartoon illustration of the interpretation of Δ*G* for two example modulation depth sensitivity functions.

3.4. Chapter Summary

In this chapter we have provided evidence that human modulation filter tuning is both carrier frequency and level dependent. Our data suggest that CMFs are tonotopic and that the human auditory system is optimized to track rapid (phonemic) modulations at high carrier frequencies and slow (prosodic) modulations at low carrier frequencies. We have suggested, based on evidence of modulation filter level dependence, that peripheral spread of excitation is likely to result in 'spread of modulation' by spread-of-carrier between CMFs. Furthermore, our data suggests systematic (tonotopic) variation in underlying cortical neuronal physiology. Our data and conclusions provide support for the cortical speech processing strategy suggested by Ding and Simon (2013) and confirmation in humans of the findings of Lakatos *et al.* (2013) in monkey CMFs. Carrier frequency and level-dependent tuning of CMFs may have implications for the cocktail party problem and appear consistent with the 'dual stream' hemispheric model suggested in music neuroimaging studies (Zatorre and Zarate, 2012). In the next chapter, the selectivity characterised in this and the previous chapter is put in the context of adaptation.

Chapter 4: **Selective Adaptation**

Adaptation to the statistical distribution of sounds has been independently reported in neurophysiological studies employing probabilistic stimulus paradigms in small mammals. However, the apparent sensitivity of the mammalian auditory system to the statistics of incoming sound has not yet been generalized to task-related human auditory perception. Here, we show that human listeners selectively adapt to novel sounds within scenes unfolding over minutes. Listeners' performance in an auditory discrimination task remains steady for the most common elements within the scene but, after the first minute, performance improves for rare (oddball) sound elements, at the expense of rare sounds that are relatively less odd. Our data provide the first evidence of enhanced coding of oddball sounds in a human auditory discrimination task and suggest the existence of an adaptive mechanism that tracks the long-term statistics of sounds and deploys coding resources accordingly.

4.1. Central Auditory Adaptation

For many species, survival depends on the ability to encode the current sensory scene with a high degree of accuracy, whilst remaining alert to novel events in the environment (Bregman, 1990; McDermott, 2009). These two demands appear in conflict in terms of their call on neural resources. Adaptation to 'enhance' representation of both common (Dean *et al.*, 2005, 2008; Watkins and Barbour, 2008; Wen *et al.*, 2009; Rabinowitz *et al.*, 2011; Sadagopan and Wang, 2008; Barbour, 2011; Jaramillo and Zador, 2011; Walker and King, 2011) and rare (Ulanovsky *et al.*, 2003, 2004; Nelken, 2004; Perez-Gonzalez *et al.*, 2005; Malmierca *et al.*, 2009; Yaron *et al.*, 2012) sounds has been reported in neurophysiological studies, seemingly in the same brain centres and employing similar probabilistic stimulus paradigms. How then does sensitivity to the statistical distribution of sounds manifest in sensitivity to both high and low probability events?

In order to assess neural sensitivity to the statistics of sounds, Dean *et al.* (2005, 2008) introduced a probabilistic paradigm in which stimulus intensities were selected according to distributions featuring low- and high-probability regions (LPRs and HPRs). We employed a similar paradigm in which listeners were presented with three variants of a stimulus, one of which occurred with high probability (80%) and the other two with low probability (10% each). Stimuli consisted of two sounds (noise bursts). One presentation of the stimulus, followed by a response, constituted a trial. After hearing the stimulus, the subject was asked to report "which sound was louder?", indicating their response by pressing 1 or 2 on a keypad. In the first experiment, the three stimulus variants differed in terms of their overall intensity (35, 55 or 75 dB SPL). In the second experiment, the three variants differed in terms of the inter-sound interval (ISI: 350, 700 or 1050 ms) and were fixed at 55 dB SPL.

4.2. Experiment 4.1: Methods

The overall method was broken down into a two-stage procedure. The first, or calibration, stage determined the just-noticeable difference (JND) for intensity for pairs of sounds at each possible intensity and ISI generating, in each case and for each listener, the intensity difference for a fixed *a-priori* probability of success in the discrimination task (~80%). The second, probabilistic, stage presented the listener with three different stimuli, each set to the sound-level JNDs determined in the calibration stage, and stimuli occurring with *a-priori* probability within a given epoch (Fig. 4.1).

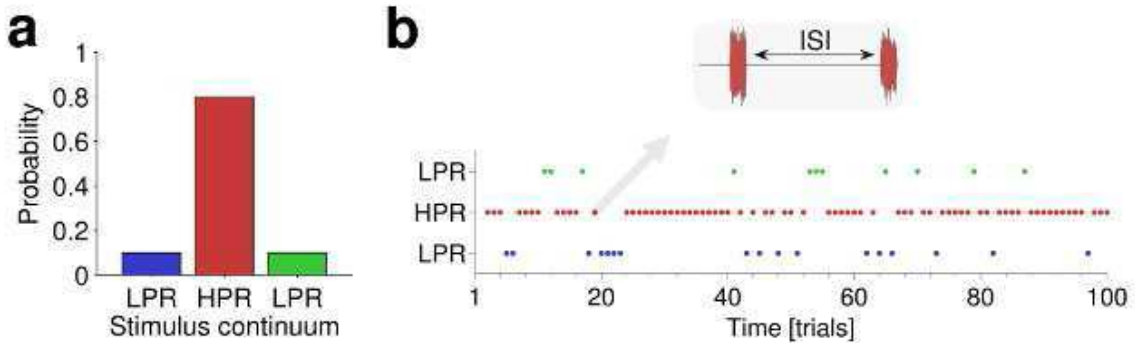

Figure 4.1: Stimulus probability. a In each of two experiments, listeners were presented with 1000 calibrated trials. Each trial was selected from three possible stimuli according to *a-priori* distributions that changed before each 100-trial epoch. The three stimuli consisted of changes in different sound features (intensity in experiment 4.1, ISI in experiment 4.2). Within an epoch, one of the three stimuli was selected with *a-priori* probability of 80% (high probability stimulus, red) and the other two versions were each selected with 10% probability (low probability stimulus, blue, green). **b** plots an example epoch consisting of 100 stimuli selected at random according to the probabilities described in panel **a**.

4.2.1. Experiment 4.1: Stimuli and task

Listeners discriminated intensity of pairs of 50 ms bursts of wideband noise (20 Hz – 20 kHz), gated with 5 ms raised-cosine ramped envelopes and separated by a silent ISI. One of the noise bursts was randomly selected to be louder than the other and the task (in each trial) was to indicate on a keypad which sound

(of the pair) was louder. Presentation of each new trial followed a subject's registration of the response to the previous trial. Directly after the response was entered, subjects were provided correct/incorrect feedback. Each noise burst was generated randomly prior to presentation. In the first experiment, the root-mean-squared (*rms*) sound pressure level (SPL) was 35, 55 or 75 dB and the ISI was fixed at 350 ms. In the second experiment, the *rms* SPL was fixed at 55 dB and the ISI was 350, 700 or 1050 ms. Noise bursts were generated digitally at 24 bit resolution. Beyerdynamic DT100 isolating headphones were used to present the stimulus (diotic) to listeners directly from a computer, at a sampling rate of 48,000 Hz.

4.2.2. Experiment 4.1: Calibration procedure.

For each of the three possible stimuli for which intensity JNDs were obtained (35 dB, 55 dB, 75 dB), an adaptive three-down one-up, two-interval forced-choice procedure was employed to estimate the point of 79.4% correct identification (Levitt, 1971). At the start of the adaptive sequence, the size of the intensity difference was set to 8 decibels (dB). Three consecutive correct responses in trials resulted in a reduction in the size of the intensity difference and one incorrect response resulted in an increase. Following a reversal (an increase in intensity difference following a decrease, or *vice versa*), the step size (starting value of 4 dB) was divided by two. Minimum step size was limited to 0.1 dB. After 20 reversals, the estimated JND was taken as the arithmetic mean of the last 10 reversals. The three runs, corresponding to the three stimuli, were conducted in a block lasting no longer than 20 minutes. Within-block run order was random. Each listener completed one block. The slowly converging adaptive procedure was designed to take around 5 minutes per run, allowing sufficient time for long-term adaptation to converge prior to the ultimate estimate of JND being acquired.

4.2.3. Experiment 4.1: Probabilistic procedure.

In the second, probabilistic, stage, listeners were presented with a block of 1000 individually calibrated stimuli (35, 55, 75 dB), where the intensity difference for each stimulus was the estimated JND obtained from the previous calibration procedure. Unbeknownst to the listeners, the 1000 trials were divided into 100-trial epochs. Within an epoch, each trial was selected from the three possible stimuli according to *a-priori* distributions (Fig. 4.1a), where one stimulus (i.e. a pair of noise bursts) was selected at 80% probability and the other two at 10% probability each (Fig. 4.1b). Over an epoch, this generated three possible distributions for the three possible stimuli: A: [10%:10%:80%], B: [10%:80%:10%] and C: [80%:10%:10%] (as depicted in Fig. 4.2a-c/4.3a-c respectively). 10 consecutive epochs were presented in a block. For each epoch, one of the three distributions was chosen with equal likelihood. This was performed in the following manner: three of each kind (A,B,C) were included plus one (of A/B/C) at random, for a total of 10 epochs. The epoch order was randomly shuffled and any permutations in which two sequential distributions of the same kind occurred (e.g., A***CC***BABACBC) were rejected and reshuffled. Each listener completed one block (of 10 epochs), taking around 30 minutes.

4.2.4. Experiment 4.2

The calibration and probabilistic procedures of experiment 4.1 were replicated for experiment 4.2, where the three possible stimuli had ISIs of 350, 700 or 1050 and stimulus level was fixed at 55 dB SPL.

4.2.5. Experiment 4.1, 4.2: Participants

Nine normal-hearing listeners participated (first experiment mean and standard deviation: 29 ± 4 years, 1 female, second experiment mean and standard deviation: 30 ± 5 years, 2 female). Seven of the listeners in experiment 4.2 also participated in experiment 4.1.

4.3. Results

In each experiment, the three possible *a-priori* distributions provide three contexts within which trials of each stimulus can be assessed. For each listener, continuous percent-correct functions, for each stimulus in each context (3x3), were calculated using a 40-trial selective (rectangular) sliding-window collapsed across epochs (N=10). These functions are plotted as mean ± standard error in the mean (SEM). Each function was tested for significant overall fluctuations in performance (*Friedman Rank Sum test,* where χ^2 is given as a measure of effect size), and for fluctuations in the difference in performance between each pair of stimuli within a given context (*Friedman Rank Sum test on the derivative*). The latter derivative test identifies fluctuations that indicate selectivity and/or prioritization between stimuli. The *Durbin-Watson test* statistic (Durbin and Watson, 1950, 1951, 1971) across all data of both experiments was close to 2 (mean: 1.93, SD: ± 0.39) indicating that correction for serial correlation was not required. In addition, we conducted permutation tests on each χ^2 statistic. For the data of each function that was tested, trial order was randomly shuffled for each listener and the respective χ^2 statistic was computed using the Friedman test. This process was repeated n = 1,000 times, in each case, and the number of permutations that resulted in χ^2 values that were equal or larger (than that of the un-permuted Friedman test) were counted (count = c) to provide an estimated P-value ($P_{est} = c+1/n+1$) which we report in place of the P-value computed in the Friedman test. Correlations, computed on the grand-average performance functions, are given with 95% confidence intervals (CI). Statistical tests that did not reach significance are denoted as not significant (*N.S.*).

4.3.1. Experiment 4.1: Results

From the calibration procedure, the mean JND (± SD) were: 2.4 ± 1.1 dB, 2.5 ± 1.1 dB and 2.6 ± 1.6 dB for the 35, 55 and 75 dB stimuli respectively. Figure 4.2 plots mean performance (±SEM) for the three calibrated stimuli (35, 55, 75 dB) within each possible context. Figure 4.2a plots performance in the three

possible stimuli when the 35 dB stimulus is selected at 80% probability. Figures 4.2b and 4.2c plot the same for the three possible stimuli when the 55- and 75 dB stimuli respectively are selected at 80% probability.

For all three high-probability stimuli, performance shows little evidence of significant fluctuation (*N.S., Friedman Rank Sum test*), suggesting that adaptation, if it occurs, is rapid for common sounds (Dean *et al.*, 2005; 2008). Indeed, it should be noted that our paradigm (including the low-pass effects of the 40-trial sliding integration window) practically precludes capture of such adaptation. In Fig. 4.2a, performance for low-probability stimuli (55 and 75 dB) is relatively steady (but lower) until about halfway through the epochs when performance for the two stimuli starts to diverge, with performance for the 55 dB stimulus declining (*N.S., Friedman Rank Sum test*), and for the 75 dB stimulus increasing ($\chi^2(59)$ 119.2, $P < 0.05$, *Permutation Test*) until it surpasses even that for the 35 dB (HPR) stimulus. Over the whole epoch, performance for low-probability stimuli at 55 and 75 dB is inversely correlated (r = -0.88, $P < 0.01$, 95% CI [-0.79, -0.92]) and diverges around the 'breakpoint' at ~30 trials: performance deteriorates for the 55 dB stimulus (r = -0.79, $P < 0.01$, 95% CI [-0.67, -0.87]) while performance for the 75 dB stimulus improves (r = 0.91, $P < 0.01$, 95% CI [0.85, 0.94]). Further evidence of selectivity/prioritization is seen by examining the derivatives. Performance for the 75 dB stimulus shows some weak evidence of changing relative to that for the 55 dB stimulus ($\chi^2(59)$ 94.7, *Friedman Rank Sum test on the derivative between the stimuli; $P < 0.1$, Permutation Test*) and relative to the 35 dB (HPR) stimulus ($\chi^2(59)$ 140.6, *Friedman Rank Sum test on the derivative of performance between the stimuli; $P < 0.05$, Permutation Test*).

In Fig. 4.2b, when the HPR corresponds to the 55 dB stimulus, performance shows little evidence of significant fluctuation for any stimulus (*N.S., Friedman Rank Sum test*). In Fig. 4.2c, when the HPR corresponds to the 75 dB stimulus, performance for the low-probability stimuli is similar to that of Fig. 4.2a. Performance for the (low probability) 35- and 55 dB stimuli is inversely correlated (r = -0.47, P = 0.02, 95% CI [-0.24, -0.65]) and splits after the breakpoint. Performance for the 55 dB stimulus deteriorates

(*N.S.*, *Friedman Rank Sum test*), while performance for the 35 dB stimulus improves (*N.S.*, *Friedman Rank Sum test*) gradually ($r = 0.63$, $P < 0.01$, 95% CI [0.45, 0.76]).

These data are consistent with the existence of an adaptive mechanism that tracks the statistics of the stimulus, refining predictions over timescales of around one minute. For the 'most odd' stimulus, when the HPR corresponds to the 35- and 75 dB stimuli performance improves (at the expense of the alternate low-probability stimulus) after around a minute, suggesting the slow build-up of oddball selectivity. When the HPR corresponds to the 55 dB stimulus (Fig. 4.2b), however, neither of the other two stimuli is 'more odd' than the other (and the 55 dB stimulus lies at the mean of the whole distribution), and overall performance is similar for all stimuli. This means that statistical evidence for stimulus prioritization is relatively weak.

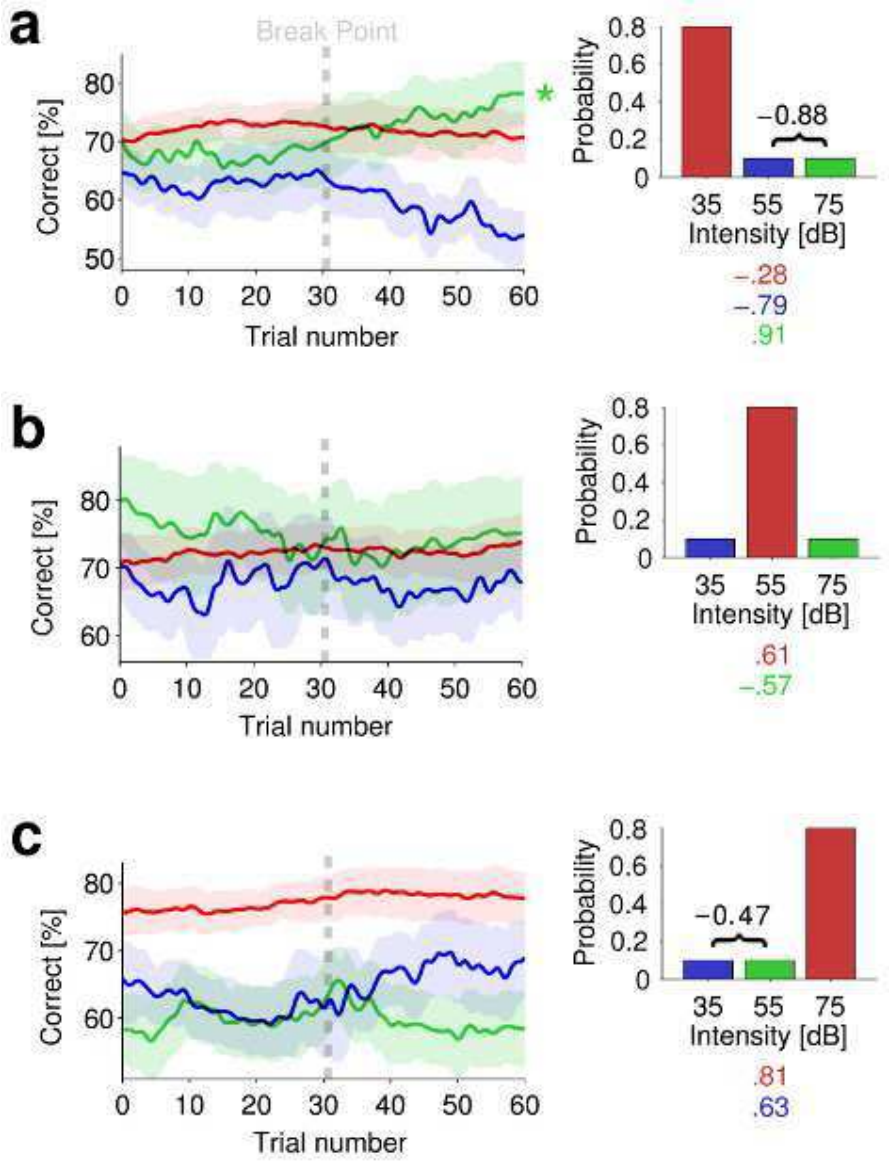

Figure 4.2: Intensity discrimination accuracy changed over time for different intensity statistics. Plot of mean (±SEM) accuracy for each stimulus (35, 55 or 75 dB) in different epochs. The colour coded correlations (*r*-values shown below each respective cartoon distribution) capture significant overall trends with time. For each epoch, correlations were also computed between the two respective low (10%) probability functions and are *r*-values are noted (in black) with bracket. Correlation values are only given where significant ($P < 0.01$). **a** plots performance in epochs where 35 dB trials occur with 80% probability. **b** plots performance in epochs where 55 dB trials occur with 80% probability. **c** plots performance in epochs where 75 dB trials occur with 80% probability. Asterisks denote significant fluctuations in performance ($P < 0.01$, *Friedman Rank Sum test*). Each trial corresponds to approximately 2 seconds (mean trial time across both experiments: 2 seconds, SD: ± 0.3).

4.3.2. Experiment 4.2: Results

From the calibration procedure, the mean JND (± SD) were 2.4 ± 0.9 dB, 2.3 ± 0.9 dB and 2.1 ± 0.4 dB, for the 350, 700 and 1050 ms stimuli respectively. Figure 4.3 plots mean performance (±SEM) for the three calibrated stimuli (350, 700, 1050 ms) within each possible context. Fig. 4.3a plots performance for the three possible stimuli when the 350 ms stimulus is selected at 80% probability. Figures 4.3b and 4.3c plot the same for the three possible stimuli when the 700- and 1050 ms stimuli respectively are selected at 80% probability.

Again, for all three high-probability stimuli, performance shows little evidence of significant fluctuation (*N.S., Friedman Rank Sum test*), suggesting that adaptation, if it occurs, is rapid for common sounds. In Fig. 4.3a, performance for low-probability stimuli (700 and 1050 ms) is relatively steady until about halfway through the epoch when the two functions diverge abruptly, with performance for the 700 ms stimulus declining (χ^2(59) 134.6, *Friedman Rank Sum test; P < 0.01, Permutation Test*), and weak evidence that the 1050 ms stimulus is increasing (χ^2(59) 84.6, *Friedman Rank Sum test; P < 0.2, Permutation Test*) until it surpasses that for the 350 ms (HPR) stimulus. Over the whole epoch, mean performance for low-probability stimuli at 700 and 1050 ms is inversely correlated (r = -0.8, P < 0.01, 95% CI [-0.7, -0.88]) and diverges around the 'breakpoint' around 30 trials. The derivative provides further evidence of this selectivity/prioritization. Performance for the 1050 ms stimulus changes with respect to that for the 350 ms stimulus (χ^2(59) 176.5, *Friedman Rank Sum test on the derivative of performance between the stimuli; P < 0.01, Permutation Test*).

In Fig. 4.3b, when the HPR corresponds to the 700 ms stimulus, performance for the low-probability stimuli (350 and 1050 ms) is positively correlated (r = 0.73, P < 0.01, 95% CI [0.58, 0.83]). It deteriorates early and then rises around a similar breakpoint to that observed in the other data. The fluctuations in performance only reach relatively weak significance for the 350 ms stimulus (χ^2(59) 101.4, *Friedman Rank Sum test; P < 0.1, Permutation Test*), offering some weak evidence of oddball effects, but are approximately paral- leled for the (correlated) 1050 ms stimulus indicating little evidence of prioritization/selectivity.

In Fig. 4.3c, when the HPR corresponds to the 1050 ms stimulus, performance for the low-probability

stimuli is again inversely correlated (r = -0.94, P < 0.01, 95% CI [-0.9, -0.96]); For the 700 ms stimulus performance deteriorates (χ^2(59) 102.5, *Friedman Rank Sum test; P < 0.05, Permutation Test*) gradually (r = -0.97, P < 0.01, 95% CI [-0.94, -0.98]), whilst there is weak evidence that performance for the 350 ms stimulus improves (χ^2(59) 82.9, *Friedman Rank Sum test; P < 0.1, Permutation Test*) with a similar gradient (r = 0.96, P < 0.01, 95% CI [0.94, 0.98]) and surpasses performance for the HPR (1050 ms) stimulus. Again, the derivatives provide further evidence of selectivity/prioritization; Performance for the 350 ms stimulus changes with respect to that for the 700 ms stimulus (χ^2(59) 135.7, *Friedman Rank Sum test on the derivative of performance between the stimuli; P < 0.01, Permutation Test*) and with respect to that for the 1050 ms stimulus (χ^2(59) 124, *Friedman Rank Sum test on the derivative of performance between the stimuli; P < 0.05, Permutation Test*). Also, there is weak evidence that performance for the 700 ms stimulus changes with respect to that for the 1050 ms stimulus (χ^2(59) 105.8, *Friedman Rank Sum test on the derivative of performance between the stimuli; P < 0.1, Permutation Test*).

Consistent with the experiment 4.1 assessing stimuli of different intensities, the inverse correlation of performance in low-probability stimuli is only evident when the high-probability stimulus is presented with either low (350 ms) or high (1050 ms) ISI. Additionally, the low-probability stimulus furthest in ISI from the high-probability stimulus ISI is enhanced after the breakpoint at the expense of the competing low-probability stimulus. This further supports the notion that the auditory system prioritizes resource allocation in favour of those low-probability sounds most different to the high-probability sounds. In both experiments, the selective enhancement of low-probability "oddball" sounds emerges around trial 30, which equates to around 60 seconds into the epoch (mean trial time: 2 seconds, SD: ± 0.3).

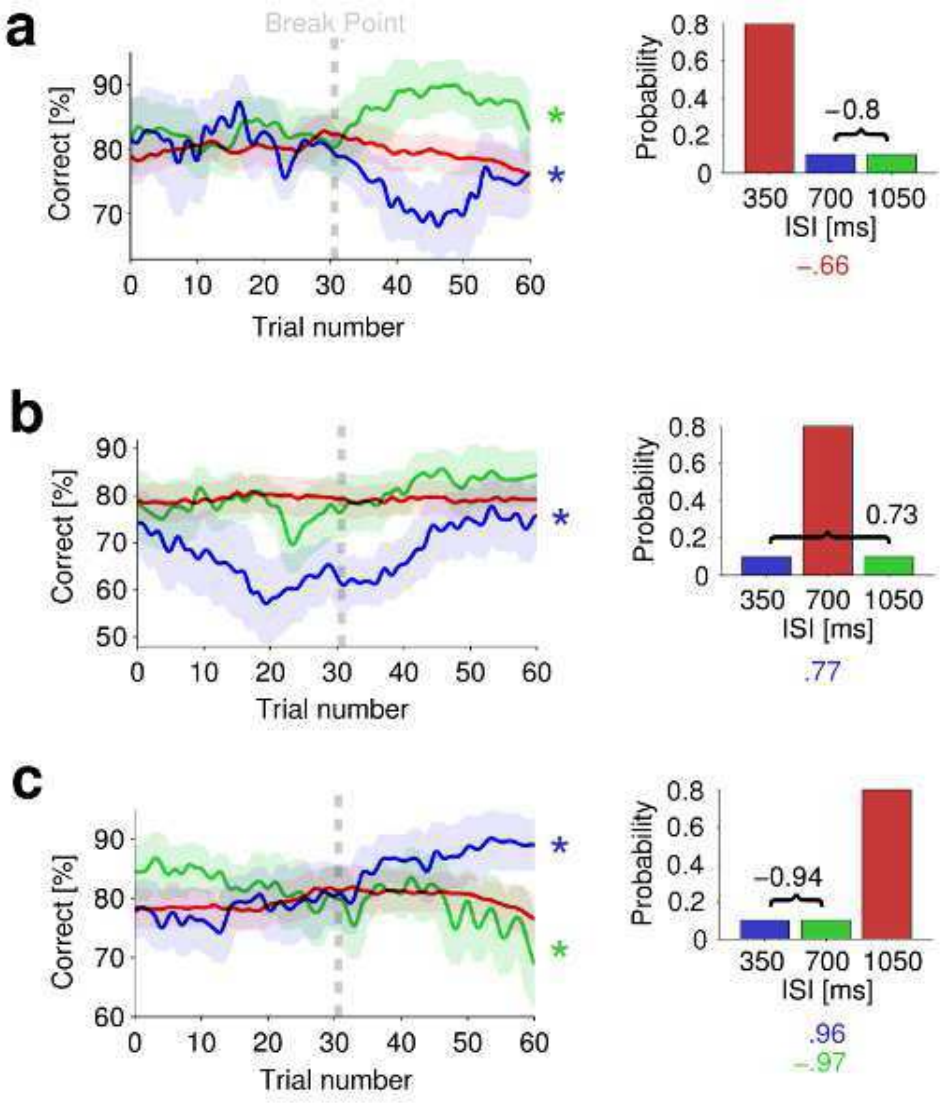

Figure 4.3a/b/c: Accuracy changed over time for different temporal statistics. Panels **a** - **c** plot mean (±SEM) accuracy for each stimulus (ISI of 350, 700 or 1050 ms) in different epochs. The colour coded correlations (*r*-values shown below each respective cartoon distribution) capture significant overall trends with time. For each epoch, correlations were also computed between the two respective low (10%) probability functions and are *r*-values are noted (in black) with bracket. Correlation values are only given where significant ($P < 0.01$). **a** plots performance in epochs where 350 ms trials occur with 80% probability. **b** plots performance in epochs where 700 ms trials occur with 80% probability. **c** plots performance in epochs where 1050 ms trials occur with 80% probability. Asterisks denote significant fluctuations in performance ($P < 0.01$, *Friedman Rank Sum test*). Each trial corresponds to approximately 2 seconds (mean trial time across both experiments 4.1 and 4.2: 2 seconds, SD: ± 0.3).

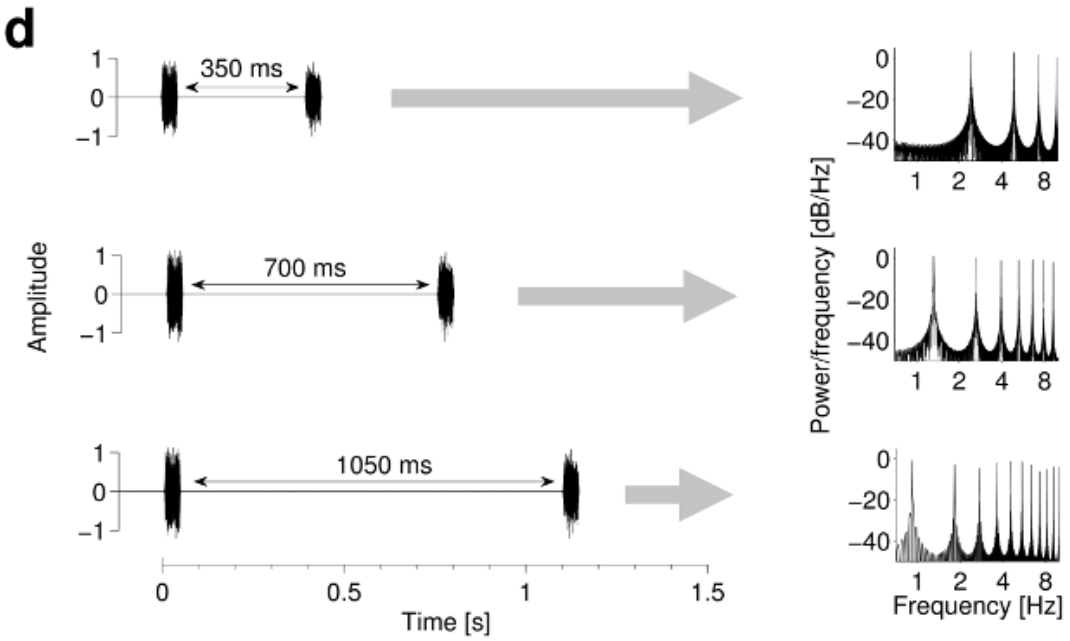

Figure 4.3d: Power spectrum versus ISI. **d** plots example waveforms for pairs of 50-ms noise signals. By varying the interval (ISI) between two sounds, we vary the effective modulation power spectrum. The left side shows the waveforms with different ISIs and the right side of the panel shows the corresponding envelope power-spectrum.

4.4. Discussion

We have demonstrated in human listeners a common strategy for processing the statistical distributions of sounds varying in intensity or timing. Sounds with the most commonly occurring intensities, or presented with the most commonly occurring intervals, are strongly represented throughout. Selective enhancement of novel events then appears to emerge after some time within the high-probability context. Discrimination performance for low-probability sounds that are most unlike the high-probability sounds is enhanced at the expense of discrimination in low-probability sounds that are most like the sounds heard with high probability. It is also striking that discrimination performance in these "oddball" low-probability sounds can surpass that of high-probability sounds (e.g., Fig. 4.3). Note too, that whilst previous reports of sensi-

tivity to "oddball" sounds indicate improved *detection* of these events (e.g., Slabu *et al.*, 2012), here we demonstrate improved *discrimination* for low-probability events.

At a phenomenological level, the adaptation evident in our data is consistent with the concept of perceptual learning (de Souza *et al.*, 2013; Skoe *et al.*, 2013). Perceptual learning is thought to reflect enhancement of perception due to synaptic plasticity (which follows practice) and hence our data may reflect rapid perceptual learning. More generally, the data are consistent with a process wherein listeners construct an internal model of the acoustic input that processes surprising, or "oddball" stimuli. Although there are several potential neural mechanisms that might underpin such adaptation, it is implied that the neural representation of the stimuli changes over time.

4.4.1. Neural mechanisms

Our data are consistent with experimental recordings from small mammals in which firing rates of auditory neurons adapt to the unfolding distributions of sound intensity (Dean *et al.*, 2005, 2008; Watkins and Barbour, 2008; Wen *et al.*, 2009; Rabinowitz *et al.*, 2011; Sadagopan and Wang, 2008; Barbour, 2011; Jaramillo and Zador, 2011; Walker and King, 2011; Ulanovsky *et al.*, 2003, 2004; Nelken, 2004; Perez-Gonzalez *et al.*, 2005; Malmierca *et al.*, 2009; Yaron *et al.*, 2012). This feature of neural coding, which emerges at the level of the primary auditory nerve, improves coding (discrimination) of the most-likely occurring intensities in a distribution of sounds (Dean *et al.*, 2008). As a population, midbrain neurons also show the capacity to accommodate bimodal (with equal probability) distributions of sound intensity (Dean *et al.*, 2005), suggesting the possibility of simultaneous adaptive coding for multiple sounds with different features. At both the midbrain (Dean *et al.*, 2008) and cortical (Ulanovsky *et al.*, 2004; Yaron *et al.*, 2012) levels, neurons demonstrate adaptation time scales on the order of hundreds of milliseconds to tens of seconds. The breakpoint in performance around 60 seconds is relatively close to the time-scale of long-term adaptation reported in these studies. This time scale is also consistent with the results of Chap-

ters 2 and 3 using slowly ramped intensity increments and with brainstem-mediated 'rapid learning' (Skoe *et al.*, 2013), suggesting a common role of long-term adaptation in humans. Ulanovsky *et al.*'s (2004) study in cats also demonstrated that cortical neurons adapt more quickly to high-probability sounds than to low-probability sounds, and that multiple timescales of 'stimulus specific' adaptation occurred concurrently. These multiple timescales appear consistent with the features of our behavioural data.

The adaptation to temporal statistics implicit in our data is less straightforward to explain, but nevertheless is consistent with recent reports implicating auditory cortex neurons in adaptive coding of temporal intervals (Jaramillo and Zador, 2011). In both cases, the timing intervals may be considered in terms of (low) modulation rates. Emerging evidence suggests auditory cortex maintains a bank of independent cortical modulation filters (CMFs), each tuned to different (low) modulation rates (Xiang *et al.*, 2013). CMFs have been implicated in speech processing (Ding and Simon, 2013) and the detection of intensity changes in Chapters 2 and 3. Contrast gain adaptation has been demonstrated in cortical neurons, whereby functions describing neuronal firing rate versus sound level show gain adjustments to best match the intensity variance of the stimulus (Rabinowitz *et al.*, 2011). Combining these two cortical processing features, by assuming that contrast and modulation processing occurs by common means, a plausible explanation for adaptation to time intervals lies in the specificity of adaptation to particular CMFs. Our temporal stimuli can be considered in terms of the statistical manipulation of modulation energy (see Fig. 4.3d) with respect to the rate at which energy is modulated. As shown in Fig. 4.3d, the ISIs of 350, 700 and 1050 ms produce energy in the envelope modulations with fundamental frequencies of around 3, 2, and 1 Hz respectively, and would, therefore, maximally excite different modulation filters. The power-spectra in Fig. 4.3d also demonstrate that the almost instantaneous envelopes generate steadily decreasing modulation energy in harmonics of the fundamental. Hence, it may be that rate-selectivity of CMFs, as proxy selectors of ISI, combined with independent CMF (contrast) adaptation, underlies the adaptive coding of temporal intervals.

Selective adaptation to oddball sounds probably involves some form of interaction between adaptive effects (Ulanovsky *et al.*, 2004; Dean *et al.*, 2008; Yaron *et al.*, 2012) and neural tuning widths on sensory continua (see O'Connell *et al.*, 2011). However, building a detailed biophysical model of this phenomenon is challenging given the paucity of relevant physiological data, and the vast range of possible circuits. Simpson *et al.*, (2014) described a phenomenological model which tended to support the idea of adaptation mediated through a sideband inhibitory influence.

4.4.2. Attention

Our listeners were instructed to attend each and every trial, and confirmed (post-test) that they made every effort to do so. The necessary attention span (around 30 minutes on average) should not tax an average adult. It might be argued that listeners' attention was captured by, or directed to the "oddball" stimulus, and that top-down processing (e.g., of salience) could mediate such "oddball" selectivity. However, it is also plausible that the well-established low-level adaptive substrates can explain the data, and provide, even, an explanation of the nature and substrates of attention itself. This would render attention deterministic, making it an involuntary statistical consequence of adaptive processing. In this scenario, 'auditory boredom' would also be a predictable and involuntary consequence of the adaptive processing. Attention has featured prominently in investigations of 'cocktail party listening'. Cortical entrainment (synchronization of neuronal duty-cycle with the envelope of the stimulus) has been suggested as one low-level substrate (Lakatos *et al.*, 2013; Ding and Simon, 2013; Zion Golumbic *et al.*, 2013). And even if entrainment is not a substrate, it is associated with and mediated by attention. Auditory neurons appear to exist in a state of perpetual oscillation, between excitatory and refractory states, known as the duty cycle (Lakatos *et al.*, 2013). Entrainment of the neuronal duty cycle to a common stimulus modulation occurs when the refractory period is brought forward in time by excitation of the neuron (also referred to as phase-reset). Therefore, low-level adaptive processes described earlier are inherently implicated in the process of en-

trainment. Extrapolating further, the suggested adaptive-statistical filtering would directly mediate entrainment and hence would mediate the putative substrate of attention.

The sensitivity to "oddball" events demonstrated here might prove useful in exploiting the structural statistics of speech and perhaps even music. Such processing could facilitate the extraction of statistically salient signals from within predictable noise (such as multi-talker babble, for example), and may even underpin higher-level statistical percepts (e.g., McDermott and Simoncelli, 2011; McDermott *et al.*, 2013). Furthermore, if such adaptive coding is a fundamental, low-level feature of the auditory system, it may be that prosody, melody and even the very structure of language and music have evolved to exploit such adaptive coding.

4.5. Chapter Summary

In this chapter we have provided direct evidence of adaptation in human auditory perception which combines the argumentation of Chapter 2 and the selectivity described in Chapter 3. We have also made the case that adaptation serves to enhance auditory representation of "oddball" sounds and have discussed some of the immediate implications for auditory perception. We have introduced a novel paradigm for studying adaptation in perception that may be applied in many conceivable permutations to further probe the interaction between selectivity and adaptation. In particular, it remains to be seen whether the same selective adaptation applies in the tonotopic (frequency) axis.

Chapter 5: General Summary

The main objectives of this thesis were to characterise selectivity and adaptation in the human auditory system, and through this characterisation to provide some evidence of adaptation in human auditory perception. Novel methods were developed and data acquired that meets these objectives. In this final chapter we document the contributions of this thesis, including novel methods and findings, and discuss these contributions in the context of the wider literature. This leads to discussion of possible directions of future research.

1.5. Contributions to Knowledge

The main contributions to the body of knowledge made in this thesis include data on selectivity for modulation rate (Chapters 2, 3, 4) and on adaptation in human auditory perception (Chapters 4). These data have resulted in the development of several testable hypotheses about auditory form and function. This thesis has also yielded novel psychophysical methods (Chapters 3 and 4) and a computational model (Chapter 2) that embody implicit hypotheses about the mechanistic nature of the auditory system.

1.5.1. Main Findings

The case for adaptation in human auditory perception has been set out along two lines. In chapter 2, psychophysical data from as far back as 1928 was accounted for by a central excitation pattern model featuring adaptation to intensity. The adaptation parameters estimated by numerical optimisation of the model are consistent with the observations of in-vivo adaptation. It was argued that this suggests that auditory intensity discrimination is limited by central auditory processing and maintained by adaptive processes. In chapter 4, psychophysical data was presented which demonstrated that listeners' auditory acuity changed over time in response to the statistics of the stimuli. This data provided evidence of a general adaptation strategy for both intensity and temporal statistics that is broadly consistent with the adaptation observed in-vivo and provides the first evidence of enhancement of human auditory perception through adaptation. Therefore we have generalised, to human auditory perception, the adaptation by auditory neurons to sound statistics reported in neurophysiological studies involving small mammals (Dean *et al.*, 2005, 2008; Watkins and Barbour, 2008; Wen *et al.*, 2009; Rabinowitz *et al.*, 2011; Sadagopan and Wang, 2008; Barbour, 2011; Jaramillo and Zador, 2011; Walker and King, 2011; Ulanovsky *et al.*, 2003, 2004; Nelken, 2004; Perez-Gonzalez *et al.*, 2005; Malmierca *et al.*, 2009; Yaron *et al.*, 2012). We have also provided some insight into the adaptive representation of common and rare sounds (Dean *et al.*, 2005, 2008; Watkins and Barbour, 2008; Wen *et al.*, 2009; Rabinowitz *et al.*, 2011; Sadagopan and Wang,

2008; Barbour, 2011; Jaramillo and Zador, 2011; Walker and King, 2011; Ulanovsky *et al.*, 2003, 2004; Nelken, 2004; Perez-Gonzalez *et al.*, 2005; Malmierca *et al.*, 2009; Yaron *et al.*, 2012).

Knowledge of the selectivity of the human auditory brain has also been extended. In chapter 2, selectivity for modulation rate was used to generalise the central auditory model and the human auditory system was shown to be insensitive to very fast and very slow modulations. In chapter 3, selectivity for modulation rate was shown to be carrier frequency dependent and it was hypothesised that the human auditory cortex features a tonotopically arranged modulation filter bank. In chapter 4, selectivity for both intensity and modulation rate was demonstrated.

5.1.2. Hypotheses

We have developed several explicit hypotheses about the auditory system. In chapter 2, it was hypothesised that central adaptation might play a critical role in human auditory discrimination. The case was made by modelling data for long-term signals which were argued to provide conditions where adaptation should have converged sufficiently that the time constants of adaptation could be neglected. The results of Chapter 2 tend to support the hypothesis that central adaptation affected intensity discrimination.

In chapter 3 it was hypothesised that peripheral (cochlear) spread of excitation could affect evidence of orthogonality of tonotopic and periodotopic axes in cortex. The data of Chapter 3 do not support the neuroimaging findings (Baumann *et al.*, 2011; Barton *et al.*, 2012) of orthogonality of tonotopic and periodotopic axes. Indeed, the spread-of-modulation hypothesis might predict that the high sound pressure levels employed in those studies could have produced sufficient peripheral spread of excitation such that any tonotopic selectivity might have been obliterated, leaving only the appearance of orthogonality. The argument for 'spread of modulation' given in Chapter 3 also has implications for the nonlinear CMF interactions described in Xiang *et al.* (2013). In particular, it remains to be seen whether these interactions

are level independent, or whether they are enhanced by increase in level (as would be predicted by the 'spread of modulation' idea of Chapter 3).

In Chapter 4 the hypothesis of Chapter 2 was extended. It was hypothesized that intensity discrimination could be affected by the statistics of the stimuli. This hypothesis was supported by the data. It was further hypothesised that sideband inhibitory networks could cause selective adaptation to rare and unusual sounds. This led to speculation on the possible roles of such statistical processing in speech and music audition.

In general, we hypothesised that psychophysical methods could provide acute data that could potentially reveal features of auditory perception unavailable to current neuroimaging methods. The findings of Chapter 3 appear to bear this out in relation to the neuroimaging studies mentioned above. The findings of Chapter 4 also suggest that the method might be sensitive enough to yield further insights that may be beyond the reach of current neuroimaging methods.

1.6. General Discussion

1.6.1. Object-based representation

The selectivity and adaptation characterised in this thesis has implications for the processes responsible for auditory object formation and for the top-down processes involved. By extending knowledge of the selectivity responsible for feature-based representation in the auditory system, we provide implications for object-based representations that appear an essential part of perception. The emergence of object-based representations in auditory cortex (Mesgarani and Chang, 2012; Pasley *et al.*, 2012; Ding and Simon, 2012, 2013; Shamma *et al.*, 2011; Teki *et al.*, 2013) suggests that the adaptation and selectivity described in Chapters 3 and 4 might have direct impact on cortical object representation. In particular, it has been suggested that sound features sharing a common temporal envelope are fused in the auditory cortex

(Shamma *et al.*, 2011; Teki *et al.*, 2013; see Bregman, 1990). Therefore, where selectivity and adaptation affect the feature-based representation they must also indirectly affect the recombination.

For example, the results of Chapter 3 might imply that, if the periodotopic map is not orthogonal to the tonotopic map, some degree of difficulty with respect to object formation (i.e., for competing objects) might be expected in situations where tonotopic/periodotopic channels interact. Also, the results of Chapter 4 might suggest that the representation of rare and unusual auditory objects might be enhanced, potentially improving the ability of temporally-coherent rare or unusual objects to be extracted from competing sounds or background noise, or potentially providing a statistical filter to remove auditory objects and features that are not salient (see Ding and Simon, 2012; 2013).

1.6.2. Speech processing

Chapter 3 made the case for the human auditory system being optimised for speech processing, demonstrating human auditory selectivity for temporal modulations with rates similar to those of human speech. This chapter provided evidence that carrier frequency and modulation rate are not independent parameters and it was suggested that this might provide a good speech processing strategy. In chapter 4, this selectivity was combined with adaptation, suggesting general mechanisms which may underpin speech processing and selectivity. Chapters 3 and 4 were discussed in the contexts of cortical speech processing and some interesting implications with respect to the neural correlates of attention were highlighted. This work may have implications for hearing-aids and/or cochlear implants.

1.6.3. Generalisation and future work

The computational model of Chapter 2 remains crude and might be extended by the use of a central auditory modulation filter bank such as that employed by Dau *et al.* (1997a/b). However, more data is required for this purpose as little is yet known of the tuning of human CMFs. This model would also be

further enhanced by the inclusion of suitable nonlinearities that would produce intermodulation interactions as characterised by Xiang *et al.* (2013), which could further aid in the fitting of the model. The model of chapter 2 also accounted for elevated increment detection JNDs in background noise as an emergent product of central adaptation, rather than as a product of peripheral (energetic) masking. Thus, a model featuring nonlinear CMFs would make predictions about JND data in background noise that could be used to validate the model. Furthermore, this model might make predictions regarding the central contribution to estimations of auditory filter characteristics using the notched noise method.

The inverted method of Chapter 3 might be useful in quantifying further aspects of the level- and frequency-dependent coding of the auditory system. In particular, the method might be applied to more complex signals such as narrowband noise and might be extended to examine the possible masking effects of background noise. Furthermore, this method and the findings of this thesis appear to have implications for auditory filter characterisation using notched noise methods (Glasberg and Moore, 1990). In particular, filter bandwidths estimated using a fixed modulation, as a function of probe frequency, might be confounded by the potential tonotopic gradient (Chapter 3) in modulation filter tuning. Further confounds might include the possibility of central adaptation caused by the notched-noise masker leading to the appearance of elevated thresholds (typically interpreted as broadening peripheral auditory filters).

The probabilistic method of Chapter 4 might be extended in various ways, including permutations on the discrimination task and stimuli. Furthermore, the probabilistic design might be adjusted to provide arbitrary stimulus distributions so as to further probe the statistical processing of the auditory system. For example, we applied a 10% probability of occurrence for 'rare' sounds, but it would be useful to know how this arbitrary low-probability affects the selectivity demonstrated in the data of Chapter 4.

More generally, the listening contexts and stimuli of the present paradigms are artificial. We have used tones and noise stimuli, with artificial presentation statistics, and presented over headphones in isolation. We have asked listeners to judge subtle, arbitrary intensity changes over blocks of repeated trials. Therefore, our paradigms have little in common with listening scenarios in the real world. Furthermore,

attention is known to mediate/moderate auditory perception (Mesgarani and Chang, 2012; Lakatos *et al.*, 2013; Ding and Simon, 2013; Zion Golumbic *et al.*, 2013). While we may assume that our listeners were attending to the stimuli, attention was not explicitly controlled in our paradigms. Therefore, it remains to be seen what generalisation of the principles demonstrated here might be seen in real world listening scenarios.

The selectivity and adaptation observed in the data presented in this thesis has been discussed in the literature contexts of both in-vivo electrophysiology and neuroimaging. The literature tends to support an interpretation of the data as characterising central neural processing. The adaptation to intensity might be localised to any stage of the auditory pathway but the temporal processing is likely localised to auditory cortex. Future work might involve neuroimaging and electrophysiology to establish the location and/or function of such processing.

Appendix A: **Time-Varying Loudness Model**

This appendix provides a condensed overview of the excitation pattern loudness model of Moore, Glasberg and Baer (1997; Glasberg and Moore, 2002). The various components of this model have been separately described in the well-known publications of Patterson *et al.* (1982), Moore (1995), Moore *et al.* (1997) and Glasberg and Moore (2002).

A.1. Introduction to Loudness Modelling

The loudness model of Moore, Glasberg and Baer (1997), later extended to include time-varying sounds by Glasberg and Moore (2002), has seen a long and fragmented development over a period of more than twenty years (Patterson *et al.*, 1982; Moore, 1995; Moore *et al.*, 1997; Glasberg and Moore, 2002), from the rounded exponential 'roex' filter defined by Patterson *et al.* in 1982 to the time-varying model of Glasberg and Moore in 2002.

A.2. The Excitation Pattern Model

Sound pressure waves pass through the outer and middle ear and enter the inner ear (cochlea), causing the basilar membrane to resonate at a given location along its length that depends on the frequency of the exciting sound. Resonance of the basilar membrane causes the displacement (shearing) of inner hair cells arranged along the basilar membrane. The extent of the shearing of each hair cell is then converted into a pulsed electrical signal by neurons attached to the hair cell. This neural representation of the pattern of resonance on the basilar membrane, caused by a given sound, is known as its excitation pattern. The electrical signal, produced by the population of neurons, is sent up the auditory nerve to the brain. This gives rise to the concept of the auditory filter, which specifies the shape of the excitation pattern for a sound of given frequency and level. To make things more complicated, there are also outer hair cells which contribute little in the way of signals sent to the brain, but which are motile and act in synchrony with the corresponding inner hair cell to amplify the basilar membrane excitation at low levels. This produces the effect of changing the shape of the auditory filter with level.

The excitation pattern model of loudness is based on the assumption that the total area of excitation along the length of the basilar membrane is integrated (on the auditory nerve) in the calculation of loudness. However, consistent with what is known of the cochlear amplifier and of neural transduction, the excitation is locally compressed before being integrated. The role of the auditory filter is to provide a summation of energy at local frequencies, where 'local' means frequencies within the auditory filter, and subsequent compression of the sum energy at the output of the auditory filter. The output of the auditory filter is known

as specific loudness. Specific loudness can also be thought of as 'loudness per filter'. This mechanism results in energetic (simultaneous) masking because the specific loudness resulting from the compressed sum of excitation at two nearby locations within a single auditory filter contributes less to overall loudness than a linear sum of the specific loudness resulting from the same excitation at two distant locations within two separate auditory filters.

A.2.1. Definitions

Loudness is the perceived intensity (I) of a sound. Intensity is defined in terms of sound pressure (x) squared;

$$I = kx^2 \tag{A.1}$$

where k is a constant that represents the specific acoustic impedance of air. To calculate I for a sound described by $x(t)$, from time t=0 to T, Eq. A.1 is then integrated over time;

$$I = \frac{1}{T} k \int_0^T x^2(t) dt \tag{A.2}$$

Intensity may then be defined in terms of a ratio, with respect to a reference (e.g., I_{ref} = 20 μW/cm^2), in decibels. This is known as the intensity *level* (L_I);

$$L_I = 10 \log_{10} \left(\frac{I}{I_{ref}} \right) \tag{A.3}$$

The use of intensity levels allows us to drop the absolute reference, and with it the *k* parameter, which simplifies the following notation.

A.3. Equivalent Rectangular Bandwidth

The equivalent rectangular bandwidth (ERB) gives a measure of auditory filter width, such that the sum excitation that falls within any given ERB will be equivalently compressed and result in an equivalent contribution to total loudness. Thus, the ERB provides a mechanism by which compression, masking and loudness are related. The mapping between frequency *f* (Hz) and ERB (Hz) shown in Fig. A.1a is achieved using the following formula (see Moore, 1995):

$$ERB = 24.7(0.00437 f + 1) \tag{A.4}$$

In order to relate ERB to frequency, the ERB number (*n*) for a given centre frequency (f_c) - as shown in Fig. A.1b - can be calculated as (see Moore, 1995):

$$n = 21.4\log_{10}(0.00437 f_c + 1) \tag{A.5}$$

Given frequency bounds defined in terms of centre frequencies between 50 – 15,000 Hz (see Moore et al., 1997), the ERB numbers of the respective upper and lower bounding auditory filters may be calculated and intervening filters specified at arbitrary ERB-scaled intervals. To this end, Eq. A.5 may be rewritten, as follows;

$$f_C = \frac{10^{(n/21.4)} - 1}{0.00437} \tag{A.6}$$

Using Eq. A.5, auditory filters at ERB intervals within the known range of the basilar membrane (50 -

15,000 Hz) may be specified for the later excitation pattern calculation and using Eq. A.6 the centre frequencies may be calculated at ERB-spaced intervals between.

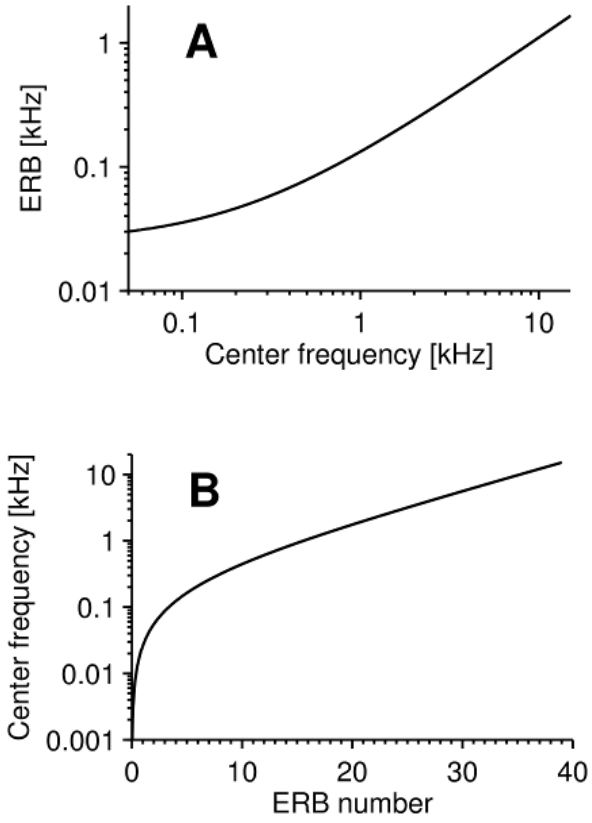

Figure A.1. ERB. A Illustration of Eq. A.4 which relates ERB to centre frequency. **B** Illustration of Eq.s A.5 & A.6 which relates centre frequency to ERB number.

A.4. Model for Steady Sounds

The first stage of the model represents the transformation of sound pressure through the outer and middle ear to the inner ear (cochlea). This transformation is represented by a fixed linear filter with a frequency dependent gain, *y*, as follows;

$$L_{I_1} = L_{I_0} \cdot y$$

(A.7)

where L_{I_0} is the input sound intensity, L_{I_1} is the intensity reaching the inner ear and *y* is the gain of the filter at that frequency. Fig. A.2 provides an illustration of the combined transfer function. Because the loudness model of Moore *et al.* is generally intended for diffuse-field sound, phase information is discarded [see Glasberg and Moore (2002) for discussion].

From this point onwards it is important to note that the input sound signal is defined as an intensity level (Eq. A.3) at a specific frequency, wherever a dB measure is used. Furthermore, excitation is defined in terms of excitation level (L_E) as an intensity ratio with respect to the excitation reference of a 1 kHz sinusoidal signal at 0 dB SPL (presented in the free field and at frontal incidence);

$$L_E = L_{I_1} - E(0_{1kHz}) \tag{A.8}$$

where $E(0_{1kHz})$ is the reference excitation level.

A.4.1. The Rounded Exponential (roex) Filter

The excitation pattern, which represents the basilar membrane response, is calculated using a set of ERB-spaced auditory filters. The auditory filter is based on the rounded-exponential ('roex') form proposed by Patterson *et al.* (1982). The roex filter is defined as;

$$w(g) = (1 + pg)e^{-pg} \tag{A.9}$$

where for a given centre frequency, f_c, the normalized frequency relationship between the f_c and a given frequency, f, (i.e., of the input signal) is given by;

$$g = |f - f_c| / f_c \tag{A.10}$$

where f_c is evaluated for a given ERB number (n) using Eq. A.6. p determines bandwidth and slope of the filter and is defined in relation to the ERB as follows (1995):

$$p = \frac{4 f_c}{ERB} \tag{A.11}$$

Larger values of p lead to more narrowly tuned filters. Thus, given an input at frequency f, $w(g)$ can be used to calculate the attenuation of the input at frequency f within the roex filter at centre frequency f_c.

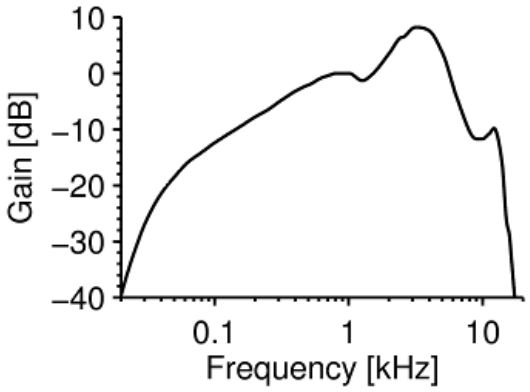

Figure A.2. Illustration of combined outer and middle ear transfer function. Note, zero dB gain at 1 kHz.

Eq. A.9 gives a symmetrical auditory filter [$w(g)$]. However, the auditory filter is known to be asymmetrical and so Eq. A.9 is broken down into two such expressions, the choice of which depends on whether the input frequency (f) is above or below the centre frequency (f_c) for the auditory filter of interest;

$$w(g) = \begin{cases} (1 + p_l g)e^{-p_l g} & for \quad f <= f_c \\ (1 + p_u g)e^{-p_u g} & for \quad f > f_c \end{cases} \tag{A.12}$$

p_l and p_u replace p to represent the parameters for input frequencies (f) below or above the centre frequency (f_c) respectively. This conditional aspect is necessary because although the auditory filter is roughly symmetrical when the excitation level per ERB is around 51 dB (Glasberg and Moore, 1990), the low-frequency 'skirt' of the auditory filter becomes less sharp with increase in level. This excitation level dependent relationship is accommodated in terms of the p_l value as follows;

$$p_l(L_E) = p_l(51) - 0.35\left(p_l(51)/p_l(51_{1kHz})\right)(L_E - 51) \tag{A.13}$$

where $p_l(L_E)$ is the value of p_l for the input excitation level of L_E, in dB, at f, and $p_l(51)$ is the value of p (Eq. A.11) at the centre frequency (f_c) for an input level of 51 dB (i.e., where the filter is symmetrical), and where $p_l(51_{1kHz})$ is the value of p_l for a 51 dB input excitation level at 1 kHz. Figure A.3 provides an illustration of the level dependent roex filter shape for excitation at 1 kHz at levels between 10 and 100 dB in 10 dB intervals.

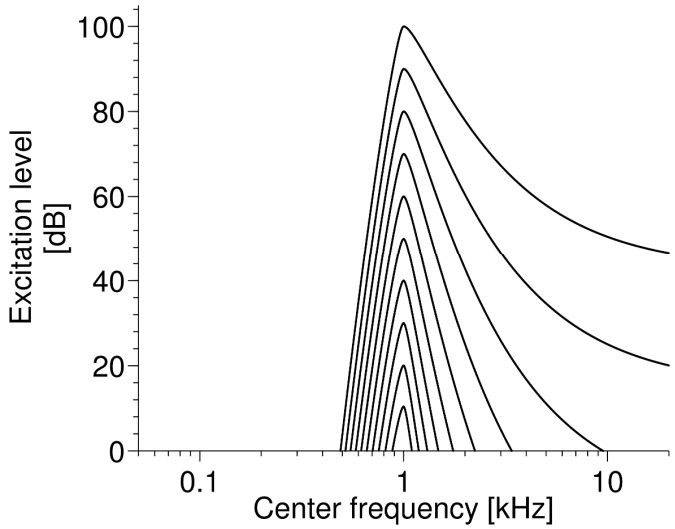

Figure A.3. Illustration of roex filter shapes (Eq. A.12) for excitation levels between 10 and 100 dB

in 10 dB intervals.

A.4.2. The Excitation Pattern

For each ERB number (*n*), the excitation pattern, *E*, is defined as the pattern of outputs from the ERB-spaced auditory filters. For a given frequency, *f*, and for an input excitation level (Eq. A.8), the excitation pattern, *E*, is then defined as:

$$E(n) = w(g(n)) \cdot L_E \tag{A.14}$$

where ERB number (*n*) is related to f_c by Eq. A.6.

A.4.3. Specific Loudness

To reflect the production of neural signals in response to inner hair cell displacement caused by excitation of the basilar membrane, the excitation pattern is transformed from excitation level into specific loudness (loudness per ERB) for the *n*th auditory filter by calculating the specific loudness in each filter according to three possible conditional expressions, which relate to the excitation level as follows in Eq. A.15 (above). Since loudness is later notated as *N*, specific loudness is notated as *N'*, to reflect the later integration (over frequency) of specific loudness to form loudness.

$$N'(n) = \begin{cases} C \cdot \left(\dfrac{2E(n)}{E(n) + T_Q(n)} \right)^{1.5} \cdot \left((G \cdot E(n) + A)^{\alpha} - A^{\alpha} \right) & \text{for } E(n) \le T_Q(n) \\ C \cdot \left((G \cdot E(n) + A)^{\alpha} - A^{\alpha} \right) & \text{for } 10^{10} \ge E(n) > T_Q(n) \\ C \cdot \left(\dfrac{E(n)}{1.04x10^6} \right)^{0.5} & \text{for } E(n) > 10^{10} \end{cases} \tag{A.15}$$

Frequency dependence (denoted with parameter *n*) refers to the *n*th auditory filter. T_Q represents the

threshold excitation in quiet and is frequency dependent as shown in Fig. 4c. The parameter *G* represents low-level gain in the cochlear amplifier, relative to the gain at 500 Hz and above, and is also frequency dependent. Note that this 'G' is not related to the 'G' of chapter 3. For a given centre frequency, f_c, *G* (in dB) is related to T_Q (in dB) with a simple subtraction;

$$G = T_Q(500) - T_Q(f_c) \tag{A.16}$$

The parameter *A* in Eq. A.15 is used to bring the input-output function close to linear around the absolute threshold, and is dependent on the value of *G* as shown in Fig. 4a. The compressive exponent α is dependent on the value of *G* as shown in Fig. 4b. At frequencies below 500 Hz T_Q rises as frequency decreases and the value ranges between 28 dB at 50 Hz and 3.7 dB at 500 Hz. Above 500 Hz T_Q is constant and equal to T_Q at 500 Hz. α is also frequency-dependent and a similar lookup table is employed such that α varies between 0.27 and 0.2, depending on the value of *G*. *C* is a constant which scales the loudness to conform to the sone scale, where the loudness of a 1 kHz tone at 40 dB SPL is 1 sone. *C* is equal to 0.047. Figure A.4d shows the result of Eq. A.16 used to transform excitation at levels between 0 and 120 dB to specific loudness for a 1 kHz signal. Finally, specific loudness is integrated, over the arbitrarily (*dn*) spaced auditory filters, between ERB numbers n_{min} and n_{max}, to produce loudness, *N*;

$$N = \int_{n_{\min}}^{n_{\max}} N'(n)dn \tag{A.17}$$

where n_{min} and n_{max} may be calculated, from centre frequencies of 50 and 15,000 Hz respectively using Eq. A.5. For a complex sound, loudness is calculated from a linear sum of excitation patterns calculated from each input sound component.

A.4.4. Energetic Masking

A formal definition of loudness allows us to derive a formal definition of energetic (simultaneous) masking with respect to two arbitrary excitation patterns for the target, $E_{t,}$ and the masker, E_m. The two excitation patterns may then be used to evaluate the degree of masking by comparing the sum of loudness for each excitation pattern alone [$N(E_m)$ + $N(E_t)$] and the loudness of the linear sum of the two excitation patterns [$N(E_m+E_t)$]. This provides a loudness ratio ($N_{masking}$, in sones);

$$N_{masking} = \frac{N(E_t + E_m)}{N(E_t) + N(E_m)} \tag{A.18}$$

A.5. Model for Time-Varying Sounds

The time-varying model (Glasberg and Moore, 2002) is an extension of the 1997 model for steady (state) sounds. In the earlier model, the sounds are defined in terms of steady sound components, which are then combined within the excitation pattern to produce an overall loudness. In the time-varying model, the excitation pattern is typically calculated, from a time-domain input signal, on an instantaneous sliding-window basis, giving a time-varying excitation pattern.

The time-varying excitation pattern is then resolved into a corresponding time-varying specific loudness function and hence is integrated to form a time-varying intermediate stage known as ‘instantaneous loudness’. Instantaneous loudness is essentially an intensity-like temporal integration of specific loudness over an arbitrarily small time interval. The ‘small’ time interval is typically on the order of 1 ms, which may be considered small with respect to the integration time constants of the auditory system (usually much longer). Thus, instantaneous loudness is calculated as ‘steady loudness’ over a very small time scale.

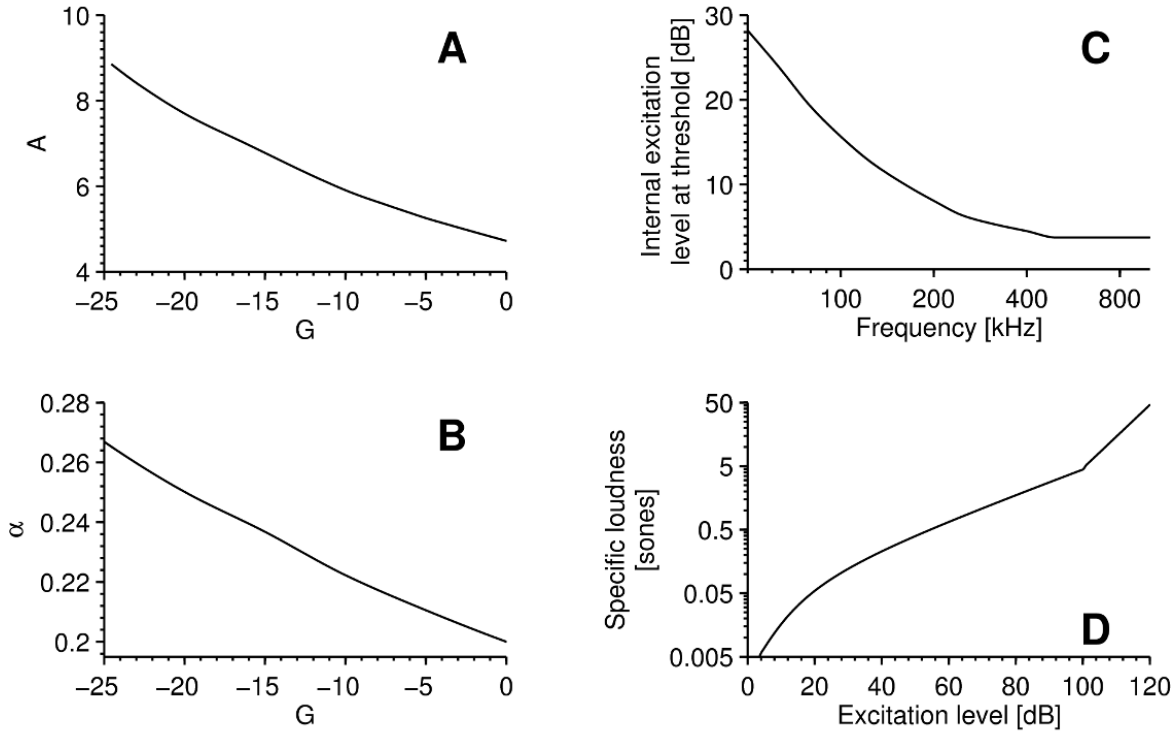

Figure A.4. Illustration of miscellaneous parameters. A Illustrating the relation between parameters A and G. **B** Illustrating the relationship between the parameters α and G. **C** illustrating the internal excitation level at threshold as a function of frequency (showing increased thresholds at low frequencies). **D** Specific loudness as a function of excitation level between 0 and 120 dB at 1 kHz, illustrating the conditional effects of Eq. A.15.

Intensity, for sound of a given integration time (Δt), is then defined in terms of an integral with respect to time (t);

$$I(t) = \frac{1}{\Delta t} k \int_{t}^{t+\Delta t} x^2(t)\,dt \tag{A.19}$$

which may again be resolved into intensity level as in Eq. A.4, and hence excitation level as in Eq. A.8, for substitution into Eq. A.15 to give Eq. A.20.

$$N'(n,t)=\begin{cases} C\cdot\left(\dfrac{2E(n,t)}{E(n,t)+T_Q(n,t)}\right)^{1.5}\cdot\left((G\cdot E(n,t)+A)^{\alpha}-A^{\alpha}\right) & \text{for } E(n,t)\le T_Q(n) \\ C\cdot\left((G\cdot E(n,t)+A)^{\alpha}-A^{\alpha}\right) & \text{for } 10^{10}\ge E(n,t)>T_Q(n) \\ C\cdot\left(\dfrac{E(n,t)}{1.04x10^{6}}\right)^{0.5} & \text{for } E(n,t)>10^{10} \end{cases} \quad \text{(A.20)}$$

Eq. A.17 is then extended to integrate the result of Eq. A.20 with respect to ERB number (n), to produce a time-varying instantaneous loudness [$N(t)$];

`

$$N(t)=\int_{n_{\min}}^{n_{\max}} N'(n,t)dn \quad \text{(A.21)}$$

A.5.1. Temporal Integration

Loudness of brief sounds increases with duration up to a limit of around 200 ms (Munson, 1947). This is known as the temporal integration of loudness. A further phenomenon captured in the time-varying loudness model is forward masking, which has a similar time scale. In order to account for these phenomena, the instantaneous loudness function is smoothed with an exponential sliding window.

To predict the decay of loudness after a sound has ceased, given an initial loudness value (N_0), the decaying value of loudness at time t may be calculated as;

$$N(t)=N_0e^{-t/\tau} \quad \text{(A.22)}$$

where τ is the time constant. This represents the decay of loudness, i.e., forward masking. To predict the accumulation of loudness with duration of a steady (fixed intensity) sound, loudness at time t is calculated as;

$$N(t) = N_{\infty}(1 - e^{-t/\tau}) \tag{A.23}$$

where N_{∞} represents the asymptotic loudness. The values of N_0 and N_{∞} may be calculated in terms of instantaneous loudness for a given signal and used to predict the effects of temporal integration.

In order to provide a time-varying output function, Eq. A.22 is re-arranged in order to relate it to the time step of the model (Δt) and used to calculate a smoothing coefficient (β);

$$\beta = e^{-\Delta t/\tau} \tag{A.24}$$

To smooth the time-varying instantaneous loudness function β is applied to calculate STL (N_{ST}) with respect to instantaneous loudness [$N(t)$];

$$N_{ST}(t) = (1 - \beta_{ST}) \cdot N(t) + \beta_{ST} \cdot N_{ST}(t - \Delta t) \tag{A.25}$$

And to calculate LTL (N_{LT}) with respect to STL;

$$N_{LT}(t) = (1 - \beta_{LT}) \cdot N_{ST}(t) + \beta_{LT} \cdot N_{LT}(t - \Delta t) \tag{A.26}$$

where

$$\tau_{ST} = \begin{cases} 22 & for \quad N(t) > N_{ST}(t-\Delta t) \\ 50 & for \quad N(t) < N_{ST}(t-\Delta t) \end{cases}$$
$$\tau_{LT} = \begin{cases} 100 & for \quad N_{ST}(t) > N_{LT}(t-\Delta t) \\ 2000 & for \quad N_{ST}(t) < N_{LT}(t-\Delta t) \end{cases}$$

(A.27)

The value of τ (and hence β) is conditional such that separate values of τ are assigned depending on whether the function is in the attack or release phase [see Glasberg and Moore (2002)]. As can be seen from the values of τ shown above, convergence is much faster for attack than for release in both cases of STL and LTL. This is intended to reflect disparity in forward and backwards masking.

Finally, Glasberg and Moore (2002) specify that the loudness of brief duration sounds (i.e., gated tones) should be calculated as the peak (maximum) value in the STL time series and that the loudness of continuous sounds (e.g., amplitude modulated tones) should be calculated as the mean (average) of the LTL time series.

A.5.2. Temporal Masking

Eq. A.18 may be extended to provide a time-varying definition of energetic masking, in terms of instantaneous loudness, as follows;

$$L_{masking}(t) = \frac{N\left(E_t(t) + E_m(t)\right)}{N\left(E_t(t)\right) + N\left(E_m(t)\right)}$$

(A.28)

However, the stated purpose of Eq.s A.25 & A.26 is to provide temporal integration (or summation) of loudness at the two respective time scales. This means that forward and backwards masking may not be quantified in terms of Eq. A.28, and are therefore outside the scope of this chapter.

A.6. Appendix A Summary

In this Appendix we have provided a condensed, practical step-by-step description of the excitation pattern loudness model which consolidates descriptions found in the multiple original articles. We have included a brief description of the function of, and rationalisation for, each modelling component.

Appendix B: ***Ethics statement***

For all listening tests described in this thesis, participants were voluntary, unpaid and gave informed verbal consent before the experiment. Participants were free to withdraw at any point. Tests were run on an ad-hoc basis. Written consent was not deemed necessary due to the low (safe) sound pressure levels employed in the test but the consenting volunteers were documented. All experimental protocols (including consent) were approved by the ethics committee of Queen Mary University of London.

Appendix C: ***Statistical methods and assumptions***

Where we present data in terms of mean and confidence intervals in this thesis, the data were checked to ensure that the data were approximately normally distributed and hence it was ensured that the measures given in this thesis are interpretable and representative. Where we employ the Friedman Test in this thesis we reasonably assume that the data are uncorrelated. In Chapter 4, where the data may not be assumed to be entirely uncorrelated, we employ a permutation test that takes into account any inherent correlations.

References

Agus TR, Thorpe ST, Pressnitzer D (2010) Rapid formation of robust auditory memories: insights from noise. Neuron 66, 610–618.

Allen JB, Neely ST (1997) Modeling the relation between the intensity just-noticeable difference and loudness for pure tones and wideband noise. J Acoust Soc Am 102, 3628–3646.

Barbour DL (2011). Intensity-invariant coding in the auditory system. Neurosci Biobehav Rev 35:2064-2072.

Baumann S, Griffiths TD, Sun L, Petkov CI, Thiele A, Rees A (2011) Orthogonal representation of sound dimensions in the primate midbrain. Nat Neurosci 14:423-425.

Barton B, Venezia JH, Saberi K, Hickok G, Brewer AA (2012). Orthogonal acoustic dimensions define auditory field maps in human cortex. Proc Natl Acad Sci USA 109:20738-20743.

Bregman AS (1990). Auditory Scene Analysis: The Perceptual Organization of Sound. (Cambridge, MA: MIT Press).

Brungart DS, Chang PS, Simpson BD, and Wang D (2006) Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. J Acoust Soc Am 120, 4007–4018.

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

Dau T, Kollmeier B, Kohlrausch A (1997a) Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers J Acoust Soc Am 102, 2892–2905.

Dau T, Kollmeier B, Kohlrausch A (1997b) Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration J Acoust Soc Am 102, 2906–2919.

Dayan P, Abbott LF (2001) Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems (MIT Press, Cambridge, Massachusetts, 2001).

de Souza ACS, Yehia HC, Sato M, Callan D (2013). Brain activity underlying auditory perceptual learning during short period training: simultaneous fMRI and EEG recording (2013), BMC Neurosci 14:8.

Dean I, Harper NS, McAlpine D (2005). Neural population coding of sound level adapts to stimulus statistics. Nat Neurosci 8:1684 – 1689.

Dean I, Robinson BL, Harper NS, McAlpine D (2008). Rapid neural adaptation to sound level statistics. J Neurosci 28:6430–6438.

Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci USA 109:11854–11859.

Ding N, Simon JZ (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728-5735.

Durbin J, Watson GS (1950), Testing for Serial Correlation in Least Squares Regression I. Biometrika 37: 409–428.

Durbin J, Watson GS (1951), Testing for Serial Correlation in Least Squares Regression II. Biometrika 38: 159–178.

Durbin J, Watson GS (1971), Testing for Serial Correlation in Least Squares Regression III. Biometrika 58: 1–19.

Durlach NI, Braida LD (1969) Intensity perception. I. Preliminary theory of intensity resolution J Acoust Soc Am 46, 372–383.

Drullman R, Fester JM, Plomp R (1994) Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am 95: 2670-2680.

Evans EF, Palmer AR (1980) Relationship between the dynamic range of cochlear nerve fibres and their spontaneous activity Exp Brain Res 40, 115–118.

Fletcher H, Munson W (1933) Loudness, its definition, measurement, and calculation J Acoust Soc Am 5, 82–108.

Fletcher H, Wegel R (1922) The frequency-sensitivity of normal ears Phys Rev 19, 553–565.

Florentine M, Buus S (1981) An excitation-pattern model for intensity discrimination J Acoust Soc Am 70, 1646–1654.

Gallun FJ, Hafter ER (2006) Amplitude modulation sensitivity as a mechanism for increment detection J Acoust Soc Am 119, 3919–3930.

Garcia-Lazaro JA, Ahmed B, Schnupp JWH (2006) Tuning to natural stimulus dynamics in primary auditory cortex. Curr Biol 16: 264-271.

Garcia-Lazaro JA, Ahmed B, Schnupp JWH (2011) Emergence of Tuning to Natural Stimulus Statistics along the Central Auditory Pathway. PLoS ONE 6(8): e22584.

Glasberg BR, Moore BCJ (1990) Derivation of auditory filter shapes from notched-noise data Hear Res 47, pp. 103–138.

Glasberg BR, Moore BCJ (2002). A model of loudness applicable to time-varying sounds J Audio Eng Soc 50, 331–342.

Glasberg BR, Moore BCJ, Peters RW (2001) The influence of external and internal noise on the detection of increments and decrements in the level of sinusoids Hear Res 155, 41–53.

Hellman WS, Hellman RP (1990) Intensity discrimination as the driving force for loudness. Application to pure tones in quiet, J Acoust Soc Am 87, 1255–1265.

Hellman WS, Hellman RP (2001) Revisiting relations between loudness and intensity discrimination J Acoust Soc Am 109, 2098–2102.

Hollander M, Wolfe DA (1973), Nonparametric Statistical Methods. New York: John Wiley & Sons, pp. 139–146.

Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. Neuroimage 50: 1202-1211.

Jaramillo S, Zador AM (2011). Auditory cortex mediates the perceptual effects of acoustic temporal expectation. Nat Neursoci 14:246-51.

Jesteadt W, Wier CG, Green DM (1977) Intensity discrimination as a function of frequency and sensation level. J Acoust Soc Am 61: 169-177.

Jepsen ML, Ewert SD, Dau T (2008) A computational model of human auditory signal processing and perception. J Acoust Soc Am 124: 422-438.

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013). The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750-761.

Lesica NA, Grothe B (2008) Efficient temporal processing of naturalistic sounds. PLoS ONE 3:e1655.

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

Long GR, Cullen JK (1985) Intensity difference limens at high frequencies. J Acoust Soc Am 78: 507-513.

Malmierca MS, Cristaudo S, Perez-Gonzalez D, Covey E (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. J Neurosci 29:5483–5493.

McDermott JH (2009) The cocktail party problem. Curr Biol 19:R1024–R1027.

McDermott JH, Simoncelli EP (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. Neuron 71:926–940.

McDermott JH, Schemitsch M, Simoncelli EP (2013). Summary statistics in auditory perception. Nat Neurosci 16:493-498.

McGill WJ, Goldberg JP (1968a) A study of the near-miss involving Weber's law and pure tone intensity discrimination. Percept Psycho-phys 4, 105–109.

McGill WJ, Goldberg JP (1968b) Pure tone intensity discrimination and energy detection. J Acoust Soc Am 44, 576–581.

Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233-236.

Miller GA (1947) Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. J Acoust Soc Am 19, 609–619.

Moore BCJ (1995) Frequency Analysis and Masking, in Hearing, BCJ Moore, Ed., (Academic Press, San Diego, California), pp. 161-205.

Moore BCJ, Glasberg BR, Baer T (1997) A model for the prediction of thresholds, loudness, and partial loudness. J Audio Eng Soc 45, 224–240.

Munson WA (1947) The growth of auditory sensation. J Acoust Soc Am 19, pp. 584-591.

Neely ST, Allen JB (1998) Predicting the intensity JND from the loudness of tones and noise in Psychophysical and physiological advances in hearing, edited by A Palmer, A Rees, Q Summerfield and R Meddis (Whurr, London), 458-464.

Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14:474–480.

O'Connell MN, Falchier A, McGinnis T, Schroeder CE, Lakatos P (2011) Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. Neuron 69:805-817.

Oxenham AJ (1997) Increment and decrement detection in sinusoids as a measure of temporal resolution. J Acoust Soc Am 102, 1779–1790.

Ozimek E, Zwislocki JJ (1996) Relationships of intensity discrimination to sensation and loudness levels: Dependence on sound frequency. J Acoust Soc Am 100: 3304-3320.

Parra LC, Pearlmutter BA (2007) Illusory percepts from auditory adaptation J Acoust Soc Am 121, 1632–1641.

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF (2012) Reconstructing Speech from Human Auditory Cortex. PLoS Biol 10(1):e1001251.

Patterson RD, Nimmo-Smith I, Weber DL, Milroy R (1982) The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. J Acoust Soc Am 72, pp. 1788–1803.

Perez-Gonzalez D, Malmierca MS, Covey E (2005). Novelty detector neurons in the mammalian auditory midbrain. Eur J Neurosci 22:2879–2885.

Peters RW, Moore BCJ, Glasberg BR (1995) Effects of level and frequency on the detection of decrements and increments in sinusoids. J Acoust Soc Am 97: 3791-3799.

Pickles JO, (2008) An Introduction to the Physiology of Hearing, 3rd ed. Emerald.

Pienkowski M, Hagerman B (2009) Auditory intensity discrimination as a function of level-rove and tone duration in normal-hearing and impaired subjects: the "mid-level hump" revisited. Hear Res 253, 107-115.

Plack CJ Gallun FJ, Hafter ER, Raimond A (2006) The detection of increments and decrements is not facilitated by abrupt onsets or offsets. J Acoust Soc Am 119, 3950-3959.

Rabinowitz NC, Willmore DB, Schnupp JWH, King AJ (2011). Contrast gain control in auditory cortex. Neuron 70:1178-1191.

Riesz R (1928) Differential intensity sensitivity of the ear for pure tones. Phys Rev 31, 867–875.

Sachs MB, Abbas PJ (1974) Rate versus level functions for auditory-nerve fibers in cats: tone-burst stimuli. J Acoust Soc Am 56, 1835–1847.

Sadagopan S, Wang X (2008) Level invariant representation of sounds by populations of neurons in primary auditory cortex. J Neurosci 28:3415-3426.

Shamma SA, Elhilali M, Micheyl C (2011). Temporal coherence and attention in auditory scene analysis. Trends Neurosci 34, 114-123.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270: 303-304.

Simpson AJR, Harper NS, Reiss JD, McAlpine D (2014) Selective Adaptation to “Oddball” Sounds by the Human Auditory System. J Neurosci 34:1963-1969.

Simpson AJR, Terrell MT, Reiss JD (2013) A Practical Step-by-Step Guide to the Time-Varying Loudness Model of Moore, Glasberg and Baer (1997; 2002), in Proc. 134th AES Conv., Rome.

Simpson AJR, Reiss JD (2013) The Dynamic Range Paradox: A Central Auditory Model of Intensity Change Detection. PLoS ONE 8(2):e57497.

Simpson AJR, Reiss JD, McAlpine D (2013) Tuning of human modulation filters is carrier-frequency dependent. PLoS ONE 8(8): e73590.

Skoe E, Krizman J, Spitzer E, Kraus N (2013) The auditory brainstem is a barometer of rapid auditory learning. Neuroscience 243:104–114.

Slabu L, Grimm S, Escera C (2012) Novelty detection in the human auditory brainstem. J Neurosci 32: 1447-1452.

Sutter ML, Schreiner CE (1995) Topography of intensity tuning in cat primary auditory cortex: single-neuron versus multiple-neuron recordings. J Neurophysiol 73: 190-204.

Teki S, Chait M, Kumar S, Shamma S, Griffths TD (2013). Segregation of complex acoustic scenes based on temporal coherence. eLife 2, e00699.

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6:391-398.

Ulanovsky N, Las L, Farkas D, Nelken I (2004). Multiple time scales of adaptation in auditory cortical neurons. J Neurosci 24:10440 –10453.

Viemeister NF (1983) Auditory intensity discrimination at high frequencies in the presence of noise Science 221, 1206–1208.

Viemeister NF, Bacon SP (1988) Intensity discrimination, increment detection and magnitude estimation for 1-kHz Tones. J Acoust Soc Am 84, 172-178.

Voss RF, Clarke J (1975) 1/F noise in music and speech. Nature 258: 317-318.

Voss RF, Clarke J (1978) 1/F noise in music: Music from 1/F noise. J Acoust Soc Am 63: 258-263.

Walker KM and King AJ (2011) Auditory neuroscience: temporal anticipation enhances cortical processing. Curr Biol 21:R251-253.

Watkins PV, Barbour DL (2008) Specialized neuronal adaptation for preserving input sensitivity. Nat Neurosci 11:1259-1261.

Watson CS, Gengel, RW (1969) Signal duration and signal frequency in relation to auditory sensitivity. J Acoust Soc Am 46: 989-997.

Wang Y, Ding N, Ahmar N, Xiang J, Poeppel D, Simon JZ (2012) Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. J Neurophysiol 107: 2033-2041.

Weber EH (1846) Der Tastsinn und das Gemeingefühl In: R. Wagner (Ed.) Handwörterbuch der Physiologie mit Rücksicht auf physiologische Pathologie 3, 481-588.

Wen B, Wang GI, Dean I, Delgutte B (2009) Dynamic range adaptation to sound level statistics in the auditory nerve. J Neurosci 29:13797–13808.

Wojtczak M, Viemeister NF (1999) Intensity discrimination and detection of amplitude modulation J Acoust Soc Am 106, 1917-1924.

Xiang J, Peoppel D, Simon JZ (2013) Physiological evidence for auditory modulation filterbanks: Cortical responses to concurrent modulations. J Acoust Soc Am 133:EL7.

Yaron A, Hershenhoren I, Nelken I (2012) Sensitivity to complex statistical regularities in rat auditory cortex. Neuron 76:603–615.

Zatorre RJ, Zarate JM (2012) Cortical processing of music. In the human auditory cortex, springer handbook of auditory research, D Poeppel, T Overath, AN Popper, RR Fay (eds.), Springer, Berlin. 261-294 (Chap. 10).

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon C, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. Neuron 77:980-991.