Individual Differences in Pitch Perception

Samuel Robert Mathias

Submitted for the degree of Doctor of Philosophy in Psychology

University of York
Department of Psychology

November 2010

Abstract

This thesis reports a number of experiments investigating individual differences in pitch perception. Experiment 1 identified otherwise normally hearing adult listeners who were relatively insensitive to the direction of small but detectable frequency changes between sequential pure tones. Following previous research, an important feature of the stimuli used in Experiment 1 was that the standard frequency was varied randomly—or 'roved'—over a wide range (400–2000.1 Hz) over trials. Subsequent psychoacoustical experiments (Experiments 2–6) revealed that the insensitivity to pitch-change direction observed in some individuals appeared to depend critically on the use of a relatively wide frequency-roving range, and the possible origins of this effect were investigated in detail. A compelling hypothesis that emerged was that roving introduced random, irrelevant frequency changes to the stimulus ensemble that interfered with pitch-direction identification (but not pitch-change detection) in some listeners. The results of Experiments 1–6 were considered under the framework of signal detection theory, and were compared to the predictions of several phenomenological models of frequency discrimination. The pre-existing models from the literature were not able to account for many aspects of the present findings without modification. The listeners taking part in the psychoacoustical experiments completed questionnaires assessing their everyday hearing experiences and musical ability. Listeners' thresholds were weakly or moderately correlated with some aspects of their self-reports, including sound segregation and musical experience/perception, and uncorrelated with other aspects, such as sound localisation and speech intelligibility. Experiment 7 used magnetoencephalography to test the hypothesis that insensitivity to pitch-change direction is associated with a greater-than-normal involvement of left-hemisphere cortical areas when listening to and identifying the direction of frequency changes. The results suggested that the extent to which the two hemispheres contributed to the recorded signals varied considerably across individuals, and contrary to the results of an earlier study, poor pitch-discrimination ability was associated with stronger right-hemisphere auditory-evoked fields.

Contents

Appendices

List of Figures

List of Tables

Acknowledgements

This document marks the end of my seven years at York, and for their friendship, support, and encouragement throughout that time I am enormously grateful to many people. Firstly, I sincerely thank my supervisor Peter Bailey; he has been a source of consistent encouragement, wisdom, and patience, and has had the stamina to read through three dissertations since my days as an undergraduate! I thank Quentin Summerfield and Tony Morland for their comments and advice as members of my research committee. I thank Quentin also for the opportunity to work with him and Peter as a summer student, now *many* years ago, and for his supervision during my MSc project. I thank the other members of the auditory group, past and present: Pádraig Kitterick, Rosemary Lovett, and Joanne Eaves. I am particularly indebted to Pádraig for the huge amount of technical help he has given me throughout the PhD. I thank Mark Hymers and Rebecca Millman for their assistance with the MEG experiment, Philip Quinlan for advice on statistics, and Christophe Micheyl for his contributions to the published material, discussion of the modelling work, and his comments on the design of the later experiments.

Of those not or no longer affiliated with the York Psychology Department, I thank my past housemates (Joe, Jim, James no. 1, James no. 2, Chris, Dave no. 1, Matt, Dave no. 2, Nic, and Ziggy) and fellow PhD students (Jodie and Claire) for their friendship. I also thank Simon (L2) for taking part without complaint in almost every experiment I've ever run. Lastly, I thank Emma and my family for their continuing love and support.

Financial Support

This PhD was supported by a studentship from the UK Biotechnology and Biological Research Council. Additional funds were provided by the guarantors of the journal *Brain* and the Experimental Psychology Society that enabled me to attend the Mid-Winter Meeting of the Association for research in Otolaryngology in Anaheim, CA (February 2010).

Declaration

This thesis comprises the candidate's own original work and has not, either in the same or different form, been submitted to this or any other University for a degree. The candidate designed all experiments and conducted all testing and analyses.

Journal Publications & Conference Presentations

Work from this thesis has appeared in the following publications and conference proceedings:

Mathias, S. R., Micheyl, C., & Bailey, P. J. (2010). Stimulus uncertainty and insensitivity to pitch-change direction. *Journal of the Acoustical Society of America, 127,* 3026–3037.

Mathias, S. R., Micheyl, C., & Bailey, P. J. (2010). Insensitivity to pitch-change direction with fixed- and roved-frequency tones. Presented at the *33rd Mid-Winter Meeting of the Association for Research in Otolaryngology*, California, USA.

Mathias, S. R., & Bailey, P. J. (2009). Pitch discrimination with roved- and fixed-frequency stimuli. Presented at the *Short Papers Meeting of the British Society of Audiology*, Southampton, UK.

Mathias, S. R., & Bailey, P. J. (2008). Insensitivity to the direction of pitch change: The role of stimulus uncertainty and correlations with self-reported hearing. Presented at the *Short Papers Meeting of the British Society of Audiology*, York, UK.

Chapter 1

Overview

1.1 General introduction

The aim of psychoacoustics is to understand the mechanisms underlying auditory perception, and to specify through experimentation the relationships between the characteristics of the sounds entering the ear and the sensations they produce. The typical approach adopted by psychoacousticians is to recruit a number of listeners and to measure their perceptions of one or several manipulations to an auditory stimulus. Usually these measurements are taken in a relatively small sample of listeners, involving several hours of testing per listener, and often after extensive practice so that performance is unlikely to improve any further ('asymptote').

This method of investigation has proven very successful, and a great deal is now known about the psychology of hearing (see Moore, 2003). However, psychoacousticians sometimes note large and often unexpected differences between the level of performance of their listeners in certain experiments (e.g., Jeffress & McFadden, 1971; Lauter, 1982, 1983; McFadden, Jeffress, & Russell, 1973; Moore, 1973a; Neff & Green, 1987; Semal & Demany, 2006). Such differences can prove difficult to reconcile with the consensus of the prevailing literature (e.g., Semal & Demany, 2006), or pose unique challenges for theories and models of auditory perception (e.g., Moore, 1973a). Sometimes researchers have suggested that what is called for in these circumstances is a systematic collection of data from a large sample of listeners spanning the range of ability (e.g., Durlach, Mason, Kidd, et al., 2003; D. M. Johnson, Watson, & Jensen, 1987). This approach would not only provide a richer profile of the individual differences, but also help to safeguard against making generalisations from results collected from a small and unrepresentative sample of listeners. It is true to say, however, that psychoacousticians have not taken this approach very often.

An important exception to the above statement is a listener's basic auditory sensitivity or hearing level, reflected in the audiogram. An audiogram is a standardised way of representing hearing level, derived by adjusting the presentation level of a pure tone with a specific frequency to a listeners' threshold for detectability (Figure 1.1). Hearing level is arguably the best understood auditory dimension along which individuals vary: standardised procedures for measuring hearing level (in dB HL) have been adopted by audiologists [British Society of Audiology (BSA), 2004], normal hearing levels across frequency are well established, and the factors influencing differences in hearing level—such as age, sex, and social factors—have been summarised comprehensively. For a review of this literature, see Davis (1995).

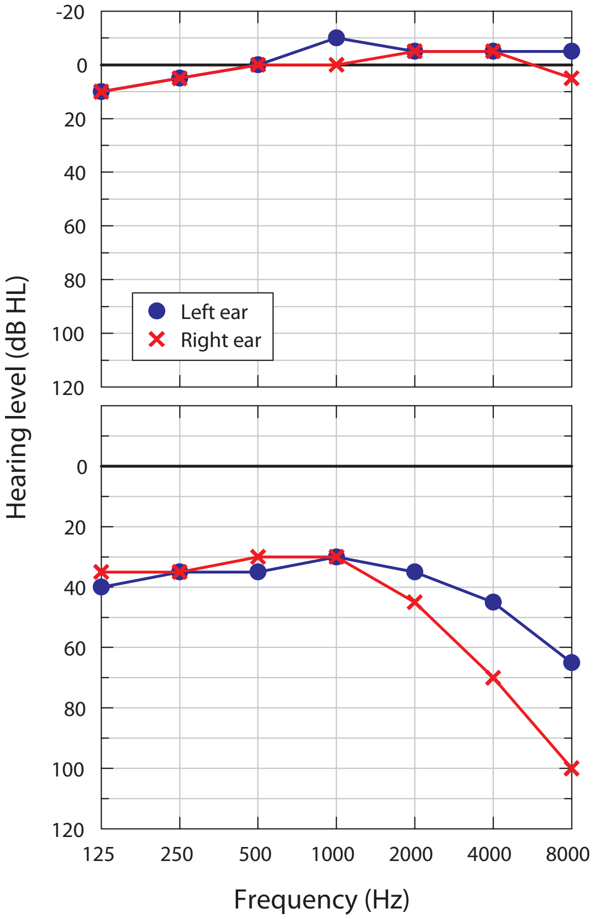


Figure 1.1. Two example audiograms. Increases in hearing level are plotted downwards, and a value of 0 dB HL represents the average hearing level for that frequency in normally hearing adults. Hearing level profiles in two listeners are shown: one who would be regarded as normally hearing (top panel) and one who would be considered to have a hearing impairment (bottom panel; a mild impairment in the left ear and a moderate impairment in the right ear).

The audiogram relies primarily on a listener's ability to *detect* a sound. However, psychoacousticians commonly require listeners to perform tasks that also require them to *discriminate* sounds based on one or more of their physical properties, such as frequency, duration, or intensity. It turns out that the audiogram is not always a good predictor of performance in other psychoacoustical experiments (e.g., Section 2.3.2). This observation has led many researchers to look for abilities other than basic auditory sensitivity that influence performance in tasks requiring discrimination (e.g., Elliott, Riach, Sheposh, & Trahiotis, 1966; Karlin, 1942; Kidd, Watson, & Gygi, 2007; Stankov & Horn, 1980). A common starting point is to hypothesise that a listener's performance on any particular auditory task is influenced both by a general ability—or 'auditory *g*'—and one or more specific abilities. Such specific abilities might include temporal acuity, governing duration-based discriminations (e.g., D. M. Green, 1971; Hirsh, 1959; C. S. Watson, 2004), and spectral resolving power, governing frequency-based discriminations (e.g., Feth & O'Malley, 1977; Moore & Glasberg, 1986). However, despite considerable effort on the part of researchers, no clear picture of the number or nature of these abilities has emerged, and no specific measure of any other auditory ability has achieved the canonical status of the audiogram as the measure of hearing performance (Section 2.3.1).

The focus of this thesis is individual differences specifically within the domain of pitch perception. In psychoacoustics, the term 'pitch' refers to a sensation rather than a physical attribute of a sound. For a pure tone, the perceived pitch is determined primarily by its frequency. Thus, if a listener is able to discriminate between two pure tones differing in frequency, it is assumed that most of the time this discrimination is based on a perceived difference in pitch (Plack & Oxenham, 2005b). The sensitivity of the human auditory system to frequency (or pitch) differences between pure tones has been measured in many classic psychoacoustical experiments (e.g., Harris, 1952b, 1952a; Moore, 1973b, 1974; Nordmark, 1968; Wier, Jesteadt, & Green, 1977). These studies all shared a similar methodology: on every trial, the listener was presented with two sequential pure tones differing slightly in frequency, and the listener indicated which of the tones was higher. The magnitude of the frequency difference was manipulated over trials, allowing the researchers to estimate the listener's threshold or difference limen for frequency (Section 2.2.4).

Listeners' thresholds for frequency discrimination can be remarkably small (e.g. Moore, 1974). However, as mentioned earlier, it is common practice in such studies to use a few highly trained listeners, or to even filter out those with either very good or very poor initial performance (e.g., Demany, 1985; Demany & Semal, 2002). When listeners are not pre-selected, frequency discrimination ability can be subject to large individual differences even after training (e.g., Amitay, Hawkey, & Moore, 2005; Semal & Demany, 2006). Clearly the experimental procedure outlined above requires listeners to not only detect a small frequency difference, but also requires them to order the tones on a scale ranging from low to high. The notion of height is central to pitch, being inherent within equivalent expressions in several languages (e.g., *hauteur tonale* in French, *Tonhöhe* in German, *visina tona* in Croatian and Serbian) and forming the basis of the current American National Standards Institute's (ANSI, 1973/1994) definition of pitch. The ANSI definition necessarily means that as soon as a pitch change is detected, its direction can be identified. While this may be true for many listeners (Jesteadt & Bilger, 1974), more recent studies have suggested that there may be important individual differences between listeners' ability to detect the presence of a small frequency difference and their ability to identify the direction of the resulting pitch change (Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Foxton, Weisz, Bauchet-Lecaignard, Delpuech, & Bertrand, 2009; Johnsrude, Penhune, & Zatorre, 2000; Semal & Demany, 2006; Tramo, Shah, & Braida, 2002). The present work follows on directly from this research.

The aim of the empirical work in this thesis was to investigate further the origins and consequences of the individual differences in pitch perception reported in particular by Semal and Demany (2006) in normally hearing listeners. This aim was approached in four ways: by replicating and extending Semal and Demany's findings through psychoacoustical experimentation; by considering the data in terms of models of frequency discrimination; by investigating whether or not the individual differences were correlated with aspects of everyday hearing; and by investigating the brain activity evoked during frequency discrimination using magnetoencephalography (MEG).

1.2 Overview of the following chapters

Chapter 2: Pitch perception and discrimination with pure tones

This chapter defines pitch and reviews the key features of pure-tone pitch perception in normally hearing listeners. The chapter covers the stimulus properties affecting pitch, frequency discrimination, the limits of pitch, and memory for pitch. The chapter goes on to discuss the results of a number of studies that have investigated individual differences in auditory perception. Not all of those studies are concerned with pitch exclusively, but are included given their relevance to the empirical work detailed in the later chapters. The chapter concludes with a selective review of studies investigating the neural correlates of pitch perception using MEG.

Chapter 3: Stimulus uncertainty and insensitivity to pitch-change direction

This chapter reports two experiments that investigated individual differences in adult listeners' sensitivity to pitch-change direction. In Experiment 1, listeners were tested using the dual-pair procedure (Semal & Demany, 2006). The experiment revealed that a number of these listeners were relatively poor at identifying the direction of small (but detectable) frequency changes between pure tones. Importantly, the standard frequencies of the tone pairs used in Experiment 1 were roved over a wide range. In Experiment 2, most of the same listeners were re-tested using stimuli that did not involve frequency roving. The key finding was that without roving, the impairments in pitch-direction identification were generally much less profound and even eliminated entirely in many listeners. Both experiments also contained conditions in which the target tones were flanked spectrally by random-frequency, pure-tone maskers. The motivation for these conditions was to determine whether pitch-change detection and pitch-direction identification abilities would be affected differently by informational masking. However, very poor performance in these conditions vitiated any useful comparisons across tasks and across experiments.

Chapter 4: The influence of frequency roving on pure-tone pitch discrimination

This chapter reports two experiments that investigated in more detail the effect of frequency roving on listeners' sensitivity to pitch-change direction. In Experiment 3, the size of the frequency-roving range was manipulated systematically over runs of trials. The data were collected from a sample of new listeners, and conditions were tested in a randomised order to compensate for potential order effects. Most conditions contained level roving in addition to frequency roving to compensate for some other possible confounds. The results supported those from Experiment 2, suggesting that impairments in pitch-direction identification depended critically on the use of a relatively wide frequency-roving range. Experiment 4 tested the same listeners with a new stimulus, in which the target frequency change occurred within a single, frequency-modulated tone rather than between two discrete tones. The starting frequencies of the new stimuli were either roved or fixed in a similar manner to the standard frequency in the previous experiments. The results revealed again that some listeners experienced difficulty identifying pitch-change direction only in the context of frequency roving.

Chapter 5: The influences of feedback and irrelevant pitch changes on pure-tone pitch discrimination

This chapter reports two experiments and a short supplementary experiment that investigated two possible explanations for the earlier findings. Experiment 5 tested the hypothesis that in previous experiments, direction-impaired listeners were able to use feedback to perform the pitch-direction identification task successfully in conditions without roving. Conditions were completed in a prescribed order that was the same for all listeners. The standard frequency was fixed in the first condition, shifted to a lower novel frequency in the second condition, and shifted to a higher frequency in the third condition; crucially, the latter two conditions were completed without feedback. The fourth and fifth conditions re-tested these low and high frequencies with feedback, and the sixth condition used frequency roving.

There was no effect of switching frequency nor of withholding feedback, suggesting that under conditions without roving, the listeners could genuinely perceive the direction of small frequency changes. Experiment 6 tested the hypothesis that roving caused irrelevant stimulus changes to occur within the stimulus ensemble, which in turn impaired the ability of some listeners to identify the direction of the relevant frequency changes. Listeners heard three tones per trial, and were instructed to name the direction of the frequency change between the second and third tones. In half of the conditions, the first tone had the same frequency as the second, and in the other half, the frequency of the first tone was random. The direction-impaired listeners were poorer at identifying pitch-change direction when the target change was preceded by an irrelevant frequency change, even though they were instructed explicitly to ignore it. However, the effect could have occurred because of a beneficial influence of the first tone when its frequency was not random. This possibility was tested in a supplementary experiment, which suggested that the results of Experiment 6 were due to a combination of beneficial and deleterious effects.

Chapter 6: Modelling listeners’ DLFs using signal detection theory

This chapter considers the key features of Experiments 1–6 in terms of phenomenological models of frequency discrimination based on signal detection theory. The chapter first outlines a standard model, which assumes that listeners’ decisions in the experiments were based on noisy sensory observations (Gaussian random variables), and that the standard deviation of the internal noise was the same under all circumstances. While the data from a few listeners in some experiments were approximately consistent with its predictions, the model could not explain three important findings: the individual differences in listeners’ sensitivity to pitch-change direction; the general effect of frequency roving; and the three-way interaction between listener, task, and frequency-roving range. The chapter discusses a number of modifications that could be made to the standard model to account for these features of the data.

Chapter 7: Correlations with self-reported everyday hearing and musical ability

This chapter investigated whether the insensitivity to pitch-change direction observed in some listeners was indicative of real-world hearing difficulties or experiences. Prior to participating in the psychoacoustical experiments, listeners completed two questionnaires. The first questionnaire required them to rate many aspects of their everyday hearing, such as their ability to understand speech in a variety of contexts, their ability localise sound sources, their ability to perceive movement, the quality and naturalness of everyday sounds, listening effort, and so on. The second questionnaire assessed their level of musical education and expertise. Exploratory factor analysis was used to group listeners' responses to similar items in the questionnaires, thereby reducing the number of statistical tests required and decreasing the likelihood of observing spurious correlations. The results indicated that, overall, listeners' thresholds were not strongly correlated with their everyday hearing. Some weak or moderate relationships were observed between listeners' self-reports and their thresholds for pitch-change detection and pitch-direction identification, particularly with items relating to music. Detection thresholds were correlated more strongly with items on the music questionnaire than identification thresholds, suggesting that it was listeners' basic sensitivity to frequency changes, rather than their ability to identify the direction of those changes, that was related to their musical experience.

Chapter 8: Neural correlates of pitch-change detection and pitch-direction identification

This chapter reports a final experiment that measured listeners' auditory-evoked fields (AEFs) with MEG. During Experiment 7, listeners heard pairs of pure tones similar to those in the earlier experiments, and completed two different tasks. One of the tasks required them to be sensitive to direction of frequency changes, whilst the other did not require this sensitivity. Listeners' AEFs were different depending on whether the pairs contained a frequency change, depending on the size of that change, and depending on the task they performed. There was considerable inter-individual variability in terms of the relative contributions of the two hemispheres to the measured signals. In some listeners, AEFs from the left hemisphere were stronger than the AEFs from the right hemisphere, whilst the reverse was true in others. An earlier study had reported that insensitivity to pitch-change direction was associated with stronger left-hemisphere AEFs. Experiment 7 found just the opposite: the right-hemisphere AEFs were stronger on average in a group of direction-impaired listeners and weaker on average in a group of unimpaired listeners.

Chapter 9: General discussion

The final chapter summarises the key findings from the thesis, and makes suggestions for future work.

Chapter 2

Literature review: Pitch perception and discrimination with pure tones

Pitch is a perceptual attribute of sound resulting from periodicity in a sounds' waveform. Pitch is an extremely complex phenomenon and is a ubiquitous element of our natural auditory environment, so it is not surprising that a great deal of scientific effort has gone into studying how pitch is perceived and how pitch is used by the auditory system. No attempt is made here to review this entire literature, since numerous volumes on the topic already exist, from such diverse fields as psychoacoustics and neuroscience (Plack & Oxenham, 2005a), phonetics and linguistics (Gussenhoven, 2004; Yip, 2002), and music cognition (Krumhansl, 1990). Instead, this chapter first provides a definition of pitch and then focuses mainly on studies that have investigated pitch using pure tones. Among other things, the chapter discusses the stimulus properties that determine pure-tone pitch, listeners' memory for pitch, and theories of how pure-tone pitch is determined in the auditory system. The chapter goes on to describe some studies that have noted large individual differences in listeners' auditory abilities, and discusses some potential factors underlying these differences. Finally, the chapter outlines some studies that have investigated the neural correlates of pitch perception using magnetoencephalography (MEG).

2.1 What is pitch?

Pitch has been defined previously by the American National Standards Institute (ANSI)—then the American Standards Association (ASA)—as 'that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale' (ASA, 1960). This definition has two important features. First, it describes pitch as an attribute of *sensation*. In other words, the term should always refer to the subjective perception experienced by the listener and not to any physical property of a sound. Pitch—like other attributes of auditory sensation such as timbre and loudness—cannot be expressed in physical units nor measured by physical means. The second feature of the definition refers to the role of pitch in *music*. There is no doubt that pitch is an essential element of Western music[1]: the relationship between the pitches of different sounds forms the basis of harmony when the sounds occur together, and melody when they occur in sequence.

This second feature of the ASA definition is slightly problematic since pitch is not confined to music. In languages such as English, pitch carries important prosodic information (see Gussenhoven, 2004), and when combined with other information, pitch can be used to identify who is speaking (e.g., Smith & Patterson, 2005). Pitch is also semantically relevant in tonal languages (see Yip, 2002). Although most psychoacousticians would agree that the production of melodies is sufficient to prove that a sound can evoke a pitch, some would not regard it as a necessary condition: pure tones with frequencies above approximately 4000/5000 Hz may vary in pitch (i.e., a listener can make higher/lower discriminations successfully) but those variations cannot be used to form recognisable melodies (Section 2.2.3). The ANSI clearly viewed the reference to music in the earlier definition as undesirable, and updated it in 1973: 'Pitch [is] that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high ... ' (ANSI, 1973/1994).

The more recent ANSI definition has also been criticised. Some have argued that it does not provide a clear distinction from other scalable attributes of sound; it could also include features such as loudness or shrillness/brightness in timbre that can be ordered reliably from low to high. Houtsma (1997) provided an amusing example of where this lack of a distinction actually caused confusion in the literature. Tanner and Rivette (1964) reported that speakers of Punjabi were unusually poor at pitch discrimination. Following common practice, the authors asked their subjects to discriminate between two tones by indicating which of the two was 'higher'. When Burns and Sampat (1980) repeated the experiment, they found performance became perfectly normal if the instructions to the subjects accounted for the fact that in Punjabi the same word is used to indicate that a sound is 'high in pitch' and 'loud'.

Another criticism of the ANSI definition is that it suggests that pitch is a purely one-dimensional quality. Although a cursory glance at music notation would appear to be consistent with this view—where the pitches of notes are represented by their vertical position on a staff—musical notes are also labelled according to their position on the chromatic scale, a repeating sequence of 12 classes (from low to high: A, A♯,

---

[1] Although the term is used occasionally in the literature, 'Western' is now something of a misnomer since clearly it is no longer confined to the western hemisphere. Much of modern popular and classical music the world over would assent—partly, at least—to the norms brought about by Western musical traditions (a history of their development is provided by P. Griffiths, 2006).

B, C, C♯, D, D♯, E, F, F♯, G, and G♯). Many researchers regard pitch height and pitch class or chroma (from Gk. *khroma*, ‘colour’) to be partially independent perceptual dimensions of pitch, since two pitches separated by 12 semitones (one octave) share the same chroma but are different in height. This and other observations from music have led many to abandon the notion of pitch as one-dimensional construct, and instead to regard pitch as a composite sensory attribute with multiple dimensions (Figure 2.1).

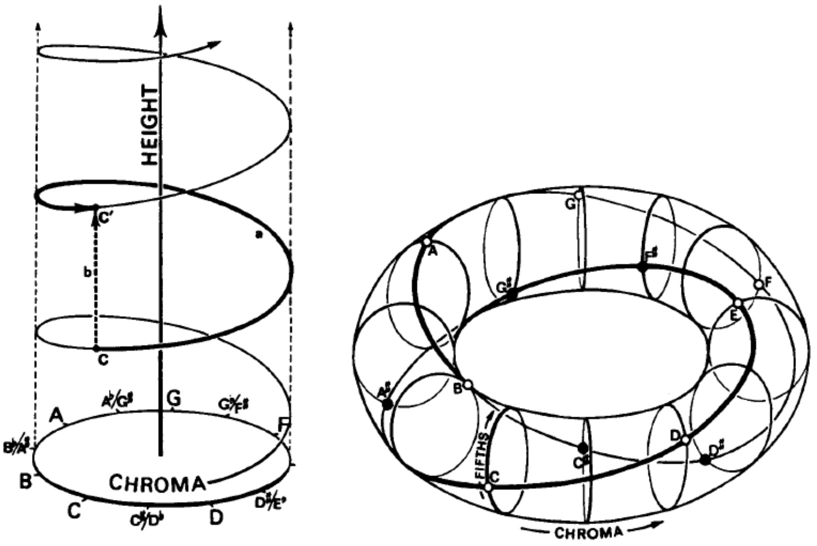


Figure 2.1. Two examples of multidimensional representations of pitch. The helical representation on the left illustrates pitch height and pitch chroma: the interval formed by two notes an octave apart (C and C′) traverses the helix completely so the notes have the same chroma (a), but also moves vertically (b). The representation on the right is a torus with a double helix, which additionally incorporates the concept of the circle of fifths. From Shepard (1982).

In short, neither the earlier nor the more recent American Standards definitions of pitch are entirely comprehensive. Consequently, this thesis follows Plack and Oxenham (2005b) in defining pitch as ‘that attribute of sensation whose variation is associated with musical melodies’. Although restrictive, this definition has two advantages over the official definitions: it provides a straightforward procedure for testing whether or not a stimulus evokes a sensation of pitch, and it dispenses with the notion of variations in pitch lying along a single dimension.

2.2 General properties of pure-tone pitch

2.2.1 What is a pure tone?

Principally, there are three types of sounds that can evoke a sensation of pitch: pure tones, complex tones, and certain kinds of noise. A ‘pure tone’ is a sinusoidal variation in pressure over time. Fourier’s theorem regards the sinusoid as the simplest oscillating function, and states that any complex periodic signal can be regarded as the sum of a number of sinusoids. This principle is illustrated for a repeating square-wave function in Figure 2.2. Thus, pure tones can be considered the building blocks of all other periodic sounds. Pure tones are also the simplest auditory stimuli to evoke a percept of pitch, making them particularly useful for psychoacousticians.

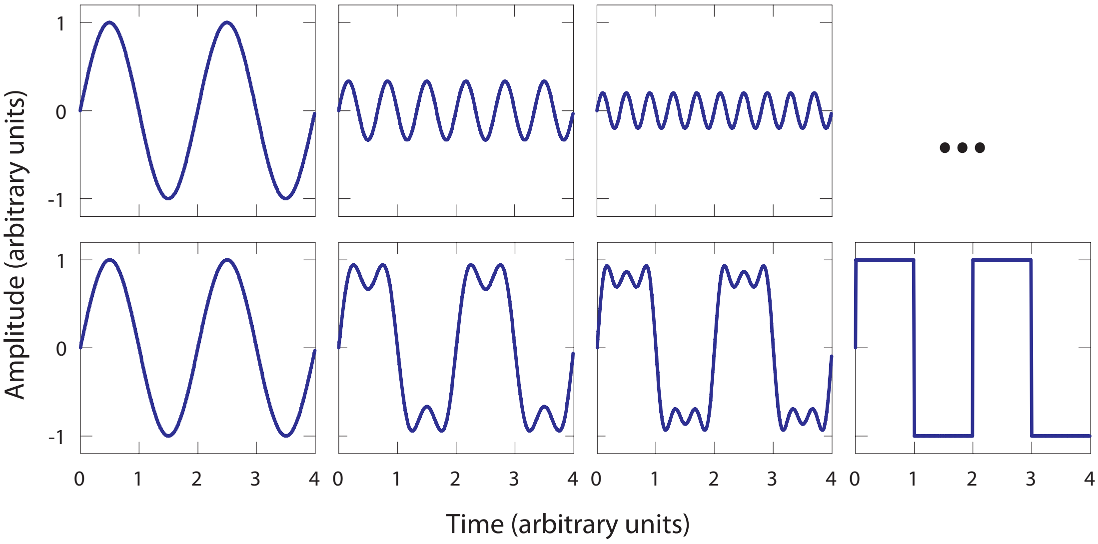


Figure 2.2. Illustration of a series of sinusoidal functions combining to create a complex square-wave function. The top three panels show (from left to right) the first, second, and third Fourier components of the square wave. The three leftmost bottom panels show the same components summed together; the rightmost bottom panel shows the result of an infinite number of summed Fourier components.

2.2.2 Stimulus properties influencing pure-tone pitch

2.2.2.1 Frequency

It has been known since at least the time of Pythagoras that the physical properties of a sound and its subjective pitch might be systematically related[2]. Although other properties also have an influence, the primary objective correlate of the pitch of a pure tone is its frequency, expressed in Hz. One method for measuring how the pitch of a pure tone depends on its frequency is to present listeners with a repeating sequence of two tones—a comparison tone and a test tone—and instruct them to adjust the frequency of the comparison tone until it subjectively sounds twice or half as high as the pitch of the test tone, the frequency of which is set by the experimenter. The classic result of pitch-halving/pitch-doubling experiments is the

---

[2] According to legend, one day Pythagoras passed blacksmiths at work and thought that the sounds emanating from their anvils being hit were beautiful and harmonious, and decided that whatever scientific law caused this to happen must be mathematical and could be applied to music. When he went back to the blacksmiths to learn how this had happened, he discovered that it was because the anvils were simple ratios of each other; one was half the size of the first, another was two-thirds the size, and so on.

mel scale. S. S. Stevens, Volkmann, and Newman (1937) used a pitch-halving method for test tones ranging in frequency from 125 to 12000 Hz. All tones were 2 s in duration and presented at 60 dB sound pressure level (SPL). A 1000-Hz tone was arbitrarily assigned a value of 1000 mels (from *mel*ody), a tone rated as half as high on average received a value of 500 mels, and so on. The results of the experiment averaged across S. S. Stevens et al.'s five listeners are shown by the dashed lines in the two panels in Figure 2.3. According to those data, halving the pitch of a pure tone is not equal to halving its frequency, and pitch in mels is not linearly related to frequency in Hz.

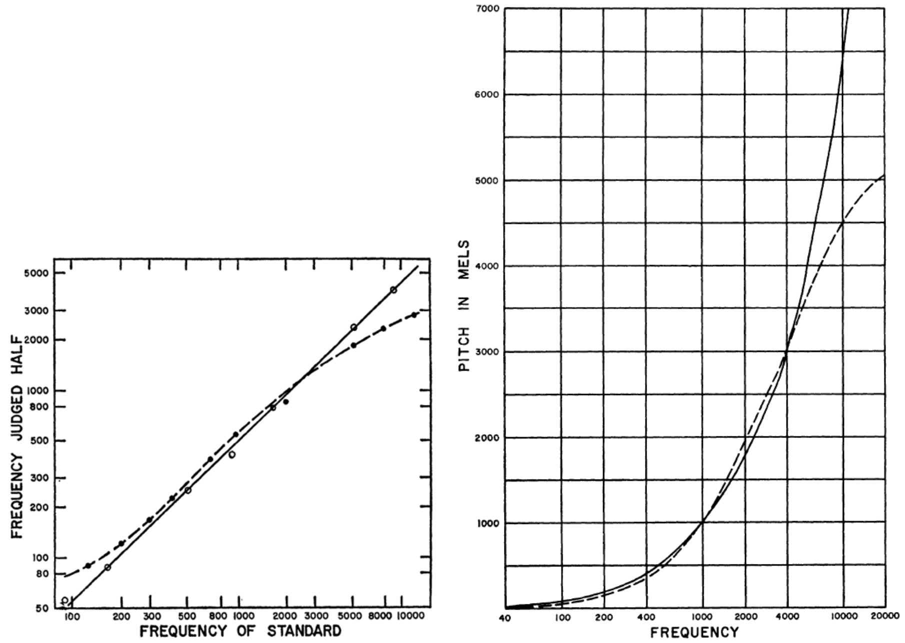


Figure 2.3. The results of two mel-scale experiments. The left-hand panel shows the standard (test) frequency plotted against the average frequency of a comparison tone judged half its frequency, both on logarithmic axes. The right-hand panel shows the test frequency (log axis) plotted against pitch in mels (linear axis). The dashed lines show the results from S. S. Stevens et al. (1937) and the solid lines show the results from Siegel (1965). From Houtsma (1997).

There are some reasons to doubt the veracity of the mel scale. Almost all of the widely used systems of tuning in modern music require two tones separated by an octave interval to have exactly a 2:1 frequency ratio[3]. The fact that very few musicians would claim that one octave sounds any larger or smaller than another is difficult to reconcile with the mel scale. Further, a replication study by Siegel (1965) found a much closer relationship between mels and frequency (those data are also plotted in Figure 2.3). Siegel argued that the discrepancies between the results of the two studies might reflect some error in apparatus or procedure, such as reverberation

---

[3] Many methods of tuning the musical scale exist, and they usually differ in their methods for dividing the octave into its 12 constituent steps (semitones). Historically, the equal-temperament scale—which divides the octave into 12 steps equally spaced logarithmically—has not been without its detractors, and many musicians, music theorists, and composers have preferred other tuning methods (e.g., just intonation). However, there is almost universal agreement on the 2:1-ratio tuning for the octave.

in the testing rooms used by S. S. Stevens et al., or a difference in the mean age of the listeners in the two studies. Although much research has investigated the effects of aging on frequency sensitivity, to my knowledge there have been no systematic studies of age-related changes in the perception of pitch intervals.

2.2.2.2 Level

Matching experiments have measured the influence of level on pitch. S. S. Stevens (1935) presented three listeners (although the data reported came from only one individual) with a repeating sequence of a comparison tone and a test tone. The tones had the same level but differed slightly in frequency, and the task was to adjust the level of the comparison tone so its pitch matched that of the test tone. When the results were plotted as equal-pitch contours, pitch increased with increasing level for tones above 1000 Hz, and decreased with increasing level for tones below 1000 Hz. S. S. Stevens' equal-pitch contours for 150-, 1000- and 8000-Hz tones are shown in Figure 2.4.

Subsequent studies have found that the magnitudes of the level-related shifts in pitch are highly variable across listeners. Snow (1936) tested nine listeners in a matching procedure using tones with frequencies below 1000 Hz. Although the negative relationship between pitch and level was replicated when equal-pitch contours were averaged across the group, two of Snow's listeners perceived no shifts at any frequency or loudness, whilst three others experienced shifts greater than 35% at the highest levels. Morgan et al. (1951) similarly noted marked individual differences in the relationship between pitch and level (25th and 75th percentiles are shown in Figure 2.4), and also between measurements taken from the *same* listener years apart.

Finally, Cohen (1961) found that much of S. S. Stevens' (1935) data could have resulted from so-called 'pitch-matching errors', the small discrepancy in the frequencies of a comparison and test tone when they were matched in pitch under equal-intensity conditions. When those errors were compensated for, the remaining shifts followed the directions of S. S. Stevens' curves but were much smaller in magnitude (2% or less; cf. Verschuure & van Meeteren, 1975). In conclusion, subsequent research has revealed the parametric effects of level on the pitch of pure tones appears to be more modest on average—and less stable across both listeners and time—than S. S. Stevens' data initially suggested.

2.2.2.3 Duration

The quality of a pure tone changes as a function of its duration. At very short durations (a few ms), a sinusoid will sound like a click without a clear pitch, but as its duration is increased its pitch becomes more easily identifiable, and the point where the transition between click and tone occurs is dependent on frequency (Doughty &

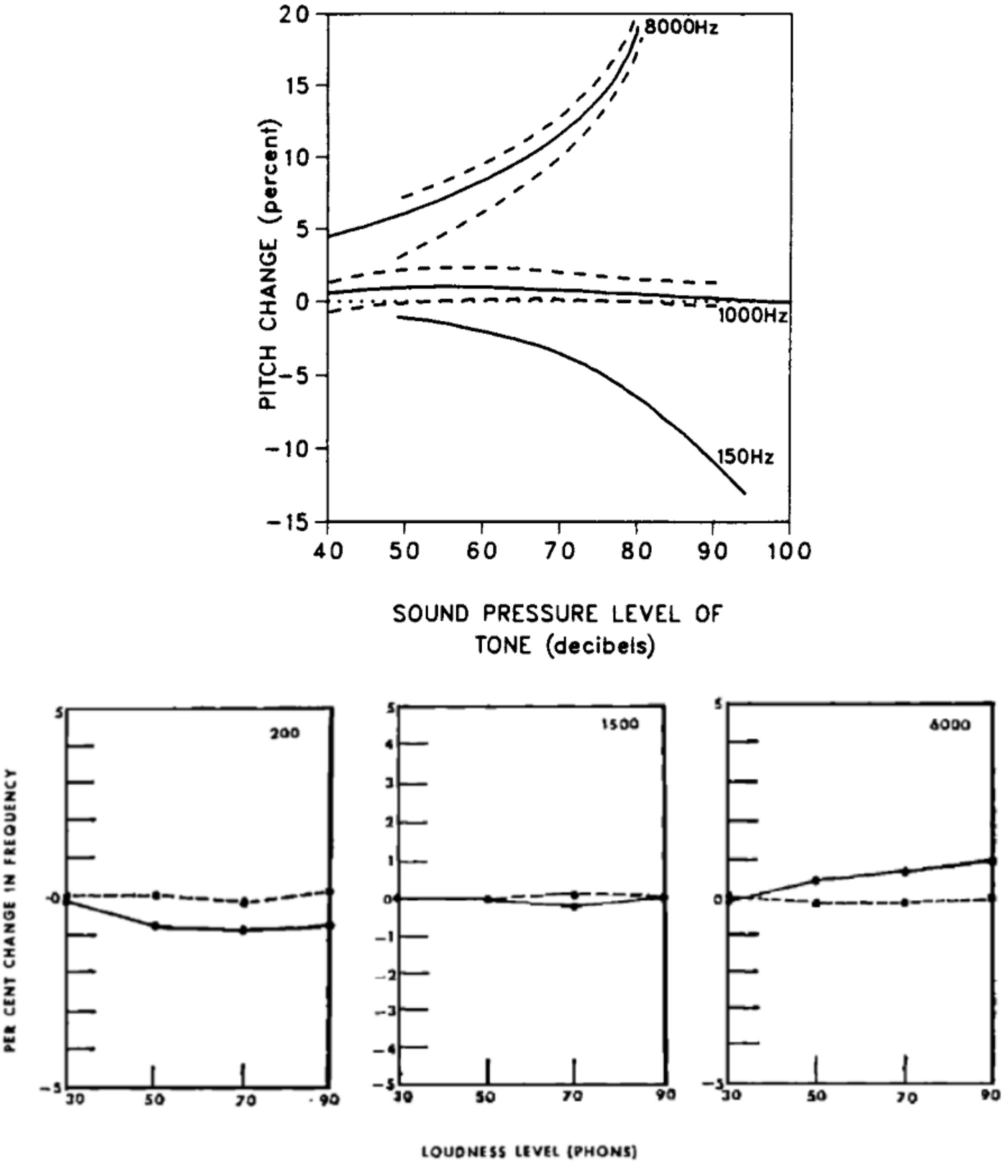


Figure 2.4. Pitch change in percent as a function of sound level for pure tones. In the top panel, the solid lines show the mean data from S. S. Stevens (1935), and the dashed lines show 25th (lower) and 75th (upper) percentiles of across the listeners tested by Morgan et al. (1951). From Houtsma (1997). The bottom panels show the data from Cohen (1961). The solid lines represent pitch-change functions and the dashed lines represent pitch-matching errors (see text) for pure tones with frequencies indicated in the top-right of each panel.

Garner, 1947). This effect is unsurprising since the width of a sound's spectrum is inversely proportional to its duration. However, the pitch of a short tone may also be different to the pitch of a longer tone with the same frequency. Doughty and Garner (1948) instructed six listeners to adjust the frequency of a comparison tone with one of six durations—6, 12, 25, 50, 100, or 200 ms—so that it matched the pitch of a 500-ms test tone with a frequency of 250, 1000, or 4000 Hz. The results showed a small but consistent trend for very short tones (6 or 12 ms) to sound lower in pitch than a 500-ms tone with the same frequency. The effect was largest at 250 Hz—approximately 4%, compared to approximately 2% for 1000 and 4000 Hz—and did not affect tones with durations of 25 ms or longer. The results also indicated an interaction between duration and level; the higher-level tones were more prone to duration-related shifts in pitch.

Somewhat similar results were reported in a matching study by W. M. Hartmann, Rakerd, and Packard (1985). The authors instructed four listeners to match for pitch all nine pairings of tones with durations 12, 25, and 50 ms at nominal frequencies of 200, 342, 584, and 1000 Hz. As in Doughty and Garner's (1947) study, shorter tones generally had a lower pitch than longer tones. However, the pattern of the results was not consistent across the group and some listeners were highly biased by the order in which the tones were presented, consistently judging whichever tone was presented second as higher in pitch, for instance. Those results indicate the influence of duration on pure-tone pitch is modest and subject to individual differences.

2.2.2.4 Envelope

The temporal envelope of a sound refers to the relatively slow changes in its amplitude over time (Figure 2.5). W. M. Hartmann (1978) found that listeners judged tones with exponentially decaying amplitudes as higher in pitch than tones with the same frequency and constant (flat) amplitudes. However, a series of experiments reported by Rossing and Houtsma (1986) suggests that there are complex interactions between the effects of frequency, duration, intensity, and envelope on pure-tone pitch. In their first experiment, four listeners (three with musical training) compared the pitch of a 40-ms test tone whose amplitude started to decay exponentially at onset to the pitch of a 40-ms comparison tone of constant amplitude. The exponential decay rate of the test tone—0.5, 1, 2, 4, or 8 dB/ms—and the intensity of the comparison tone—70, 80, 90, or 100 dB SPL—were the experimental variables, and the test frequency ranged from 200 to 3200 Hz. Their second experiment was identical to the first except that the amplitude of the test tones exponentially increased over time. The results of both experiments supported the findings of W. M. Hartmann (1978): a decaying tone had a higher pitch and an increasing tone had a lower pitch, on average, than an equivalent tone with flat amplitude. However, the frequency and the overall level of the tones influenced the strength of the effects; the shifts in pitch were larger at higher levels.

Follow-up experiments also identified a mediating role of duration, suggesting that envelope-induced pitch effects are also linked to changes in the *average* intensity of the tones over time.

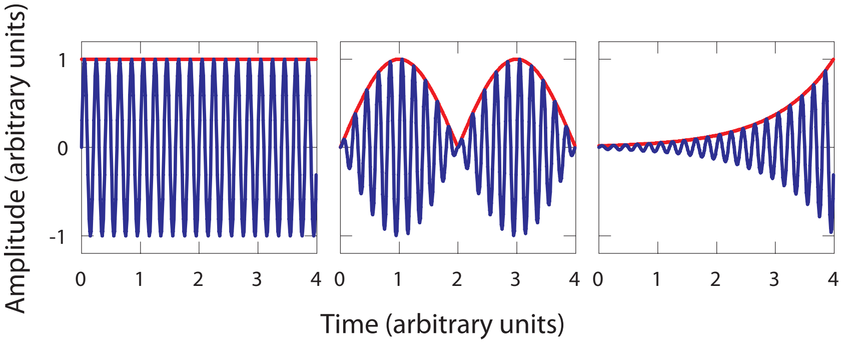


Figure 2.5. Three pure tones with different amplitude modulations (AMs). From left to right: a tone with a constant (unmodulated) amplitude, a sinusoidal AM tone, and a tone with an exponentially rising amplitude. The tones all have the same carrier frequency, so could be said to share the same temporal fine structure (blue lines) but differ in their temporal envelope (red lines).

2.2.2.5 Adaptation

Adaptation (or sometimes ‘fatigue’) refers to a range of effects—in both vision and audition—where a decline in the response of the sensory system is observed after sustained stimulation. A number of studies have shown that prior stimulation by an adapting tone alters the pitch of a test tone. The phenomenon was first reported by von Békésy (1960) who exposed listeners to an adapting pure tone of 94 dB SPL for 2 min. The pitch shift was bi-directional, in the sense that an adapting tone with a frequency lower than the following tone resulted in an upward shift in pitch, and vice versa. Similarly, Christman (1954) presented five listeners with 85-dB tones with durations of either 1 or 2 min and with frequencies of between 400 and 800 Hz. The adapting tones were always presented to the listeners’ left ears, and were followed by a test tone of 600 Hz to their left ears, and then a comparison tone presented to their right ears. Consistent with von Békésy’s data, when the adapting tone was lower in frequency, the test tone was perceived as having a higher pitch, and the opposite (although slightly stronger) effect was produced by an adapting tone with a higher frequency. Subsequent experiments have revealed that the effect of the adapting tone on pure-tone pitch is reduced as the silent interval (ISI) between it and the test tone is increased (Christman & Williams, 1963; Rakowski & Hirsh, 1980), such that the effect is negligible with ISIs longer than 100 ms. The effect is also ear-specific, with an adapting tone having virtually no effect on the perceived pitch of a tone presented to the contralateral ear (Larkin, 1978).

2.2.3 Limits of pure-tone pitch

Pure tones begin to sound ‘tonal’ or ‘pitch-like’ when their frequencies are above approximately 30 Hz (Guttman & Pruzansky, 1962). However, the lower limit of pitch for pure tones has not attracted much interest in psychoacoustics, possibly because for such low frequencies to be audible, tones must be presented at extremely high levels[4]. The upper limit of pure-tone pitch depends on which definition of pitch is adopted. If the ANSI (1973/1994) definition is used, then the upper limit can be regarded as highest frequency that a listener can reliably discriminate; this could include frequencies above 12000 Hz, depending on procedure and the listeners tested. However, the definition adopted in this thesis provides an alternative test of whether or not a particular sound lies within the limits of pitch: can it produce a melody? The fourth octave above middle C (C8, approximately 4186 Hz) forms the practical upper limit of pitch in music: it is the highest note on a standard 88-key piano and the highest note on any other orchestral instrument (the piccolo) is around this value. Concordantly, a number of scientific observations suggest that the upper limit of musical pitch for pure tones lies between 4000 and 5000 Hz. Bachem (1948) instructed listeners with absolute pitch—a rare ability of a person to identify or recreate a musical note without an external reference—to name (as musical notes) pure tones ranging in frequency, presented individually in a random order. His listeners tended to judge all frequencies above a certain point (usually around 4000 Hz) as having the same pitch. In another study (Ohgushi & Hatoh, 1989), 63 possessors of absolute pitch were similarly instructed to identify pure tones ranging in frequency from 1000 to 15000 Hz. For the 30 best listeners, Figure 2.6 shows the percentage of subjects who correctly identified each note. The results show a sharp drop in accuracy starting at A♯7 (approximately 3729 Hz) and ending at F8 (approximately 5588 Hz). Although the dotted line shows local peaks in accuracy above F8, from this note up to B9, the mean accuracy is close to chance level (8.3%). The results suggest that the upper limit of absolute pitch identification is around the same value as the upper limit of melody.

Several other studies provide evidence for an upper limit of musical pitch in musicians without absolute pitch. W. D. Ward (1954) presented listeners with a comparison tone and a test tone on every trial, and instructed them to adjust the frequency of the test tone so it was an octave above the comparison. The listeners were unable to octave-match the test tone to comparison tones above 2700 Hz,

---

[4] Harmonic complex tones are much more suitable for this purpose, since one of their key features is that they usually evoke a sensation of pitch that corresponds to the frequency of the fundamental (F_0) rather than any of the higher-frequency harmonics, even if the stimulus has been high-pass filtered so that the F_0 component itself has been removed (the phenomenon of the missing fundamental). Therefore, the lower limits of pitch can be studied using complex tones without confounds associated with very low-frequency energy (see Krumbholz, Patterson, & Pressnitzer, 2000; Pressnitzer & Patterson, 2001).

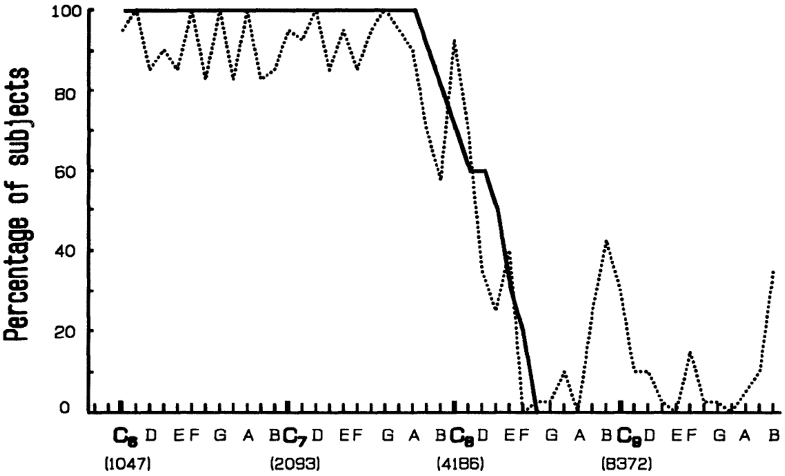


Figure 2.6. The upper limit of musical pitch for pure tones in musically trained listeners. The dotted line represents the percentage of 30 listeners with absolute pitch who correctly identified different musical notes (Ohgushi & Hatoh, 1989). The solid line represents the percentages of subjects whose mean adjustments were higher than the starting frequencies of the tones (Semal & Demany, 1990). From Semal and Demany (1990).

suggesting an upper limit of musical pitch below 5400 Hz. Attneave and Olson (1971) presented listeners with short, random pure-tone melodies and instructed them to reproduce the melody as faithfully as possible in a different frequency region (transposition) by adjusting the frequency of a comparison tone. In their two musically trained listeners[5], the ability to transpose melodies broke down badly for frequencies higher than 4000 Hz. Finally, Semal and Demany (1990) presented 10 musically trained listeners with sequences of two pure tones. The second tone was always higher than the first by a constant musical interval, but the starting frequency of the first (lower) tone was always randomly roved. Listeners transposed the sequence upward or downward in frequency until the pitch of the second (higher) tone was just above the upper limit of musical pitch. The results of this experiment are also plotted in Figure 2.6, and match very closely Ohgushi and Hatoh's (1989) results from listeners with absolute pitch.

In summary, the studies mentioned above suggest that there is an upper limit of musical pitch for pure tones somewhere between 4000 and 5000 Hz. Those studies could be criticised because all of them based their conclusions on data from musically trained listeners. Musicians obviously have much more experience of using pitch information than do non-musicians, and the limits of pitch might be quite different for listeners without such training—studies investigating the effects of psychoacoustical and musical training are discussed in Sections 2.2.5.3 and 2.3.3.1. A second discussion point concerns the origin of the correlation between the results

---

[5] Some listeners without formal musical training were also tested, but their performance was highly 'idiosyncratic' and difficult to interpret. It is very likely that Attneave and Olson's (1971) task would be extremely difficult for listeners without musical training.

of those experiments and musical tradition. It could be that music has developed because of sensory constraints, or it could be that the upper limit of pitch originates from exposure to the norms of music. As will be discussed in the Section 2.2.6, there is considerable evidence that tones below and above the 4000/5000 Hz boundary are coded physiologically by different mechanisms, supporting the former alternative.

2.2.4 Frequency discrimination

By measuring the smallest discriminable difference in frequency between two pure tones—referred to as the difference limen for frequency (FDL or DLF)—researchers can investigate sensitivity to pitch whilst manipulating the physical dimensions of the stimuli. A common method of measuring a listener's DLF involves presenting two tones differing slightly in frequency. The order of the higher and lower tones is randomised, and the listener indicates which of the tones is higher in pitch. The DLF for a certain percentage of correct responses over a run of such trials can be estimated using an adaptive procedure, in which the difference in frequency between the two tones is reduced as the listener makes a correct response and increased as the listener makes an incorrect response (the staircase method; e.g., Levitt, 1971; Kaernbach, 1991b). Alternatively, researchers can use the method of constant stimuli. This method involves recording the percentage of correct responses for a number of frequency differences, fitting a psychometric function to the data, and obtaining the magnitude of the frequency difference predicted by the curve for the desired percentage correct. DLFs are commonly plotted either in terms of the absolute frequency difference between the tones (ΔF, in Hz), or as a proportion or percentage of the test frequency $[100 \cdot (\Delta F/F)]$.

A great deal of previous research has measured DLFs in normally hearing listeners (e.g., Emmerich, Ellermeier, & Butensky, 1989; Harris, 1952a, 1952b; Henning, 1970; Moore, 1973a, 1974; Moore & Glasberg, 1989; Nelson, Stanton, & Freyman, 1983; Nordmark, 1968; Rosenblith & Stevens, 1953; Sek & Moore, 1995; Wier et al., 1977). A striking feature of the results of those experiments is that DLFs can be very small. For instance, Moore (1974) measured three listeners' DLFs for 200-ms tones at various test frequencies. For tones at a frequency of 1000 Hz, the mean absolute DLF across the three listeners was 1.8 Hz, a relative DLF of less than 0.2%. For comparison, the frequency difference between a 1000-Hz tone and another higher in frequency by one semitone (the smallest pitch interval used in most music) is over five times larger than the DLF measured in Moore's listeners. Thus, at certain frequencies—provided the tones have a sufficient duration and level—normally hearing listeners are usually remarkably sensitive to frequency differences.

2.2.4.1 The effect of frequency on DLF

DLFs vary as a function of the standard frequency of the test tones. Wier et al. (1977) measured DLFs in four listeners for 500-ms tones with frequencies ranging from 200 to 8000 Hz. Based on their results and those from other studies, Wier et al. estimated that the logarithm of the absolute DLF is related linearly to the square root of the test frequency. Those data are summarised in Figure 2.7. As pointed out by Moore (2003), the theoretical significance of this relationship is still not clear. Moore (1973a) measured three listeners' DLFs for tones (with various durations) with frequencies from 250 to 8000 Hz, and plotted relative rather than absolute DLF against frequency. The results revealed a 'u'-shaped relationship: relative DLFs were small for frequencies between 500 and 2000 Hz—0.13 to 0.15% for one listener—and increased dramatically for frequencies below and, particularly, above this range. For instance, the relative DLF at 8000 Hz was 1.21% for one listener, over eight times larger than his relative DLF at 2000 Hz. Very similar results were reported in a follow-up study by Sek and Moore (1995), which also measured three listeners' DLFs for tones over a range of frequencies. Those studies indicate that the auditory system is most sensitive to frequency differences in a relatively narrow band of frequencies (500–2000 Hz) and poorer at frequency discrimination beyond this range.

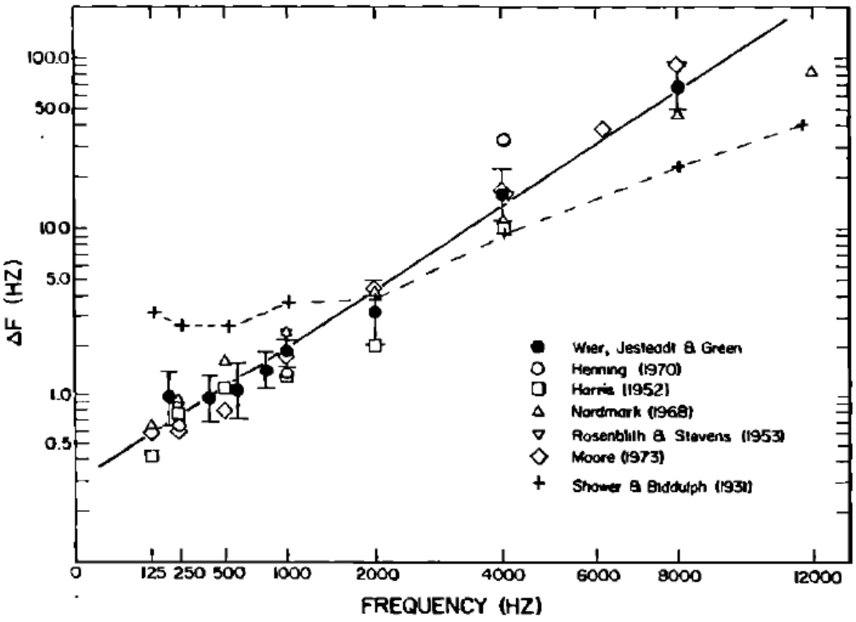


Figure 2.7. Summary of the results of several studies measuring DLFs as a function of frequency. All studies except for Shower and Biddulph (1931) measured DLFs using pure tones. From Wier et al. (1977).

2.2.4.2 The effect of duration on DLF

DLFs also vary as a function of duration, being larger for short tones and smaller for longer tones. As well as measuring DLFs at different frequencies, Moore (1973a) also manipulated the duration of the tones from 6.25 and 200 ms. The results, summarised for one listener in Figure 2.8, not only illustrate a main effect of duration, but also indicate that there is an interaction between duration and frequency on DLF: the reduction in DLF with increasing duration *decreases* with increasing frequency, up to around 4000 Hz. Since the strength of the pitch evoked by a pure tone is also dependent to some extent on an interaction between its duration and frequency (Section 2.2.2.3), a plausible explanation is that relative DLF is dependent simply on the number of periods in the tones to be discriminated. However, Moore's data do not entirely support this explanation because although this may be the case for tones below 4000 Hz, the interaction is not monotonic and the duration effect increases again for tones with frequencies above this value.

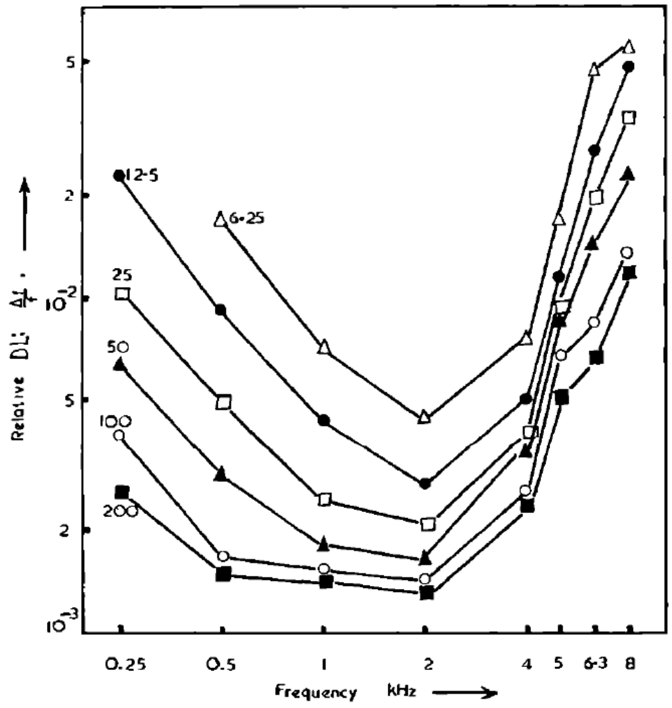


Figure 2.8. DLF for one listener as a function of frequency and duration, in ms. From Moore (1973a).

2.2.4.3 The effect of level on DLF

DLFs also vary with level, although this variation is more modest than variations with frequency and duration. Harris (1952b) measured three listeners' DLFs for tones at different frequencies and at different levels of perceived loudness, measured in phons; the phon scale is a frequency-dependent equiloudness metric that corresponds to dB SPL for a 1000 Hz tone. The mean relative DLF across the listeners decreased with increasing loudness in a monotonic fashion from 5 to 30 phons. For 1000 Hz tones, relative DLF was approximately 0.6% at 5 phons and close to 0.1% at 30 phons. Harris also concluded that the general relationship between DLF and loudness was approximately constant across the frequencies tested (60–4000 Hz). However, two subsequent studies have reported an interaction between frequency and level on DLFs (Nelson et al., 1983; Wier et al., 1977). Both studies found that decreases in sensation level (0 dB SL being a listener's pure-tone detection threshold) increased DLFs more for low-frequency tones than for high-frequency tones. For example, as they increased sensation level from 10 to 40 dB, Wier et al. found a decrease DLFs from 4.3% to 0.5% at 200 Hz, and from 1.5% to 0.9% at 8000 Hz.

2.2.4.4 Frequency and level roving

An added complication when considering the effects of level on DLFs is that a difference in frequency between two tones can often carry a concomitant difference in level or loudness. Differences in level could arise because of limitations in the particular stimulus-delivery apparatus used by the researchers (for instance, a non-flat frequency response of the headphones), and differences in loudness could arise because of the properties of the auditory system. In such cases, listeners might be able to use this additional information to improve task performance. One method of compensation is to introduce random variations in the level of the tones both between and within trials. This technique is commonly called roving (or jittering, if the dimension is varied over a relatively narrow range).

Henning (1966) measured a number of listeners' DLFs for tones with level roving: the level of one of the tones on each trial was fixed at approximately 77 dB SPL, and the other was attenuated by a random amount ranging from 8 to 20 dB. DLFs were measured for 250-ms tones with frequencies between 1000 and 15000 Hz. For tones below 4000 Hz, the results indicated little difference between DLFs measured for roved-level stimuli and DLFs measured in previous studies using fixed-level stimuli. However, a modest increase in DLFs due to level roving was observed by Emmerich et al. (1989), who measured DLFs for tones fixed in level and for tones roved over a 12-dB range in the same listeners. In a third study, Moore and Glasberg (1989) measured fixed- and roved-level DLFs, this time with a level-roving range of 6 dB. The increase in thresholds due to level roving was more modest than that observed by Emmerich et al., and was not statistically significant. In a fourth study, Dai, Nguyen, and Green (1995) measured DLFs for fixed-level tones and for tones roved in level over almost 20 dB. Despte marked individual differences, the mean DLFs across all listeners were almost identical for the fixed and level-roved tones. Taken together, those results show that at most DLFs measured with level roving are only marginally larger than those measured without level roving, at least for tones below 4000 Hz. This suggests that in previous experiments listeners based their discriminations primarily on cues associated with tone frequency rather than on any level and/or loudness cues.

In contrast to the modest effects of level roving, a number of previous studies have demonstrated that frequency roving is particularly deleterious to DLFs. Harris (1952a) measured DLFs in two groups of listeners: the first group (32 listeners) were tested with a fixed standard frequency of 1000 Hz, and the second group (12 listeners) were tested with a roved standard that was randomly varied between 950 and 1050 Hz. The mean DLF across the roved-standard group was approximately 30% larger than the mean DLF across the fixed-standard group[6]. However, the procedure used in that study was unusual because Harris compared DLFs across large, unequal groups, and the listeners were not trained extensively before testing. Using the method of constant stimuli, Jesteadt and Bilger (1974) measured sensitivity in d' to a 6-Hz frequency difference in the same four listeners. The standard frequency of the tones within a run of trials was either fixed at 1000 Hz, roved from 980 to 1010 Hz, or roved from 795 to 1260 Hz. Performance was best on average for the fixed standard ($d' = 2.95$), worse for the narrowly roved standard ($d' = 2.34$), and worst overall for the more widely roved standard ($d' = 1.84$).

Two more studies employed frequency roving over much wider ranges. Demany and Semal (2005) measured d' in four listeners for frequency differences between two very short tones (always containing either 6 or 30 cycles). In one experiment, the relative frequency difference between the tones was fixed on every trial, but the absolute frequencies of the tones themselves roved over the range 400–2400 Hz. In a second experiment, the frequency of the second tone in each trial was fixed at 1000 Hz. Prior to the main experiments, the frequency difference corresponding to a d' of 2.0 was estimated in each listener. The results revealed a considerable effect of frequency roving. On average, the relative frequency difference had to be set approximately 2.3 times larger in the first experiment than in the second experiment in the 30-cycle conditions, and approximately 4.4 times larger in the 6-cycle conditions. Amitay et al. (2005) measured DLFs in three groups of listeners. The groups differed in the training they received: the first was trained using stimuli in which frequency was fixed; the second with stimuli roved from 900 to 1100 Hz; and the third with stimuli roved from 570 to 2150 Hz. In their final training sessions, the mean relative DLF in the fixed-stimuli group (approximately 0.4%) was smaller than the mean DLF in the roving-stimuli groups. However, in that study the mean relative DLF was approximately 1% in the narrow-roving group, and 0.8% in the wide-roving group. The authors suggested that this result indicated a nonlinear relationship between relative DLF and the size of the frequency-roving range, but inspection of the variance between listener's DLFs within each of their groups reveals

---

[6] That study actually measured DLFs whilst also varying the duration of the silent interval separating the tones in each trial, and it did not measure the all of the same durations in both groups. The figure of 30% was derived by comparing the conditions with the shortest duration used to test both groups (3 s).

that the individual differences were substantially larger in the roving-stimuli groups than in the fixed-stimuli group. This finding suggests that the influence of frequency roving on listeners' DLFs is subject to considerable individual differences, with some listeners performing more poorly because of frequency roving than others.

2.2.4.5 Automatic frequency discrimination

A remarkable finding in the literature suggests that it is possible to discriminate a difference in frequency between two pure tones even when the frequency of the first is not perceived consciously. Demany and Ramos (2005) presented listeners with sums of five synchronous pure tones (inharmonic chords) separated by frequency intervals that varied randomly between 6 and 10 semitones. Since the components of the chords were presented at the same level and were not separated in space, listeners generally could not hear out the individual components and reported hearing the chords as single objects. On each trial, a silent ISI and an isolated comparison tone followed each chord. In the 'up/down' condition, the isolated tone was a semitone above or below (at random) one of the three middle components of the chord (at random again), and the task was to judge whether the tone was higher or lower in frequency than the chord component. In the 'present/absent' condition, the comparison tone was either identical to one of the three middle components or halfway between two components, and the task was to judge whether it was present in the chord or not. Figure 2.9 shows the results obtained from 11 listeners in the up/down and present/absent conditions. In the present/absent condition, performance was generally poor; this reflects the fact that although the chord components were certainly resolved in the listeners' auditory systems, it was essentially impossible to hear them out individually. The components were fused at a central level of the auditory system, and for this reason they produced on each other an informational masking effect (Section 2.3.2). By contrast, performance in the up/down condition was much better and d' was greater than 1 in most of the listeners tested. This somewhat counter-intuitive result suggests that listeners are able to perceive a frequency or pitch change without necessarily perceiving the frequency or pitch of both tones. A follow-up experiment contained a third condition—'present/close'—which was identical to the present/absent condition except that when the tone was not present in the chord, it was 1.5 semitone above or below (at random) one of the three intermediate components. Figure 2.9 also shows the results from that experiment. Performance was better in the present/close condition than the present/absent condition, contrary to the predictions of a model based on signal detection theory (D. M. Green & Swets, 1966; Macmillan & Creelman, 2005), which assumed that it would be more difficult to discriminate 'close' and 'present' trials simply because they were more similar than 'absent' and 'present' trials. The authors posited an alternative theory, wherein listeners based their decisions on the output of

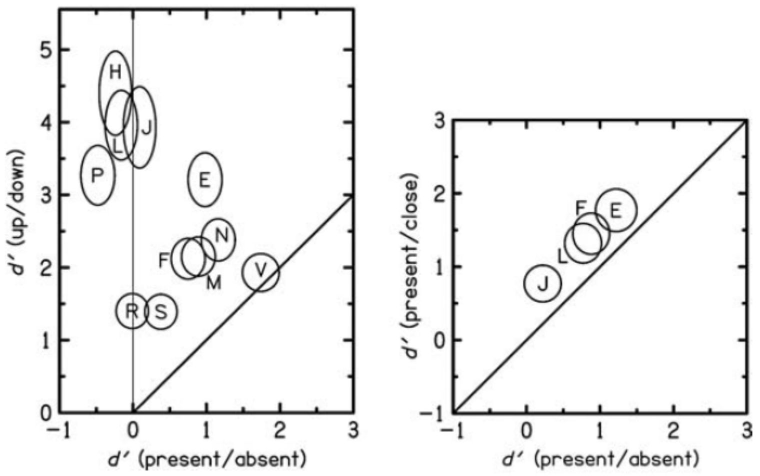


Figure 2.9. Results from Demany and Ramos (2005). The left-hand panel shows d' for 11 listeners in the present/absent and up/down conditions. Each ellipse or circle is centred on the d' values measured in the two conditions in a given listener, and its surface represents a 95% confidence area. Oblique lines indicate where the ellipses could be centred if d' were identical in the two conditions. The right-hand panel shows d' for four listeners in the present/absent and present/close conditions.

automatic frequency-shift detectors (FSDs). Some FSDs are activated only by upward shifts, whilst others are activated only by downward shifts, and within each subset, an FSD responds more strongly to small shifts (e.g., 1-semitone) than to larger shifts. When both sets of FSDs are active simultaneously—as was presumably the case in each experimental condition—the perceived shift is in the direction corresponding to the sum of the total activation. Since an 'absent' trial would have activated both sets of FSDs equally, this sum would be zero, or close to zero assuming internal noise. Subsequent experiments revealed that the hypothesised FSDs worked effectively over a range of silent ISIs separating the chord and comparison tones, and when the second rather than the first tone was buried in the chord. More recently, Demany, Pressnitzer, and Semal (2009) investigated the tuning properties of FSDs using the up/down paradigm and found that d' as a function of the frequency difference could be described accurately by a scaled gamma probability density function (Figure 2.10). This analysis suggests that FSDs respond maximally to frequency shifts of around 120 cents.

Note that in Demany and Ramos' (2005) original investigation, although all of the listeners tested in the first experiment were able to perform above chance in the up/down condition, three listeners were also tested but dropped before the experiment proper. An unpublished experiment I conducted as an undergraduate tested 48 untrained listeners in the basic up/down paradigm and variants of it that contained fewer chord components and wider frequency intervals between the components. Only about one-third of those tested achieved d' scores greater than

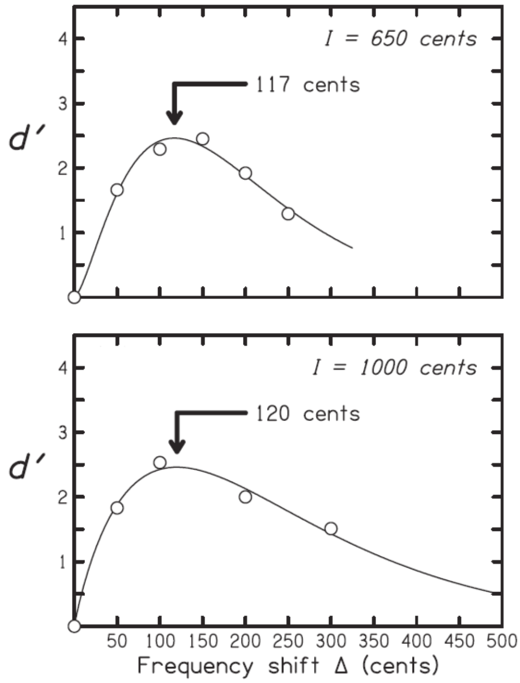


Figure 2.10. Tuning properties of FSDs. The circles are the average values of d' obtained from seven listeners at various sizes of frequency difference between the target chord component and comparison tone (x-axis), and when the components of the chord where separated by 600- (top panel) or 1000- (bottom panel) cent intervals. The curves are best-fitting functions, and the frequency differences where the functions reach their maxima are noted. From Demany, Pressnitzer, and Semal (2009).

1 in the original five-component condition[7]. These observations cannot be taken as evidence against the *existence* of FSDs, but do suggest that not everyone is able to utilise the output of FSDs in this kind of experiment.

2.2.5 Memory for pitch

In most of the experiments described so far in this chapter, a silent ISI of a few hundred ms separated the tones in each trial, and the listeners must have obviously retained in some form of memory information about the first tone over this period for successful task performance. In a model initially proposed to account for aspects of intensity perception, Durlach and Braida (1969) suggested that there are two strategies in which memorised information can be used: a sensory-trace mode and a context-coding mode. In sensory-trace mode, a percept is compared directly to the memory trace of the previous percept. In context-coding mode, a percept is compared with a set of memory traces of previous percepts, including possibly quite ancient traces, and the outcome of the comparison is represented by a verbal label (e.g., ‘halfway between x and y’). The authors pointed out that context coding is

---

[7] One of those who was unable to perform well in this condition was the author, despite having had the most practice in the procedure and—as demonstrated in later chapters of this thesis—having very small DLFs.

necessarily used in tasks that require absolute judgments about single stimuli, but may also be used in discrimination tasks if it is profitable to do so. For instance, context coding would presumably be the most efficient strategy if a listener has to make a same/different judgment on two stimuli separated by a 24-hour ISI, since a verbal label can be perfectly memorised for a long time. Conversely, the sensory-trace mode would be more efficient in a pitch-discrimination experiment involving wide frequency roving, whereby only a comparison to the last trace in memory would be useful.

Setting aside issues of listening strategy, what is the nature of the auditory 'trace' described by Durlach and Braida (1969)? Considerable prior research has studied auditory memory and in particular memory for pitch[8]. This work has suggested that traces are retained in numerous qualitatively different stores depending on the time over which the information is required, and that there are separate and independent stores for different attributes of auditory sensation, including pitch.

2.2.5.1 Pre-perceptual memory

There is compelling evidence that in the visual domain there are at least two types of short-term sensory memory: a short-lived but high-capacity iconic memory that is tied to spatial position and very sensitive to masking; and a more enduring but limited-capacity short-term visual memory that is not tied to spatial position and less sensitive to masking (e.g., Phillips, 1974). A phenomenon called auditory backwards recognition masking (ABRM) suggests—somewhat controversially—that a similar distinction also exists in audition. In his initial experiment, Massaro (1970a) instructed three listeners to identify as high or low a 20-ms pure tone of either 870 Hz (the high tone) or 770 Hz (the low tone). On each trial, one of the two test tones was presented, and followed by a 500-ms tone of 820 Hz after an ISI ranging from 0 to 500 ms. Before data collection, the listeners were trained in the task for about 15 h. The results (shown in Figure 2.11) indicate that performance improved steadily as the ISI increased from approximately 40 ms to approximately 250 ms, and then plateaued.

Since the test tones were always clearly audible in Massaro's (1970a) experiment, poor performance at small ISIs was attributed to a difficulty recognising their pitch and not simply to a difficulty detecting them. This was confirmed by Bland and Perrott (1978), who contrasted the effects of ABRM on two different tasks. One was the same as Massaro's original task, and in the other a single 10-ms test tone was presented in half of the trials and listeners indicated whether this tone was present or absent. The results were qualitatively different in the two tasks, and suggested that ABRM did not affect the latter at all. According to Massaro and colleagues, the results from ABRM experiments provide evidence for two short-term auditory memory stores (Massaro, 1972; Massaro & Loftus, 1996). When a listener hears a

---

[8] Probably due in part to the obvious role of memory for pitch in music (see Deutsch, 1999).

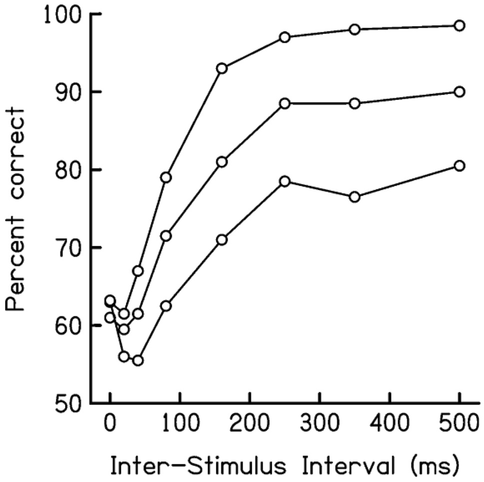


Figure 2.11. Accuracy of correctly identifying a pure tone as high or low in pitch as a function of the ISI in an ABRM experiment (Massaro, 1970a). The lines present data from three different listeners. From Demany and Semal (2008).

short sound, a representation of the sound is created in the auditory equivalent of iconic memory, referred to as pre-perceptual auditory memory (PPAM). A sound begins to be encoded into this memory system at its onset (or soon thereafter) and is 'read out' progressively into the more enduring short-term auditory memory (STAM) store. After approximately 250 ms the transfer to STAM is more or less complete. If a second sound is presented less than 250 ms after the first, the analysis of the first is interrupted and cannot be resumed because it has been replaced by the second in PPAM.

Subsequent experiments have suggested that other perceptual judgments in addition to pitch—including loudness, duration, timbre, spatial position, and speech distinctions—are also affected by ABRM (see Massaro & Loftus, 1996). There is a trend in those studies for ABRM to plateau at approximately 250 ms, suggesting that PPAM is a general auditory memory module and is irrespective of the particular features of the stimulus being encoded. If this inference were true, however, then the properties of the masking tone should also have little effect on ABRM. Sparks (1976) found that ABRM did not occur when the masker and test tones were different in frequency by more than 10 Hz, or when noise maskers (broadband or band-limited to the spectral region of the test tones) were used, implying that ABRM occurs only when the masking sound is perceptually similar to the test tones. Moreover, a number of researchers have argued that the type of psychoacoustical procedure used to study ABRM can influence the results in a manner inconsistent with the notion of a basic PPAM store (C. S. Watson, Kelly, & Wroton, 1976; Yost, Berg, & Thomas, 1976).

Another criticism of studies that support ABRM is that the subjects were usually required to make absolute judgments. On each trial, the percept evoked by the test tone had to be compared to one or more representations in a longer-term memory store. Therefore, the deleterious effect of the masker may have occurred while the test tone was being analysed in PPAM, or while the percept (stored in STAM) was being compared to the long-term representations and a decision was being made. To remove this ambiguity, Massaro and Idson (1977) tested listeners in the traditional, two-alternative forced-choice (2AFC) paradigm: two successive tones differing in frequency were presented in each trial, and listeners indicated whether the second was higher or lower than the first. The tones were 20 ms in duration and separated by a variable ISI. In this situation, again, performance increased as the ISI increased, up to at least 250 ms. Finally, the study by Demany and Semal (2005) described earlier supports the existence of PPAM, but suggests that its capacity may not be fixed at 250 ms. In their experiments, the two tones presented on each trial (either 6 or 30 sinusoidal cycles) were separated by an ISI that varied between runs of trials. When the frequency of the first tone was roved, d' increased rapidly with increasing ISI, then decreased more slowly. The ISI for which d' was maximal was not the same for the two classes of stimuli: the optimal ISI was approximately 400 ms for the 6-cycles tones and approximately 1 s for the 30-cycles tones.

2.2.5.2 Short-term memory

If two tones are separated by a relatively long ISI—for instance, 5 or 10 s—frequency discrimination is more difficult than if the ISI is shorter, particularly if frequency roving is used (Bull & Cuddy, 1972; Harris, 1952a). The degradation observable with silent ISIs and a roving procedure is of interest because in this case the influence of context coding is probably minimal, and so the data are likely to reflect the decay of the sensory trace in STAM. Kinchla and Smyzer (1967) modelled this decay as a ‘random walk’ process, wherein the trace is randomly modified at a constant rate. After a time t, therefore, the trace differs from the original representation by a quantity that can be considered as a Gaussian random variable with a mean of 0 and a variance proportional to t. Although more recent studies have suggested that sensory decay is not, as the model suggests, independent of the properties of the stimulus (Demany, Montandon, & Semal, 2005), and that the process may not be random at all (Gold, Murray, Sekuler, Bennett, & Sekuler, 2005), Kinchla and Smyzer (1967)’s model remains a convenient approximation for the decay of pitch in STAM.

There appears to be little benefit of rehearsal during a silent ISI on the retention of a pitch trace in STAM. If listeners are instructed rehearse overtly through humming, frequency discrimination actually becomes slightly worse than if the listeners were silent during the ISI (Kaernbach & Schlemmer, 2008; Massaro, 1970b). One study suggested that covert rehearsal might improve retention, since performing a task that prevented rehearsal over the silent ISI was deleterious to performance (Keller, Cowan, & Saults, 1995). However, those results are problematic because the listeners had to depress buttons during the distracting task, which presumably produced audible sounds. Demany, Clément, and Semal (2001) found that when listeners were instructed to make any one of three perceptual discriminations between tones in a trial (frequency, modulation frequency, or level), providing a visual cue for the particular dimension to be discriminated during the ISI (before the listener could otherwise determine which dimension this was) did not improve performance. Those results suggest that retention is unaffected by rehearsal or attention after the formation of the trace. However, results from other experiments show that frequency discrimination can be improved if attention is directed towards a target before or during its presentation. For instance, Demany, Montandon, and Semal (2004) showed that retention is improved if during the presentation of the target, attention is focused on that tone rather than an irrelevant tone presented simultaneously. In that case, the focus of attention on the target presumably led to a more accurate representation of its pitch in STAM.

Experiments wherein relatively long ISIs were filled with intervening sounds suggest that there may be a special store for pitch information within STAM. In a seminal experiment, Deutsch (1970) instructed 12 listeners to indicate whether two pure tones selected from a chromatic musical scale (a standard and a comparison) and separated by a 5-s ISI were the same or different in pitch. The task was trivial for the listeners when the ISI was silent. However, accuracy was much poorer on average when six extra random tones (chromatic notes selected from the same octave) were presented during the ISI. When spoken numbers (1 to 12) replaced the intervening tones in other conditions, performance was again close to perfect, regardless of whether or not the listeners additionally had to recall the numbers they heard during the ISI. This result suggests that there is a specific store for pitch in STAM, which is not degraded generally by other acoustic information. In Deutsch's original experiment, care was taken so that none of the random intervening tones was ever the same as the standard or the comparison tones. In a long series of follow-up experiments, Deutsch and colleagues manipulated the properties of the intervening sequence, controlling among other things whether or not the intervening tones contained the standard or the comparison or both, and where within the random sequence they occurred. The results are described in detail in Deutsch (1999), and are essentially consistent with two ideas: that pitch traces stored in STAM are attenuated but never destroyed completely by the subsequent presentation of other pitches, and that a trace is attenuated maximally by other pitches that are close to but different from it. These rules can lead to complex patterns of inhibition and disinhibition. For instance, the trace of one intervening tone may inhibit the trace of another intervening tone, in turn *dis*inhibiting the trace of the standard and improving performance (Deutsch & Feroe, 1975).

If a pitch-specific store exists in STAM, the other perceptual properties of the intervening sounds should have little impact over the retention period. A number of studies have suggested that this is indeed the case. Semal and Demany (1991) tested eight listeners in a paradigm similar to the one used by Deutsch (1970), but manipulated the timbre of the intervening tones: in some conditions they were simply pure tones (the same as the standard and comparison tones), and in others they were harmonic complex tones comprising the second to ninth harmonics of a fundamental frequency. The type of intervening tone did not significantly affect the results. In a follow-up study, Semal and Demany (1993) also found that STAM for pitch was similarly unaffected by differences in the level and the amplitude envelope of the intervening tones. Those results strongly infer the existence of a pitch-specific STAM module (see also Krumhansl & Iverson, 1992).

An alternative interpretation of the results discussed above is that STAM may not have a unique store for pitch, and instead that STAM is able to encode and retain information successfully from any one particular dimension. This view is supported by experimental data suggesting that at least some aspects of timbre can be retained in a similar way to pitch, with no deleterious effect of intervening tones with different pitches (Starr & Pitt, 1997). An experiment conducted by Clément, Demany, and Semal (1999) sheds some light on this hypothesis. The authors compared the decays of pitch and loudness traces by first measuring listeners' difference limens for frequency and level using pairs of pure tones separated by a silent ISI of 500 ms. The corresponding physical differences in cents or dB were then consistently presented with four values of ISI: 0.5, 2, 5, and 10 s. In the case of level discrimination, d' markedly decreased when the ISI increased from 0.5 to 2 s, but was not further reduced when the ISI was longer. In the case of frequency discrimination, the decline of performance as a function of ISI duration was significantly less abrupt. The divergence suggests that pitch and loudness are processed in separate modules of auditory memory. However, this result could have arisen because STAM is for some reason particularly poor at retaining loudness information. To my knowledge, experiments have not been conducted to determine if memory for loudness is robust against the effects of intervening tones, and the decay rates of traces for aspects of timbre have not been ascertained.

2.2.5.3 Learning

A listener's performance in any psychoacoustical task typically improves during the initial stages of the experiment, and is generally not at asymptote even after a period of practice sufficient to ensure that they are completely familiar with its requirements. In many circumstances, performance continues to improve slowly over many hours of listening. Robinson and Summerfield (1996) referred to this as perceptual learning, as opposed to simply procedural learning. For frequency discrimination using pure tones, learning is strongly frequency-specific. Demany and Semal (2002) trained eight listeners—two were discarded because their performance was very good initially—in a frequency-discrimination task. The standard frequency was always 3000 Hz during training, and tones were only presented monaurally. After approximately 11000 trials, the standard frequency of the tones was switched to either 1200 or 6500 Hz. It was hypothesised that the training might generalise to 1200 Hz more effectively than to 6500 Hz, since the lower frequency might be encoded by the same mechanism that would encode a tone of 3000 Hz (Section 2.2.6). However, analysis of listeners' regression slopes—indicative of the rate of learning over time—indicated that new learning had occurred when the frequency was switched; in other words, learning had *not* generalised from 3000 Hz to either 1200 or 6500 Hz (see also Demany, 1985; Irvine, Martin, Klimkeit, & Smith, 2000; Wright, 1998). By contrast, when the frequency was kept the same but the tones were presented to the other ear, new learning did not occur. In a second experiment, listeners were trained using a standard tone that was a harmonic complex with a missing fundamental corresponding to either 100 or 500 Hz, and tested after training using pure tones with the same pitch. Learning did not generalise from complex to pure tones. Those results suggest that learning is possibly also timbre-specific[9].

What mechanism could account for long-term perceptual learning? A number of studies from the animal physiological literature suggest that the training process can modify the neural correlates of the stimuli. For instance, Fritz, Shamma, Elhilali, and Klein (2003) trained ferrets to detect a pure tone with a specific frequency among sounds with broadband spectra, and assessed simultaneously the spectro-temporal response field of neurons in the animals' primary auditory cortex. They found that the behavioural task swiftly modified neural response fields, in such a way as to facilitate perceptual detection of the target tone. Some of the changes in receptive fields persisted for hours after the end of the task. In another study, Recanzone, Schreiner, and Merzenich (1993) trained owl monkeys in a frequency-discrimination task using pure tones with a fixed standard frequency. Discrimination performance improved, and as in human data, this improvement was limited to the frequency region of the standard. The authors then found that in the monkeys' primary auditory cortices, the training affected the neural responses to pure tones. In particular, the cortical area responding to the standard frequency was abnormally large. A control monkey passively exposed to the same stimuli did not exhibit this widening. Those results suggest that long-term training results in the physiological tuning of the auditory system to the test stimuli in fully developed adults (cf. Brown, Irvine, & Park, 2004).

Functional imaging studies provide evidence for similar cortical plasticity through training in humans. Cansino and Williamson (1997) tested three listeners in a

---

[9] It is not clear if the timbre-specificity came about because the complex tones lacked the fundamental. It could be that for learning to be effective, listeners must be exposed to energy at the test frequency, rather than the training and subsequent test stimuli simply being similar in timbre. Complex tones with their fundamental, manipulated in timbre in other ways (e.g., the relative amplitude of the odd- and even-numbered harmonics, or the number of unresolved harmonics), might provide a means to test this hypothesis.

frequency-discrimination task using pure tones whilst recording the extra-cranial magnetic fields evoked by the stimuli using magnetoencephalography (MEG; see Section 2.4). In the early stage of testing whilst performance was at chance level, no significant differences in the average evoked fields were found between correct and incorrect responses. When performance improved to at least 75% of trials correct, the pattern of the magnetic field differed significantly between correct and incorrect responses during the first 70 ms following the onset of the second tone, more than 100 ms prior to listeners indicting their judgment by a button press. Similarly, Menning, Roberts, and Pantev (2000) measured the magnetic fields evoked by stimuli in an oddball paradigm. Listeners heard long sequences of pure tones comprising 70% standards (1000 Hz) and 30% deviants (1020 Hz), making a response whenever they heard one of the deviant tones. The listeners were trained on the task every day for three weeks, and were scanned using MEG on the first session, the middle session, the last session, and again three weeks after the last session of training had ended. The detection of a deviant stimulus amongst a sequence of standards often creates a particular kind of pattern in neurophysiological data called mismatch negativity (MMN; Section 2.4.2.4). The MMN from the deviant stimuli increased in amplitude during training, but had decreased by three weeks after training. The results suggest a reorganisation of the cortical representation for the trained frequencies. Similar plasticity through training has even been observed in the frequency-following response, a putative measure of auditory processing in the brainstem using electroencephalography (EEG; Carcagno & Plack, 2011).

2.2.6 Place versus temporal mechanisms

Classically, there have been two theories of how information about sound frequency is coded by the auditory system. Since frequency is the primary determinant of pure-tone pitch, it is tempting to refer to these as theories of pitch. However, these theories are not applicable to complex tones or periodic noise, whose spectra do not necessarily contain energy at the frequencies corresponding to their pitches. The challenge of producing a universal pitch theory or model is formidable considering the disparate nature of pitch-evoking sounds. For a thorough review of both historical and contemporary pitch models, see de Cheveigné (2005).

The cochlea is the earliest point at which frequency selectivity is apparent in the auditory system. Sound energy at different frequencies leads to different patterns of excitation along the basilar membrane due to the changing physical properties along its length (Figure 2.12), and the precise excitation pattern could be used to determine tone frequency. This is called place theory, or sometimes rate-place theory since in terms of neural activity frequency is represented by the firing rate of neurons responding to excitation at different places along the basilar membrane. A number of early models have attempted to use place theory to account for the results of studies of pitch matching and frequency discrimination (e.g., Maiwald, 1967; Whitfield, 1967; Zwicker, 1970; Zwicker & Fastl, 1990). Arguably the most comprehensive of those models was described by Zwicker (1970). Zwicker's model posits that any kind of change in an auditory stimulus is detected because of the concomitant change in the level of excitation (in dB) anywhere along the basilar membrane. In the case of a frequency difference, detection will take place at the point of steepest slope on the low-frequency edge of the excitation pattern. Thus, a 100-Hz tone is perceived as different to a 200-Hz tone because their excitation levels differ at this position on the basilar membrane.

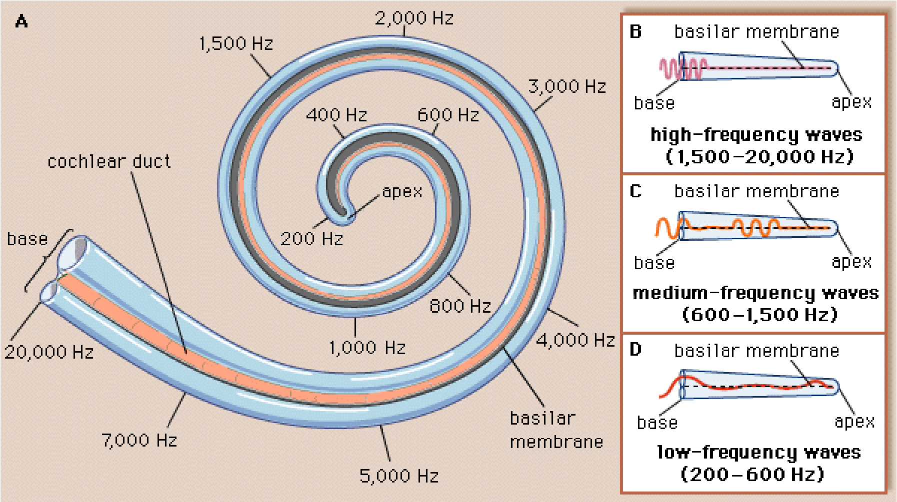


Figure 2.12. Schematic illustration of the analysis of sound frequencies by the basilar membrane. (a) The basilar membrane becomes progressively wider and more flexible from the base of the cochlea to the apex. As a result, each area of the basilar membrane vibrates preferentially to a particular sound frequency. (b) High-frequency sounds cause maximum vibration of the area of the basilar membrane nearest to the base of the cochlea; (c) medium-frequency waves affect the centre of the membrane; (d) and low-frequency waves preferentially stimulate the apex of the basilar membrane.

A number of authors have argued that place theory cannot explain many of the effects discussed earlier in this chapter. For instance, Moore (1973a) argued convincingly that place theory fails to account for the parametric effects of frequency and duration on DLF for tones below 5000 Hz. Short-duration tones present a challenge for place models because their spectra are wider than the spectra of longer tones. For very short tones, the slope of their spectra may be less steep than the slope of their excitation patterns, placing a physical limitation on the best performance that can be expected from a place model. Based on the assumption that a listener can detect a 1-dB change in excitation level, Moore calculated that values of $d \cdot \Delta F$ smaller than or equal to 0.24 should not be observed, where d is the duration of the tone in s and ΔF is the change in frequency in Hz between two tones corresponding to a change in basilar excitation of 1 dB. The values $d \cdot \Delta F$ Moore obtained using short-duration tones below 5000 Hz were considerably smaller than 0.24, suggesting that his listeners could not have been using the slope of the excitation pattern on the basilar membrane as a basis for discrimination.

An alternative to place theory is that frequency is represented in the timing pattern of the neural impulses evoked by the sound. This is based on the principle of phase locking, the property of some neurons to fire preferentially at a certain phase of a stimulus. Phase locking has been observed using electrophysiological recording techniques in the sensory systems of many different animals (see Carr, 1993). Phase locking to auditory stimuli occurs in the neural fibres of the auditory nerve (e.g., D. J. Anderson, Rose, Hind, & Brugge, 1971; Gleich & Narins, 1988; Hill, Stange, & Mo, 1989; D. H. Johnson, 1980; Kettner, Feng, & Brugge, 1985; Köppl, 1997; Manley, Köppl, & Yates, 1997; Palmer & Russell, 1986; Sachs, Woolf, & Sinnott, 1980; Salvi, Saunders, Powers, & Boettcher, 1992; Smolders & Klinke, 1986; Weiss & Rose, 1988). Thus, the frequency of a pure tone whose frequency is in the range over which phase locking occurs could be determined exclusively by the temporal firing pattern of the auditory nerve fibres.

For frequencies above 4000–5000 Hz, DLFs are influenced by duration and level in a manner seemingly consistent with place theory, and auditory-nerve recordings indicate that the ability of a fibre to phase lock to a pure tone breaks down at high frequencies (Figure 2.13). In many animals, phase locking is virtually non-existent past 4000 Hz, although there is considerable variability between species. This suggests that there may be an effective upper-frequency limit of temporal coding. A number of the other studies discussed in this chapter lend credence to the idea that there is a switch in the way frequency is determined at around 4000–5000 Hz. First, frequency discrimination deteriorates dramatically as frequency is increased above approximately 4000 Hz (Moore, 1973a; Sek & Moore, 1995). Second, the effect of tone duration on DLF increases as the frequency is raised above 4000–5000 Hz (Moore, 1973a). Third, absolute pitch, the perception of musical melody, and musicians' ability to recognize musical intervals breaks down above 4000 Hz (Section 2.2.3).

In summary, there is strong evidence that the representation of pure-tone frequency is not governed by the same mechanism across the entire audible spectrum. A boundary appears to exist somewhere around 4000–5000 Hz, such that frequencies below this boundary are determined by the timing of phase-locked auditory nerve fibres, and frequencies above this boundary are determined by the pattern of excitation on the basilar membrane. This value corresponds reasonably closely to the point at which pure tones can no longer carry musically relevant pitch information, and it is tempting to conclude from this that musical pitch is determined purely by temporal coding. There are some reasons for caution, however. The finding that DLFs are affected by level at all complicates a purely temporal account of pitch perception, and it is possible that place and temporal information are combined in some way to

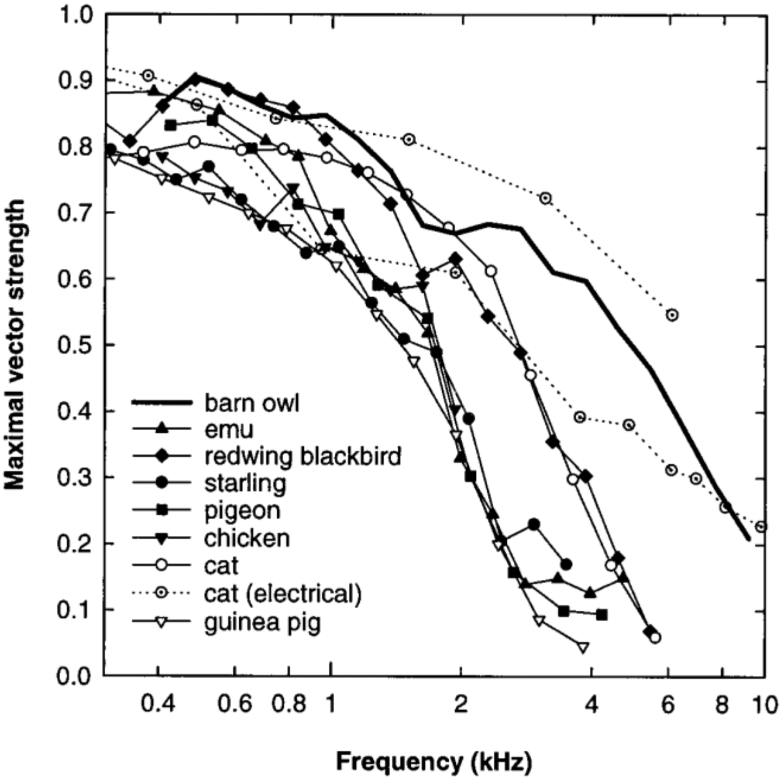


Figure 2.13. A comparison of auditory nerve phase-locking ability (measured in terms of 'maximal vector strength') as a function of frequency in several avian and mammalian species. Data were taken from the following sources: barn owl, from Köppl; emu, from Manley et al. (1997); redwing blackbird, from Sachs et al. (1980); starling, from Gleich and Narins (1988); pigeon, from Hill et al. (1989); chicken, from Salvi et al. (1992); cat, from D. H. Johnson (1980); guinea pig, from Palmer and Russell (1986). Two additional sets of electrophysiological data are shown auditory nerve fibres in the cat in response to direct electrical stimulation (R. Hartmann & Klinke, 1987; Dynes & Delgutte, 1992). From Köppl (1997).

produce a sensation of pitch[10].

2.3 Individual differences

Studies that have investigated individual differences in auditory perception fall into two categories: those that have taken a broad-brush psychometric approach and those that have concentrated on measuring performance in a single auditory task. The former approach has a very long history in psychology, most notably in the areas of intelligence (e.g., Neisser et al., 1996; Spearman, 1904; Thurstone, 1938) and personality (e.g., Digman, 1990). The goal of those experiments was usually to determine the number and nature of independent dimensions—or factors—along which individuals vary, and they involved running large numbers of subjects on batteries comprising many tests and analysing the results using data-reduction techniques. Some studies of this kind have investigated individual differences within auditory perception specifically, and have usually contained at least one test of frequency or pitch discrimination. Such studies using either one of two auditory test batteries are discussed below. The approach of measuring one auditory dimension at a time has been more favoured by psychoacousticians; some of those studies are particularly relevant to the experiments outlined later in the thesis and are discussed in detail here.

2.3.1 The psychometric approach

2.3.1.1 Seashore measures of musical talent

The first large-scale psychometric test battery designed to measure individual differences in audition was devised and developed in the early part of the 20th century by Seashore (1919a, 1919b, 1938). Seashore's interest was in music rather than auditory perception, and the primary motivation behind his psychometric work was to screen for musical aptitude in children. His idea was to have children complete an extensive battery of measures covering the gamut of skills necessary for the appreciation and production of music. Their performance on each measure would be assigned a percentile rank based on normative data, and a 'talent chart' would outline their particular musical strengths and weaknesses [e.g., 'Rosabelle is not of the musical type of mind. She has an average sense of pitch and an inferior sense of intensity, but a rather good sense of time ...'; Seashore (1919b, p. 20)].

All of the Seashore measures were made available commercially, in some cases along with the custom-built apparatus required to administer them (Figure 2.14).

---

[10] Evidence to support this latter point comes from a study using so-called 'transposed' stimuli, which present the temporal information of low-frequency sinusoids to locations in the cochlea tuned to high frequencies (Oxenham, Bernstein, & Penagos, 2004).

The protocol and stimuli (recorded on phonograph record) for five of the perceptual measures were distributed together (Seashore, 1919a). In the first measure—'the sense of pitch'—listeners heard pairs of pure tones and were required to judge whether the second was higher or lower than the first. In the second—'the sense of intensity'—listeners judged whether the second tone was quieter or louder than the first. In the third—'the sense of time'—listeners heard three clicks marking two silent intervals, and judged whether the second interval was longer or shorter. In the fourth—'the sense of consonance'—listeners heard two instances of two pure tones presented synchronously, and judged whether the second instance sounded better or worse (in terms of the two tones sounding smooth or blended as one). In the fifth—'tonal memory'—listeners heard two melodic sequences of pure tones, the second sequence differing from the first in only one note (mistuned up or down), and identified the position of the deviant in the sequence. Additional measures were added later (Seashore, Lewis, & Saetveit, 1960), including another time measure, in which listeners heard two tones different in their duration and judged whether the second was longer or shorter. In a new rhythm measure, listeners heard two tapped rhythmic phrases and judged whether they were the same or different. In a quality discrimination measure, listeners heard two complex tones with same fundamental frequency and duration, and judged whether they were the same or different in timbre (relative amplitude of the higher harmonics). The tests all used a method of constant stimuli rather than measuring thresholds directly, and the magnitude of the stimulus difference to be discriminated was gradually reduced over a fixed number of trials to yield a range of difficulty levels.

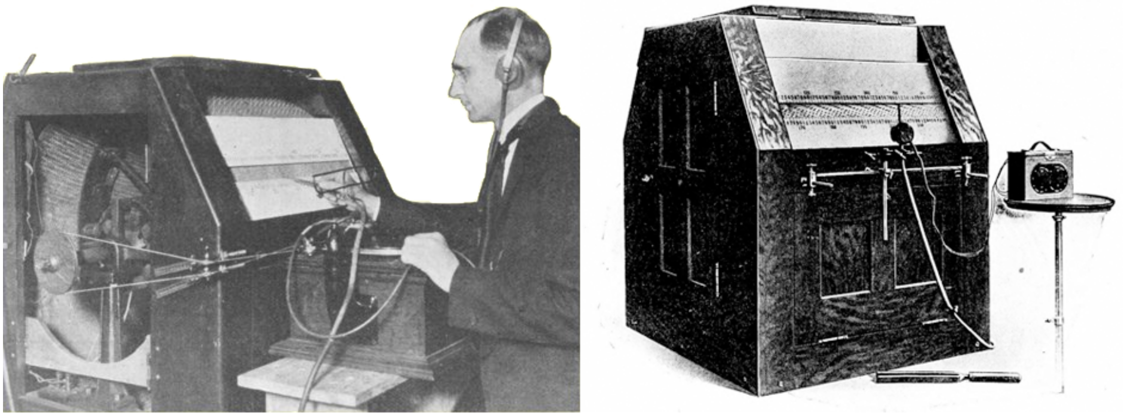


Figure 2.14. Seashore at work (left) with a tonoscope (shown alone on the right), used to analyse the pitch of sounds recorded on disk. From Cary (1922) and Seashore (1919b).

Seashore's approach was atomistic—he clearly viewed each test as measuring an independent perceptual skill. However, he himself never subjected the data to factor analysis. McLeish (1950) tested 100 undergraduate or postgraduate psychology students from the University of Leeds on six of the measures (tonal memory, pitch, consonance, intensity, the first time measure, and rhythm). The tests were modified slightly, in that each contained a 'don't know' response option. Frustratingly little detail was given about how or where the tests were administered, but presumably testing was carried out in groups with the stimuli delivered over loudspeaker from a gramophone record. No practice trials were reported. Exploratory factor analysis revealed two factors: each of the measures loaded on the first factor, and the tonal memory and rhythm measures loaded negatively on the second factor. Closer inspection of the results reveals that the zero-order correlations between performances on the measures were all weak or moderate. The time and rhythm measures correlated moderately with each other ($r = .43$) but only very weakly or not at all with the other measures ($-.12 \leq r \leq .28$). Tonal memory correlated moderately only with pitch and consonance ($r = .43$ and $r = .42$, respectively), and intensity correlated moderately only with pitch ($r = .43$). The listeners also completed some other non-Seashore musical tests (not factor-analysed) and were subjected to a general intelligence test (Cattell, 1949); scores on the Seashore measures on average correlated moderately with performance on the other musical tests but did not correlate with intelligence scores. McLeish took his results to indicate the existence of a general auditory processing or musical ability that was independent of general test taking.

A stronger factor analysis study was actually completed much earlier (Karlin, 1942)[11]. Karlin tested 200 high school pupils from Chicago in a comprehensive 33-test battery, which included all of the Seashore measures described above, plus tests of pure-tone masking, speech intelligibility, musical tempo, male and female voice recognition, and memory for visual images. There were four tests of pitch discrimination in particular: the Seashore measure using pure tones, a test using complex tones, a test using voices, and a test using pure tones with very short durations. Including multiple items designed to measure the same hypothesised construct is very desirable in factor analysis. Performance in proportion correct (PC) was calculated for each item. Listeners were tested in groups of about 25 in a quiet classroom, the auditory stimuli were played from a phonograph record, and responses were made with pen and paper. Since psychometric testing was very popular in the American education system at the time, the pupils' IQ-test scores were available from the school's records and were included in the analysis. The results revealed nine factors (only eight of which could be interpreted by the author) which appeared to be independent since performance on items from one factor on average correlated only very weakly with performance on items from other factors. Of particular relevance to this thesis was that all of the tests of pitch discrimination loaded together onto a single factor regardless of the stimulus, along with the Seashore timbre and tonal memory measures. The correlation between performances on those tests implies that they may rely on overlapping mechanisms; this makes

---

[11] The work was completed as part of Karlin's PhD thesis under the supervision of Thurstone.

sense since they all require a representation of the spectral content of sound in some form or another. The lack of correlations with the other measures—including those measuring masking, timing, visual attention, and speech intelligibility—suggests that spectral discrimination is a relatively independent auditory ability. Consistent with the results of McLeish (1950), neither age nor IQ was a good predictor of performance on any of the tests.

A third study was conducted using 241 adult male prisoners from Colorado State Penitentiary (Stankov & Horn, 1980). The authors' 44-test battery contained the Seashore measures and those from two more musical batteries (Drake, 1954; Wing, 1962), multiple measures of the intelligibility of speech degraded in numerous ways, masking, and memory span for numbers and digits. In the tests that did not involve speech, only the Seashore measures used pure tones; the others used piano tones instead. The listeners were tested in groups of eight to 55 at a time, and stimuli were presented via a tape player. The factor analysis revealed seven factors. The Seashore pitch measure using pure tones loaded onto a factor that included other tests logically involving the discrimination of pitch differences over a short period of time (e.g., indicating whether a sequence of three tones was descending or ascending) or pitches heard simultaneously (e.g., identifying the number of notes within a chord). Unlike Karlin's (1940) factors, however, the Seashore pitch measure did not load on the same factor as the Seashore timbre or tonal memory measures. Instead, those measures loaded on a factor that seemed to be related to comparing three or more sounds in one sequence to three or more sounds in another sequence, indicative possibly of a more long-term auditory memory store (cf. Section 2.2.5). Again, the Seashore measures of pitch, timbre, and tonal memory appeared to be independent of performance on the timing and rhythm measures, and independent of other tasks such as speech intelligibility. A major difference between the results of the study by Stankov and Horn and the previous ones is that IQ score obtained through prison records correlated moderately with some of the auditory factors. However, these correlations are difficult to untangle because the study was conducted with adults from an obviously unrepresentative sample of the general population, and IQ correlated with musical experience.

In summary, the use of the same set of measures in three different factor analysis studies allows their implications for pitch perception to be compared relatively easily. Whereas McLeish's (1940) analysis did not separate pitch perception from the other measures in the Seashore battery, both Karlin (1942) and Stankov and Horn (1980) suggest that there are some sharp divisions in the latent factors underpinning performance on different auditory tasks. Sensitivity to spectral properties—as indexed by pitch and timbre discrimination—may be separate and distinct from speech intelligibility, resistance to masking, and sensitivity to loudness and duration. There are some important discrepancies between their results, however, particularly concerning the role of latent factors that are not auditory-specific, such as those indexed by general IQ tests. All three studies tested large numbers of listeners from specific populations (schoolchildren, students, or prison inmates), several at a time in large rooms, with pre-recorded stimuli presented over loudspeakers. It is therefore reasonable to suspect that although the authors would have attempted to minimise issues like reverberation and background noise, the data collected were considerably less precise than if they had been collected in a laboratory. There is also of course a trade-off between the total number of listeners and tests included in an experiment and the amount of data collected from each individual. The number of trials completed in per test was usually very low (20 to 50) and fewer data per listener would obviously result is a less accurate representation of their ability on any particular task, and less control over learning and practice effects. In light of these issues, then, some differences in their results are to be expected.

2.3.1.2 Tests of basic auditory capabilities

In a review of early factor analysis studies, D. M. Johnson et al. (1987) made the argument for a modern standardised battery of tests. From the tests described in that study, C. S. Watson and colleagues derived a subset of eight tests of basic auditory capabilities (TBAC; first reported use by C. S. Watson, Johnson, Lehman, Kelly, & Jensen, 1982). The TBAC included seven discrimination tests. Each used a three-interval oddity procedure: a standard stimulus was followed by two test stimuli, and listeners judged which of the two was different from the standard. In the pitch-discrimination test, listeners heard a 250-ms, 1000-Hz standard tone presented at 75 dB SPL, and the frequency difference to be discriminated ranged from 2 to 256 Hz. The other discrimination tests measured loudness and duration using isolated pure tones, rhythm using a sequence of 40- to 90-ms pure tones, loudness using tones buried in the centre of a nine-tone pattern, temporal order using pairs of short-duration tones flanked by other tones, and temporal order using syllables rather than tones. In the eighth test, listeners heard nonsense syllables in cafeteria noise and were instructed to identify the sound by choosing one of three written alternatives. For each test, PC was recorded.

In a study testing 52 normally hearing university students, B. U. Watson (1991) found that scores on six of the eight tests (not pattern discrimination or syllable intelligibility) correlated moderately with scholastic aptitude test scores for mathematics ($.30 \leq r \leq .42$). Three of the eight tests (frequency, intensity, and duration discrimination with pure tones) also correlated with WAIS–R IQ scores (Wechsler, 1981) in a group of 24 listeners with different learning disabilities ($.34 \leq r \leq .57$). This finding is at odds with the results from the psychometric experiments using the Seashore battery (see above), and should be treated with caution because scatter plots revealed that the data contained bivariate outliers: given the relatively small number of listeners tested by B. U. Watson, the outliers could have skewed the results substantially. In a later study, B. U. Watson and Miller (1993) tested 94 undergraduates—24 of whom met the criteria for reading disability—on the TBAC along with other tests of speech perception, long-term and short-term memory capacity, and reading skill. The data were analysed using a structural equation modelling approach and most of the TBAC were not included in the final model. However, inspection of the correlation matrix reveals that with the exceptions of the syllable-discrimination test and the syllable-temporal-order test, performance on the TBAC was uncorrelated with any other measure (the eight tests all correlated positively with each other, $0.35 \leq r \leq 0.65$). In two studies testing hearing-impaired older adults (Humes & Christopherson, 1991; Christopherson & Humes, 1992), performance on the TBAC was found to be uncorrelated with hearing loss, and was able to account for a small amount of the variability in a speech identification task using multiple regression (although pure-tone audiometry alone accounted for 72% of the variance). Taken together, the results suggest that the TBAC is measuring abilities that are more or less unrelated to extrinsic factors.

Two studies explored the factor structure of the TBAC (?, ?, ?). Surprenant and Watson carried out a factor analysis study with 93 normally hearing young adults. The authors tested the listeners on the basic TBAC plus three speech discrimination tests in which consonant-vowels, words, or sentences were presented at various signal-to-noise ratios. The analysis revealed three factors: a non-speech discrimination factor including most of the TBAC measures, a speech identification factor including the three new tests as well as the original syllable test, and a temporal-order discrimination factor including syllable- and tone-order subtests. When the factor analysis was repeated with 45 listeners for whom SAT scores and audiometric data were available, four factors were reported: both the speech and non-speech factors were again observed, hearing thresholds loaded on to their own factor, and both verbal and math SAT scores loaded on the temporal order factor (cf. B. U. Watson, 1991).

Finally, the most recently published use of the TBAC by Kidd et al. (2007) has been the largest single factor analysis of auditory abilities, involving 340 listeners. In this experiment, the basic TBAC was expanded by a further 11 tests. The extra tests required listeners to discriminate different perturbations of noise (sinusoidal AMs and ripples), detect and discriminate the length of silent gaps within noise, and identify word, sentence, nonsense-syllable, and environmental sounds in noise. The analysis revealed four factors. One factor—labelled ‘loudness and duration’—included the discrimination tests for loudness, duration, rhythm, 200-Hz AM, and (counter-intuitively) the original TBAC syllable identification test. A second factor included the discrimination of slower AMs (≤ 60 Hz) and gap-in-noise detection. A third factor—‘familiar sounds’—included all of the new identification tasks. The final factor—'pitch and time'—included the pitch, embedded tone, and both temporal order tasks from the original TBAC, plus the new gap- and ripple-discrimination tasks. The authors concluded that their most significant finding was that of an independent ability for recognising familiar sounds; such ability might logically require efficiently retrieving information stored in long-term memory, an efficient strategy for matching sounds to their correct representations, and being able to focus on the most informative spectral/temporal aspects of the stimulus. Interpreting the remaining factors is more difficult. The authors suggest that rather than revealing a distinction between spectral and temporal abilities, the three factors reveal three distinct types of temporal processing abilities. For instance, the 'loudness and duration' factor may reflect the ability to detect overall changes in energy or duration, but it is not clear why such an ability would be involved in syllable identification or rhythm discrimination, variations in which can be independent of overall energy and overall duration. In short, the analysis performed by Kidd et al. suggests that recognising familiar sounds including speech is influenced by processes and abilities that are relatively distinct from those involved in low-level stimulus feature discrimination, probably because the former relies to a much greater extent on long-term memory representations. When the stimuli and tasks are not semantically or linguistically relevant, the distinctions lie along dimensions that are difficult to interpret.

2.3.2 Informational masking

A specific area of the psychoacoustics literature that has discussed individual differences in detail involves the phenomenon of informational masking (IM). The term—first coined by Pollack (1975)—refers to a decrement in listener performance (e.g., elevated thresholds or smaller d') caused by irrelevant sounds whenever the underlying cause of the masking is not energetic in nature. 'Energetic' in this context refers to the type of masking that can been modelled accurately as a function of the amount of masker energy falling close to the frequency of the target stimulus, such as noise within a critical band surrounding a pure tone (e.g., Fletcher, 1940). Thus, IM could be said to occur whenever the amount of masking observed in an experiment is greater than predicted by such a traditional masking model. A number of recent commentaries have questioned the usefulness of the concept of IM, suggesting that the term is too broad (Kidd, G. Jr., Mason, Richards, Gallun, & Durlach, 2008) and even whether the distinction between energetic and informational is appropriate at all (Durlach, 2006). Despite these definitional issues, IM experiments have revealed interesting patterns of individual differences relevant to the experiments reported in this thesis.

The archetypal IM experiment was described by Neff and Green (1987). The authors measured thresholds for detecting a pure tone in three listeners. The listeners judged which of two temporal intervals contained a pure-tone target, which always had a duration of 200 ms and which was fixed in frequency throughout a run of trials at either 200, 1000, or 5000 Hz. In some conditions the intervals were silent except for the target, in others the intervals also contained 60-dB SPL broadband noise, and in others the intervals contained a fixed number of additional pure tones whose summed level was 60 dB. The frequencies of the additional tones were randomly selected between intervals from the range 0–5000 Hz (Figure 2.15). The conditions with the additional tones produced more masking on average than the conditions containing additional noise—sometimes in excess of 50 dB. This result was unexpected, since on a given trial there was only a very small probability that any of the additional tones would fall close to the frequency region occupied by the target, so the amount of masking predicted by traditional models was small. Subsequent IM experiments have reported very similar results when the target was surrounded by a wide protected spectral zone (e.g., 160 Hz) designed to minimize energetic masking effects (e.g., Durlach, Mason, Shinn-Cunningham, et al., 2003; Durlach et al., 2005; Kidd, G. Jr., Mason, Deliwala, Woods, & Colburn, 1994; Neff, 1995; Neff & Callaghan, 1988; Neff, Dethlefs, & Jesteadt, 1993; Oh & Lutfi, 1998; Oxenham, Fligor, Mason, & Kidd, G. Jr., 2003; Wright & Saberi, 1999).

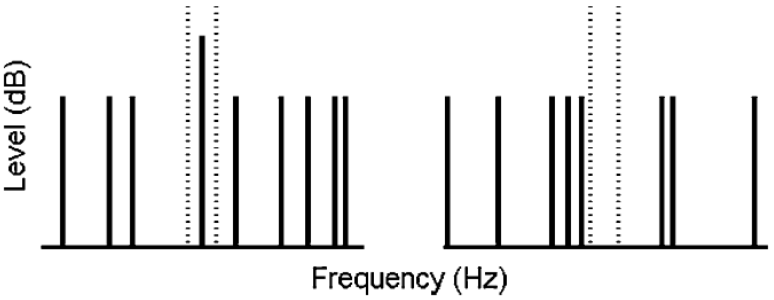


Figure 2.15. An illustration of an example trial from a standard IM experiment. The two panels are the first and second intervals in the trial, respectively, and each line represents a pure tone. The target (in the first interval) is the taller solid line in the centre of the spectrum, while the dashed lines indicate the boundaries of the protected region where masker components are not allowed to fall. From Kidd, G. Jr. et al. (2008).

Neff and Green (1987) concluded that the source of this striking effect must have been the inherently greater degree of uncertainty during trials containing random-frequency tone maskers, since in a second experiment, fixing the frequency of the additional tones between intervals within a trial greatly reduced the amount of masking (see also Neff & Callaghan, 1988). The amount of masking was largest when there were 10–20 additional tones per interval, and masking was smaller when there were more tones (50, 100, or 200). A quantitative uncertainty-based account of IM was provided by Lutfi and colleagues (Lutfi, 1993; Oh & Lutfi, 1998). According to their component-relative entropy (CoRE) model, in an IM experiment a listener focuses his or her attention on a relatively wide frequency band that includes the signal frequency, and makes a response based on whichever interval in a trial contains the most energy within the band. If the listening band is larger than the protected zone, the masker tones would essentially add noise to the decision variable and IM would be observed. The amount of IM was modelled as proportional to the variability in the output of a bank of weighted bandpass filters—whose equivalent rectangular bandwidths were the same as those of auditory filters measured psychophysically (Glasberg & Moore, 1990)—with the weights applied to each filter representing the listening band. The CoRE model allows for the additive influences of energetic and informational effects to be teased apart in an IM experiment, and it has been able to account for the results obtained in many earlier and more recent IM studies, including the non-monotonic relationship between IM and the number of masker components (Figure 2.16).

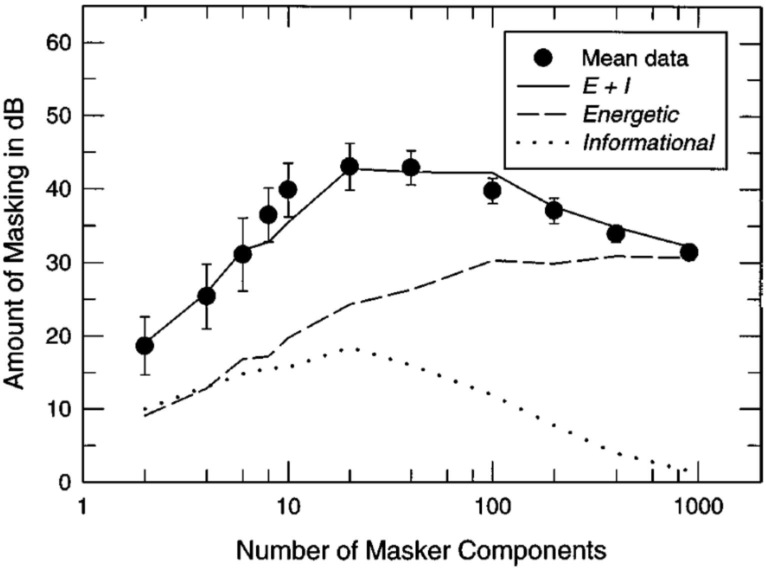


Figure 2.16. Typical results from an IM experiment. The number of frequency components in the masker is plotted horizontally, and the amount of masking is plotted vertically. Group mean data are plotted as filled symbols. The solid line shows the predicted thresholds based on the CoRE model, while the dashed and dotted lines are estimates of the amounts of energetic and IM, respectively. From Oh and Lutfi (1998).

Other authors have claimed that uncertainty may be neither necessary nor sufficient to produce IM (Durlach, Mason, Kidd, G. Jr., et al., 2003). Some studies have found that IM occurs even when the frequencies of the maskers are fixed for a run of trials (Wright & Saberi, 1999; Durlach et al., 2005), and that IM can be reduced substantially if the maskers and target are made perceptually dissimilar in some way (Durlach, Mason, Shinn-Cunningham, et al., 2003; Kidd, G. Jr. et al., 1994; Oh & Lutfi, 2000). These findings suggest that it might be better to describe IM as occurring when listeners are not able to segregate the target from the background, and that increases in uncertainty simply increase the probability that a listener will not achieve segregation on any given trial.

Individual differences in IM are huge. One listener may be highly susceptible to IM, whilst another may be able to detect a target tone amongst pure-tone maskers almost as well as they would in quiet. In a meta-analysis of several previous studies, Neff and Dethlefs (1995) found that the standard deviation of 28 listeners' thresholds when intervals contained two pure-tone maskers was 17 dB SPL. By contrast, the standard deviation for thresholds with a 60-dB broadband noise masker over the same number of listeners was only 2 dB. That study suggests that the two types of masking are controlled by different mechanisms; listeners' thresholds with pure-tone and noise maskers were not correlated (r = .04), and the variability in thresholds over the group decreased as the number of pure-tone maskers increased (i.e., the proportion of masking attributed to IM gave way to energetic masking). The CoRE model provides a convenient explanation for what high- and low-threshold listeners are doing differently in IM experiments. By either altering the weights or the number of summated auditory filters, the CoRE model can predict large individual differences with reasonable accuracy. High susceptibility to IM is caused by a 'holistic' listening strategy—in which listeners monitor a non-optimally wide frequency band—whilst a more 'analytic' strategy—in which the listener is able to ignore a greater amount irrelevant sound energy falling spectrally far from the target—is less susceptible to IM. The spread of the data from the listeners considered by Neff and Dethlefs suggests that individual differences in IM can be characterised as a continuum, with individuals at one end listening more holistically, and those at the other end listening more analytically.

The holistic/analytic characterisation has proven popular in IM research, but what causes a listener to be more holistic or analytic? Neither susceptibility to energetic masking nor quiet thresholds seem to influence susceptibility to IM (e.g., Neff et al., 1993). Neff, Kessler, and Dethlefs (1996) suggested that sex differences might play an important role. In a re-examination of their 49 listeners, they found that females were far more likely to exhibit large amounts of IM than males, despite no significant difference between their thresholds in quiet (sex differences for noise maskers were not reported). Because musicianship logically relies in part on an ability to listen analytically, Oxenham et al. (2003) hypothesised that a greater proportion of musicians would fall towards the analytic end of the continuum. In their first experiment, thresholds were measured in 12 trained musicians and 12 non-musicians. Although there was little difference between the groups in a condition that produced relatively small amounts of IM, the musicians exhibited much smaller thresholds than had the non-musicians in a condition with a greater contribution of IM. Listener sex, age, quiet threshold, and auditory filter shape (measured with a follow-up notched-noise experiment) could not account for the group difference. There is somewhat mixed evidence for a role of training in IM: some studies have shown little improvement in their listeners' masked thresholds over time (Neff & Green, 1987; Neff et al., 1993), whereas others note moderate improvements (Oh & Lutfi, 1998). In either case, there is no evidence to my knowledge of such profound improvements that would change an individual previously considered a holistic listener into an analytic listener. Although psychoacoustic experiments are sometimes very long, an effect of even greater amounts of practice can never be ruled out. The role of musicianship in IM may therefore reflect musical experience being a form of extremely long-term training.

The CoRE model is unable to account for the role of perceptual segregation in IM, so consequently it is not compatible with any individual differences observed in the effectiveness of efforts to reduce similarity. However, researchers have noted that these differences appear to be much less profound than those that were observed for uncertainty effects (at least in adults, cf. J. W. Hall, Buss, & Grose, 2005). Neff (1995) compared the relative release from IM due to reducing uncertainty and reducing similarity by shortening the duration of the signal. Correlations between listeners' thresholds in conditions of maximal versus reduced uncertainty were weaker than correlations between conditions with longer versus shorter signals. In other words, reducing stimulus uncertainty by creating temporal asynchrony between the signal and maskers was more consistently effective than reducing masker uncertainty. A similar conclusion was reached by Durlach, Mason, Shinn-Cunningham, et al. (2003). The conclusion that similarity-based release from IM is more consistent across listeners cannot be drawn firmly because the hypothesis has not received much attention, but it is perhaps relevant that individual differences are rarely commented upon in the auditory scene analysis literature (e.g., Bregman, 1990).

Evidence so far strongly supports the idea that IM is a non-peripheral, attention-based effect. On this basis, we might predict hearing loss to have little impact on IM. Micheyl, Arthaud, Reinhart, and Collet (2000) measured IM in monaural listening conditions in normally hearing listeners and listeners with symmetric or asymmetric hearing impairment. Once the reduced audibility of the stimuli was compensated for, the normally hearing and hearing-impaired listeners did not differ significantly in their susceptibility to IM overall. However, for listeners with asymmetric hearing loss, more IM tended to occur in the ear with the greater impairment. Similar results were reported by Kidd, G. Jr., Arbogast, Mason, and Walsh (2001) in an experiment that measured the degree of similarity-based release from IM in hearing-impaired and normally hearing listeners. Their results suggested that, due to poorer peripheral analysis (e.g., Micheyl & Carlyon, 1998), the hearing-impaired listeners were less able to take advantage of perceptual segregation cues.

Conflicting results were reported by Alexander and Lutfi (2004), who tested hearing-impaired and normally hearing listeners using several degrees of masker uncertainty. In that study, the hearing-impaired group appeared to show less IM than the normally hearing group. The authors explained this result in terms of the SL of the masker components: for hearing-impaired listeners, maskers had lower SLs at the auditory filters, so produced less ensemble variance (uncertainty). This conclusion appears to be supported by a follow-up study in which normally hearing listeners' masked thresholds were small when tested using maskers with SLs more comparable to the hearing-impaired group. It is difficult to reconcile this finding with the previous IM studies that generally show *worse* (not better) performance for hearing-impaired listeners. To estimate the contribution of energetic masking, Alexander and Lutfi used a condition in which maskers were not roved over a run of trials. This may have over-estimated the contribution of energetic masking, given that several authors have observed large IM effects with minimal stimulus uncertainty. Clearly, more research is needed to explain the interaction between hearing impairment and susceptibility to IM.

2.3.3 Individual differences in DLFs

2.3.3.1 Musicianship

Musicianship might be expected to have a spectacular impact on performance in basic experiments involving pitch or frequency discrimination. Indeed, musically trained individuals make excellent test subjects for psychoacousticians since they are more likely to already understand complex acoustic terminology, require less training, and find rigorous experimentation less tedious. However, surprisingly few studies have actually compared musicians' and non-musicians' DLFs. Spiegel and Watson (1984) tested 30 members of a symphony orchestra and a group of non-musician controls drawn opportunistically from among the graduate students, researchers, and staff at the authors' institution. The groups were approximately but not formally matched in gender and age. Among other tests, DLFs were measured using a same-different task: on each trial, listeners heard either two identical 300-ms pure tones (separated a 500-ms ISI) or two different tones, and the instructions were to indicate whether the two tones were different or not. The average relative DLF was 0.37% in the musician group, and 0.68% in the non-musician group. The difference between the means was smaller than Spiegel and Watson had expected, and the groups showed considerable overlap; approximately half of their non-musicians had DLFs that were in the same range as the musicians. This result suggests that musicians and non-musicians do not differ as greatly in their sensitivity to frequency as might be expected superficially. A possible limitation of the study is that several members of the control group had taken part in similar experiments before, and this prior experience is likely to have improved their DLFs (Section 2.2.5.3).

The importance of controlling for practice effects was highlighted in another study by Kishon-Rabin, Amir, Vexler, and Zaltz (2001). The authors measured DLFs in 16 musicians—including classically trained musicians and those with backgrounds in contemporary music (e.g., rock or jazz)—and 14 non-musicians, none of whom had any prior experience in psychoacoustical experiments. Again, the non-musicians had DLFs that were on average approximately twice as large as the musicians, but after just 1 h of practice the non-musicians had improved to the level initially obtained in the musicians (during this time, the musicians had also improved slightly). In a third study, Micheyl, Delhommeau, Perrot, and Oxenham (2006) compared the influences of musical and psychoacoustical training on DLFs directly. The study tested 30 classically trained musicians and 30 non-musicians. The authors used very stringent selection criteria: the musicians all had at least 10 years of experience and were members of the National Superior Music Conservatory of Paris. The listeners' left and right ears were tested monaurally in the traditional 2AFC paradigm using both pure and harmonic complex tones. The results of the experiment are shown in Figure 2.17. On average across all conditions and runs, the DLFs of the musicians (0.13%) were more than six times smaller than those of the non-musicians (0.86%). The much larger group difference in this study was attributed to the selection criteria. The results also indicated that the classical musicians achieved asymptotic performance within a relatively short time: in the first hour of testing, the musicians already had relative DLFs lower than 0.2% on average, and no significant improvement was observed after the second testing block. In a second experiment, non-musicians were tested over a much longer training period. After between 4 and 8 h of practice, the non-musicians obtained on average DLFs as small as those of the musicians from the first experiment. In summary, the results suggest that musicianship does not confer a particular advantage for fine-grained frequency discrimination other than providing a form of relatively inefficient training, since the group differences can be eliminated when all the listeners are trained for a few hours in a psychoacoustical procedure.

The above finding from a small number of psychoacoustical studies sits in stark contrast to the enormous literature using functional imaging to compare pitch processing in musicians and non-musicians. Those studies have often found substantial differences in the pattern of neural activity elicited in the brains or brainstems of musicians compared to non-musicians during even passive listening to pitch-evoking sounds (e.g., Besson, Faita, & Requin, 1994; Crummer, Walton, Wayman, Hantz, & Frisina, 1994; Koelsch, Schroger, & Tervaniemi, 1999; Pantev et al., 1998; Pantev, Engelien, Candia, & Elbert, 2001; Tervaniemi, Rytkonen, Schroger, Ilmoniemi, & Naatanen, 2001; Tervaniemi, Just, Koelsch, Widmann, & Schroger, 2005; Schlaug, 2001; Wong, Skoe, Russo, Dees, & Kraus, 2007). What could account for this discrepancy? It is true to say that in most of those studies the non-musician controls were untrained, or at least not trained so that their performance was in line with their musical counterparts. Since psychoacoustical training elicits changes in cortical activity in and of itself (Section 2.2.5.3), it could be argued that many of

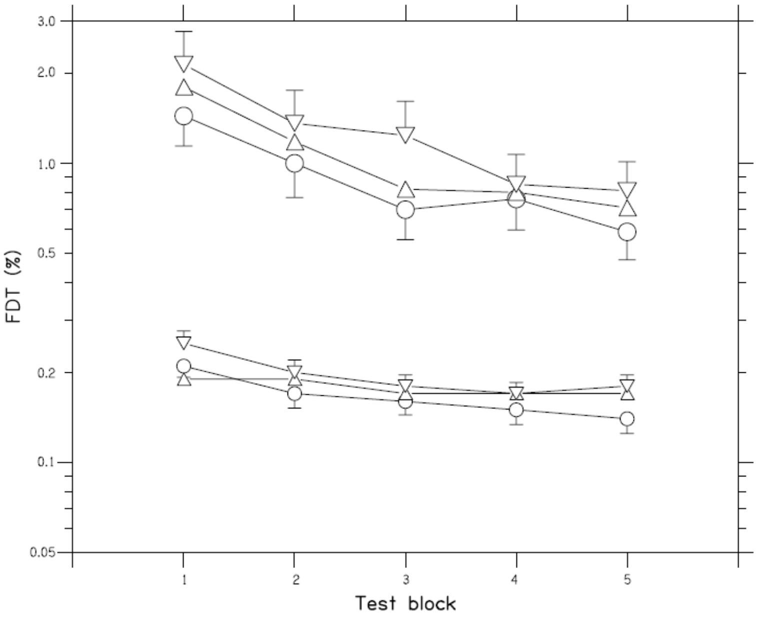


Figure 2.17. DLFs as a function of time in musicians and non-musicians. Each symbol represents the geometric mean DLF for a group of listeners (lower symbols, musicians; upper symbols, non-musicians). The different symbols represent different stimulus conditions, and the abscissa of each symbol represents the run of trials over which the DLF was obtained. From Micheyl et al. (2006).

those results were influenced by the fact that the authors were comparing trained and untrained listeners[12]. Another explanation is that frequency discrimination is not actually very 'musical', since the intervals to be discriminated are usually much smaller than a semitone, and test frequencies do not normally correspond to notes on a musical scale. Musicianship certainly confers advantages in other experiments that are more musical in nature, such as those involving the recognition of specific musical notes or intervals (e.g., Attneave & Olson, 1971), and those wherein the task is particularly challenging, such as in an IM experiment (Oxenham et al., 2003). In the latter case, successful performance (small detection thresholds in conditions containing maskers) is assumed to be achieved through an analytic listening strategy that some listeners may simply never adopt or never be able to use efficiently. Skilled musicians—who are used to listening to music analytically—may be able to apply their skills in an experiment with a strong IM component, resulting in smaller thresholds. It remains an open question whether psychoacoustical training can eventually neutralise the influence of musicianship in these other tasks. For a detailed examination of the role of musical experience—in terms of both formal training and passive enculturation—in different perceptual tasks, see Bigand and Poulin-Charronnat (2006).

---

[12] A functional magnetic resonance imaging (fMRI) study by Gaab and Schlaug (2003) claimed to have compared cortical responses during two pitch-discrimination tasks in performance-matched musicians and non-musicians, but the authors provided neither any performance data nor any details of training in the paper.

2.3.3.2 Detecting and identifying the direction of frequency changes

The classic experiments described in Section 2.2.4 required listeners to be—and indeed demonstrated that they were—sensitive to the direction (up or down) of very small frequency changes between temporally discrete pure tones separated by a silent ISI. However, four more recent studies have suggested that in certain circumstances some listeners can detect the presence of a small frequency change, but are unable to identify the direction of the resulting percept (Foxton et al., 2004, 2009; Johnsrude et al., 2000; Tramo et al., 2002; Semal & Demany, 2006).

Johnsrude et al. measured 45 listeners' DLFs in two frequency-discrimination tasks using pure tones. In the 'pitch direction' task, the listeners heard two pure tones—the first tone always had a frequency of 1000 Hz and the second always had a different frequency—and judged the direction of the change on each trial. In the 'simple discrimination' task, the listeners again heard two tones on each trial, but this time the second tone was equiprobably identical to or higher in frequency than the first tone (1000 Hz), and the listeners made same/different judgements. The cohort consisted of 14 neurologically normal listeners and 31 patients who had undergone surgical resection from either the right or left temporal lobe for the relief of intractable epilepsy. The clinical patients were split into four groups determined by which hemisphere had the lesion and whether or not the lesion encroached the first transverse temporal gyrus (or Heschl's gyrus, HG). The results (Figure 2.18) revealed that the patients with right-lateralised lesions that encroached HG had larger DLFs in the pitch-direction task than did the controls and the other patient groups, but there were no significant differences between the groups' DLFs for the simple-discrimination task. The authors took this result to suggest that successfully identifying the direction of frequency changes recruits brain areas that are not necessary for simply detecting those changes (Section 2.4.3). There are two limitations of that study. First, during the simple-discrimination task, listeners were only ever presented with frequency changes in one direction, which could have made the task easier than if the changes were equiprobably in either direction. Second, the DLFs measured by the authors were all very large, even those in the normal controls (approximately 3.5% on average; cf. Section 2.2.4). The results of the experiment could have been quite different if the listeners had been sufficiently well practised at both of the tasks.

Similarly, Tramo et al. (2002) reported DLFs measured using both a direction-identification task and a same-different task in five clinical patients with different cortical lesions and in eight normal controls. One patient, MHS, had extensive lesions that included HG in both hemispheres. DLFs for the direction-identification task and the same-different task—measured at 250, 500, 1000, and 2000 Hz—were much larger in MHS that the others tested, and MHS's DLFs were significantly larger in the former task than the latter. By contrast, there were no significant differences between the

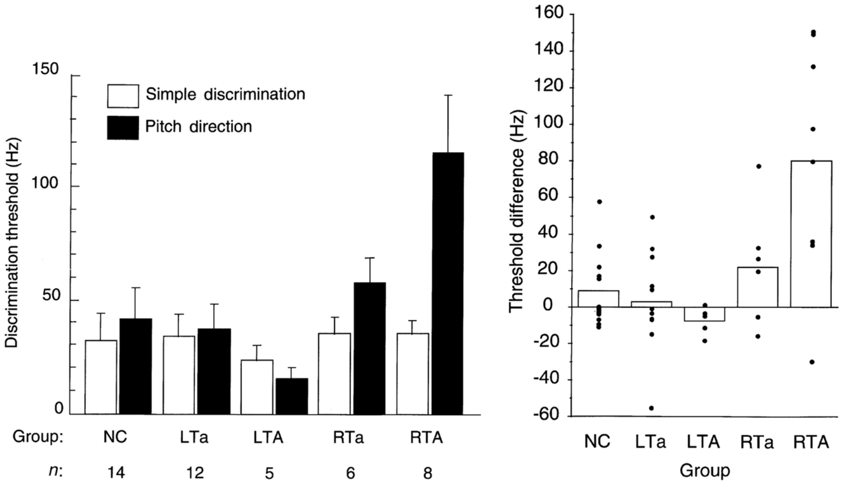


Figure 2.18. DLFs for two frequency-discrimination tasks in clinical patients and normal controls. In the left-hand panel, mean DLFs measured in two different tasks (different bars) and in five groups are shown: normal controls (NC), and patients with left-lateralised or right-lateralised lesions that encroached (LTA, RTA) or did not encroach (LTa, RTa) HG. Error bars are standard errors. In the right-hand panel, the bars represent the difference between the DLFs in the two conditions and the points represent the data from individual listeners. From Johnsrude et al. (2000).

DLFs measured in the two tasks in the other listeners.

A behavioural dissociation between frequency-change detection and frequency-direction identification has also been observed in listeners with congenital amusia. Amusia is the name given to a collection of difficulties that are characterised by a general failure to perceive music normally, affecting an estimated 4% of the population (Cuddy, Balkwill, Peretz, & Holden, 2005; Peretz, Champod, & Hyde, 2003; Sloboda, Wise, & Peretz, 2005). A huge literature on amusia has emerged in recent years, investigating the differences in the perceptual and cognitive abilities of amusics compared to controls (e.g., Bella & Peretz, 2003; Foxton et al., 2004; Gosselin, Jolicoeur, & Peretz, 2009; Hyde & Peretz, 2003, 2004; Jones, Zalewski, Brewer, Lucker, & Drayna, 2009; Loui, Guenther, Mathys, & Schlaug, 2008; Peretz et al., 2002; Pfeuty & Peretz, 2010; Tillmann, Peretz, Bigand, & Gosselin, 2007; Tillmann et al., 2010), its neural correlates both functional (e.g., Hyde, Zatorre, & Peretz, 2010; Moreau, Jolicoeur, & Peretz, 2009; Peretz, Brattico, & Tervaniemi, 2005) and anatomical (e.g., Hyde et al., 2007), and its possible genetic underpinnings (e.g., Peretz, Cummings, & Dube, 2007; T. D. Griffiths, 2008). Foxton et al. (2004) tested 10 adult listeners who reported lifelong difficulties in listening to music and 10 controls matched for time spent in education, and the groups were formally performance-dissociated using a standardised test battery for amusia (Peretz et al., 2003). DLFs were measured in three different conditions. In the 'segmented task' condition, the listeners heard two pairs of pure tones in each trial. The tones within each pair were 250 ms in duration and separated by a 100-ms ISI. One of the pairs in a trial contained two tones differing in frequency, and the listeners judged which pair this was. In a 'glide task' condition, the stimuli and task were identical except that the 100-ms within-pair ISIs were no longer silent: a single, 600-ms pure tone replaced the pair containing the same tones, and the two tones in the other pair were connected by a linear frequency glide. In a 'direction task', each trial contained two tone–glide–tone stimuli—one downward glide and one upward glide—and the listeners judged which interval contained the upward glide. In each condition, the frequency of the standard tone (or the first plateau in the glide stimuli) was 500 Hz. Details about training or the number of trials completed per listener were not reported. On average, the amusic group had larger DLFs than the control group in all three tasks, and their DLFs in the direction task were the largest overall.

Semal and Demany (2006) used an elegant dual-pair paradigm to measure DLFs for frequency-change detection and frequency-direction identification in nine listeners under directly comparable stimulus conditions. In their first experiment, the listeners heard two pairs of pure tones in each trial. The tones were all 250 ms in duration and presented at 65 dB SPL, separated by a 250-ms within-pair silent ISI and a 700-ms between-pair ISI. In one pair the tones were identical, and in the other pair the tones differed in frequency. In the 'detection' task, listeners indicated which pair (first or second) contained the difference in frequency. In the 'identification' task, listeners indicated the direction (up or down) of the difference. Listeners' detection DLFs (DDLFs) and identification DLFs (IDLFs) were estimated based on 50 threshold estimates per condition using an adaptive procedure measuring 75% correct on the psychometric function, and every trial was followed by feedback. The results are shown in the top-left-hand panel of Figure 2.19. In their three listeners who had the smallest DLFs overall, their IDLF was on average approximately 1.5 times smaller than their DDLF[13]. In three more listeners, IDLFs and DDLFs were similar. In the remaining three listeners, their IDLF was considerably larger than their DDLF, indicating that they were relatively insensitive to the direction of small but detectable frequency differences. The results of Semal and Demany's first experiment cannot be attributed to practice effects: each listener had completed a very large number of trials in both tasks, so the DLFs measured were extremely stable (in fact, the size of each symbol in the top-left-hand panel of Figure 2.19 exceeds two geometric standard errors), and subsequent re-examination of the data as a function of time did not reveal substantial amounts of learning over the course of their acquisition. In a follow-up experiment, the listeners who were poor at making absolute judgements about the direction of frequency change (up or down) were also poor when they had to make relative judgements (e.g., describing two upward frequency changes as similar). The discrepancy between detection and identification abilities observed for frequency in those listeners also did not generalise to other sound dimensions, such as level or the rate of AM.

---

[13] This counter-intuitive aspect of Semal and Demany's (2006) data is explained in Chapter 6.

What makes Semal and Demany's (2006) finding particularly intriguing is that their listeners, unlike those in the previous studies, appeared to be audiologically and neurologically normal. The elevation in IDLFs—though similar in magnitude to that observed by Johnsrude et al. (2000) and Foxton et al. (2004)—was not symptomatic of any obvious real-world problems. An important difference between Semal and Demany's study and those involving clinical patients and amusics is that the former employed frequency roving—on each trial, the frequencies of the first tones in each pair were randomly varied over a wide range (400–2400 Hz). As discussed in Section 2.2.4.4, a number of previous studies have demonstrated that frequency roving is deleterious to normally hearing listeners' DLFs. This raises the question of whether the specific deficit observed for some of Semal and Demany's listeners would occur under conditions in which the stimuli are not roved in over such a wide range.

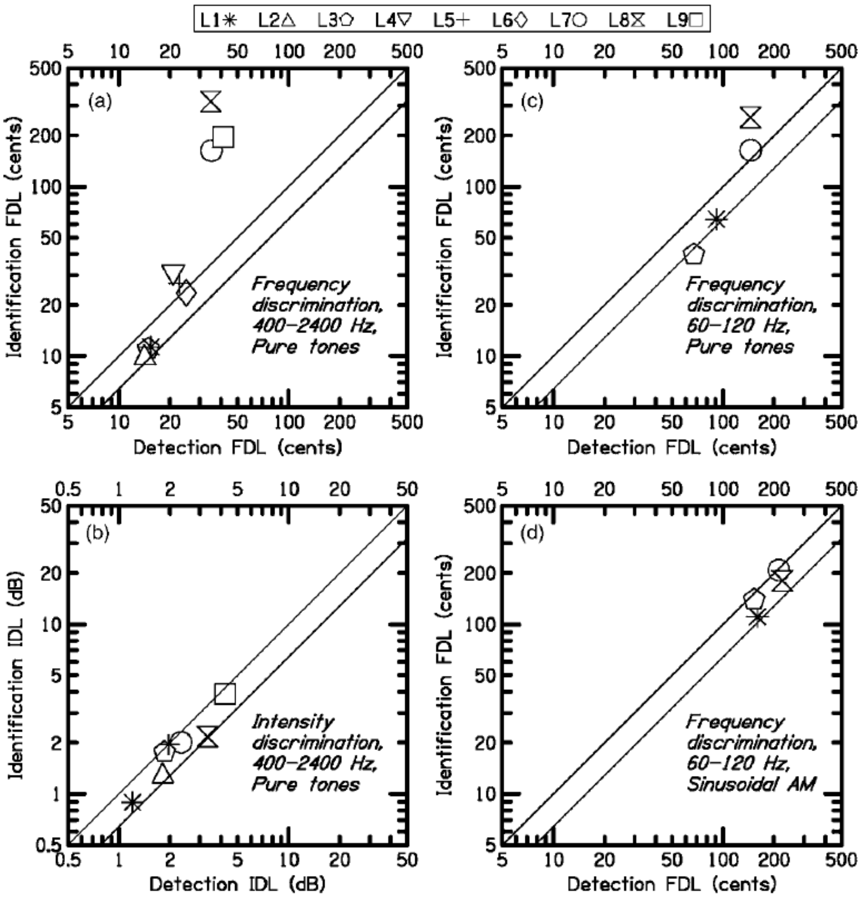


Figure 2.19. Data from Semal and Demany (2006). The top-left-hand panel shows the DDLFs and IDLFs of the nine listeners tested in the first experiment, measured in musical cents and scaled logarithmically. The bottom-left-hand panel shows detection and identification difference limens for level for seven listeners. The top-right-hand panel shows DLFs measured for pure tones roved over the range 60–120 Hz. The bottom-right-hand panel shows DLFs for sinusoidal AMs. In each panel, the lower oblique line represents a prediction of signal detection theory, described in detail later.

2.4 MEG studies of pitch perception

2.4.1 What is MEG?

MEG is a functional imaging technique. Unlike fMRI, which measures of localised changes in cerebral blood-oxygen level, MEG—along with its progenitor, electroencephalography (EEG)—provides a relatively more direct measure of the electrical activity generated by populations of neurons within the brain. Whereas EEG uses electrodes placed directly onto an individual's scalp to measure changes in electrical current flow, MEG uses super-conducting quantum interface devices (SQUIDs) to measure extra-cranial magnetic fields. The SQUIDs are usually positioned a few cm from an individual's head and operate at temperatures close to 0 °K through immersion in liquid helium inside a cryogenic vessel (a dewar). Extra-cranial magnetic fields are extremely weak in comparison to environmental sources, so MEG scanners are almost always housed inside a magnetically shielded room (Figure 2.20). For a detailed review of the basic principles of MEG, see Hämäläinen, Hari, Ilmoniemi, Knuutila, and Lounasmaa (1993).

Figure 2.20. An MEG scanner. The left-hand panel shows outer door of the magnetically shielded room containing the scanner. The right-hand panel shows the scanner with a subject (the supervisor) and operator (the author). The cylinder on the left of the picture is the dewar containing the SQUIDS. The operator is marking locations on the subject's head with a Polhemus digitisation system, which provides information on the subject's head shape and position relative to the sensors.

The often-cited major strength of MEG and EEG is their relatively fine temporal resolution. The blood-oxygen-level dependent (BOLD) response measured by fMRI is sluggish relative to the cerebral processes with which it is associated: after the onset of a sensory stimulus, the BOLD response takes several seconds to peak, which is obviously a very long time after the stimulus has been perceived (see Jezzard, Matthews, & Smith, 2003). MEG and EEG are not limited theoretically in this way; their temporal resolution is limited only by their sampling rate and their baseline sensitivity to the neural activity under investigation. As a result, MEG and EEG have the potential to resolve cerebral processes in the order of milliseconds, rather than in the order of seconds as with fMRI.

The precise source of the MEG signal is believed to be the intracellular current flow in the dendritic trunks of pyramidal neurons (see Okada, 2003). The magnetic field evoked by a single neuron is infinitesimal; tens of thousands of neurons with similar orientations must be active simultaneously to produce a magnetic field measurable with MEG, and because there may be neighbouring areas of current flow oriented in the opposite direction, hundreds of thousands could actually be necessary (Okada, 1983). The spatial resolution of MEG is coarser that that of fMRI, but is finer than that of EEG. However, reconstructing the sources of cerebral activity from the magnetic field patterns observed with MEG (the inverse problem) is extremely difficult. Helmholtz (1853) showed that a current distribution inside a conductor cannot be retrieved uniquely from knowledge of the magnetic field outside the conductor; in terms of MEG, knowledge of the time-varying magnetic fields outside the human head is not sufficient to establish the exact pattern of neural activity that created those fields. In other words, there is *no unique solution* to the inverse problem. Consequently, source analysis in MEG requires several assumptions regarding the conductor (the head) and the sources themselves. The most commonly adopted technique is to model cerebral activity as one or several equivalent-current dipoles within a spherical conductor (see Baillet, Mosher, & Leahy, 2001), with the strength, orientation, and position of the dipole(s) determined by iterative processes so that they match as closely as possible the pattern observed at the level of the sensors (Figure 2.21). Although many other approaches exist, almost all of the studies described in this section have used the standard dipole approach to implicate the brain areas involved in auditory processing.

If a dipole is oriented perfectly radially with respect to the surface of a spherical conductor, it lies along a line that passes through the centre of the conductor and consequently does not produce an external magnetic field (Sarvas, 1987). This feature makes MEG preferentially sensitive to sources with partially tangential orientations (Hämäläinen et al., 1993; Mosher, Leahy, & Lewis, 1999). Moreover, MEG is most sensitive to sources closest to the sensors; the probability of a source being detected is inversely proportional to its depth (Hillebrand & Barnes, 2002). Therefore, MEG is relatively insensitive to deep sources, such as those emanating from the midbrain or brainstem, and most sensitive to tangentially oriented, superficial sources, such as those emanating from the sulci of the grey matter. Fortunately for auditory researchers, the auditory cortex occupies a band of tissue along the superior border of the temporal lobe just below the lateral (or Sylvian) fissure, largely involving HG and

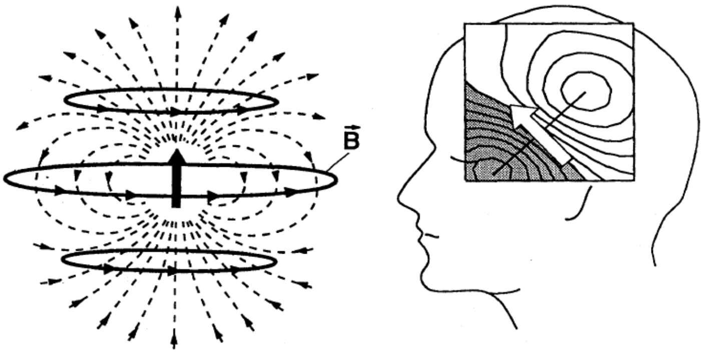


Figure 2.21. Illustration of the magnetic field generated by an equivalent-current dipole. The left-hand panel depicts the dipole (arrow), the volume currents (dashed lines), and the resultant magnetic field ($\vec{B}$). The right-hand panel shows an example of a magnetic field pattern measured outside the head from a single dipole located beneath the mid-point of the line joining the extrema of the field. From Hämäläinen et al. (1993).

the posterior two-thirds of the superior temporal gyrus (Figure 2.22).

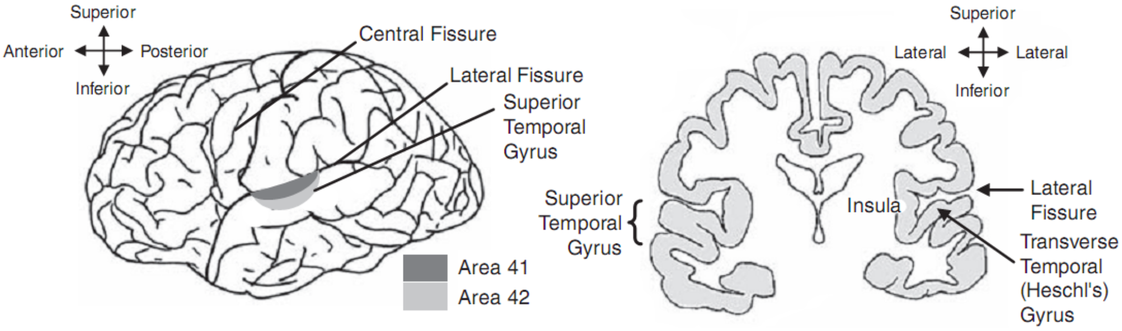


Figure 2.22. Lateral (left) and coronal (right) representations of the human brain illustrating the locations of the auditory areas. From Gelfand (2010).

2.4.2 Auditory evoked fields

The traditional approach to investigating functional processes with EEG and MEG has been to present listeners with stimuli many times, to record the resultant electrical or magnetic field signals, and then to average those signals across all the repetitions of a particular stimulus or across a class of similar stimuli. This approach is said to reveal the activity *evoked* by the stimuli. In auditory research, the different components or deflections within the average waveform are called event-related or auditory-evoked fields (AEFs). The equivalent components in EEG data are called event-related potentials (ERPs). In EEG research, different ERPs are labelled first by a letter (N or P) indicating their polarity at a particular electrode or group of electrodes—usually located on the vertex of the skull, since this is where the clearest auditory responses are often observed—and then by a number or lower-case letter referring to their latency. For instance, N100 or N1 refers to a negative-going ERP occurring approximately 100 ms after the onset of a stimulus. In MEG research, AEFs are often similarly labelled to promote compatibility across studies, but the suffix ‘m’ is usually added. The effect of head tissue is smaller in MEG than in EEG, meaning that the MEG signal is generally clearest in the sensors closest to the activity. Because the polarity of a bilateral cerebral source is usually reversed across hemispheres, the use of the ‘N’ or ‘P’ prefix is usually less appropriate for AEFs than for ERPs. Some researchers therefore replace N or P with an ‘M’. Thus, the magnetic equivalent of the N1 could be called the N100m, N1m, or M100.

2.4.2.1 Early responses

Despite the relative insensitivity of MEG to deep sources, some studies have observed the neural activity of structures in the brainstem in response to sounds (e.g., Erné & Hoke, 1990; Erné, Scheer, M., Pantev, & Lütkenhöner, 1987; Iramina & Ueno, 1995; Lütkenhöner, Lammertmann, Ross, & Pantev, 2000; Parkkonen, Fujiki, & Mäkelä, 2009). For instance, Parkkonen et al. scanned seven listeners with MEG and EEG concurrently whilst they passively listened to 0.6-ms clicks generated by a piezoelectric crystal. When the data were bandpass filtered between 180 and 1000 Hz, the authors found a period of increased activity in the click condition relative to a silent baseline condition, with peaks between 2 and 7 ms after the onset of the click, in both the MEG and EEG data. Standard dipole-fitting algorithms failed to localise all of these peaks, but the involvement of at least three different sub-cortical structures was implicated; a peak at approximately 2.5 ms localised to the auditory nerve, a peak at 3 ms localised to the cochlear nucleus, and a peak at 5.8 ms localised to an area close to the lateral lemniscus.

When Parkkonen et al. (2009) bandpass filtered their same MEG data between 30 and 300 Hz, they found peaks in activity that followed the brainstem responses and localised to the cortex. An AEF was observed peaking at approximately 17 ms in five of the seven listeners, and another and at approximately 30 ms was observed in two of the listeners. Other studies have also found AEFs in response to clicks and other sounds around 10, 17 to 19, 30, 50, and 70 ms after stimulus onset (e.g., Chait, Simon, & Poeppel, 2004; Hashimoto et al., 1995; Kuriki, Nogai, & Hirata, 1995; Levänen & Sams, 1997; Lütkenhöner, Krumbholz, Lammertmann, et al., 2003; Mäkelä, Hämäläinen, Hari, & McEvoy, 1994; McEvoy, Mäkelä, Hari, Hämäläinen, & Hari, 1994; Pantev et al., 1995; Pelizzone et al., 1987; Scherg, Hari, & Hämäläinen, 1989; Yoshiura, Ueno, Iramina, & Masuda, 1995; Yvert, Crouzeix, Bertrand, Seither-Preisler, & Pantev, 2001)—these are collectively referred to as middle-latency AEFs. Middle-latency AEFs are weaker and more capricious than later AEFs: not all of middle-latency AEFs are observed in every MEG study or in every listener within each study. When they are observed, the earliest middle-latency AEFs occurring at or before approximately 20 ms after stimulus onset localise to the antero-medial part of HG and are generally regarded as representing the initial activity of human primary auditory cortex. The later deflections tend to localise more laterally, such as towards antero-lateral HG or the superior temporal gyrus.

2.4.2.2 M100

The middle-latency AEFs are followed by a deflection at approximately 100 ms (M100). Because of its prominence and robustness across listeners and stimuli, the M100 has been the most investigated AEF. The M100 is observed in response to virtually any auditory stimulus, and a commonly held hypothesis is that it reflects the processes of detecting the onset of sensory input, although it is not clear why such an onset detector would operate so late in the processing stream. Dipole-source analysis usually localises the M100 to the planum temporale (e.g., Lütkenhöner & Steinsträter, 1998), however it is likely that the deflection is a composite response with contributions from multiple different brain areas that are active at or around the same latency (Lütkenhöner, Lammertmann, & Knecht, 2001; Lütkenhöner, 2003; Lütkenhöner, Krumbholz, & Seither-Preisler, 2003; Näätänen & Picton, 1987; Woods, 1995). Not all of the functions encapsulated by the M100 are necessarily strictly auditory in nature: there is evidence that its EEG equivalent (the N1) involves frontal and premotor brain areas (Näätänen & Picton, 1987). The amplitude, latency, and spatial parameters of the M100/N1 are widely exploited experimental variables, and have been shown to be sensitive to temporal and spectral properties of sounds (e.g., Cansino, Williamson, & Karron, 1994; Kuriki & Murase, 1989; Lütkenhöner, 2003; Lütkenhöner et al., 2001, 2000; Lütkenhöner & Steinsträter, 1998; Pantev, Eulitz, Elbert, & Hoke, 1994; Pantev, Hoke, Lehnertz, & Lütkenhöner, 1989; Pantev et al., 1988; Pantev, Hoke, Lütkenhöner, Fahrendorf, & Stöber, 1990; Roberts, Ferrari, & Poeppel, 1998; Roberts & Poeppel, 1996; Romani, Williamson, & Kaufman, 1982; Rosburg, Haueisen, & Sauer, 2002; Seither-Preisler, Krumbholz, & Lütkenhöner, 2003; Seither-Preisler, Patterson, Krumbholz, Seither, & Lütkenhöner, 2006a; Stufflebeam, Poeppel, Rowley, & Roberts, 1998), the context in which sounds are presented (e.g., Salajegheh et al., 2004), and a listener's state of attention (e.g., Poeppel et al., 1996).

Several authors have claimed that the location of the M100 may reveal tonotopic or even periodotopic organisation within the cortex. Pantev et al. (1988) presented four listeners with 500-ms pure tones with frequencies of 250, 500, 1000, 2000, and 4000 Hz. The depth of the dipole sources of the M100 increased monotonically with increasing frequency, suggesting the existence of a logarithmically spaced tonotopic map with more lateral dipole locations for low frequencies and more medial locations for high frequencies. Very similar observations were made before by Romani et al. (1982), although for steady-state responses to AM tones rather than transient AEFs. However, although M100 tonotopy has been corroborated in some studies

(Cansino et al., 1994; Kuriki & Murase, 1989; Lütkenhöner & Steinsträter, 1998; Pantev et al., 1994, 1989, 1990), it has not been replicated in others (Roberts & Poeppel, 1996; Stufflebeam et al., 1998). Tonotopic maps certainly exist in human auditory cortex—there is clear evidence for them in non-human animals and in humans using other methods (e.g., Howard III et al., 1996; Merzenich & Brugge, 1973)—so this discrepancy is probably due to the substantial complexity of the neural processing underlying the M100. Moreover, there are at least two cortical tonotopic maps in the cat (e.g., Kaas & Hackett, 2000; Merzenich & Brugge, 1973; Morel, Garraghty, & Kaas, 1993) and a study using very high-resolution fMRI suggests the existence of two mirror-symmetric tonotopic maps sharing a low-frequency border in human primary auditory cortex (Formisano et al., 2003). The issue is further complicated when one considers the influence of individual differences on the M100. Lütkenhöner, Krumbholz, Lammertmann, et al. (2003) measured AEFs in 10 listeners to 500-ms pure tones with frequencies 250, 500, 1000, and 2000 Hz. About 250 stimuli of each type were presented in random order in four MEG sessions per listener. M100 dipole-source locations showed a high intra-individual reproducibility, but also a substantial inter-individual variability. In most cases, either the dipole location exhibited no significant frequency dependence at all, the dipoles for the four frequencies did not show orderly alignment, or the data disagreed with the single-dipole model. In the few cases showing an arrangement of dipoles consistent with the assumption of an orderly cortical tonotopic map, the most relevant coordinate varied from subject to subject. Thus, the M100 may simply be too complex or broad a response to represent sufficiently the locus of a single tonotopic map.

Another MEG study by Pantev et al. (1988) suggested that the dipole location of the M100 might reflect its pitch rather than its frequency. The authors presented six listeners with three different stimuli: a 250-Hz pure tone, a 1000-Hz pure tone, and complex tone composed of the fourth-to-seventh harmonics of 250 Hz (each tone was 500 ms in duration). The third stimulus produces a strong pitch at 250 Hz, but contains no energy below 1000 Hz; to ensure that the perceived pitch was not due to combination tones resulting from nonlinear interaction in the auditory periphery, the 250-Hz region was masked by simultaneously presenting a continuous narrowband noise centred at 250 Hz. The reasoning was that if M100 location is dependent on frequency, the location of the M100 to the third tone would be closer to that of the second tone, but if M100 location is dependent on pitch, it would be closer to that of the first tone. The results supported the latter hypothesis. However, Pantev et al.'s third stimulus *did* contain energy at 250 Hz. A later MEG study by Langner, Sams, Heil, and Schulze (1997), which used complex tones without noise maskers, suggested instead that the periodicity (F_0 frequency) of a complex tone is represented in the M100 on a topographic map lying orthogonal to the usual tonotopic map observed with pure tones. However, as with M100 tonotopy, M100 periodotopy may also be subject to considerable individual differences: a third MEG study using complex tones by Lütkenhöner (2003) failed to replicate the results from either Pantev et al. or Langner et al..

Although the results reported by Lütkenhöner and colleagues (Lütkenhöner, 2003; Lütkenhöner, Krumbholz, Lammertmann, et al., 2003) question the validity of using the dipole locations of the M100 to study tonotopy or periodotopy, M100 amplitude and latency may still be useful dependent variables. Indeed, there is some evidence that the latency of the M100 is sensitive to F_0 frequency. Seither-Preisler et al. (2006a) presented seven listeners with harmonic complex tones whose spectra fell within the range 800–5000 Hz and with F_0-frequencies corresponding to 100, 200, 400, and 800 Hz. The harmonics of the tones were all either in sine phase, in alternating phase (even harmonics $+90^\circ$ relative to the odd harmonics), or random phase. With every doubling of F_0, the latency of the peak decreased by approximately 2.8 ms. By contrast, the relative phases of the harmonics had no effect on M100 latency, suggesting that the M100 reflects perceptual processes occurring after these features have been accounted for. The decrease in peak latency with increasing pitch could have occurred because a temporal mechanism could extract higher pitches more rapidly than lower pitches.

2.4.2.3 Pitch-specific AEFs

Even if the latency of the M100 varies with pitch, the fact that it also varies with intensity and spectral composition means that any components of the M100 truly associated with pitch are fundamentally confounded with components that reflect responses to other stimulus features. This limitation prompted Krumbholz, Patterson, Seither-Preisler, Lammertmann, and Lütkenhöner (2003) to develop a new MEG protocol using iterated rippled noise (IRN). IRN is created by delaying a sample of random noise by d ms, adding it back to the undelayed noise, and repeating the process n times. IRN is 'noisy' but evokes a sensation of pitch corresponding to $1/d$ kHz. IRN has proven to be a very useful stimulus in pitch research, since highpass filtering IRN at an appropriate frequency can remove any spectral peaks associated with its periodicity, and the strength of the pitch percept can be manipulated parametrically by manipulating the value of n (Yost, 1996a, 1996b). Another advantage for the purposes of MEG is that a transition from noise to IRN or vice versa can be made without creating a change in spectrum or overall sound level. In their first experiment, Krumbholz et al. presented seven listeners with stimuli consisting of 2 s of noise followed by 510 ms of IRN with varying values of d and n. The normal M100 was observed after the onset of the stimulus, and a second prominent AEF—a pitch-onset response (POR)—was observed approximately 150 ms after the transition to IRN. The latency of the POR peak increased with increasing values of d (i.e., decreased with increasing pitch), and its amplitude increased with increasing n

(i.e., pitch strength). The authors repeated their experiment with stimuli containing the opposite transition: the amplitude and latency of the M100, this time in response to IRN, did not vary systematically with d or n (cf. Seither-Preisler et al., 2006a), and no AEF was observed after the transition from IRN to noise.

Dipole-source modelling of the POR attributed its 'centre of gravity' to the mediolateral part of HG. On average, the POR dipole was 12.4 mm more anterior, 6.0 mm more medial, and 10.9 mm more inferior than the M100 dipole (Figure 2.23). This result suggests that the POR and M100 are to some extent generated by different brain areas, with the POR receiving greater contributions from lateral HG and the M100 receiving greater contributions from planum temporale. However, in another MEG study, reducing the time between stimulus onset and the point of transition from noise to IRN resulted in a decrease in the amplitude of the POR (Seither-Preisler, Krumbholz, Patterson, Seither, & Lütkenhöner, 2004). This suggests that that the two AEFs are not entirely distinct: the response to energy onset appeared to cause a degree of refractoriness in the POR, implying at least one neural generator common to both the POR and M100.

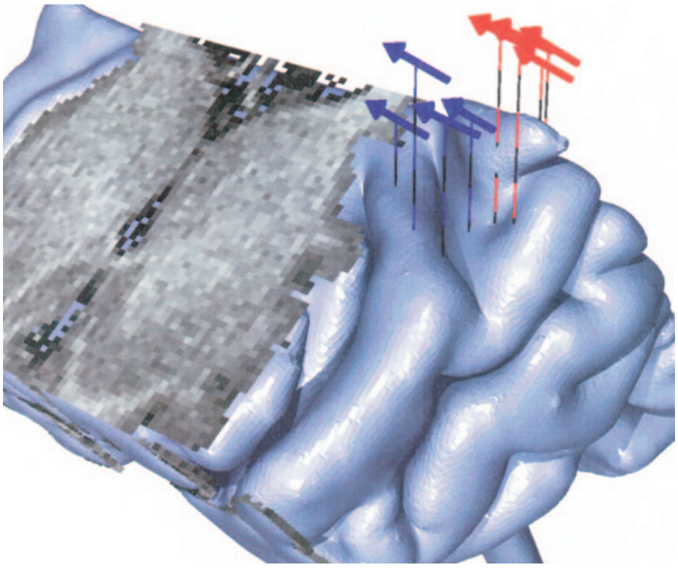


Figure 2.23. Source locations of the POR (blue) and the M100 (red) for a single listener, estimated from four measurement sessions and projected into a three-dimensional reconstruction of the listener's left temporal lobe. The dipoles are shifted upwards by 3 cm from the actual position of the dipole to prevent them from being partially hidden under the cortical surface. Each colour bar on the vertical source markers is 5 mm in height. From Krumbholz et al. (2003).

Both the M100 and the POR are transient AEFs that appear to be sensitive to the onset of certain acoustic features. However, there is also evidence for sustained AEFs that continue for as long as a sound evokes a percept of pitch. Gutschalk, Patterson, Rupp, Uppenkamp, and Scherg (2002) presented 12 listeners with regular and irregular click trains at three sound levels. Along with the normal transient peaks observed after sound onset, the click trains also evoked a sustained AEF starting at approximately 300 ms. The sustained AEF rose to a steady state around 400 ms, where it stayed until the click trains ended. Dipole-source modelling revealed that the AEF could be explained by the combination of an anterior source located in lateral HG and a posterior source in planum temporale. The strength of the anterior source was largest for the regular click train, which evoked a sensation of pitch, but was insensitive to sound level. By contrast, the posterior source varied with level but was insensitive to temporal regularity (i.e., whether a sensation of pitch was produced or not). This double dissociation suggests that lateral HG—close to but not the same location implicated by the dipole-location of Krumbholz et al.'s (2003) original POR—is involved in the processing of continuing pitch. A follow-up study used a transition paradigm with click trains (Gutschalk, Patterson, Scherg, Uppenkamp, & Rupp, 2004) to effectively remove the posterior source of the sustained AEF; now the transient and sustained AEFs both localised exclusively to lateral HG (similar results were reported with IRN by Seither-Preisler et al., 2003).

Krumbholz et al.'s (2003) transition paradigm essentially decomposes the M100 by removing the components of the AEF related to the onset of acoustic energy, with the inference that those remaining components are related to the perception of pitch. If this inference is true, then the results of that and subsequent studies suggest that lateral HG is particularly important for pitch, at least with IRN and click-train stimuli. Further evidence of lateral HG as a 'pitch centre' in non-primary auditory cortex comes from fMRI experiments. Following on from an earlier study that used positron emission tomography (PET; T. D. Griffiths, Büchel, Frackowiak, & Patterson, 1998), Patterson, Uppenkamp, Johnsrude, and Griffiths (2002) contrasted the BOLD response when listeners heard IRN with the BOLD response when they heard broadband noise. The authors reported that lateral HG was selectively more responsive to IRN (replicated by D. A. Hall & Plack, 2009; Warren, Uppenkamp, Patterson, & Griffiths, 2003). In that contrast, the pitch of the IRN was fixed throughout the block; random and melodic variations in pitch over time did not elicit any greater activity of lateral HG, but instead activated areas higher in the auditory processing stream including planum temporale and planum polare (see also T. D. Griffiths et al., 1998; Puschmann, Uppenkamp, Kollmeier, & Thiel, 2010; Warren et al., 2003). Those studies suggest that there is a hierarchy of pitch processing within human cortex, with activity moving antero-laterally as the hierarchy proceeds.

It is reasonable to expect a true pitch centre to respond similarly to all stimuli that evoke a sensation of pitch, provided they are matched for pitch strength. Indeed, a landmark study by Bendor and Wang (2005) found evidence for pitch constancy in the auditory cortex of the marmoset. The authors identified using single-unit recordings neurons that responded maximally to a specific pitch over a range of different stimuli, including pure tones, harmonic complex tones, and IRN. The neurons were located mostly in an area of non-primary auditory cortex near the antero-lateral border of the primary auditory area (Figure 2.24). Does lateral HG in human cortex contain homologous neurons? Chait, Poeppel, and Simon (2006) presented 20 listeners with stimuli containing transitions from broadband noise to Huggins pitch—produced by presenting the same broadband noise to both ears except for a narrow frequency band which is interaurally decorrelated, creating a sensation of pitch corresponding roughly to the centre of the decorrelated band—and transitions from noise to noise containing a pure tone. Both types of transition, which are perceptually similar but physically very different, evoked a prominent AEF approximately 150–200-ms after transition onset. Their dipole-source locations were very similar, and very similar to the transition responses reported by Krumbholz et al. (2003) and Gutschalk et al. (2004). These results suggest the POR is a general pitch mechanism that is insensitive other stimulus properties. The fMRI evidence for pitch constancy in lateral HG is less clear cut. D. A. Hall and Plack (2007, 2009) have claimed that the area does not respond robustly to pitch-evoking stimuli; they suggest that activity in lateral HG might instead reflect non-pitch-related features of IRN. However, Puschmann et al. (2010) recently found just the opposite: lateral HG was responsive to Huggins pitch and tone-in-noise stimuli (see also Penagos, Melcher, & Oxenham, 2004; Schönwiesner & Zatorre, 2008).

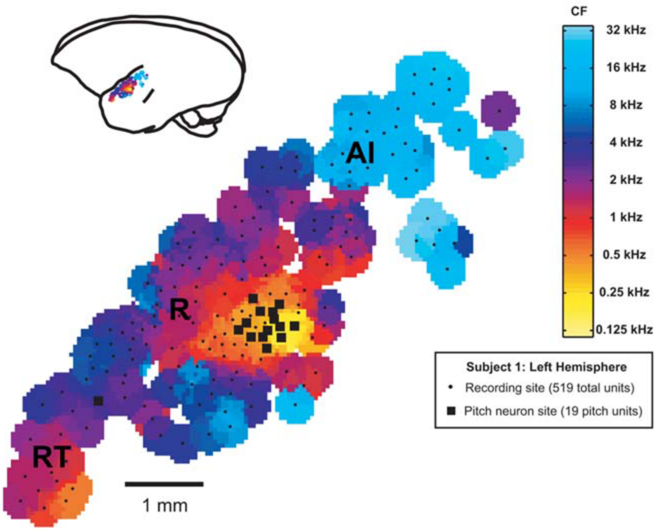


Figure 2.24. Location of the pitch area in marmoset auditory cortex. The coloured area shows a tonotopic map from the left hemisphere of one marmoset. Pitch-selective neurons (black squares) were found clustered near the antero-lateral border of primary auditory cortex (AI). Frequency reversals indicate the borders between AI and the rostral field (R), and AI and the rostral temporal field (RT). From Bendor and Wang (2005).

A study by Chait, Poeppel, de Cheveigné, and Simon (2007) called in question whether transition AEFs are truly in response to pitch. The authors presented 24 listeners with 1440-ms stimuli. In the constant condition (C), the stimulus was a continuous pure tone with a fixed frequency and level (randomly selected per stimulus). In the random condition (R), the stimulus was a sequence of short, temporally contiguous pure tones, each with a random frequency. The third and forth conditions consisted of transitions between the constant and random stimuli (CR and RC, respectively), both of which evoked AEFs that were delayed relative to the point of transition. The AEFs observed in the CR and RC conditions differed considerably from each other (Figure 2.25). CR transitions evoked M50 and M150 AEFs, whereas RC transitions evoked a single AEF with a latency of approximately 150 ms. Chait et al. argued that the RC transitions evoked a POR-like AEF even though both the C and R stimuli evoked very clear pitches individually. Both RC and noise-to-pitch transitions are similar in the abstract sense that they go from relative disorder to relative order; for instance, in IRN, 'order' is introduced by a constant regular-interval noise sample. Thus, the POR may be a generic response reflecting the detection of a temporal edge rather than anything specific to pitch. However, Chait et al. did not attempt to localise the sources of their responses. It remains to be determined whether a noise-to-pitch transition AEF has the same neural generators as a transition AEF with no associated pitch onset or pitch change.

2.4.2.4 MMN

The MMN is a prominent AEF or ERP elicited by an infrequent stimulus interspersed among a sequence of standard stimuli. The basic auditory MMN paradigm involves presenting listeners with a sequence of sounds, most of which are identical (the

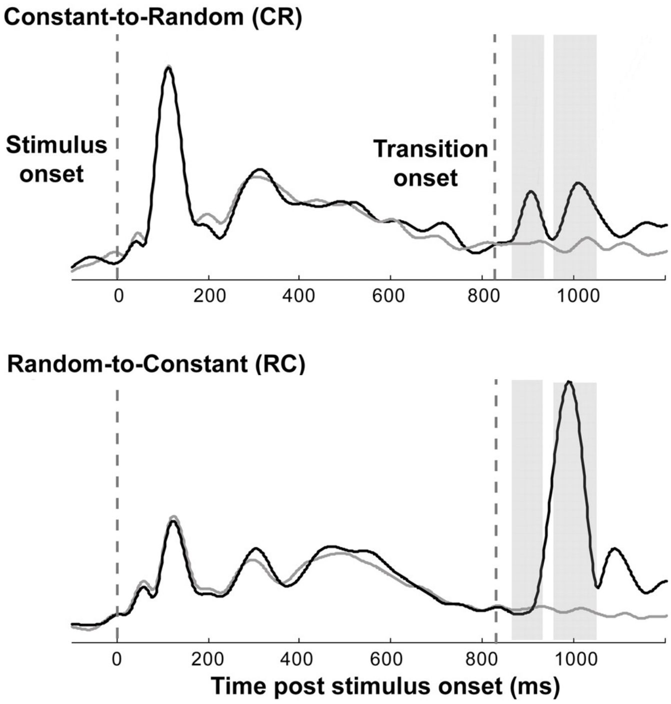


Figure 2.25. Transition responses. The black lines show the average AEFs to the CR (top panel) and RC (bottom panel) stimuli, and the grey lines show the average AEFs to the C (top panel) and R (bottom panel) stimuli. From Chait et al. (2007).

standard) but are sometimes perceptibly different (the deviant). The averaged waveform evoked by the standard sound is then subtracted from that evoked by the deviant sound, leaving a ‘difference wave’. This difference wave usually contains a peak, the MMN, around 100–200 ms after stimulus onset. An MMN can be evoked by virtually any deviation, including a difference in frequency (e.g., Hari et al., 1984; Jacobsen & Schröger, 2001), level (e.g., Lounasmaa, Hari, Joutsiniemi, & Hämäläinen, 1989), perceived spatial location (e.g., Paavilainen, Karlsson, Reinikainen, & Näätänen, 1989), and timbre (e.g., Tervaniemi, Winkler, & Näätänen, 1997). The standard stimuli do not necessarily all need to be identical; an MMN can also be evoked by stimuli that deviate from a rule, for example by the repetition of a tone in a sequence of tones descending in frequency (Tervaniemi, Maury, & Näätänen, 1994). The MMN has been described as memory-based, since it appears to depend on the presence of a memory trace representing the repetitive aspects of the standards (Cowan, Winkler, Teder, & Näätänen, 1993; Sams et al., 1985). The MMN is also described as pre-attentive, since no behavioural task is required to elicit it (e.g., Takegata, Paavilainen, Näätänen, & Winkler, 2001). Due to its relevance to central auditory processing, memory, attention, and executive control, the MMN has proven very popular among researchers and has been the most studied AEF and ERP after the M100 and N1. For a more or less up-to-date review of the MMN literature, see Näätänen, Paavilainen, Rinne, and Alho (2007).

It could be argued that because the auditory MMN is a correlate of neural mechanisms involved in *general* change detection, its utility as a tool for studying pitch perception is limited. However, there is some evidence that there are specific and specialised MMN generators. Using EEG, Paavilainen, Alho, Reinikainen, Sams, and Näätänen (1991) found that the scalp topographies of the MMNs evoked by changes in frequency, level, and duration are not identical, suggesting that these three kinds of change are processed by at least partly distinct neural populations. Subsequently, a number of EEG and MEG studies using dipole modelling have reported differences in the location of different MMNs in the range of a few millimetres and/or differences in their dipole orientations (Frodl-Bauch, Kathmann, Möller, & Hegerl, 1997; Giard et al., 1995; Levänen, Hari, McEvoy, & Sams, 1993; Rosburg, 2003). Analogous results were also reported by Molholm, Martinez, Ritter, Javitt, and Foxe (2005) in a study using fMRI. Further support comes from experiments in which deviations occurred on two acoustic dimensions simultaneously: the MMN evoked by a two-dimensional change—such as in frequency and interaural time delay, frequency and duration, or duration and level—is approximately equal to the sum of the MMNs elicited by its one-dimensional components, exactly as if each of the combined one-dimensional components elicited its own MMN (e.g., Schröger, 1995). Taken together, those findings suggest the basic MMN, like the basic M100, is made up of contributions from different specific brain areas depending on the kind of change that is detected.

2.4.3 AEFs in listeners with poor IDLFs

This final section of the literature review discusses in some detail a single MEG study that is very closely related to the empirical work reported in Chapter 8 (Foxton et al., 2009). As reviewed in Section 2.3.3.2, under certain circumstances, some listeners can detect the presence of a small frequency change between temporally discrete pure tones, but are not able to identify its direction. Two studies that found such a dissociation tested listeners with brain damage (Johnsrude et al., 2000; Tramo et al., 2002). In both studies, the listeners who were poorest at frequency-direction identification all had lesions that encroached right-lateralised HG. Based on those findings, Foxton et al. reasoned that successfully identifying the direction of small frequency changes might rely on processes within right-lateralised brain areas that are not involved in simply detecting those changes. However, as shown by Semal and Demany (2006), the same dissociation can be observed also in listeners with no audiological or neurological disorder. Foxton et al. hypothesised that these otherwise normally hearing listeners experience a diminished contribution of their right auditory cortices—or conversely, an increased contribution of their left cortices—during frequency or pitch discrimination.

Prior to their first experiment, Foxton et al. (2009) measured DLFs in 28 normally hearing listeners using two tasks. The stimuli in the first task were 100-ms pure tones with different frequencies connected with a 300-ms linear frequency glide. The listeners heard one upward-gliding sound and one downward-gliding sound, both of the same magnitude, separated by a 500-ms gap, and in a random order. The listeners decided whether the first or the second sound glided upwards in pitch. In the second task, the listeners heard one steady 500-ms pure tone and one sound that glided in pitch, which was randomly an upward or a downward glide. Again, the two sounds were presented in a random order and were separated by 500 ms, and the listeners decided whether the first or the second sound contained the glide. In both tasks, the stimuli were centred on a common frequency per trial, which was one of seven values between 513 Hz and 595 Hz. The authors selected two groups of seven from the 28 individuals tested: a low-threshold group, whose DLFs in the pitch-direction task were in the range 8–25 cents, and a high-threshold group, whose DLFs in the pitch-direction task were much larger, in the range 1.4–2.5 semitones. In the pitch-change detection task, the DLFs were in the range 4–45 cents in the low-threshold group, and in the range of 27–85 cents in the high-threshold group. The groups did not differ markedly in terms of age or gender, and although none was a professional musician, the low-threshold group had received more musical training than the high-threshold group.

In the first MEG experiment, the 14 listeners heard glide stimuli similar to those presented in the behavioural sessions. Glides were either small (1.2 semitones) or large (9 semitones), upward or downward in direction, and had centre frequencies roved over the range 513–595 Hz. After each sound, the listeners decided whether the direction of the glide matched that of an arrow presented on a monitor. The MEG data were analysed using a 'regional-source montage' approach, which was intended to capture the activity from two coarsely defined areas of cortex, one in each hemisphere. A regional source is simply two dipoles with the same spatial location and orthogonal orientations averaged together[14]. The mean waveforms for the different hemispheres, glide sizes, and groups are shown in the top panels of Figure 2.26. The figure shows that all of the waveforms contained a prominent M100, which tended to be larger in the left source than in the right in both groups, and a less prominent deflection occurring later, which also tended to be larger in the left source. The authors subjected the peak of the M100 to an analysis of variance (ANOVA), and found it did not differ reliably between the two groups or the two sources. The authors entered the mean source dipole moments over four time windows (250–500, 500–750, 750–1000, 1000–1250 ms) into another ANOVA, and found a significant three-way interaction between source hemisphere, glide size, and group. The bottom panel of Figure 2.26 illustrates the interaction. From 250 to 1250 ms after stimulus onset (i.e., during and immediately after the listeners heard a frequency glide), the high-threshold group showed significantly greater activity in the left source during both small and large glides. Although the same trend was evident in the low-threshold group, it was only significant during the large glides. Foxton et al. took their results to suggest that their high-threshold listeners used sub-optimal, left-hemisphere brain regions to solve the pitch-direction task, which in turn limited their DLFs.

On a superficial level, Foxton et al.'s (2009) conclusions could be considered in keeping with a popular hypothesis of the specialised roles of the two hemispheres during auditory perception (see Zatorre, 2003; Zatorre & Gandour, 2008). The hypothesis states that the left-hemisphere auditory cortex has an anatomical structure that allows acoustic information to be transmitted and processed more quickly than in the right hemisphere, resulting in the left hemisphere having greater sensitivity to a sound's temporal aspects than the right hemisphere, but a relatively coarser representation of its spectrum. Conversely, the right hemisphere has a complementary structure that means it has coarser temporal but finer spectral resolution than the left hemisphere.

There is some evidence to support the existence of anatomical differences between the auditory cortices in both hemispheres. In a volumetric MRI study of 40 individuals, Penhune, Zatorre, MacDonald, and Evans (1996) found that there was reliably more white matter projecting to left HG than to right HG. Similarly, B. Anderson, Southern, and Powers (1999) found that the axons of these neurons were more thickly myelinated in the left hemispheres of eight cadavers. These studies lend

---

[14] In EEG, a regional-source montage would require three orthogonally oriented dipoles. However, in MEG, since any medial components of a dipole fitted within a spherical conductor are null, two dipoles are sufficient.

support to the hypothesis because it is well established in neurophysiology that more thickly myelinated fibres transmit neuronal impulses more quickly (see also Galuske, Schlote, Bratzke, & Singer, 2000; Hutsler & Gazzaniga, 1996; Sheldon, 1981). Further support for the idea of hemispheric specialisation in auditory processing comes from functional imaging studies. Using PET, Zatorre and Belin (2001) measured the changes in cerebral blood flow in response to sequences of pure tones that alternated in pitch by one octave at different temporal rates. As the speed of the alternation increased, so did the neural response in auditory cortices in both hemispheres, but the magnitude of this response was significantly greater in the left than in the

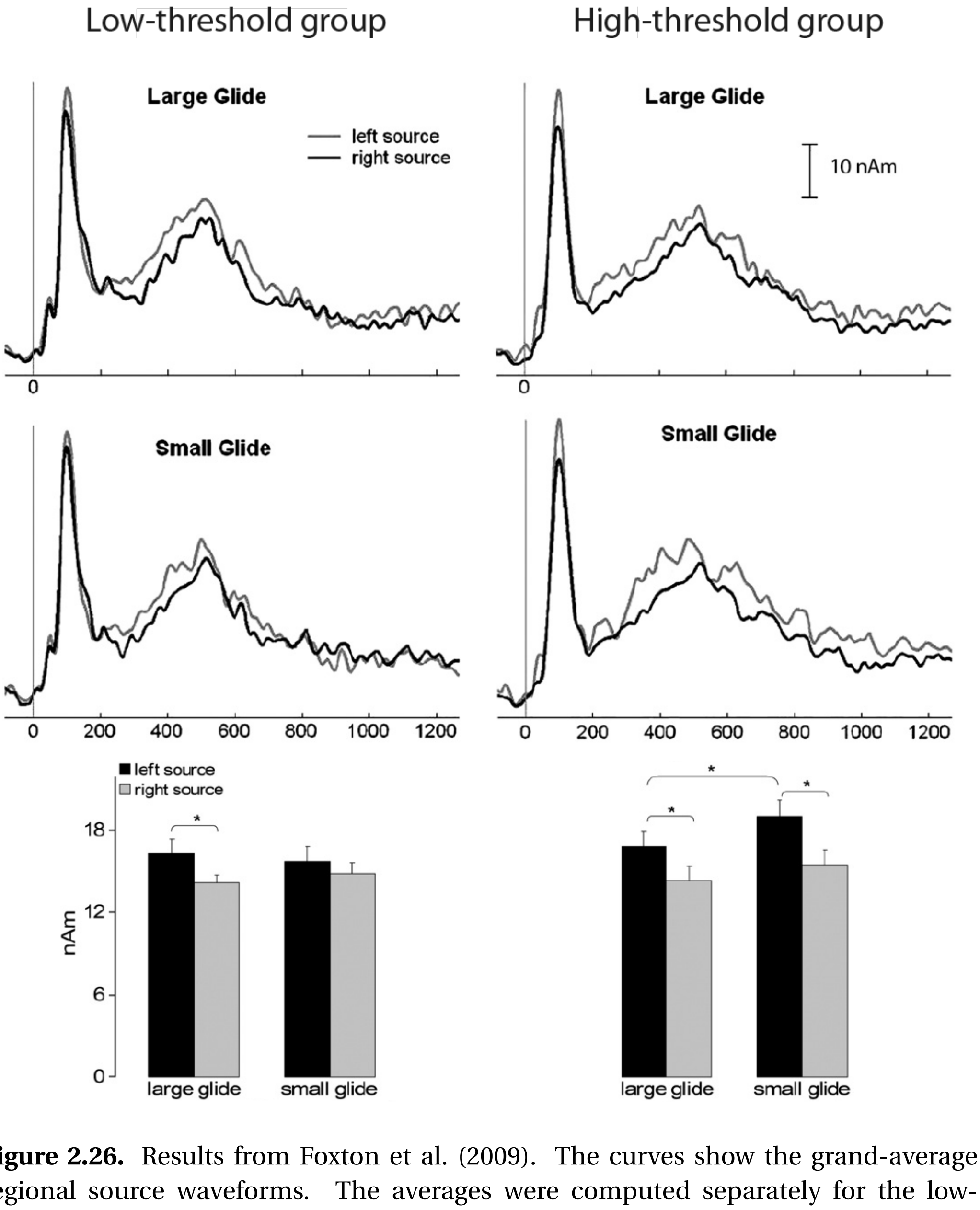


Figure 2.26. Results from Foxton et al. (2009). The curves show the grand-average regional source waveforms. The averages were computed separately for the low-threshold and high-threshold groups (left and right, respectively) and for the large and small pitch glides (top and bottom, respectively). The bars show the mean source-waveform amplitudes, collapsed across 250–1250 ms. The error bars show 1 standard error of the mean, and the asterisks represent significant differences.

right. Conversely, when the speed of the alternations was fixed but the number of tones sampled within the octave increased, the activity of the right hemisphere increased more profoundly (replicated with fMRI by Jamison, Watkins, Bishop, & Matthews, 2006). A similar result was reported by Schönwiesner, Rubsamen, and von Cramon (2005), who used noise bands that systematically varied in their spectral width and their temporal rate of change. The increasing temporal rate of change elicited increases in activity from lateral portions of HG that were more consistent in the left that the right, and increasing spectral rate of change elicited increases that were more consistent in the right (see also Boemio, Fromm, Braun, & Poeppel, 2005; Liégeois-Chauvel, Giraud, Badier, Marquis, & Chauvel, 2003; Liégeois-Chauvel, Graaf, Laguitton, & Chauvel, 1999; Okamoto, Stracke, Draganova, & Pantev, 2009).

Assuming that the right-hemisphere auditory cortex does have superior spectral resolution than the left-hemisphere cortex, how could this explain the results of Foxton et al. (2009) and Johnsrude et al. (2000)? One could argue that because the hypothesis predicts that poorer right-hemisphere function leads to generally poorer spectral resolution, it would affect both the detection and the identification of frequency changes equally. It should be noted that there is considerable scepticism concerning the basic temporal/spectral distinction, since the functional imaging literature does not support it unequivocally. Studies have failed to find a consistent left- or right-hemisphere advantage for fast AM rates (Giraud et al., 2000), for fast stimulus presentation rates (Harms & Melcher, 2002), for the introduction of rapid frequency modulation (D. A. Hall et al., 2002), or for the introduction of increases in the n parameter of IRN (T. D. Griffiths et al., 1998; T. D. Griffiths, Uppenkamp, Johnsrude, Josephs, & Patterson, 2001). The evidence for a right-hemisphere advantage in the perception of melody later on in the auditory-processing stream is arguably more consistent (e.g., Patterson et al., 2002; Scott, Blank, Rosen, & Wise, 2000; Zatorre & Belin, 2001), but the perception of pitch in some stimuli, such as IRN, is likely to based purely on temporal information. Thus, as pointed out by several authors, it is simply not meaningful to consider 'temporal' and 'spectral' as delineating the ends of a dimension that affords rapid temporal resolution at one end and pitch processing at the other (Patterson et al., 2002; Scott & Wise, 2004). Another important point is that studies have also found that the degree of hemispheric asymmetry observed for a particular stimulus can change depending on the task listeners perform. For instance, the activity of right-hemisphere areas increases if listeners are required to discriminate a frequency change compared to if they are required to passively listen to the same stimuli (e.g., Brechmann & Scheich, 2005; Zatorre, Evans, & Meyer, 1994), or if they are required to discriminate its phonetic structure instead (e.g., Zatorre, Evans, Meyer, & Gjedde, 1992). In other words, the asymmetry of processing in auditory cortex is influenced by both stimulus and task characteristics. Thus, a refinement of the basic hypothesis might be to consider the right hemisphere as specialised for the perception of variations in pitch over time, containing mechanisms that allow a listener to determine the direction of small pitch changes quickly and accurately. Foxton et al.'s (2009) conclusion is more consistent with this refinement.

There are a number of issues with Foxton et al.'s (2009) interpretation of their data. First, although their high-threshold group had a more profound inter-hemispheric asymmetry in AEF strength, the pattern of their data was not qualitatively different from that of the low-threshold group. The latter group also had a stronger left-hemisphere source on average, but the asymmetry was weaker. Therefore, what Foxton et al. found was effectively a correlation between the extent of a listeners' inter-hemispheric auditory-processing asymmetry and their ability to determine frequency-glide direction. The authors recorded performance accuracy during the MEG session, but they did not perform any brain–behaviour correlations that would have strengthened their conclusion.

A second issue is that, since the listeners were always required to identify the direction of the glides they heard, the group difference could be related to the mechanisms involved in pitch discrimination, the detection of a frequency glide, the mechanisms associated with the presence of sound energy, or simply a general difference in the strength of the MEG signal emanating from the left and right hemispheres. The authors would undoubtedly support the first option, but crucially they did not report the strength of the interaction between source hemisphere and listener group on the amplitude of the M100. This statistic would have been informative because the M100 occurred before the listeners were required to identify the direction of a glide. Even if the interaction was not significant, there was clearly a trend for a greater leftward asymmetry in the amplitude of the M100 in the high-threshold listeners than in the low-threshold listeners. This observation weakens Foxton et al.'s (2009) argument that the differences they found were related to pitch-direction identification per se. It is also not clear why there was a general leftward asymmetry in their data at all since it is assumed that the right hemisphere has the more important for role in pitch discrimination.

A third issue is that the listeners did not have much experience with frequency or pitch discrimination before taking part in the experiment. DLFs were measured prior to the MEG session, but at most the listeners heard only 180 glide stimuli before their AEFs were measured. DLFs measured with such a small number of trials might be quite unreliable, and with more testing, the large performance difference between Foxton et al.'s (2009) groups might have diminished.

In an EEG experiment, Pardo and Sams (1993) found that an MMN was evoked when a series of pitch glides contained occasional deviant glides in the opposite direction the standard. Other studies have shown that an MMN is also elicited for pairs of discrete tones that violate an otherwise regular feature of ascending or descending pitch (Paavilainen, Saarinen, Tervaniemi, & Näätänen, 1995; Paavilainen, Jaramillo, Näätänen, & Winkler, 1999; Saarinen, Paavilainen, Schröger, Tervaniemi, & Näätänen, 1992). In their second experiment, Foxton et al. (2009)'s same listeners heard a series of pitch glides designed to elicit an MMN: the standards were all 1.2-semitone glides in the same direction and a roving centre frequency, and the deviant was a 1.2-semitone glide in the opposite direction. During the scan, the listeners attended to a silent movie. Behavioural data collected at the end of the recording session revealed significant a group difference in the ability to detect the deviants, with the low-threshold and high-threshold groups detecting on average 90% and 61% of the deviants, respectively. An MMN was observed in both groups, but despite the considerable group difference in detecting the deviants behaviourally, the trend for a smaller MMN in the high-threshold group was not significant. The authors suggested that this trend reflects the fact that the high-threshold group were less able to recognise the deviants as different from the standards, and the failure to reach statistical threshold resulted from a lack of power.

2.5 Chapter summary

Pitch-evoking sounds are an essential part of our natural auditory environment, and our auditory systems are well equipped to deal with pitch information. The current consensus is that pure-tone pitch below 4000/5000 Hz is derived through timing-based mechanisms that utilise the phase-locking properties of neurons in the auditory nerve, rather than through the place of maximal excitation on the basilar membrane (although tonotopic organisation is preserved up into the cortex). These mechanisms are evidently very efficient; experiments have demonstrated that normally hearing listeners are usually sensitive to tiny changes in frequency between pure tones, and are able to store pitches effectively in memory. Factor analysis studies have suggested that listeners' pure-tone frequency-discrimination abilities correlate with their abilities to discriminate other aspects of a sound's spectrum, but are at best only weakly correlated with other auditory abilities—such as speech perception and resistance to masking—and uncorrelated with non-auditory abilities. Research suggests that the influence of musicianship on DLFs is surprisingly modest and can be diminished with sufficient psychoacoustical training, although the difference between musicians and non-musicians is more profound in other, more complex auditory tasks. One source of inter-subject variability relevant to the work conducted in this thesis is that some listeners are insensitive to the direction (up or down) of frequency changes between pure tones. This dissociation was first observed in special groups (amusics and those with brain damage) but more recently has been found in listeners from the general population. A hypothesis that has emerged from the literature is that pitch-direction identification relies more heavily on areas in the right hemisphere of the brain, and research suggests—tentatively—that listeners who are relatively insensitive to pitch-direction may be relying on complementary and less efficient mechanisms in their left hemispheres.

Chapter 3

Experiments 1 and 2: Stimulus uncertainty and insensitivity to pitch-change direction

3.1 Rationale

The following chapter reports two psychophysical experiments measuring listeners' difference limens for frequency (DLFs). In the classic experiments discussed in Section 2.2.4, listeners were presented with two successive pure tones on each trial. These tones differed slightly in frequency, and the listeners indicated which tone had the higher pitch. This kind of task requires listeners to not only detect the small difference in frequency, but also requires them to be sensitive to its direction, either up or down. The DLFs measured in those experiments were typically very small (e.g., Moore, 1974). However, a series of more recent experiments has suggested that in certain circumstances some listeners can detect the presence of a small frequency difference, but are unable to identify the direction of the resulting percept (Section 2.3.3.2).

Of the psychoacoustical studies demonstrating a dissociation between DLFs for frequency-change detection and DLFs for frequency-direction identification, only one (Semal & Demany, 2006) used listeners selected from the general population. Listeners in previous studies either had unilateral or bilateral temporal lobe lesions which encroached Heschl's gyrus (HG; Johnsrude et al., 2000; Tramo et al., 2002) or suffered from congenital amusia (Foxton et al., 2004). By contrast, Semal and Demany's listeners appeared to be audiologically and neurologically normal. An important difference between Semal and Demany's study and those involving clinical patients is that the former employed frequency roving—on each trial, the frequencies of the first tones in each pair were randomly varied over a wide range (400–2400 Hz). This raises the question of whether the specific deficit observed for some of Semal and Demany's listeners would occur under conditions in which the stimuli are not roved in frequency over a wide range.

A number of previous studies have demonstrated that frequency roving is deleterious to normally hearing listeners' DLFs (Section 2.2.4.4). In particular, the study by Amitay et al. (2005) suggests that frequency roving increases DLFs overall, but to different extents for different listeners. Since that study only measured DLFs using a task requiring sensitivity to frequency-change direction, and DLFs with and without roving were not measured in the same listeners, it is not known if the influence of roving on DLFs is greater in individuals who have difficulty identifying the direction of frequency changes.

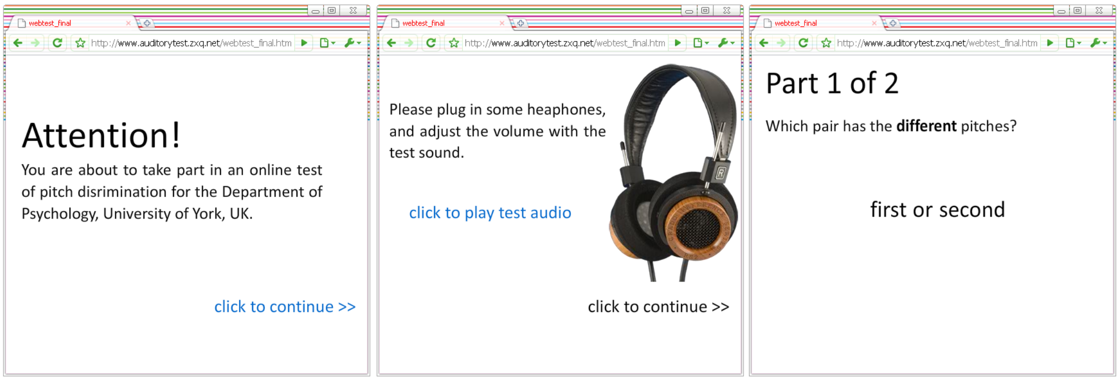
There are several possible reasons why frequency roving might cause larger DLFs. One possibility is that the absence of roving allows listeners to forgo a direct comparison of the tones within each trial. Instead, the repeated presentation of the standard frequency could allow listeners to build a relatively more precise and stable referent in long-term memory, and the tones within each trial could be compared to the referent rather than to each other. Another possibility is that frequency roving increases stimulus uncertainty. Uncertainty is thought to be a key feature of experiments that have observed large amounts of informational masking (IM; Section 2.3.2). Models of IM (Lutfi, 1993; Durlach et al., 2005) attribute the decrements in signal detection to the increase in ensemble variance at the output of auditory channels; essentially, listeners' thresholds increase as the variability in the stimuli over trials increases. Whether larger DLFs in frequency-discrimination tasks are caused by a change in listening strategy or by increased uncertainty—or for some other reason—is not presently clear.

The experiments described in this chapter aimed to investigate the role of stimulus uncertainty in some listeners' insensitivity to the direction of frequency changes between pure tones. I follow Semal and Demany (2006) in assuming that the primary cue listeners used in these experiments was pitch, and therefore that, although cues other than pitch may have sometimes been available, it is appropriate to refer to them as pitch discriminations; other possible cues on which the discriminations could have been based are discussed in Section 3.4. As a preliminary step, a short pre-test experiment was developed in order to screen for listeners who might experience a particular difficulty perceiving pitch-change direction. Experiment 1 proper aimed to replicate and extend the findings from Semal and Demany's main experiment. As in that study, 'detection' DLFs (DDLFs) and 'identification' DLFs (IDLFs) were measured in a cohort of mostly pre-selected listeners using a dual-pair paradigm with frequency roved both between and within trials. The experiment also contained additional conditions that included task-irrelevant, pure-tone 'maskers'; the purpose of these conditions was to determine whether or not increased stimulus uncertainty brought about by IM would affect

DDLFs and IDLFs differently. In Experiment 2, most of the same individuals were retested in a similar set of conditions, this time without frequency roving.

3.2 The pre-test

The pre-test was very similar to the one used by Semal and Demany (2006) to select listeners for their original experiments (personal communication with the authors). There were two versions of the pre-test. The first version was administered in large lecture rooms to undergraduate year groups as a whole, whilst the students remained in their seats after the end of a lecture. An announcement was made at the beginning of the lectures, explaining that a short psychological experiment concerning hearing would be conducted at the end. The students were given the opportunity to leave before the test started. The sounds were delivered via the loudspeaker system, visual instructions were delivered via the over-head projector, and responses were made on pre-printed response sheets. The online version was advertised via email, and contained step-by-step visual instructions. Listeners were required to connect headphones to their computer and to adjust their volume settings to a comfortable level using a test tone. Responses were made with mouse clicks. Screenshots of the online pre-test are shown in Figure 3.1.



Attention!

You are about to take part in an online test of pitch disrimination for the Department of Psychology, University of York, UK.

click to continue >>

Please plug in some headphones, and adjust the volume with the test sound.

click to play test audio

click to continue >>

Part 1 of 2

Which pair has the **different** pitches?

first or second

Figure 3.1. Screenshots at various stages of the online version of the pre-test. Listeners were instructed to attach headphones and to enter a valid email address, and responses were made by clicking the text (blue when hovered over). The test can be found at http://www.auditorytest.zxq.net/.

Both versions of the pre-test contained the same 60 trials, summarised in Table 3.1. The stimuli and tasks were identical to those in the conditions without maskers in Experiment 1: there were two pairs of pure tones per trial, one of which contained a frequency change (ΔF, expressed in musical cents); the first 30 trials were ‘detection’ trials and in the last 30 were ‘identification’ trials (explained in greater detail in Section 3.3.1.2). Only the identification trials required listeners to be sensitive to the direction of the frequency difference. The pre-test took around 8 min.

Table 3.1. Trials in the pre-test.

Trials	Task	ΔF (cents)
1–10	detection	150
11–20	detection	50
21–30	detection	25
31–40	identification	600
41–50	identification	60
51–60	identification	30

In total 672 individuals completed one or other pre-test. In each individual, the proportion correct (*PC*) of identification trials was subtracted from the *PC* of detection trials. A positive score indicated that the listener might have a relative difficulty with pitch-direction identification. Listeners for the main experiments were therefore chosen from the 119 individuals (17.71%) whose identification *PC* was smaller than their detection *PC*, and for whom this difference exceeded one standard deviation of the mean difference across the whole group (Figure 3.2).

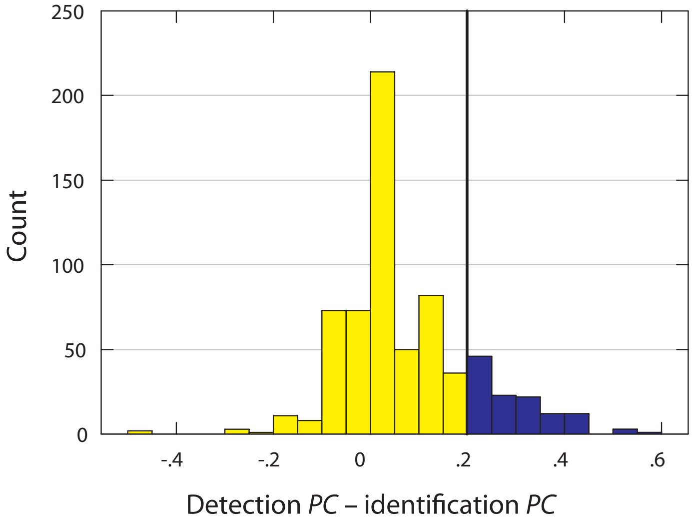


Figure 3.2. Histogram showing the distribution of scores on the pre-test. Listeners for the main experiments were selected from the right-hand side of the vertical reference line (blue bars), which represents one standard deviation above the mean difference score (.20).

3.3 Experiments 1 and 2

3.3.1 Methods

3.3.1.1 Listeners

Sixteen listeners (13 females; 12 right-handed; age range 18–26 years) took part in Experiment 1, of whom 13 also took part in Experiment 2 (11 females; 10 right-handed; age range 18–24 years). In what follows, individual listeners are identified by number (L1, L2, etc). L1 was the author, and only L1–3 had any prior experience in psychoacoustical experiments involving frequency and/or pitch discrimination. L4–16, who were paid for their participation, were undergraduates selected using the pre-test. L1–3 took part in the online version of the pre-test and their results did not indicate a difficulty with pitch-direction identification. Listeners varied in their musical experience, but none was a professional musician or performed music regularly.

Hearing levels were measured for frequencies between 250 and 4000 Hz in octave steps using pure tone audiometry. For all listeners except three, hearing levels did not exceed 20 dB HL at any frequency. For L4, hearing level was 25 dB HL for 250 and 500 Hz in the right ear, and 30 dB HL for 250 and 500 Hz in the left ear. For L12 and L16, hearing levels were 25 dB HL for 250 and 500 Hz in their right ears[1]. For all listeners, worst-ear averages over all frequencies were smaller than 20 dB HL.

3.3.1.2 Stimuli

All the tones in Experiments 1 and 2 consisted of 250-ms sinusoids presented at a nominal sound pressure level (SPL) of 60 dB, with random starting phases and 20-ms cosine-squared on/off ramps. The stimuli were generated digitally and delivered diotically through headphones (Sennheiser HD580) using a 24-bit digital-to-analogue converter (LynxONE) at a sampling rate of 22500 Hz.

In both experiments, each trial contained four successive tones arranged temporally into two pairs, with a 700-ms silent interval (ISI) between pairs, and a 250-ms ISI between the tones in each pair (Figure 3.3). In half of the conditions in both experiments, all four tones per trial were isolated pure tones. In one of the pairs, selected randomly, the second tone was equiprobably higher or lower in frequency than the preceding tone by an amount (ΔF) expressed in musical cents (1 cent equals 1/100th of a semitone or 1/1200th of an octave). In the other pair, both tones were identical. In Experiment 1, the frequencies of the first tones in each pair were independently drawn from a rectangular distribution defined on a log-frequency (cents) scale with limits of 0 and 3102 cents above 400 Hz (400 and 2400.1 Hz). In Experiment 2 there was no frequency rove—the first tone in each pair always had the same frequency, 979.8 Hz (1551 cents above 400 Hz), which corresponds to the centre of the roving range in Experiment 1.

In the other half of the conditions in both experiments, the stimuli were identical except for the addition of four random-frequency tones per trial. The first tone in each pair—the ‘target’—was presented with two spectrally flanking and temporally

---

[1] Since all the stimuli in the experiments were clearly audible to all listeners, that it is unlikely that the minor audiometric abnormalities observed in some individuals influenced the results. Therefore, I have chosen to include the data from all listeners in the thesis.

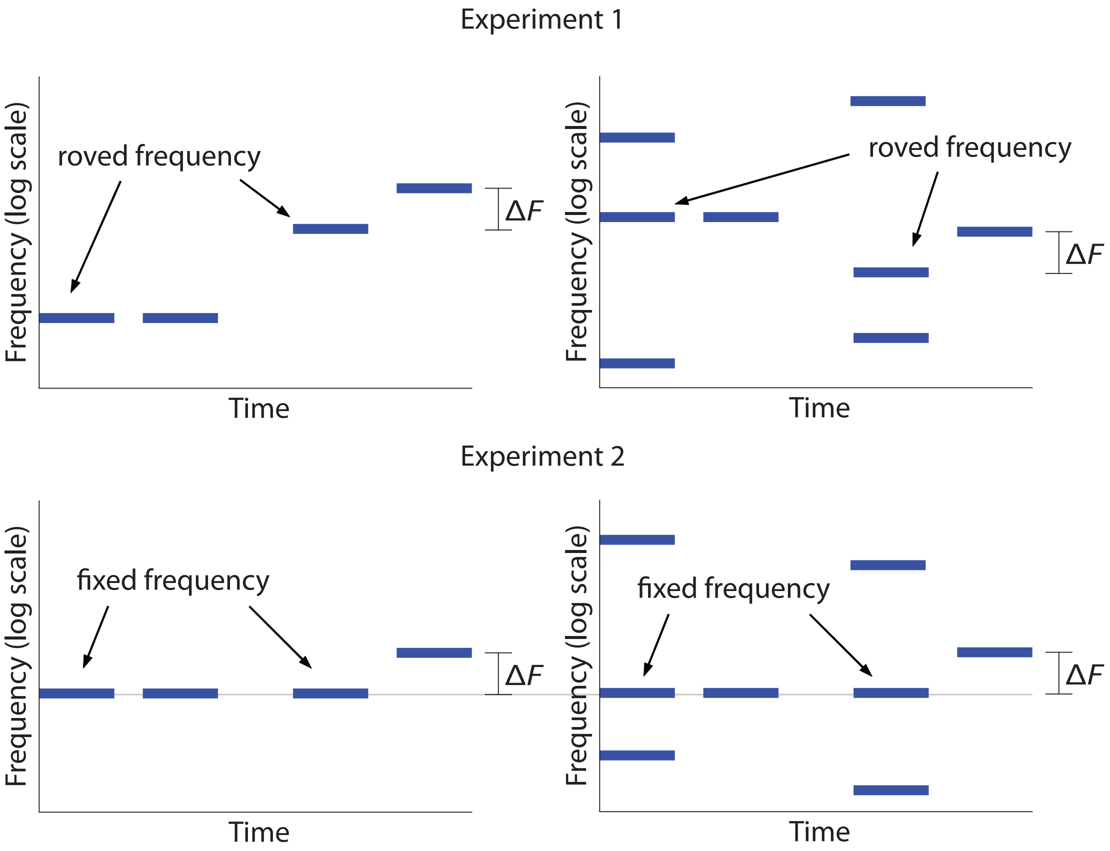


Figure 3.3. Illustrations of example trials in Experiments 1 and 2. Blue lines represent sinusoids. The top panels show trials in Experiment 1 and the bottom panels show trials in Experiment 2. The left- and right-hand panels show trials without and with additional maskers, respectively. The grey lines in the bottom panels represents the standard frequency (979.8 Hz) used in Experiment 2. In all four examples, the correct responses would be ‘second’ (for detection) or ‘up’ (for identification).

synchronous pure tones, referred to hereafter as ‘maskers’. The frequencies of the maskers were always randomly drawn from a rectangular log-frequency distribution, but constrained so that one masker was always higher and one was always lower than the target, and that each was separated from the target by at least 650 cents. The limits of the masker-roving range were 550 and 4952 cents above 200 Hz (274.9 and 3493.7 Hz). So in these conditions, the first tone in each pair could be considered an inharmonic chord with three sinusoidal components. In one of the pairs, selected randomly, the isolated tone occurring after the chord was equiprobably higher or lower in frequency than the central chord component by ΔF. In the other pair, the isolated tone was identical to the central component. Note that in Experiment 2, although the target and isolated tones were not roved in frequency, the maskers were still roved over the 274.9–3493.7 Hz range.

3.3.1.3 Procedure

Prior to testing, hearing levels were measured and each listener completed two questionnaires: the speech, spatial, and qualities of hearing scale questionnaire (SSQ; Gatehouse & Noble, 2004) and a bespoke questionnaire about their musical experience and expertise. Following these assessments, listeners took part in Experiment 1. DLFs were measured using two different kinds of stimuli (see above) and two tasks—making four conditions in total—in separate runs of trials. When the stimuli did not contain maskers, the detection task required listeners to indicate which tone pair contained the frequency change (first or second), and the identification task required listeners to indicate the direction of the change (up or down) that occurred in the pair containing different tones, without specifying which pair this was. When the stimuli contained maskers, the tasks were the same except listeners made their comparisons between the central chord component and the isolated tone: in the detection task, they indicated which pair contained the frequency change; and in the identification task, they indicated the direction of the change. Listeners first completed a few trials in each condition to familiarise themselves with the stimuli and procedure. Testing was carried out individually in a sound-attenuating booth (IAC) in three sessions lasting approximately 1 h each, on different days. Responses were made using a keyboard, and were followed by visual feedback and a 600-ms pause before the start of the next trial.

The magnitude of ΔF in a run of trials was set initially to 100 cents, and was manipulated within the run using a weighted one-up one-down adaptive procedure that estimates DLFs corresponding to 75% correct on the psychometric function (Kaernbach, 1991b). Up to the fifth reversal in the direction of the staircase, ΔF was decreased by a factor of $\sqrt[3]{2.25}$ following a correct response and increased by a factor of 2.25 following an incorrect response. At the fifth reversal onwards, down and up step sizes were $\sqrt[3]{1.5}$ and 1.5, respectively. In each run, ΔF was limited to a maximum of 600 cents; this limit was imposed so that in the conditions containing maskers, the frequency of the isolated tone could never exceed that of the highest or lowest maskers. A run ended after the 14th reversal. DLF for a run was defined as the geometric mean of all ΔF values used from the fifth reversal onward. During testing, instructions for the current task were presented on the computer screen, together with the current trial number and current ΔF to the nearest integer. The condition alternated from one run to the next, with the order counterbalanced across listeners. Twenty DLFs were measured in each condition and in each listener.

Thirteen of the 16 listeners who had taken part in Experiment 1 subsequently took part in Experiment 2. The procedure of this experiment was similar to that of Experiment 1, except that 15 DLFs were measured in each condition and in each listener, using the non-roved stimuli described above. Testing was carried out in two 1-h sessions on different days.

3.3.2 Results

3.3.2.1 Experiment 1: Wide frequency roving without maskers

DLFs in the conditions without maskers in Experiment 1 are shown in the top-left-hand panel of Figure 3.4, in which listeners' geometric mean DDLFs and IDLFs are plotted as abscissae and ordinates, respectively. Mann–Whitney tests with a Bonferroni-corrected significance criterion ($\alpha = .003125$) were used to compare DDLFs and IDLFs on a listener-by-listener basis. The results indicated that L7–14—whose data form a visible cluster in the panel—all had significantly larger IDLFs than their DDLFs ($U \leq 24.00,\ z \leq -4.76,\ p < .001,\ r \leq -.75$). IDLF was also significantly larger than DDLF in L6, whose data fell outside of this cluster, but the effect was considerably smaller for this listener ($U = 76.00$, $z = -3.35$, $p < .001$, $r = -.53$) than in L7–14. The data from these individuals, hereafter referred to as 'direction-impaired' listeners, are shown as red symbols in Figure 3.4.

As in Semal and Demany's (2006) study, impairments in pitch-direction identification observed here persisted after several hours (in this case, approximately 3 h) of task practice with feedback after each trial. Comparisons between the DLFs measured in direction-impaired listeners in the two studies should be made with some caution, because those measured here may have been limited by the 600-cents cap imposed on ΔF, which was not present in Semal and Demany's study. The DLFs were also based on fewer threshold estimates in this experiment (20 per listener per condition) than in Semal and Demany's study (50 per listener per condition).

None of the other listeners showed a difference between their DDLF and IDLF that was significant at the corrected level ($U \geq 95.00,\ z \geq -2.84,\ p \geq .004,\ r \geq -.45$). L1 and L2, who achieved the smallest DLFs overall, had IDLFs that were approximately 5 to 6 cents smaller than their DDLFs. Though counterintuitive, smaller IDLFs than

DDLFs were also observed in some listeners by Semal and Demany (2006), and this pattern of performance is broadly consistent with the predictions of a model they propose in which both the detection and identification tasks are limited by the same internal noise (Section 6.2). For L3–5, IDLFs were slightly larger than their DDLFs. Semal and Demany also observed this pattern of performance, and overall DLFs in L1–5 are reasonably similar to those of their counterparts in that study.

Two listeners did not fit neatly into any of the groups identified by Semal and Demany (2006). L15 and L16 had very large DDLFs (143 and 253 cents respectively),

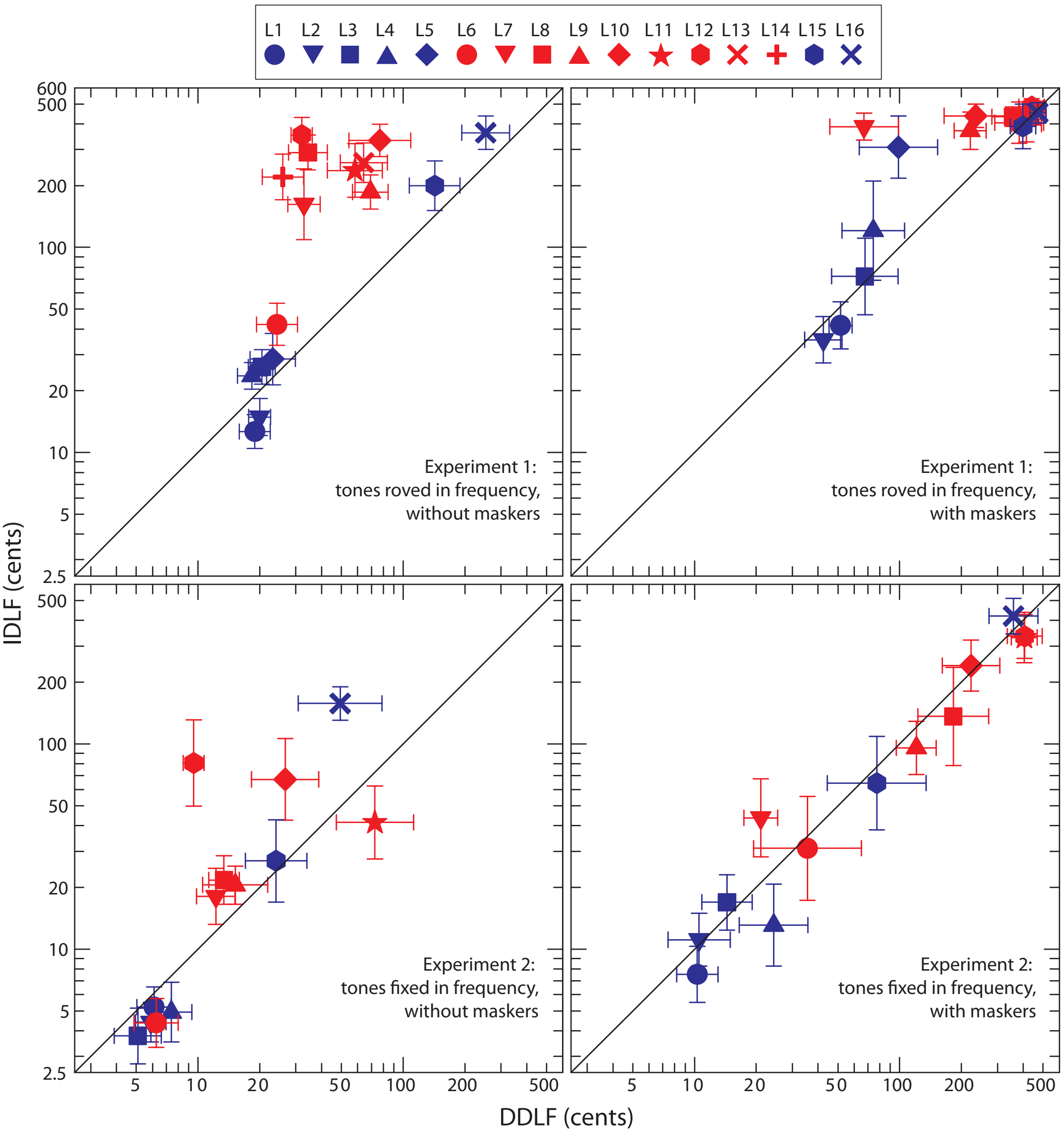


Figure 3.4. Results of Experiments 1 and Experiment 2. Each listener is represented by a unique symbol and number. Listeners represented by red symbols are those deemed to be direction-impaired listeners in Experiment 1 (see text). The abscissa and ordinate of each symbol represents that listener's geometric mean DDLF and IDLF, respectively, with error bars representing geometric 95% confidence intervals, on logarithmic axes. Error bars are smaller than the symbol if not shown. The diagonal line in each panel represents the locus of equivalent performance.

yet relatively their IDLFs were not elevated significantly. This combination of large IDLF and large DDLF in the same individual was not observed for the smaller sample of listeners tested by Semal and Demany. L15 exhibited a bias for better performance on both tasks when the frequency difference was in the second pair, responding correctly on only 62% of first-pair trials compared to 85% of second-pair trials. Contrastingly, L16 was not able to perform either task at the accuracy targeted by the adaptive procedure (75% correct), with an overall accuracy of 63% correct across all trials. The DLFs measured in this listener are therefore likely to be an underestimate of her true thresholds, probably brought about by the use of the 600-cents limit on ΔF.

3.3.2.2 Experiment 1: Wide frequency roving with maskers

The top-right-hand panel of Figure 3.4 shows listeners' DDLFs and IDLFs in the conditions containing maskers in Experiment 1. The addition of the maskers resulted in much larger DLFs overall, with each listener's DDLF and IDLF being larger when the stimuli included maskers than when no maskers were present. Mann–Whitney tests with a Bonferroni-corrected significance criterion were used again to compare listeners' DDLFs and IDLFs in these conditions. The tests revealed significantly larger IDLFs than DDLFs in L5, L6, L9, and L10 ($U \leq 76.00,\ z \leq -3.30,\ p \leq .001,\ r \leq -.52$), but not in L8 and L11–14 ($U \geq 152.50,\ z \geq -1.28,\ p \geq .20,\ r \geq -.20$). It seems unlikely that the listeners deemed direction-impaired using the stimuli without maskers were *less* impaired when maskers were introduced, and more likely that their DLFs with maskers were underestimated considerably in the present experiment due to the limit on ΔF. L5 is the only listener to have a significantly larger IDLF than DLF with maskers but not without maskers.

3.3.2.3 Experiment 2: No frequency roving without maskers

DLFs in the conditions without maskers in Experiment 2 are shown in the bottom-left-hand panel of Figure 3.4. Overall, these DLFs were smaller than the DLFs measured in the equivalent conditions in Experiment 1. Bonferroni-corrected Mann–Whitney tests ($\alpha = .003846$) were again used to compare listeners' DDLFs and IDLFs. Four listeners had significantly larger IDLFs than their DDLFs (L8, L10, L12 and L16; $U \leq 42.00,\ z \leq -2.92,\ p < .003,\ r \leq -.46$), and none of the remaining listeners exhibited a significant difference between their DDLF and IDLF ($U \geq 58.00,\ z \geq -2.26,\ p \geq .02,\ r \geq -.36$).

Of the nine listeners who had significantly larger IDLFs than DDLFs in the conditions without maskers in Experiment 1 (L6–14), only three showed a similar effect in Experiment 2 (L8, L10 and L12). For L6, L7, L9 and L11, IDLFs were no longer significantly larger than their DDLFs at the corrected level. L16 was the only listener whose IDLF was significantly elevated relative to their DDLF in Experiment 2 and not in Experiment 1, probably as a result of her thresholds being underestimated in Experiment 1. L13 and L14 were unavailable for re-testing in Experiment 2.

3.3.2.4 Experiment 2: No frequency roving with maskers

DLFs in the conditions with maskers in Experiment 2 are shown in the bottom-right-hand panel of Figure 3.4. Bonferroni-corrected Mann–Whitney tests revealed that L7 was the only listener to have a significantly larger IDLF than DDLF in these conditions ($U = 30.00,\ z = -3.05,\ p < .01,\ r = -.48$). For all of the others, their IDLF was not significantly different from their DDLF ($U \geq 52.00,\ z \geq -2.51,\ p \geq .01,\ r \geq -.40$). The DDLFs and IDLFs measured in L10, L12, and L16—for whom the comparison between DLFs measured with and without maskers would have been most informative—were all close to 600 cents and therefore probably underestimated. Thus, it is not clear whether increasing stimulus uncertainty by introducing an element of IM increased or decreased a listener's relative insensitivity to pitch-change direction.

3.3.2.5 Influence of roving on I/D ratios without maskers

A listener's relative ability to identify pitch-change direction can be quantified by dividing their IDLF by their DDLF. Figure 3.5 (left-hand panel) shows the geometric mean 'I/D ratios' (in conditions without maskers) for the group who took part in both experiments. A paired-samples t-test performed on the log-transformed I/D ratios confirmed that they were reliably smaller in Experiment 2 for the group as a whole [$t(12) = 2.81,\ p < .05,\ r = -.63$]. The effect remained significant when only the listeners who showed the pitch-direction impairment in either or both experiments (L6–12 and L16) were included in the analysis [Figure 3.5, right-hand panel; $t(7) = 2.63,\ p < .05,\ r = -.70$].

3.3.2.6 Susceptibility to IM

A listener's relative susceptibility to IM can be quantified also by dividing their DLF measured with maskers by their DLF measured without maskers. However, since DLFs with maskers were probably underestimated for at least some listeners in both Experiments 1 and 2, an analysis of these ratios was unlikely to have been very informative and was not performed.

3.4 Discussion

The following discussion will focus primarily on the results from those conditions in which maskers were not included. An unpublished experiment by the author had suggested that adding additional maskers would result in larger DLFs due to IM

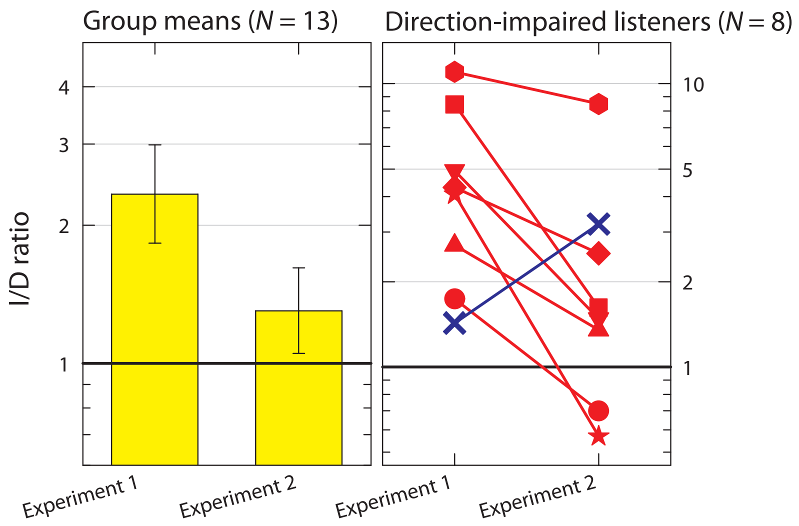


Figure 3.5. I/D ratios for the conditions without maskers in Experiments 1 and 2. The left-hand panel shows geometric mean ratios for the group taking part in both experiments, with bars representing 1 geometric standard error of the mean, on a logarithmic axis. The right-hand panel shows the I/D ratios for the eight individuals showing the pitch-direction impairment in either or both experiments, with symbol and colour combinations consistent with Figure 2. Solid horizontal lines in both panels represent an I/D ratio of 1 (i.e., equal DDLFs and IDLFs).

(Section 2.2.4.5). Since IM is often described as an effect of stimulus uncertainty (Section 2.3.2), and the influence of roving on frequency discrimination could be due to uncertainty also, the present experiments were designed to explore the possible interactions between susceptibility to IM, the use of frequency roving, and the task listeners performed. However, the IM effect in Experiments 1 and 2 was much stronger than initially expected, and the DLFs measured in many listeners in the experiments were probably underestimated. This would have vitiated any conclusions drawn from the results regarding the influence of IM.

The observed tendency for some listeners to be poorer at pitch-direction identification than pitch-change detection has been suggested by Foxton et al. (2004) to be a characteristic of people with congenital amusia. Although the present study did not involve testing for amusia in a standardised way (cf. Peretz et al., 2003), our questionnaires did not reveal any gross differences between direction-impaired and unimpaired listeners in terms of their musical ability or experience, and none of the listeners reported any difficulty in appreciating music (Chapter 7). The result that relative impairments in pitch-direction identification tended to disappear when fixed-frequency stimuli were used also suggests that the impairment observed here is different in origin to that seen in patients with cortical lesions (Johnsrude et al., 2000; Tramo et al., 2002).

All of the direction-impaired listeners tested in Experiments 1 and 2—and their counterparts in Semal and Demany's (2006) study—were female[2]. It is not clear whether this finding reflects a real gender difference in normally hearing listeners' sensitivity to pitch-change direction, or whether it is simply the result of a sampling bias. The proportion of female Psychology undergraduates is typically very high, so it is very likely that the majority of the individuals tested using both the present pre-test and the one by Semal and Demany (personal communication) were female. However, gender information was not recorded explicitly by either pre-test, which prevents making an accurate estimate of the expected male/female ratio within the sample of direction-impaired individuals.

Consistent with previous studies (Amitay et al., 2005; Demany & Semal, 2005; Harris, 1952b; Jesteadt & Bilger, 1974), Experiment 2 revealed that when frequency roving was removed, DLFs were usually reduced considerably. One possible reason why DLFs were smaller in the absence of roving is that it was no longer necessary for listeners to compare directly the pitch of the two tones within each pair for correct task performance. Instead, in Experiment 2 listeners could compare each tone individually against an internal representation of the standard tone (979.8 Hz), which occurred three times on every trial. In a theory of intensity discrimination, Durlach and Braida (1969) referred to these two strategies as sensory-trace and context coding, respectively. The repeated presentation of the standard tone in Experiment 2 may have provided listeners with the opportunity to build an internal referent that was relatively more precise and stable than the evanescent echoic memory trace of the tones in Experiment 1. Thus, in Experiment 2 the variance of the internal noise associated with context coding would be less than that associated with trace coding, which could explain why DLFs were generally smaller in that experiment. However, Experiment 2 also revealed that impairments in pitch-direction identification relative to pitch-change detection were reduced when roving was not used. Since context coding does not eliminate the need for the listener to be sensitive to pitch-change direction, it is unclear why a switch from trace coding to context coding should improve thresholds so much more in the identification task than in the detection task.

Another interpretation of the results is that removing frequency roving decreased stimulus uncertainty. The effects of uncertainty have been modelled previously in different ways for different psychophysical experiments, for example by varying the width of listening bands in probe-signal detection (Schlauch & Hafter, 1991), as the number of noisy templates in visual discrimination (Pelli, 1985), or as the ensemble variance at the output of auditory channels in informational masking (Lutfi, 1993; Durlach et al., 2005). However, in general, more uncertainty is thought to lead to an increase in internal noise and consequently poorer task performance. In Experiments 1 and 2, any model of performance that assumes that the effects of roving are due to stimulus uncertainty would additionally have to assume that, in

---

[2] This is also true for Experiments 3–5 in later chapters of the thesis.

some listeners, uncertainty results in a larger proportional increase in internal noise in the identification task than in the detection task. As with the coding-strategy hypothesis, why this would be the case is not presently clear.

While removal of the frequency roving is the most likely explanation for the reduction in IDLFs between Experiments 1 and 2, other explanations are conceivable. Since Experiment 1 was always performed first, the improvement in DLFs could reflect learning. This potential confound was addressed in Experiment 3 (Section 4.2), in which a new sample of direction-impaired listeners completed runs of fixed- and roving-frequency conditions in random order.

Other possible explanations for the differences between Experiments 1 and 2 follow from the fact that the nominal level of the tones in the experiments was fixed. Listeners might have exploited this feature of the stimuli in two ways. First, as discussed by Semal and Demany (2006), differences in the output level of a single auditory channel could provide a cue other than pitch for successful task performance. This strategy would be useful in the detection task, but for it to be useful in the identification task, the listener would have to choose an appropriate channel. The optimal frequency of the channel would have to be consistently higher or lower than the frequency of the first tone in each pair, so that the direction of the excitation-level change co-varied consistently with the direction of the frequency change. This listening strategy might have been more difficult to achieve when the stimuli were roved in frequency than when they were fixed, because in the former case it would require constant re-selection of the listening channel, and in the latter case the same channel could be used for all trials. Listeners may have also used differences in loudness associated with changes in frequency (cf. Mauermann, Long, & Kollmeier, 2004; Moore, Glasberg, Low, Cope, & Cope, 2006) as a cue for successful task performance. Equal-loudness contours over the range 400–2400 Hz are not monotonic (Suzuki & Takeshima, 2004), so for tones roved over this range the direction of the loudness cue would not always be concordant with the direction of a frequency change. If tones were fixed in frequency, however, loudness-change direction and frequency-change direction could covary systematically because the standard tone would occupy a single point on the loudness-frequency function. Use of either or both of the strategies highlighted above could have contributed to the smaller DDLFs than IDLFs in Experiment 1, and to the reduced differences between DLFs for the two tasks in Experiment 2. This issue was also investigated in Experiment 3 described the next chapter.

Chapter 4

Experiments 3 and 4:
The influence of frequency roving on pure-tone pitch discrimination

4.1 Rationale

The results of Experiments 1 and 2 suggest that the relative insensitivity to pitch-change direction observed in some otherwise normal listeners depended critically on the use of stimuli that were roved over a wide frequency range. The following chapter reports two further experiments that aimed to confirm and extend this conclusion. As before, both the experiments measured listeners' DDLFs and IDLFs. There were two principal aims of Experiment 3. The first aim was to test if the previous results could be replicated whilst controlling for the potential confounds highlighted in Section 3.4. A new sample of direction-impaired listeners completed runs of different conditions in random order, and most of the conditions included random changes in level from tone to tone. The second aim was to test whether the impairment in pitch-direction identification depends critically on the use of a very wide roving range (3102 cents, as in Experiment 1), or whether significant impairments can also be observed with smaller roving ranges.

In a series of experiments, Demany, Carlyon, and Semal (2009) measured DDLFs and IDLFs for discrete and continuous frequency changes. They found that listeners had smaller DLFs in both tasks when tones within a pair were connected by a frequency glide than when they were separated by a silent ISI (cf. Sek & Moore, 1999; Lyzenga, Carlyon, & Moore, 2004). Although the data reported in that study only included listeners who did not experience difficulty identifying pitch-change direction, the authors mentioned in a footnote that four additional listeners had larger IDLFs than DDLFs in both the discrete and continuous conditions. This observation suggests that the impairment in pitch-direction identification is not specific to the use of temporally discrete tones. However, the experiments in that study always involved wide frequency roving. Experiment 4 assessed whether the effects of frequency roving on DLFs and I/D ratios observed in the previous experiments were specific to the particular stimuli employed, or generalised to a second stimulus arrangement in which the target frequency change was continuous.

4.2 Experiment 3

4.2.1 Method

4.2.1.1 Listeners

The 10 listeners who took part in Experiment 3 formed two groups: 'experts' and 'novices'. The expert group consisted of four listeners (4 males; 3 right-handed; age range 24–61 years). The group included L1 and L2 from the previous experiments, and two new listeners, L17 and L18. All of the experts had taken part in many previous psychoacoustical experiments involving frequency and/or pitch discrimination. The novice group consisted of six new listeners (L19–L24; 6 females; 4 right-handed; age range 18–22 years). The novices were all selected from the cohort who had completed the pre-test in exactly the same manner as the listeners in the previous experiments. None of the novice group had prior experience in such experiments. As before, none of the listeners in either group was a professional or practising musician.

4.2.1.2 Stimuli and procedure

The new listeners first underwent the same preliminary assessments as in Experiments 1 and 2 (questionnaires and audiometry). For five of the new listeners, hearing levels did not exceed 20 dB HL at any frequency. For the remaining listener (L19), hearing level at 250 Hz was 30 dB HL in one ear. For all listeners, worst ear averages over all frequencies were smaller than 20 dB HL.

As in Experiments 1 and 2, in the main part of the Experiment 3 listeners heard two pairs of tones on every trial. The frequency of the first tone in each pair was either fixed or randomly selected from a rectangular probability distribution. There were four levels of frequency-roving range: a *wide* range of 3102 cents (400–2400.1 Hz, as in Experiment 1), a *medium* range of 310 cents (895.9–1071.6 Hz), a *narrow* range of 31 cents (988.9–971.4 Hz), and *none*, corresponding to the absence of frequency roving (as in Experiment 2). For each roving range, the level of each tone in every trial was randomly set within a 7-dB range (±3.5 dB around 60 dB SPL). To assess if these level differences influenced DLFs, two more stimulus arrangements without frequency roving were included: in the first, all the tones were presented at 60 dB SPL (as in Experiments 1 and 2), and in the second, tone level was roved between pairs but was fixed for the two tones within each pair. None of the conditions contained maskers.

In each of the six stimulus conditions, 20 DDLFs and 20 IDLFs were obtained from each listener, resulting in 12 conditions and 240 threshold measurements per listener in total. The DLFs were measured over 11 experimental sessions, each taking approximately 1 h, on different days. The initial session involved gathering the questionnaire and audiometric data, followed by a short practice block. The remaining sessions contained two experimental blocks, and each block contained one run of each condition in a random order.

The adaptive procedure in Experiment 3 was similar to the one used previously, but included the following modifications. If listeners made an error within the first three trials, two additional first-phase reversals were added to the run; in such cases, the measurement phase started on the seventh rather than the fifth reversal. The maximum allowed ΔF was 2400 cents rather than 600 cents. If listeners did not make an error within the first three trials, each adaptive run lasted for 12 rather than 14 reversals. Again, the DLF was defined as the geometric mean of all values visited during the measurement phase.

4.2.2 Results

4.2.2.1 Effects of level roving

DLFs in the conditions in which level was roved and test frequency was fixed are shown in Figure 4.1. The left- and the right-hand panels show the data in the expert and novice groups, respectively. These data were analysed using two repeated-measures analyses of variance (ANOVAs) with task (detection, identification) and level roving (none, between pairs, between and within pairs) as factors, and the log-transformed DLFs measured in each group as the dependent variable. In these and the following ANOVAs reported in throughout thesis, Greenhouse-Geisser corrections have been applied where the assumption of sphericity (determined with Mauchly's test) was violated. Neither ANOVA revealed a significant main effect of task [experts: $F(1,3) = 9.86,\ p = .05,\ \eta^2 = .77$; novices: $F(1,5) = 5.85,\ p = 0.06,\ \eta^2 = .54$], although in the experts the trend for IDLFs to be smaller than DDLFs was close to being significant. Level roving also had no significant effect in the expert group [$F(2,6) = 1.72,\ p = .26,\ \eta^2 = .37$]. In the novice group, a main effect of level roving was observed [$F(2,10) = 12.29,\ p < .01,\ \eta^2 = .71$]. Planned comparisons revealed that this effect was driven by a difference between the two forms of roving: roving level between and within pairs resulted in significantly larger DLFs than roving level between pairs only [$F(1,5) = 14.60,\ p < .05,\ r = .86$]. No significant interaction between task and roving condition was found in either group [experts: $F(2,6) = 0.51,\ p = .63,\ \eta^2 = .15$; novices: $F(2,10) = 3.62,\ p = .07,\ \eta^2 = .42$].

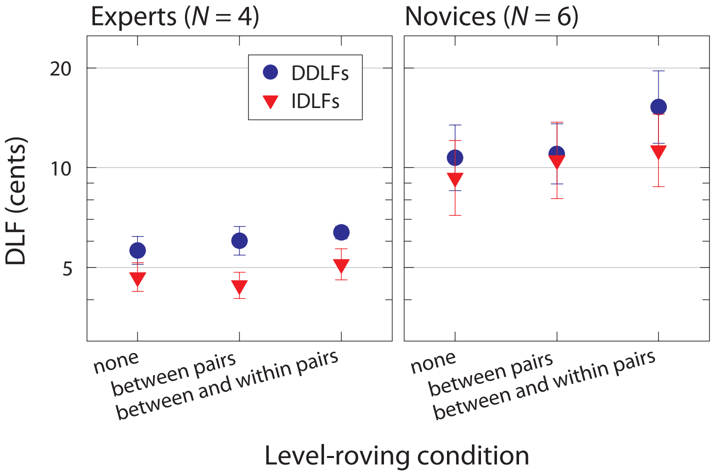


Figure 4.1. Results of Experiment 3 for expert (left-hand panel) and novice (right-hand panel) listener groups in conditions with level roving and without frequency roving. The ordinate of each symbol represents that group's geometric mean DLF in that condition, on logarithmic axes and with bars representing 1 geometric standard error.

4.2.2.2 Effects of frequency roving

Figure 4.2 shows DLFs and I/D ratios in the conditions in which the frequency of the tones was roved and in the corresponding no-rove control. All these conditions involved level roving both within and between pairs in a trial. These data were analysed using two repeated-measures ANOVAs with task (detection, identification) and frequency-roving range (none, narrow, medium, wide) as factors, and the log-transformed DLFs from each group as dependent variable. In the expert group, the results showed significant main effects of roving range [$F(3,9) = 44.47,\ p < .001,\ \eta^2 = .94$] and task [$F(1,3) = 12.84,\ p < .05,\ \eta^2 = .81$], but no significant interaction between the factors [$F(3,9) = 1.15,\ p = 0.38,\ \eta^2 = 0.28$]. This corresponds to the observation that DLFs in both tasks increased with roving range, and that IDLFs were generally smaller than DDLFs. Planned comparisons revealed that the narrow roving range yielded larger DLFs than the no-roving condition [$F(1,3) = 92.07,\ p < .01,\ r = .98$], and that the medium roving range yielded larger DLFs than the narrow roving range [$F(1,3) = 44.71,\ p < .01,\ r = .97$]. DLFs did not differ significantly between medium and wide roving.

There was a significant main effect of frequency-roving range on DLFs in the novice listeners [$F(1.20, 6.00) = 34.11,\ p < .05,\ \eta^2 = .87$]. Planned comparisons revealed that each increment in roving range yielded significantly larger DLFs [$F(1,5) \geq 9.43,\ p < .05,\ r \geq .81$]. Although there was no significant main effect of task [$F(1,5) = 3.45,\ p = .12,\ \eta^2 = .41$], there was a significant interaction between the factors [$F(3,15) = 16.13,\ p < .001,\ \eta^2 = .76$]. The result of this interaction is seen more clearly in the I/D ratios plotted in the lower-right panel of Figure 4.2: the novice listeners became poorer at identification than at detection—and therefore more direction-impaired—as the frequency-roving range increased. I/D ratios generally only exceeded 1 in the medium and wide roving ranges.

4.2.3 Discussion

The first aim of Experiment 3 was to replicate the results of Experiments 1 and 2 whilst controlling for the potential confounds highlighted in Section 3.4, namely order effects and level cues. The experiment revealed relative impairments in pitch-direction identification in a new group of listeners who completed conditions in a random order. This result suggests that it is unlikely that the reason listeners were less impaired in Experiment 2 was that it was performed after Experiment 1.

The effects of level roving on frequency discrimination have been explored in previous studies, most of which found that roving did not increase DLFs significantly (Dai et al., 1995; Henning, 1966; Moore & Glasberg, 1989). The non-significant (experts) and modest (novices) effects of level roving found in Experiment 3 are perhaps not surprising given that the level-roving range used (7-dB) was close to the one used by Moore and Glasberg (see also Henning, 1966). Importantly, although the novice listeners had slightly elevated DLFs when level was roved between and

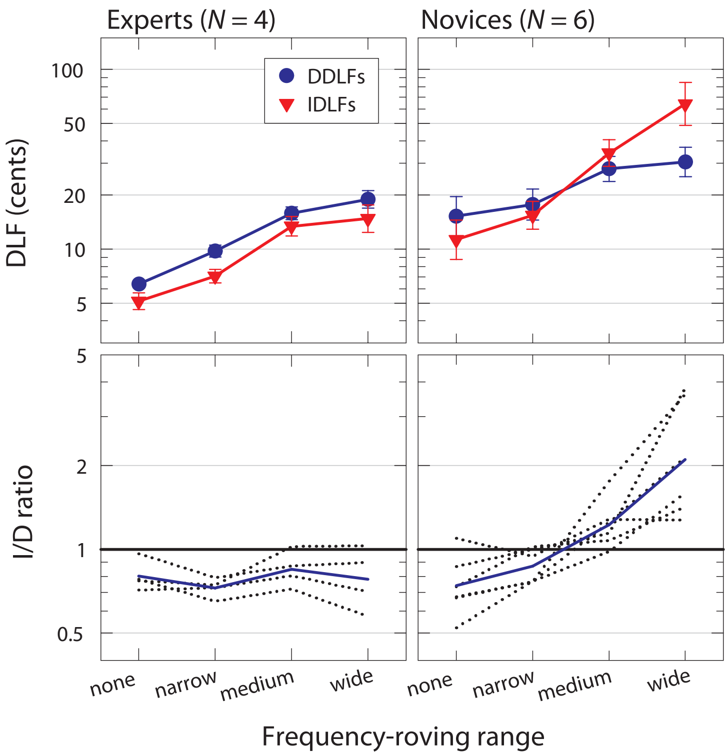


Figure 4.2. Upper panels: same as Figure 4.1, for the no-rove condition and the three conditions in which test frequency was roved, in each case including level roving both within and between tone pairs in a trial. Lower panels: each dotted line represents a single listener's I/D ratios, and the thicker lines represent the geometric group mean I/D ratios, on logarithmic axes. The horizontal reference lines represent an I/D ratio of 1.

within pairs, their IDLFs were not larger than their DDLFs in these conditions, which did not involve any frequency roving. This finding is consistent with the results of Experiment 2, and suggests strongly that the differences between the results of Experiments 1 and 2 are not likely to be due to the use of cues other than pitch, which were available in Experiment 2 but not in Experiment 1.

The second aim of Experiment 3 was to determine whether the I/D ratios measured in direction-impaired listeners depends on relatively *wide* frequency roving. The results revealed a monotonic increase in the average I/D ratio with increasing roving range in the novice group (L19–L24). In other words, as the roving range progressively widened, the ability of direction-impaired listeners to identify pitch-change direction became poorer. I/D ratios generally only exceeded 1 in the medium or wide roving ranges. It is important to point out that, overall, the impairments in pitch-direction identification observed in L19–L24 in this experiment turned out to be less severe than those observed in the previously pre-selected group (L4–16). In the wide frequency-roving range conditions in Experiment 3, L19–L24 had a group geometric mean I/D ratio of 2.10, compared to 3.48 in L6–14 in Experiment 1. Moreover, unlike L10, L12, and L16 in Experiment 2, none of the listeners in Experiment 3 had elevated IDLFs in conditions without frequency roving. Because Experiment 3 was considerably longer than Experiments 1 and 2 combined (11 versus 3 h), the smaller I/D ratios could have come about because of more task practice; Semal and Demany (2006) noted that the pitch-direction impairment improved over time in one of their three direction-impaired listeners. Although the severity of selective impairments in pitch-direction identification—and the dependence of these impairments on frequency-roving range—varied appreciably across listeners in Experiments 1–3, overall the results show that such impairments are far more likely to be observed under conditions in which the standard frequency of the stimuli is roved over a wide range than under conditions in which this frequency is fixed.

4.3 Experiment 4

4.3.1 Method

All the listeners from Experiment 3 were retested after the final session of the previous experiment. The stimuli in Experiment 4 were similar to those used above, except that the pair of same-frequency tones was replaced by a constant-frequency tone, and the pair of different-frequency tones was replaced by two 250-ms steady-state plateaux connected by a frequency sweep (Figure 4.3). The sweep was a half-cycle sinusoidal frequency-modulation function with no change in amplitude and with phase continuity. The two tones had durations of 750 ms and were separated by a 700-ms silent ISI. Thus, instead of hearing four discrete tones on every trial, listeners heard two tones, with one containing a continuous frequency sweep. The sweep could occur in either the first or the second tone, and could be upward or downward. The level of each tone was independently varied over a 7-dB range centred on 60 dB SPL[1]. DDLFs and IDLFs were measured using two stimulus conditions: the starting frequency of the tones was fixed at 979.8 Hz for a run of trials, or roved independently from 400–2400.1 Hz. Six DLFs were measured per condition in each listener.

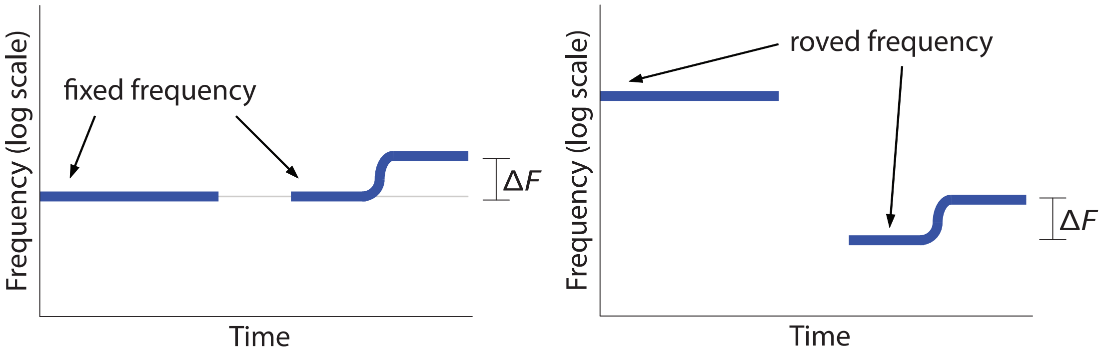


Figure 4.3. Illustrations of example trials in Experiment 4. Blue lines represent sinusoids. The left-hand panel shows a trial in which the starting frequency of the tones was fixed (grey line representing 979.8 Hz), and the right-hand panel shows a trial in which the starting frequency of the tones was roved. In both examples, the correct responses would be 'second' (for detection) or 'up' (for identification).

4.3.2 Results and discussion

The group geometric mean DDLFs and IDLFs (and the corresponding I/D ratios) measured in Experiment 4 are shown in Figure 4.4. A repeated-measures ANOVA with the experts' DLFs as the dependent variable, and with frequency roving (roving, no roving) and task (detection, identification) as factors, revealed a main effect of frequency roving [$F(1,3) = 43.72$, $p < .05$, $\eta^2 = .94$]. As in Experiment 3, the effect of task was not significant [$F(1,3) = 0.76$, $p = .45$, $\eta^2 = .20$], but IDLFs were on average lower than DDLFs. There was no significant interaction between the factors [$F(1,3) = 0.36$, $p = 0.86$, $\eta^2 = .12$]. The results of the ANOVA are consistent with the results from Demany, Carlyon, and Semal (2009), and indicate that in the experts, both detecting and identifying the direction of continuous frequency sweeps are affected by frequency roving similarly.

A corresponding ANOVA for the novice group revealed significant main effects of frequency roving [$F(1,5) = 179.46$, $p < .001$, $\eta^2 = .97$] and task [$F(1,5) = 10.14$, $p < .05$, $\eta^2 = .67$]. As in Experiment 3, there was also a significant interaction between the factors [$F(1,5) = 9.22$, $p < .05$, $\eta^2 = .65$]. These results suggest that roving mediated

---

[1] The level roving used in this experiment could not control for potential loudness or excitation-pattern cues—as in Experiment 3—because the level was always constant within each tone. However, the results of Experiment 3 suggest that it was very unlikely that listeners actually made use of such cues.

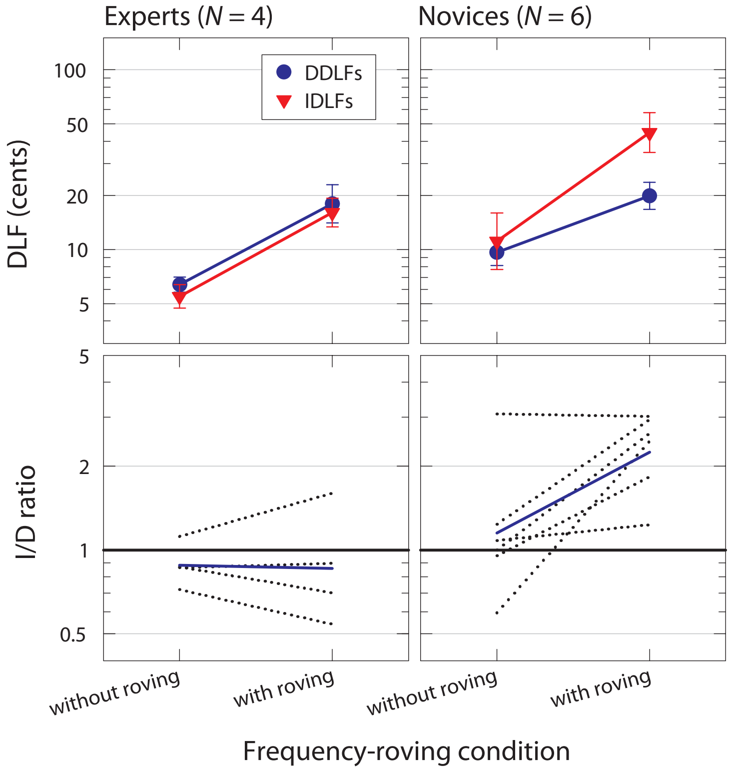


Figure 4.4. Same as Figure 4.2, for the conditions in Experiment 4.

the relative difficulty in pitch-direction identification using continuous stimuli. In this experiment, the geometric mean I/D ratio observed in the novice group with roved stimuli was 2.25, which is similar to the ratio observed in the same listeners in the maximum-roving condition in Experiment 3 (2.10).

4.4 Chapter discussion

Experiment 3 confirmed and extended the main finding from Experiments 1 and 2. The relative insensitivity to pitch-change direction observed in some normal listeners is related critically to the frequency-roving range of the test tones, with those listeners' pitch-direction identification abilities being poorer, on average, with wider roving. Experiment 4 required listeners to discriminate pitch changes using a different stimulus with a continuous rather than a discrete frequency change, and the results suggested that frequency roving similarly affected their DLFs. Section 3.4 outlined two possible reasons why DLFs might increase with increasing roving range. One possibility was that listeners could have switched from a context-coding strategy, which allowed them to compare the tones to a relatively precise and stable referent, to a sensory-trace strategy, in which echoic traces of the tones within each trial were compared to each other (Durlach & Braida, 1969). The other possibility was that stimulus uncertainty elevated DLFs. Although both are plausible reasons for a *general* increase in DLFs with frequency roving, neither provides a convincing explanation for the observed interactions between task and roving range. The following section proposes three possible explanations for the source of the interactions.

4.4.1 Frequency-shift detectors

The first explanation is based on the idea of frequency-shift detectors (FSDs). As outlined in Section 2.2.4.5, evidence for FSDs comes from two studies (Demany & Ramos, 2005; Demany, Pressnitzer, & Semal, 2009) demonstrating that listeners are able to identify the direction of a frequency change between pure tones even when the pitch of the first tone was made imperceptible by burying it within a random, inharmonic chord. In their discussion, Semal and Demany (2006) proposed that there are two mechanisms of pitch discrimination: an efficient mechanism that is sensitive to the direction of small changes and is based on the output of FSDs, and a less-efficient mechanism that is not sensitive to direction and is not based on FSDs. The authors suggested that FSDs do not respond to small shifts in the brains of direction-impaired listeners, and therefore that these listeners are relying on the less-efficient mechanism when discriminating small frequency changes. However, the result that insensitivity to pitch-change direction is greatly reduced or eliminated when wide frequency roving is not used suggests either that this explanation is incorrect, or that when the stimuli are not roved widely in frequency, a listener's ability to identify pitch-change direction no longer relies on FSDs. As already mentioned, when the standard frequency of the tones is fixed, listeners may be able to use context coding instead of sensory-trace coding (Durlach & Braida, 1969). If FSDs are necessary for successful task performance only when listeners use sensory-trace coding, less or no impairment would be observed with fixed-frequency stimuli.

The FSD-based explanation makes several important assumptions that need to be tested. The first is that the hypothesised FSDs actually exist. Physiological studies have found that the auditory cortex in nonhuman animals contains neurons that selectively respond to either rising or falling frequency glides (e.g., Whitfield & Evans, 1965; Zhang, Tan, Schreiner, & Merzenich, 2003), and that the firing rates of some neurons in response to a target tone are affected by the frequencies of preceding tones, even when the tones are separated by up to 500-ms silent ISIs (e.g., Weinberger & McKenna, 1988; McKenna, Weinberger, & Diamond, 1989). According to the FSD hypothesis, these neurons may respond in a similar manner when the preceding tones are rendered imperceptible, either by burying them within a chord or by some other method. To my knowledge, this prediction has not yet been tested.

Of course, FSDs might not be reducible to single neurons and might consist instead of neuronal assemblies. If this is the case, then evidence could come from human functional imaging techniques. A future experiment could modify Demany and Ramos' (2005) original up/down paradigm for use in such an experiment. For instance, if the isolated comparison tone in every trial were also buried within an inharmonic chord, informational masking (IM) would have an effect on both target tones and the upward or downward change might not be perceived. (Care would have to be taken to ensure that no other upward or downward frequency shifts were created by the new chord components.) Presumably, FSDs would still operate under these conditions because they do not rely on the conscious perception of pitch, thus their activity might still be observed by functional imaging. The activity could be in the form of mismatch negativity (MMN), since the characterisation of FSDs as part of a mechanism for automatically detecting or discriminating auditory change is similar, at least superficially, to the characterisation of the generators of the MMN. To date, only psychoacoustical studies using the paradigm described by Demany and Ramos have supported the idea of FSDs.

The second assumption is that a change in coding strategy occurs between roved and non-roved frequency discrimination. One way to test this assumption would be to use a procedure similar to the one described by Clément et al. (1999), which measures DLFs or accuracy for pitch discrimination as a function of the silent ISI separating the tones on each trial. If listeners are able to use a referent stored in long-term memory when the stimuli are not roved in frequency, then their DLFs or performance in conditions with a fixed standard should be less affected by increases in ISI duration than their DLFs or performance in conditions with a roved standard. Note that the FSD-based explanation does not necessarily predict a difference between direction-impaired and unimpaired individuals in this kind of experiment: direction-impaired listeners may be just as able to switch from sensory-trace to context coding. The final assumption—perhaps the most difficult to test—is that only sensory-trace coding relies on the use of FSDs.

4.4.2 Learning using feedback

A second possible explanation for the findings is that, when performing the identification task with stimuli that were not roved widely in frequency, direction-impaired listeners were able to make more effective use of the detailed feedback provided after every trial[2]. According to this explanation, direction-impaired listeners could not perceive a small pitch change as upward or downward per se in those conditions, but could detect a difference between tone pairs containing upward and downward changes because they perceived the second tone as different in the two cases. Owing to the provision of feedback after each trial, listeners could have learned to correctly label the two different-sounding cases as either 'up' or 'down', resulting in smaller DLFs. This strategy requires listeners to be able to *generalise* across the instances of

---

[2] Credit is given to Catherine Semal—who acted as a reviewer on the submitted paper detailing Experiments 1–4—for originally suggesting the learning hypothesis. I also thank her for a preliminary suggestion of how to test the hypothesis, which influenced the design of the Experiment 5.

either category, so performance would be worse if the tones forming the up and down cases varied widely across trials. This could explain why IDLFs were considerably elevated relative to DDLFs when the frequency-roving range of the stimuli was wide.

The above explanation is consistent with the results from the studies of perceptual learning in frequency discrimination (Section 2.2.5.3). Those studies suggest that learning is strongly frequency-specific: when listeners hear fixed-frequency tones, learning occurs at that frequency and does not generalise across a wide frequency band. Conversely, this observation predicts that less learning should occur if listeners hear roved-frequency tones, because there is necessarily less exposure to any particular frequency. Support for the prediction comes from the experiment by Amitay et al. (2005), in which less learning occurred on average in the two groups of listeners trained with roved-frequency tones than in the group trained with fixed-frequency tones.

If the learning hypothesis is correct, in conditions without wide frequency roving, direction-impaired listeners might show an initial insensitivity to pitch-change direction that decreases over trials. This was investigated by examining how the DLFs measured in the listeners who were direction-impaired in Experiment 1 but not in Experiment 2 (L6, L7, L9 and L11) changed over time. Figure 4.5 shows threshold estimates for those listeners in Experiment 2 (conditions without maskers) as a function of run number, both individually and as an average 'de-meaned' DLF (see caption). Although L9 appears to have shown some systematic improvement in the identification task over the course of the experiment, the data overall do not strongly support the idea that the listeners were initially insensitive to pitch-change direction and then gradually learned to use feedback for more successful performance. Although this observation does not support the learning hypothesis, it does not rule it out completely because the listeners may have learned to use feedback in Experiment 2 within the first run of trials. Experiment 5 in the following chapter provided a stronger test of the hypothesis by measuring direction-impaired listeners' DLFs at different standard frequencies, both with and without feedback.

4.4.3 Task-irrelevant pitch changes

A third explanation for the results is that direction-impaired listeners were more easily confused or distracted by the irrelevant—and sometimes much larger—pitch changes that occurred between the last and the first tones from adjacent trials, or between pairs within a trial. In order to test this hypothesis, a trial-by-trial re-analysis of the data from Experiment 1 was performed. For each listener and each task (detection and identification), trials completed during the course of the experiment were sorted into two bins depending on whether or not the direction of the frequency change between the second and third tones in the trial (the between-pair frequency change) was in the same direction as the target (within-pair) frequency change on that same trial. Proportion correct (*PC*) was recorded for each bin. The results are shown in the top panel of Figure 4.6. In the detection task, the congruence of between- and within-pair changes did not have a significant effect on performance [direction-impaired listeners: $t(8) = -1.53$, $p = 0.17$, $r = .48$; unimpaired listeners: $t(4) = -1.33$, $p = .26$, $r = .55$]. However, in the identification task, direction-impaired listeners performed more poorly on trials in which the direction of the between-pair change was opposite to that of the within-pair change [$t(8) = 2.87$, $p < .05$, $r = .71$]. This effect was not found for the unimpaired listeners in that experiment [$t(5) = 0.59$, $p = 0.59$, $r = .28$]. The same pattern was observed when this analysis was performed on the data from the wide frequency-roving conditions in Experiment 3—frequency-direction incongruence also resulted in poorer performance in identification trials for the novice listener group [Figure 4.6, bottom panel; $t(5) = 5.73$, $p < .05$, $r = .93$]. This exploratory analysis provides support for the interference hypothesis, which is also examined in the next chapter.

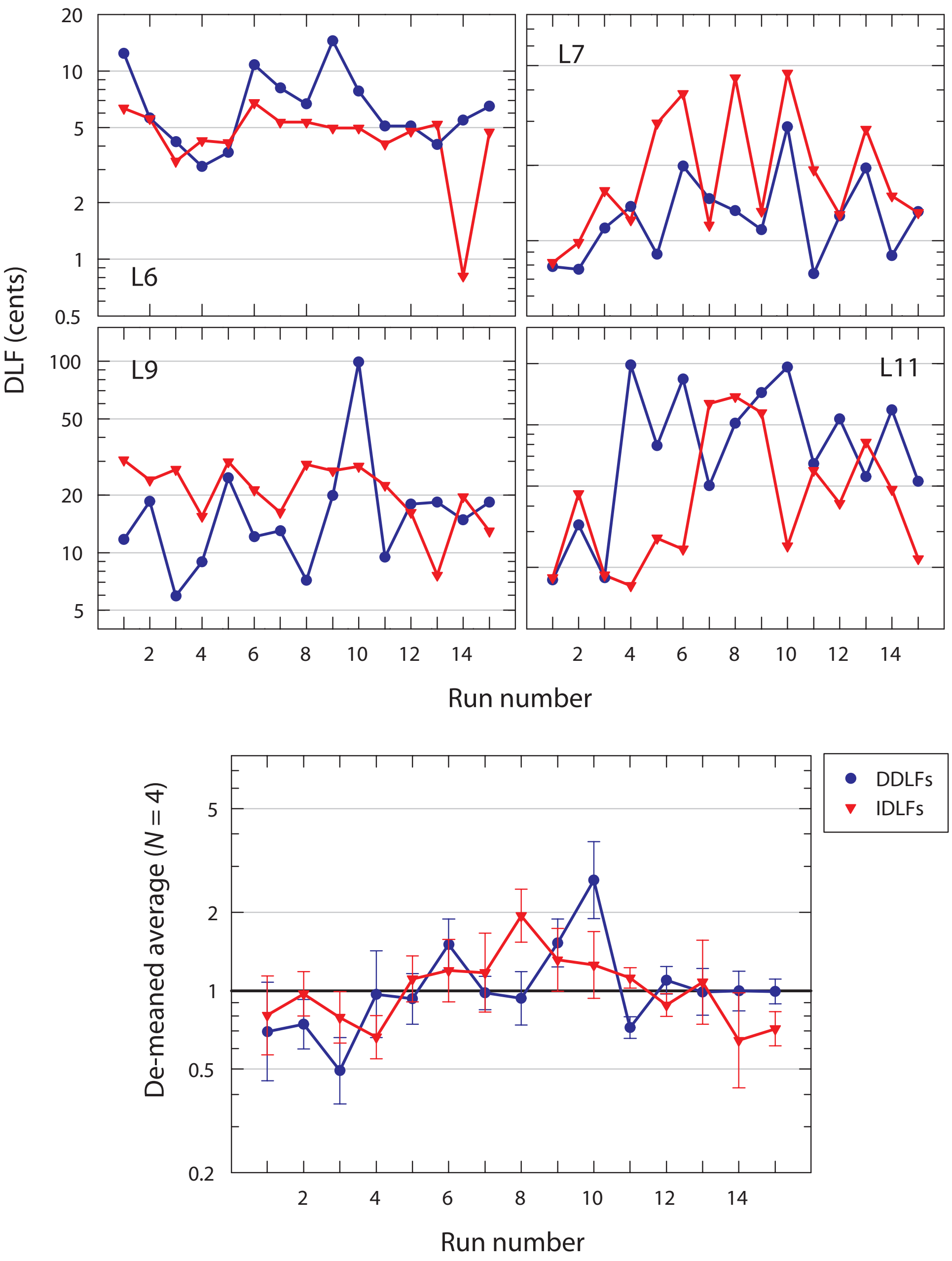


Figure 4.5. Learning effects in Experiment 2 (conditions without maskers). The top four panels show the DLFs measured in L6, L7, L9, and L11 as a function of run number. Each listeners' final DDLF and IDLF were computed by taking the geometric mean of these thresholds. To examine the overall learning trend in the four listeners, the DLF of each individual and each run was 'de-meaned' by dividing it by the corresponding final DLF. The bottom panel shows the geometric group mean of this score as a function of run number, with error bars representing 1 geometric standard error.

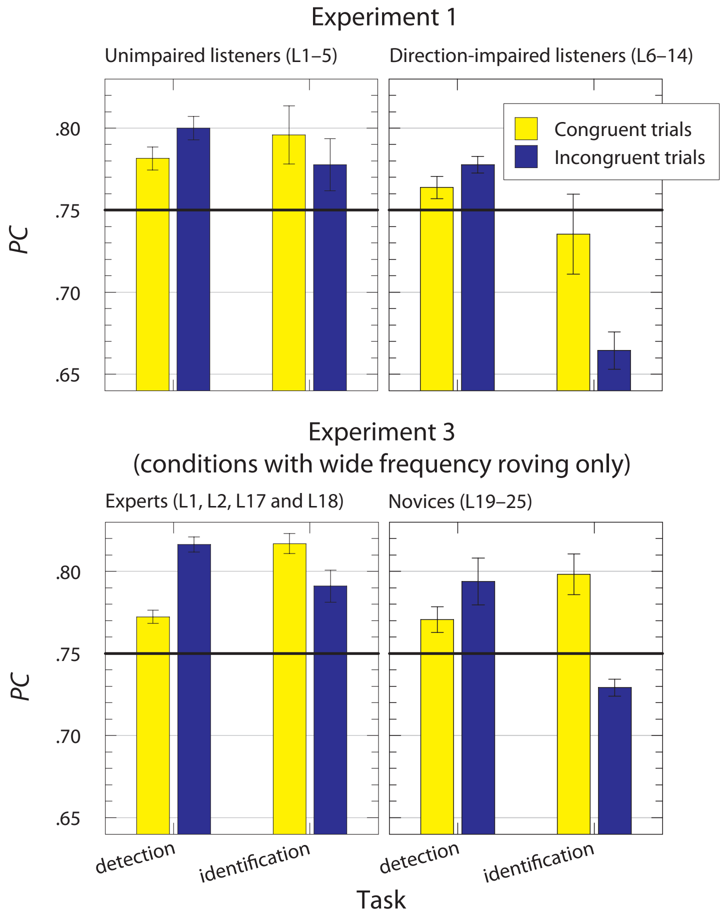


Figure 4.6. Results of the trial-by-trial analysis performed on the data from Experiment 1 (top panel) and the wide-roving conditions in Experiment 3 (bottom panel). Bars represent the group mean *PC* in the unimpaired listeners (left-hand panel) and in the direction-impaired listeners (right-hand panel). *PC* is shown separately for congruent and incongruent trial types, and for trials in the detection and identification tasks. Error bars represent 1 standard error of the mean. Horizontal reference lines represent 75% correct, which was measured by the adaptive routine in both experiment experiments. Data from L15 and L16 are excluded from the top panel because of those listeners' unusual performance.

Chapter 5

Experiments 5 and 6: The influences of feedback and irrelevant pitch changes on pure-tone pitch discrimination

5.1 Experiment 5

5.1.1 Rationale

In Section 4.4, three possible explanations were put forward for the critical role of frequency roving in some individuals' insensitivity to pitch-change direction. Experiment 5 aimed to test the second of these competing explanations, hereafter referred to as the 'learning hypothesis' (Section 4.4.2). On each trial in the experiment, listeners heard two pure tones differing in frequency, and the frequency of the first tone (the standard) was fixed within most of the adaptive runs. After four successive adaptive runs, the standard was shifted to a different frequency. According to the learning hypothesis, this manipulation would require direction-impaired listeners to re-learn to label upward and downward changes at the new frequency, resulting in larger DLFs. In some conditions in the experiment, the trials after a shift in standard frequency were not followed by feedback. Under these circumstances, the learning hypothesis predicts that the DLFs measured in direction-impaired listeners would be even larger because they should not be able to label upward or downward changes successfully.

5.1.2 Method

5.1.2.1 Listeners

Six listeners took part in Experiment 5. All of the listeners had taken part in at least two of the previous experiments reported in the thesis: L1, L2, and L3 (two male; one left-handed; aged 25–26 years) were unimpaired expert listeners, and L8, L9, and L12 (all female; two right-handed; age range 18–20 years) were direction-impaired listeners. Thus, it was expected prior to their taking part in the present experiment that the latter three listeners would experience difficulty identifying the direction of small frequency changes using roved-frequency tones, and that this difficulty would be less pronounced for fixed-frequency tones.

5.1.2.2 Stimuli and procedure

In each trial in Experiment 5, listeners heard a pair of 250-ms pure tones. The levels of the tones were randomly set within a 7-dB range centred on 60 dB SPL. The two tones in a trial were separated by an ISI of 250 ms, and differed in frequency by ΔF, expressed in cents. The direction of the frequency change was equiprobably upward or downward, and listeners were instructed to name the direction of this change on each trial.

Unlike the previous experiments—in which runs of each condition were completed in a counterbalanced or random order—in Experiment 5 all the listeners completed the runs in the same order, summarised in Table 5.1. There were six 'phases' in the experiment, each containing four runs, and within each run the value of ΔF was tracked using an adaptive procedure that estimates DLFs corresponding to 75% correct on the psychometric function (Section 4.2.1.2). In the first five phases (runs 1–20), the frequency of the first tone was fixed at a specific value—either 0, 1551, or 3102 cents above 400 Hz (400, 979.8, or 2400.1 Hz)—that switched every phase. In the final phase (runs 21–24), the frequency of the first tone was roved over the 3102-cents range used in Experiment 1 (400–2400.1 Hz). In phase 1 (runs 1–4) and phases 4–6 (runs 13–24), as in all the previous experiments, responses were followed by feedback and a 600-ms pause before the start of the next trial. In phases 2 and 3 (runs 13–20), visual feedback was omitted, and the next trial started 600 ms after the response. Twenty-four DLFs were measured per listener in total. Testing was carried out individually in a sound-attenuating chamber, and the listeners completed the experiment in a single session lasting approximately 40 min.

5.1.3 Results and discussion

The results of Experiment 5 are shown in Figure 5.1, in which listeners' geometric mean DLFs in each of the six phases are plotted as ordinates. The data were

Table 5.1. Details of the runs of trials in Experiment 5.

		Standard frequency		
Phase	Runs	cents above 400 Hz	Hz	Feedback after trials?
1	1–4	1551	979.8	yes
2	5–8	0	400.	no
3	9–12	3102	2400.1	no
4	13–16	0	400.	yes
5	17–20	3102	2400.1	yes
6	21–24	roved	roved	yes

subjected to a mixed ANOVA with phase (1–6) as a within-subjects factor, group (unimpaired, direction-impaired) as a between-subjects factor, and listeners' log-transformed geometric mean DLFs as the dependent variable. The ANOVA revealed significant a main effect of phase [$F(5,20) = 4.29$, $p < .01$, $\eta^2 = .51$], no main effect of group [$F(1,4) = 5.39$, $p = .08$, $\eta^2 = .57$], and a significant interaction between phase and group [$F(5,20) = 3.78$, $p < .05$, $\eta^2 = .49$]. Inspection of the figure suggests that the interaction was driven by the direction-impaired listeners in particular having considerably larger DLFs during phase 6 than during phases 1–5. Planned comparisons were used to investigate whether the DLFs measured in the three direction-impaired listeners during phase 1 were different to those measured in phase 2, whether those in phase 2 were different to those in phase 3, and so on. None of these comparisons revealed statistically significant differences [$F(1,2) \leq 5.41$, $p \geq .15$, $r \leq .85$], except for the comparison between phases 5 and 6 [$F(1,5) = 858.73$, $p < .01$, $r = .999$].

The learning hypothesis predicts that direction-impaired listeners would have to re-learn to label upward and downward pitch changes whenever the standard frequency of the tones is shifted by a sufficient amount, and that re-learning would be prevented if these trials are not followed by feedback. In the present experiment, however, neither switching to a novel standard frequency nor withholding feedback had a reliable effect on the DLFs measured in the group of three direction-impaired listeners. Demany and Semal (2002) trained eight listeners in a frequency-discrimination task using tones whose standard frequency was always 3000 Hz. After approximately 11000 trials, the standard frequency of the tones was switched to either 1200 or 6500 Hz. The authors found that learning did not generalise from 3000 Hz to either novel frequency, which were different from the original standard by 1586.3 and 1338.6 cents, respectively. In the present experiment, the standard was shifted by 1551 cents between phases 1 and 2, and by 3102 cents between phases 2 and 3. It is therefore unlikely that the null result came about because the shifts in standard frequency were not large enough to prevent the generalisation of learning.

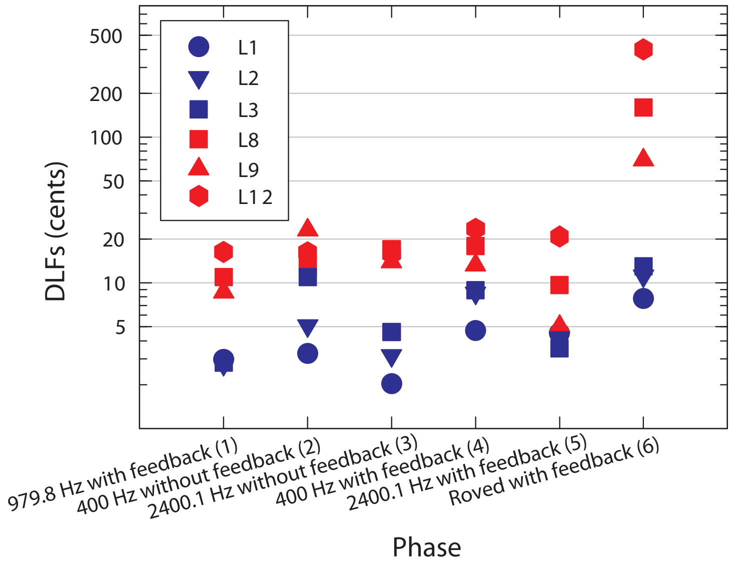


Figure 5.1. Results of Experiment 5. The ordinate of each symbol represents an individual's geometric mean DLF, on logarithmic axes, for the six phases in the experiment. Symbol shapes and colours are consistent with those in Figure 3.4.

The null result is also unlikely to have been caused by a lack of power, since the use of wide frequency roving in the final phase of the experiment again caused a dramatic increase in DLFs. Moreover, the fact that fixed-frequency DLFs remained relatively small even without feedback (phases 2 and 3) suggests that the direction-impaired listeners could genuinely perceive the direction of small frequency changes under these conditions. For these reasons, the learning hypothesis does not provide a plausible explanation for the earlier findings.

Are the results of the present experiment consistent with another explanation for the role of frequency roving posited in Section 4.4.3, namely the 'interference hypothesis'? As discussed earlier, preliminary support for the hypothesis came from trial-by-trial re-analyses of the data from Experiments 1 and 3, which found that the direction-impaired listeners were poorer at the identification task when trials contained incongruent frequency changes. Such incongruent changes could not occur within trials in Experiment 5, but could occur between trials (i.e., the change between the last tone of the previous trial and the first tone of the current trial). For each listener and each phase, trials completed during the course of Experiment 5 were sorted into two bins depending on whether or not the direction of the preceding between-trial change was in the same direction as the within-trial change. The bins were collapsed across phases 1–5 since none of these conditions involved frequency roving. The results (Figure 5.2) show that the direction-impaired listeners were much worse when a trial was preceded by an incongruent pitch change than when a trial was preceded by a congruent change. Importantly, this effect was the strongest in the context of frequency roving (phase 6). Thus, the results of Experiment 5 provide more evidence for the interference hypothesis.

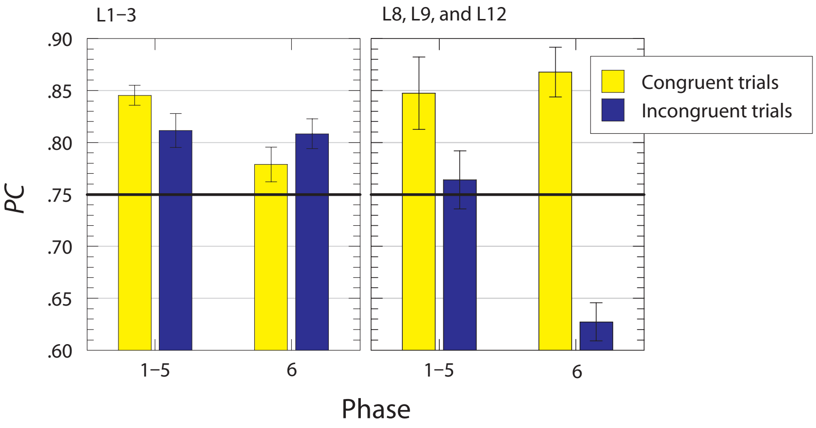


Figure 5.2. Same as Figure 4.6, except for the trial-by-trial analysis performed on the data from Experiment 5 (see text).

5.2 Experiment 6

5.2.1 Rationale

Although the trial-by-trial analyses of Experiments 1, 3, and 5 all support the interference hypothesis, they all lack strong lacked stastical power because they were unable to control for the number of trials in each bin and the value of ΔF on any given trial. Furthermore, although it is plausible that the irrelevant changes led to more errors on incongruent trials, it is also plausible that direction-impaired listeners simply responded with the irrelevant direction whenever they could not perceive the direction of the target change. In other words, the previous trial-by-trial analyses can not be used to establish whether the incongruence effects in the previous experiments were a *cause* or a *consequence* of poor pitch-direction identification.

Experiment 6 aimed to provide a stronger test of the interference hypothesis by measuring DLFs using a new procedure. Listeners heard three tones on each trial, and were instructed to identify the direction of the frequency change that occurred between the second and third tones. The frequency-roving range of the tones was manipulated over runs of trials in a similar way to in Experiment 3. The main experimental manipulation, however, was the frequency of the first tone in each trial relative to the frequency of the second. In half of the conditions, the first tone always had the same frequency as the second tone, and in the other half of the conditions, the first was roved in frequency independently on the second tone. According to the interference hypothesis, direction-impaired listeners should not be able to ignore the additional, irrelevant change occurring under the latter conditions, and thus should have larger DLFs when the first tone in every trial is random in frequency.

The duration of the ISI between the first and second tones on each trial in Experiment 6 was sufficiently long to ensure minimal effects of both forward masking and recognition masking (Section 2.2.5.1). The effect of a preceding tone on pure-tone pitch discrimination has been investigated in several previous studies. In one condition from an experiment by Mondor, Breau, and Milliken (1998), 12 listeners heard two pure tones on each trial. Each tone was equiprobably 555 or 869 Hz, and the silent ISI was 150, 450, or 750 ms. The listeners judged whether the second tone was high (869 Hz) or low (555 Hz). The authors were concerned primarily with reaction times, but accuracy data were also reported. Listeners made slightly more errors when the first tone was different to the second—these trials contained an irrelevant pitch change of approximately 776.2 cents—but only at the shortest ISI; when the ISI was 450 or 750 ms, the listeners actually made fewer errors when the first tone was different. However, it is difficult to draw firm conclusions from the results of that experiment, because the listeners completed only about 24 trials of each type, and the differences in PC probably amounted to one or two trials (see also Mondor & Breau, 1999).

Two more studies have investigated the effects of a preceding tone on pure-tone pitch discrimination. Ruusuvirta (2000) presented 14 listeners with sequences of five 30-ms pure tones and instructed them to judge whether the fifth tone was higher or lower than the fourth. The fourth tone always corresponded to C6 (1046.5 Hz), and the fifth was up or down in frequency by 7, 14, or 21 Hz from C6 (11.6–35.1 cents). In one condition, the first three tones were all also C6, and in others, the tones were always above or always below C6. The results showed that listeners experienced systematic response biases that were essentially the same as the effect of incongruence observed in the present experiments: listeners made more errors when the frequency difference between the third and fourth tones was in the opposite direction to the difference between the fourth and fifth tones, and fewer errors when the two differences were in the same direction. In another experiment by Ruusuvirta, Wikgren, and Astikainen (2008), listeners heard two tones per trial, and the effects of tones from the previous trial were investigated. Although the authors did not interpret their results this way, precisely the same biases were observed. Importantly, neither of these experiments reported differences in overall accuracy. In other words, although their responses were biased toward one direction or the other by preceding irrelevant pitch changes, the listeners were not *less accurate* when an irrelevant change preceded a revelant change than when there was no irrelevant change. Since those experiments did not measure listeners' DLFs, it is not known whether Ruusuvirta and colleagues tested any listeners who had difficulty identifying pitch-change direction.

5.2.2 Method

5.2.2.1 Listeners

Experiment 6 involved three experts (L1–3; one female; one left-handed; age range 25–26 years) and a new group of 11 potentially direction-impaired novices (L25–36;

10 female; one left-handed; age range 18–30 years) selected using the pre-test. None of the novices had prior experience in psychoacoustical experiments, and none was a professional or practising musician.

5.2.2.2 Stimuli and procedure

Prior to taking part Experiment 6, the new listeners underwent several stages of preliminary assessment. Questionnaire and audiometric data were collected as before, which revealed that none of the listeners had hearing levels exceeding 20 dB HL at any of the frequencies tested. After these assessments, the new listeners also took part in a shortened version of Experiment 1. The stimuli and procedures of the shortened version were the same as those described in Sections 3.3.1.2 and 3.3.1.3, except that the conditions containing maskers were omitted, the tones in every trial were roved in level over a 7-dB range centred on 60-dB SPL, and the modifications to the adaptive procedure described in Section 4.2.1.2 were implemented. Twenty DDLFs and 20 IDLFs were measured per listener over two sessions, lasting approximately 1 hr each[1].

In the main part of Experiment 6, listeners heard three tones on each trial. The tones were separated by 250-ms ISIs and were roved in level. In all of the conditions, the frequency of the third tone was always equiprobably upward or downward from the frequency of the second tone by ΔF. The second tone was either fixed in frequency or randomly roved over one of the three frequency-roving ranges used in Experiment 3: *narrow* (988.9–871.4 Hz), *medium* (895.9–1071.6 Hz), or *wide* (400-2400.1 Hz). The frequency of the first tone was either the same as that of the second tone, or randomly and independently drawn from the wide roving range (Figure 5.3). There were eight conditions in the experiment in total. The listeners were instructed to identify the direction of the change between the second and third tones, and to ignore always the first tone. In each of the eight conditions, five DLFs were obtained from each listener, resulting in 40 threshold measurements in total. The adaptive runs were completed in random order over two sessions on different days, each taking around 1 hr.

5.2.3 Results and discussion

5.2.3.1 Novice listeners' DLFs

Consider first the data shown in the top-right-hand panel of Figure 5.4, in which the group geometric mean DLFs measured in the 11 novices are plotted as ordinates. The log-transformed DLFs were subjected to a repeated-measures ANOVA with

---

[1] The additional assessments were made for the purposes of later correlation analyses (see Chapter 7), but also provided the listeners with a few hours of practice prior to taking part in the main experiment.

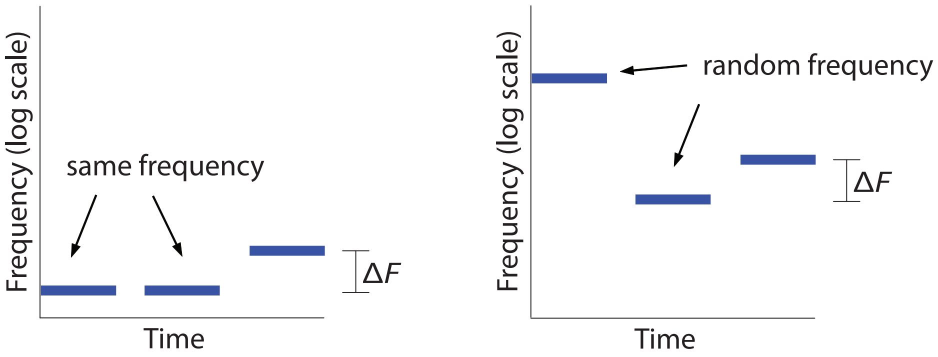


Figure 5.3. Illustrations of example trials in Experiment 6. Blue lines represent sinusoids. The left-hand panel shows a trial in which the first tone is the same frequency as the second. The right-hand panel shows a trial in which the frequency of the first tone is random; this created an irrelevant pitch change within the trial. In both examples, the correct responses would be 'second' (for detection) or 'up' (for identification).

frequency-roving range (none, narrow, medium, wide) and the relationship between the first and second tones in each trial (same, random) as within-subjects factors. The ANOVA revealed both main effects to be significant [roving range: $F(1.49, 14.93) = 63.02$, $p < .001$, $\eta^2 = .86$; relationship between the first and second tones: $F(1, 10) = 195.17$, $p < .001$, $\eta^2 = .95$]. There was no significant interaction between the two factors [$F(1.58, 15.79) = 3.19$, $p = .08$, $\eta^2 = .24$].

The main effect of the relationship between the first and second tones is illustrated in the bottom-left-hand panel of Figure 5.4. The panel shows individual and group mean 'random/same ratios', derived by dividing a listener's DLF measured when the first tone was random by their corresponding DLF measured when the first and second tones had the same frequency. The geometric mean of this ratio across all novice listeners and conditions was 2.63. In other words, when the first tone was random in frequency, the novice listeners had DLFs that were on average over two-and-a-half times larger than they were when the first tone was the same as the second in frequency.

5.2.3.2 Expert listeners' DLFs

The top-left-hand panel of Figure 5.4 shows the group geometric mean DLFs measured in the three experts taking part Experiment 6. The ANOVA for this group revealed that there was a significant main effect of roving range [$F(3, 6) = 7.82$, $p < .05$, $\eta^2 = .80$]. The main effect of the first tone failed to reach the threshold for statistical significance [$F(1, 2) = 12.83$, $p = .07$, $\eta^2 = .87$], but there was a significant interaction between the two factors [$F(3, 6) = 12.38$, $p < .01$, $\eta^2 = .86$].

As in Experiment 3, increasing the frequency-roving range within a run of trials resulted in a moderate increase in the expert listeners' DLFs, but this increase was smaller than in the novices' DLFs. The individual and group mean random/same

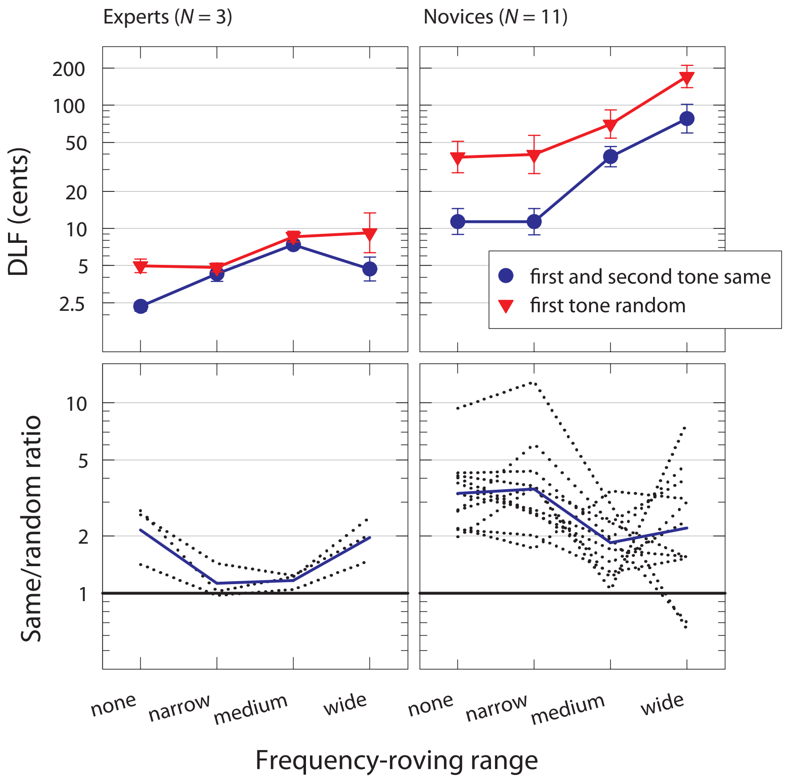


Figure 5.4. Top panels: Results of Experiment 6 for expert (left-hand panel) and novice (right-hand panel) listeners. The ordinate of each symbol represents that group's geometric mean DLF for that condition, on logarithmic axes and with bars representing 1 geometric standard error. Bottom panels: each dotted line represents a single listener's random/same ratios (see text), and the thicker lines represent the geometric group mean ratios, on logarithmic axes. The horizontal reference lines represent a random/same ratio of 1.

ratios are shown in the bottom-left-hand panel of Figure 5.4. Although the main effect was not significant, the mean ratios shown in the panel are all positive, indicating that DLFs were larger when the first tone in each trial was random in frequency. For the narrow- and medium-roving ranges, the ratios are close to 1 (i.e., little or no effect of the first tone being random or the same as the second tone), but in the fixed-frequency and widely-roved conditions, the ratios are larger. Why this pattern of data appears in the experts is not clear, and may be the result of chance variation. Importantly, the geometric mean ratio across all three experts and conditions was 1.53, which is smaller than the value observed in the novices. This result suggests that the experts were less affected than the novices by the introduction of an irrelevant pitch change to the stimulus ensemble.

5.2.3.3 Effects of incongruence

According to the interference hypothesis, a preceding irrelevant pitch change should be the most deleterious to the performance of direction-impaired listeners when it is incongruent with the following relevant change. To test whether this prediction was true in Experient 6, a trial-by-trial re-analysis of the data was performed. For each listener and each condition in which the first tone had a random frequency, trials were sorted into two bins depending on whether or not the direction of the irrelevant change (between the first and second tones) was in the same direction as the relevant change (between the second and third tones). The results (Figure 5.5) show that the novice listeners were more greatly affected by incongruence than the experts, consistent with the interference hypothesis.

5.2.3.4 Relationship between random/same ratios and the effect of roving

If insensitivity to pitch-change direction is a consequence of greater susceptibility to interference from irrelevant pitch changes, then the effects of the two experimental manipulations used in Experiment 6 should be related. To determine the extent to which listeners were affected by roving, their geometric mean DLF measured with widely roved tones (condition wherein the first and second tones had the same frequency) was divided by their corresponding geometric mean DLF measured with fixed-frequency tones. To determine the extent to which listeners were affected by irrelevant pitch changes, their geometric mean random/same ratios across the narrow- and medium-roving conditions were calculated. When both scores were log-transformed, there was a strong positive correlation between them ($N = 14$, $r = .74$, $p < .01$; untransformed scores $r = .74$, $p < .01$; see Figure 5.6). This finding suggests that both effects could have been caused by greater susceptibility to interference in listeners who experience difficulty identifying pitch-change direction.

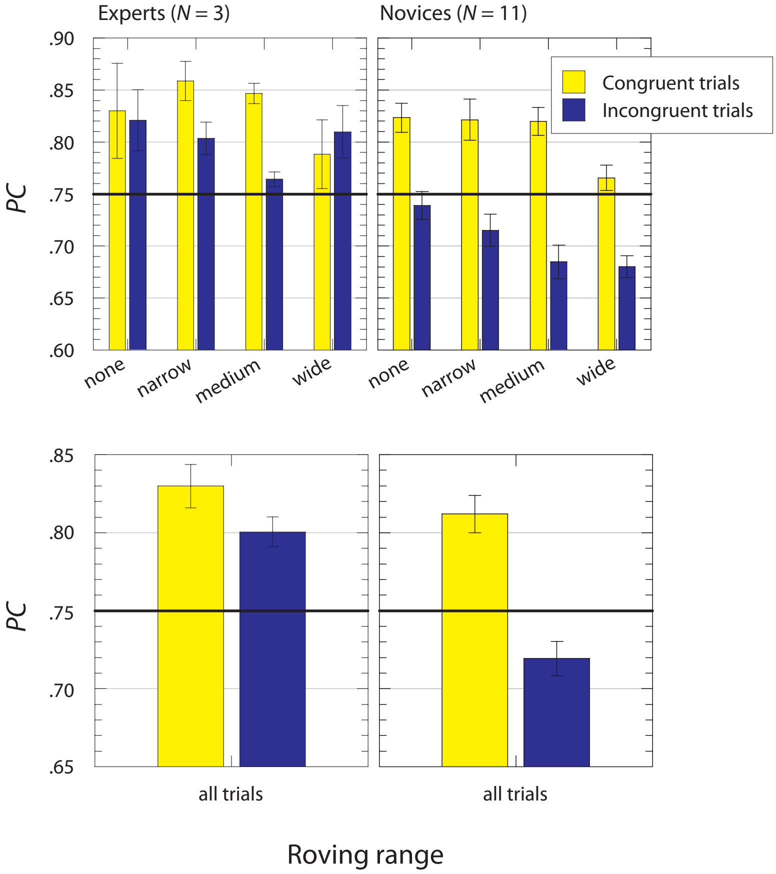


Figure 5.5. Same as Figures 4.6 and 5.2, except for the trial-by-trial analysis performed on the data from Experiment 6 (see text).

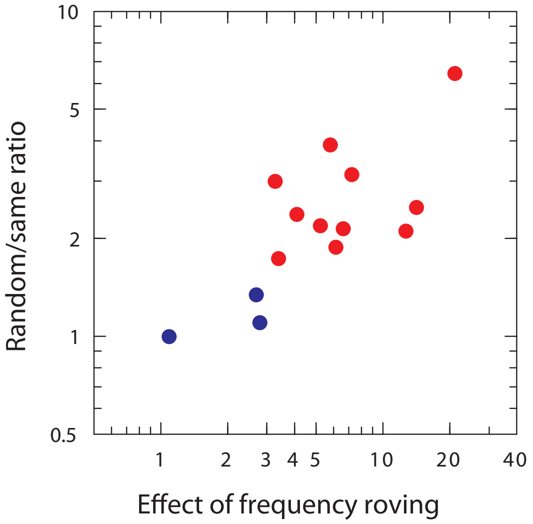


Figure 5.6. The relationship between listeners' random/same ratio and the effect of roving on DLFs (see text) in Experiment 6, on logarithmic axes. The blue symbols represent the three experts and the red symbols represent the 11 novices.

5.2.3.5 Supplementary experiment: Cuing versus interference

An alternative explanation for the results of Experiment 6 is that when the first tone in each trial was the same frequency as the second, the first conferred a general advantage to frequency discrimination. This explanation cannot be discounted based on the data in Figure 5.4.

After completing Experiment 6, four of the novice listeners (L25–28) took part in a short supplementary experiment. There were two conditions in the experiment, in which 10 DLFs were measured per listener in a single session. In one condition, listeners heard three tones per trial. The first tone was always the same frequency as the second, the second tone was roved widely in frequency, and the second and third tones differed in frequency by ΔF. The other condition was identical except that the first tone in each trial was replaced by silence. Listeners were told that trials would contain only one frequency change, and were instructed to name its direction.

The DLFs measured in the supplementary experiment are shown in Figure 5.7. A paired-samples t-test performed on listeners' log-transformed DLFs revealed that they were reliably larger when the first tone was omitted [$t(3) = -5.17,\ p < .05,\ r = -.96$]. This result is consistent with the idea that frequency discrimination is improved when the standard is preceded by a tone with the same frequency. Also plotted in Figure 5.7 are the same listeners' DLFs from a condition in Experiment 6 in which both the first and second tones in each trial were randomly and independently roved over a wide frequency range. These DLFs are larger than the DLFs from the conditions in the supplementary experiment, although the difference between the two-tone condition and the condition from Experiment 6 failed to reach the threshold for two-tailed statistical significance [$t(3) = -3.04,\ p = .06,\ r = -.91$]. A reasonable account of the results reported above is that in Experiment 6, the first tone was able to both improve frequency discrimination when it was the same frequency as the standard *and* worsen discrimination when it was random in frequency (discussed in more detail below).

5.3 Chapter discussion

The experiments reported in this chapter aimed to evaluate two explanations for the role of frequency roving in some listeners' insensitivity to pitch-change direction. The results of Experiment 5 did not support the idea that direction-impaired listeners were relying on feedback—instead suggesting that those listeners could genuinely perceive the direction of small pitch changes—in conditions without roving. Experiment 6 aimed to test the hypothesis that direction-impaired listeners are poorer at pitch-direction identification because they are confused or distracted by additional, irrelevant pitch changes. The results of Experiment 6, along with those

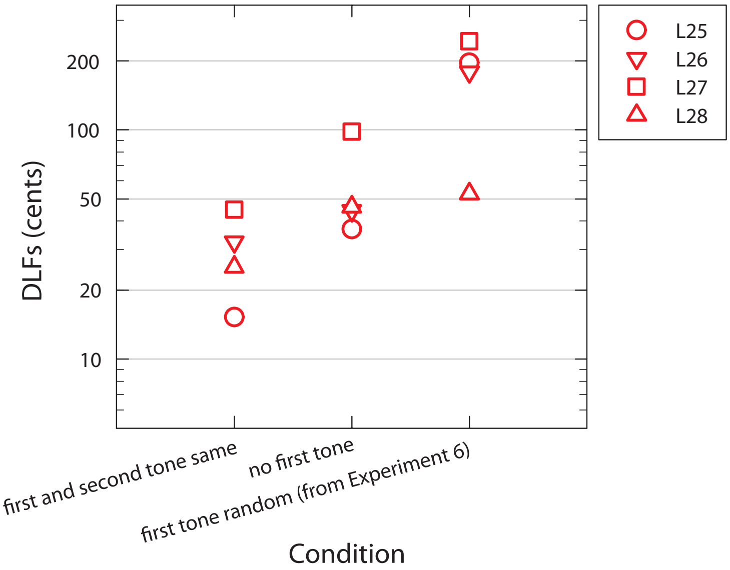


Figure 5.7. Results from four novice listeners in the supplementary experiment and in a single condition from Experiment 6 (see text). Each listener is represented by a unique unfilled symbol. The ordinate of each symbol represents the geometric mean of either five or 10 DLFs on logarithmic axes.

of a supplementary experiment, to some extent supported the hypothesis. Listeners' DLFs were larger when trials contained an additional, irrelevant frequency change, and this effect was more pronounced in the novices than it was in the experts. Moreover, there was a strong positive correlation between the degree to which a listener's DLFs were elevated by roving and the degree to which they were elevated by the random-frequency preceding tone. The correlation implies that the two effects are related, and possibly controlled by similar underlying processes.

The results of the supplementary experiment raise further questions about the role of task-irrelevant frequency changes in pitch discrimination. The first tone in each trial appeared to both reduce DLFs when it was the same frequency as the second, and increase DLFs when its frequency was random. The repetition of the standard frequency could improve DLFs by several conceivable mechanisms. The first tone could have acted as a cue for selective attention, providing listeners with an opportunity to focus on the standard frequency. It is well established that prior stimulation of a target frequency using a pure tone can improve—among other things—signal detection (e.g., T. J. Green & McKeown, 2001; Greenberg & Larkin, 1968; D. M. Johnson & Hafter, 1980; Macmillan & Schwartz, 1975), duration discrimination (e.g., Mondor & Bregman, 1994), level discrimination (e.g., L. M. Ward & Mori, 1996), and concurrent sound segregation (e.g., Demany et al., 2004) at that frequency. In these experiments, performance is sometimes modelled in terms of a listening band, which both enhances the representation of auditory information falling within it and attenuates the representation of information falling outside its limits.

It is reasonable to expect prior stimulation of the standard frequency to improve a listener's accuracy or thresholds for pure-tone frequency discrimination, although direct evidence of this is actually rather limited (Mondor et al., 1998; Mondor & Breau, 1999; Ruusuvirta, 2000; Ruusuvirta et al., 2008). In Experiment 6 and the supplementary experiment, in addition to providing an attention cue, the repetition of the standard frequency could have provided multiple opportunities for listeners to form an accurate representation of its pitch in memory (multiple 'looks'). To my knowledge, no studies have yet attempted to establish whether listeners can use a repeated standard tone either as an attention cue or for multiple looks—or both simultaneously—in frequency discrimination. One way to answer this question might be to employ narrowband noise as a precursor to each trial. On each trial, a listener hears two pure tones differing in frequency and is instructed to name the direction of the change. The standard tone—whose frequency would be roved over a wide range—would be preceded by a burst of noise, band-limited so that its width corresponds to that listener's DLF. The experimental manipulation would be the centre frequency of the noise: in one condition, the centre frequency would be the same as the standard in each trial, and in another, the centre frequency would be random. Under the former condition, the narrow-band noise might be useful as an attention cue, but not for multiple looks. Control conditions in which a different task was employed (e.g., duration discrimination) would have to be included to establish the efficacy of the noise as a cue.

It is important to point out that even if the listeners in the present experiments were able to make effective use of a helpful preceding tone, this does not necessarily weaken the argument for the interference hypothesis. In the supplementary experiment, the repetition of the standard frequency could have suppressed an interference effect of the tones from the preceding trial, in which case the results would be entirely consistent with the hypothesis. Another possibility is that all of the listeners tested were able to make use of a repeated standard frequency, but that the direction-impaired listeners in particular experienced a considerably stronger deleterious effect when the first introduced an irrelevant pitch change. This refinement of the interference hypothesis requires further exploration.

Chapter 6

Modelling listeners' DLFs using signal detection theory

6.1 Introduction

Signal detection theory (SDT) provides a highly flexible framework for describing and studying the decision-making processes of individuals in a psychological experiment. SDT was devised originally by Tanner and Swets (1954) as an alternative to older approaches to psychophysics, referred to collectively as threshold theories. Threshold-theory approaches consider the internal representations (observations) underlying an individual's decisions in an experiment as being finite states; for instance, 'up', 'down', or 'no change' in an experiment involving frequency discrimination. Consequently, the individual's sensitivity to a particular stimulus manipulation is determined by the probability that the manipulation leads to an observation of the appropriate state. The key innovation of Tanner and Swets' theory was to consider observations as instead lying along continua in the form of Gaussian distributions. Therefore, sensitivity is determined by the probability that the observation came from the appropriate *distribution*, and can be quantified as the distance between the means of the distributions relative to their standard deviations (d'). In their seminal textbook, D. M. Green and Swets (1966) presented a generalised version of the theory, which provided methods for estimating d' along with indices of response bias (e.g., c'; β) for various widely used experimental paradigms. For an in-depth description of SDT and its applications, see Macmillan and Creelman (2005).

In the following chapter, the framework of SDT is applied to the findings of the experiments described earlier in the thesis. The chapter first outlines a simple model of frequency discrimination proposed originally by Semal and Demany (2006). The model can be considered to represent the standard set of predictions for an ideal listener in the basic dual-pair paradigm used in their experiments[1]. The chapter then

---

[1] The term 'ideal listener' is avoided in the rest of this chapter because the objective here is not to detail how listeners should ideally behave in the experiments, but to provide a quantitative explanation

considers some of the key features of the data from Experiments 1–6 in terms of SDT, and discusses the challenges these features pose for the model. As will be discussed, the standard model does not provide a convincing explanation of the results without considerable modification. The chapter includes a number of suggestions for how future models might overcome these challenges.

6.2 The standard (CVD) model

Semal and Demany's (2006) model has been formulated in two previous publications (Micheyl, Kaernbach, & Demany, 2008; Semal & Demany, 2006), but it is also described here given its relevance to the thesis. The model posits a synthetic listener, Λ, for whom decisions in an experiment involving frequency discrimination are based on the echoic memory traces of internal pitch observations. On each trial, Λ makes one pitch observation per tone. Each observation is modelled as a quantity, x_i , which is linearly related to the frequency of the tone plus a random perturbation. Formally,

$$x_i = \gamma f_i + n_i, \tag{6.2.1}$$

where γ represents a constant scaling (or gain) factor, f_i represents the frequency of one of the tones in the trial (generated by the experimental procedure), and n_i represents an instance of a random variable with a Gaussian probability density function, with a mean of 0 and a standard deviation of σ. The value of i denotes the temporal order of the variables, so x_1 is the observation corresponding to the first tone in the trial, and f_1 is the frequency of the first tone. For simplicity and with no loss of generality, I assume hereafter that γ is always equal to 1; this allows all the remaining variables in the model to be expressed in the same units, namely musical cents.

Semal and Demany (2006) describe Λ as using a differencing decision strategy. In a two-interval, two-alternative forced-choice (2I-2AFC) paradigm such as those used in Experiments 5 and 6, Λ calculates the difference between the two sensory observations ($x_2 - x_1$), and then responds 'up' or 'down' depending on whether the sign of this difference is positive or negative. Of particular relevance to the following discussion is how Λ behaves in the dual-pair paradigm used in Experiments 1–3[2]. On detection trials in such an experiment, Λ compares the differences between observations within each stimulus pair (i.e., the difference between x_1 and x_2 to the difference between x_3 and x_4) and responds with whichever difference yields the larger absolute value. On identification trials, Λ responds 'up' or 'down' depending on

---

for the results from real listeners, some of whom clearly did not behave in an ideal way. I use 'synthetic listener' instead.

[2] Possibly also in Experiment 4, assuming that continuous frequency changes are perceived in the same manner as discrete frequency changes.

whether the sign of the larger absolute difference is positive or negative. The decision processes for the three examples are illustrated in Figure 6.1.

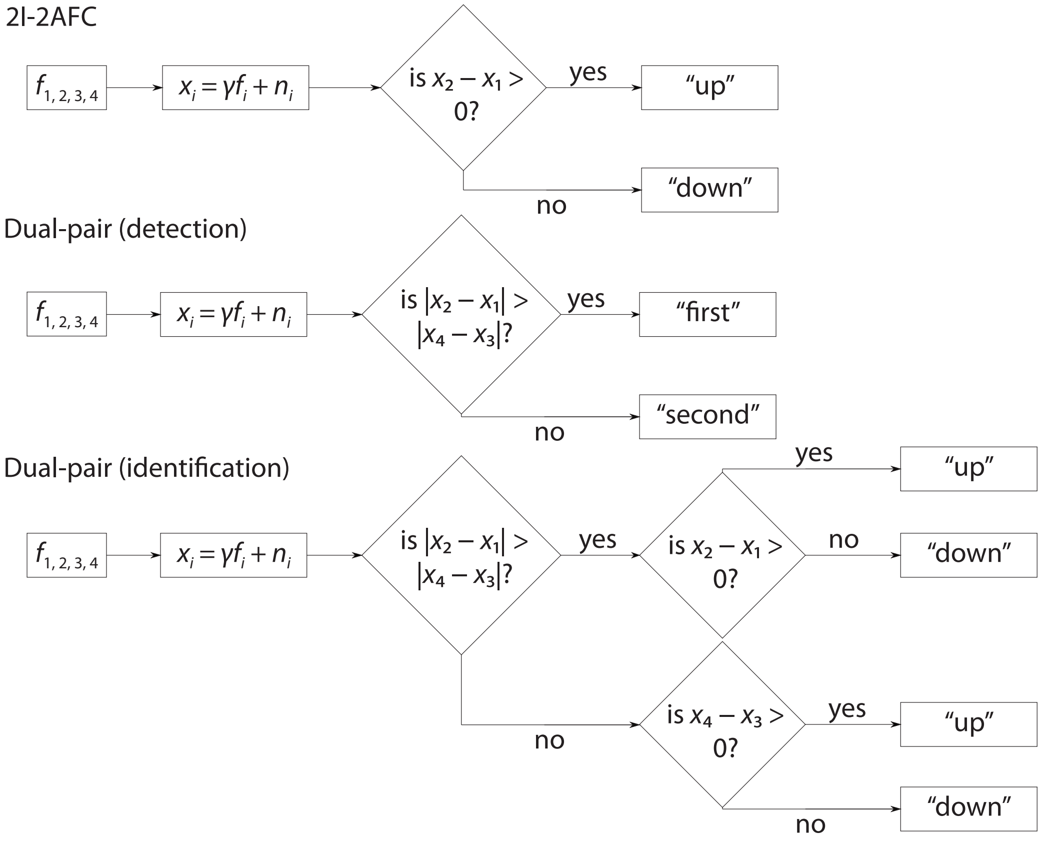


Figure 6.1. The decision-making processes for Λ under the CVD model. The top flowchart shows the process for trials in a 2I-2AFC experiment. The middle and bottom flowcharts show the processes for detection and identification trials in a dual-pair experiment, respectively.

In summary, there are two key features of the model proposed by Semal and Demany (2006). The first is that sensitivity in any experiment or experimental condition is limited solely by internal sensory noise. Since γ is set to 1, this noise is modelled as a standard Gaussian function with a single parameter, σ, which is assumed to be constant across all stimuli and all experimental manipulations. The second feature is that Λ always employs a differencing strategy. For these reasons, I refer to the model as the constant-variance differencing (CVD) model.

6.3 Implications from Experiments 1–6

The findings from the experiments reported in earlier chapters of this thesis present major challenges for the standard CVD model of frequency discrimination. The first feature is that listeners varied in their relative sensitivity or insensitivity to pitch-change direction. For example, in Experiment 1—which measured DDLFs and IDLFs using the same dual-pair paradigm used by Semal and Demany (2006)—listeners' I/D ratios ranged from 0.67 to 11.01. Whereas some listeners were considered direction-impaired and exhibited IDLFs that were elevated relative to their DDLFs, others without such impairment had IDLFs that were equal to or smaller than their DDLFs.

According to the CVD model, how should listeners perform in the standard dual-pair paradigm? A simulation procedure was used to generate DLFs for Λ under the constraints of the model. For integer values of σ between 1 and 100 cents, a DDLF and an IDLF were derived by running Λ through the adaptive procedure that predicts thresholds corresponding to 75% correct on the psychometric function. The adaptive procedure was the same as that used to estimate DLFs in human listeners, except that each adaptive run contained 10000 second-phase reversals. As before, DLFs were defined as the geometric mean of all values of ΔF visited during the second phase. The results (shown in Figure 6.2) revealed that the I/D ratio was constant and approximately 0.64 irrespective of the value of σ. Thus, under the traditional constraints of SDT, if DDLFs and IDLFs are limited only by the same internal sensory noise, IDLFs should always be smaller than DDLFs.

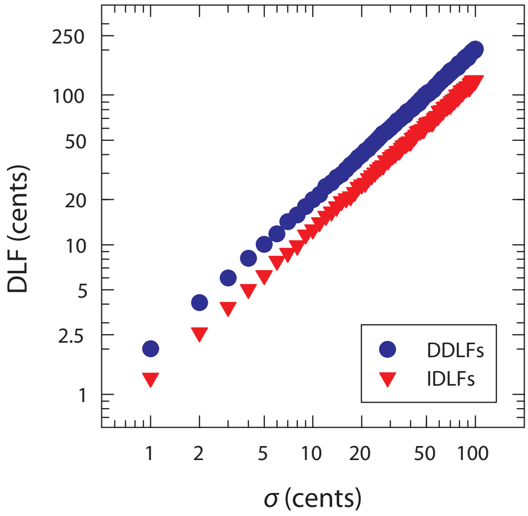


Figure 6.2. Simulated DDLFs and IDLFs for Λ under the CVD model for integer values of σ between 1 and 100 cents, on logarithmic axes.

A theoretical investigation by Micheyl et al. (2008) demonstrated that the same prediction could be derived mathematically. The authors provided equations defining the relation between d' and the maximum proportion of correct trials for Λ under the CVD model in the dual-pair detection and identification tasks. Following Macmillan, Kaplan, and Creelman (1977), in the detection task:

$$PC_{\mathrm{CVD_D}} = \left[\Phi\left(\frac{d'}{2}\right)\right]^2 + \left[1-\Phi\left(\frac{d'}{2}\right)\right]^2, \tag{6.3.1}$$

where $PC_{\mathrm{CVD_D}}$ represents the proportion of correct responses in the detection task under the CVD model, and Φ represents the cumulative standard normal distribution function. The converse equation, which gives d' as a function of PC, is:

$$d' = 2\Phi^{-1}\left(\frac{1}{2} + \sqrt{\frac{PC_{\mathrm{CVD_D}}}{2} - \frac{1}{4}}\right), \tag{6.3.2}$$

where Φ^{-1} represents the inverse cumulative standard normal function. According to this equation, when $PC_{\mathrm{CVD_D}}$ equals .75—the probability of correct responses at which DLFs were measured in the present experiments—d' is approximately equal to 2.10. The corresponding equations in the identification task are:

$$PC_{\mathrm{CVD_I}} = \Phi\left(\frac{d'}{2}\right), \tag{6.3.3}$$

and

$$d' = 2\Phi^{-1}\left(PC_{\mathrm{CVD_I}}\right). \tag{6.3.4}$$

Incidentally, Equation 6.3.4 is identical to that defining the relation between d' and PC for an unbiased observer in the single-interval Yes-No (YN) paradigm (Macmillan & Creelman, 2005). The equation demonstrates that the value of d' corresponding to a correct-response probability of .75 in the identification task is approximately equal to 1.34. Thus, Equations 6.3.2 and 6.3.4 validate the results of the simulation reported above: listeners should have I/D ratios of 0.64.

The prediction of the CVD model is clearly not generally correct considering the variability in real listeners' I/D ratios. Figure 6.3 is a copy of Figure 3.4, with additional oblique lines in each panel representing the loci of performance predicted for Λ under the model. Inspection of the figure shows that in Experiments 1 and 2 (conditions with and without maskers), most of the listeners had I/D ratios that were considerably larger than 0.64. Only two listeners in Experiment 2 had ratios smaller than 0.64, and in both cases their geometric 95% confidence intervals overlapped with the lower oblique line in the panel, suggesting that their I/D ratios were not reliably smaller than this value.

That real listeners' I/D ratios were never reliably smaller than 0.64 is consistent with the results reported by Micheyl et al. (2008). In that study, the authors tested 11 trained listeners in frequency, level, and amplitude-modulation-rate discrimination. In each case, thresholds for unsigned change detection and change-direction identification were measured using the dual-pair paradigm. I/D ratios were found to lie between 0.64 and 1 for all three discriminations in all 11 listeners. The authors pointed out that modifications to the standard CVD model could predict ratios between 0.64 and 1 by changing the way in which the sensory observations are generated. For instance, if observations were discretised rather than allowed to vary along continua, ratios would be closer to 1[3]. Importantly, all of the models

---

[3] Discretisation of observations into the smallest possible number of categories would constitute a

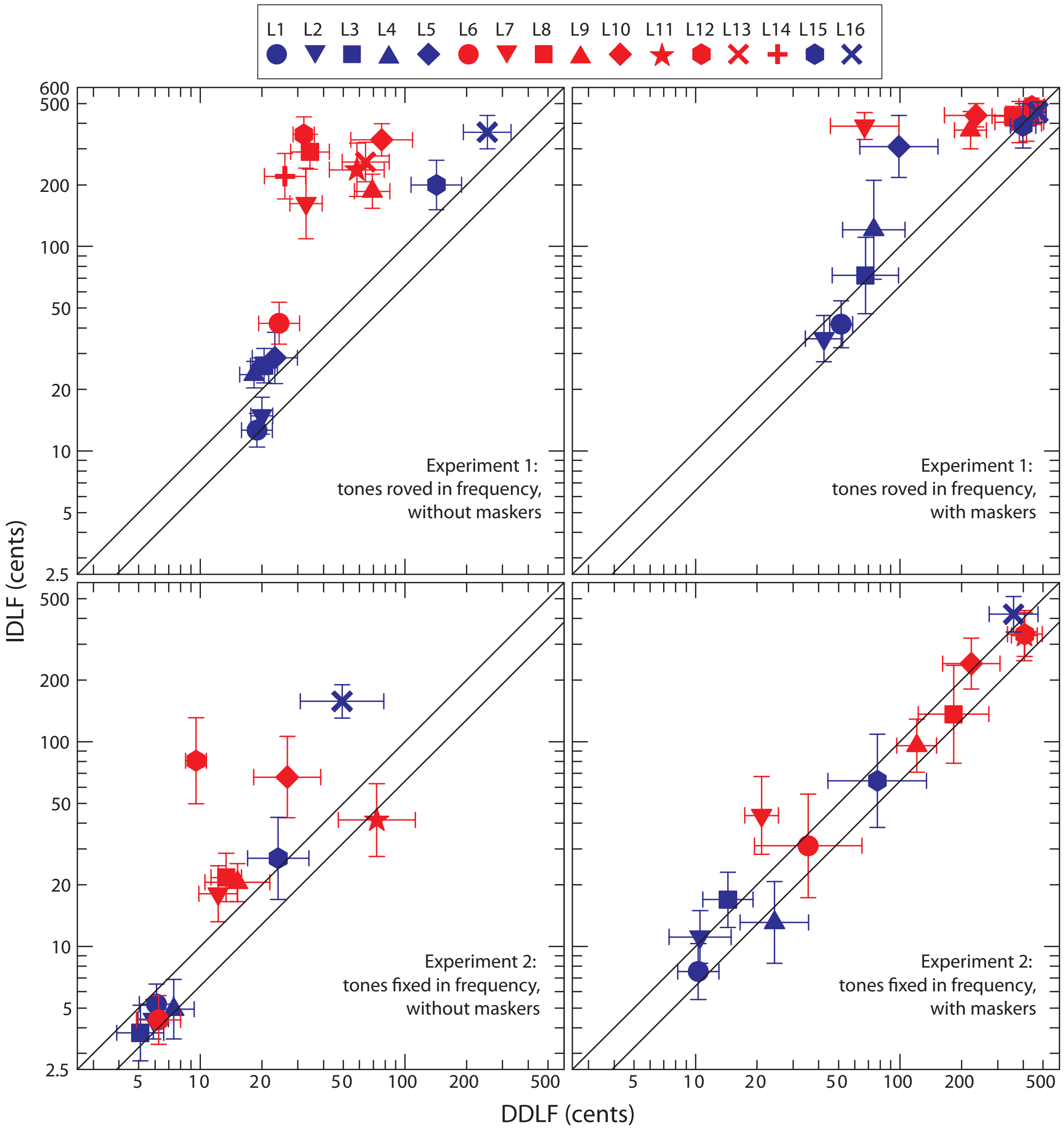


Figure 6.3. Same as Figure 3.4, except with the loci of performance predicted by the CVD model added to each panel (lower oblique lines).

considered by Micheyl et al. assume, like the CVD model, that the same internal noise limits detection and identification performance, and none of them predict I/D ratios larger than 1. Thus, the first major challenge posed by the findings of the present experiments is to explain why, in many listeners, IDLFs are *larger* than DDLFs.

The second feature of the data that poses a challenge to models of frequency discrimination is that there was a strong general tendency for DLFs to increase with increasing frequency-roving range, even in the listeners who experienced no difficulty identifying pitch-change direction. For instance, in the expert listeners from Experiment 3, DDLFs and IDLFs were both elevated by roving. On average, narrow roving increased the experts' DDLFs by approximately 1.53 times, medium roving by approximately 2.48 times, and wide roving by approximately 3.02 times. The corresponding factors for IDLFs were approximately 1.38, 2.65, and 3.01 times. These results are consistent with those of other studies measuring listeners' DLFs with fixed- and roved-frequency tones, which all found that roving was deleterious to performance (Section 2.2.4.4).

The CVD model—which always uses a differencing decision strategy—cannot account for changes in DLFs as a function of frequency-roving range. Mathematically, the application of roving corresponds to adding the same quantity drawn from a random distribution to the two observations within a trial (in a 2I-2AFC experiment) or within a pair (in a dual-pair experiment). This quantity cancels out when one takes the difference between the observations, yielding no difference in predicted DLFs whether roving is applied or not. Some authors have argued that under conditions without roving, the differencing strategy is not optimal (Macmillan & Creelman, 2005; Noreen, 1981). 'Optimal' in this sense means that a strategy achieves the maximum PC for any value of d'. It is plausible, therefore, that a model of frequency discrimination could account for the influence of frequency roving if it uses a decision strategy that is more efficient (i.e., obtains a greater PC) with fixed-frequency stimuli.

In two theoretical investigations, Micheyl and colleagues proposed alternative decision rules for Λ in the dual-pair paradigm (Micheyl & Messing, 2006; Micheyl & Dai, 2009). Under this new formulation, Λ bases its decisions on the outcome of statistical likelihood ratios. On each trial, likelihood ratios are determined for each of the alternatives (given the observations), and Λ selects whichever alternative yields the largest likelihood ratio (see also Noreen, 1981). Unlike the differencing strategy, this approach is more efficient when the observations within a pair are statistically independent. When the standard frequency is roved, these observations are correlated because roving adds the same random quantity to both. The strength of the correlation is proportional to the variance in this quantity, thus the efficiency of the strategy goes down as the roving range is widened. In other words, the strategy is most efficient when roving is not applied. This alternative formulation—referred to as

---

threshold model, which predicts I/D ratios of 1.

the constant-variance independent-observations (CVIO) model—predicts increases in DLFs with increasing frequency-roving range, whilst maintaining the assumption that the internal sensory noise remains the same under all conditions.

Equations for calculating likelihood ratios and for applying decision rules for Λ under the CVIO model have been provided by Micheyl and Messing (2006) and Micheyl and Dai (2009). However, these equations are extremely complex and are not reproduced here; there is a more convenient way to determine whether the model provides a reasonable account for the present findings. Using simulations and integration by parts, the authors were able to calculate values of PC and d' for the dual-pair detection and identification tasks when the observations within a pair are completely independent. The functions are shown as dashed curves in Figure 6.4. These curves indicate that when the maximum probability of a correct response is .75, the values of d' are approximately equal to 1.74 in the detection task and 1.30 in the identification task[4]. The authors also demonstrated that when the observations within a pair are highly correlated, the CVD and CVIO models are mathematically equivalent. Therefore, with highly correlated observations, the predicted values of d' are approximately equal to 2.10 in the detection task and 1.34 in the identification task. These curves are also plotted in Figure 6.4.

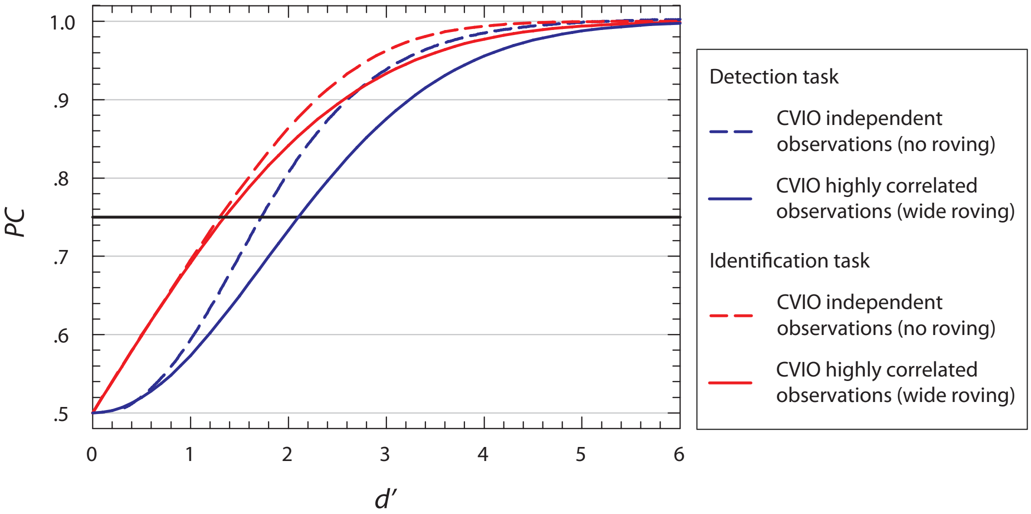


Figure 6.4. The relation between PC and d' for Λ in the dual-pair paradigm under the CVIO model. The horizontal line represents $PC = .75$.

On the basis of the values just presented, it can be seen that according to the CVIO model, listeners' DDLFs and IDLFs measured with roved-frequency tones should be no more than approximately 1.21 and 1.03 times larger, respectively, than their DDLFs and IDLFs measured with fixed-frequency tones. These predictions are considerable underestimations of the true effects of roving observed in real listeners who do not experience difficulty identifying pitch-change direction (e.g.,

---

[4] Only the latter of the two predicted values of d' was expressly stated in Micheyl and Dai (2009). The former value was calculated here using Micheyl and Messing's (2006) original MATLAB code.

the experts in Experiment 3). Neither the CVD model nor the CVIO model provide a convincing explanation for the effect of roving in the present experiments: even if real listeners are able to make use of an independent-observations decision strategy, their sensitivity to frequency changes under conditions with roving must be additionally limited by at least one other factor.

Another feature of the data, which presents arguably the most pressing challenge for models of frequency discrimination, is that the direction-impaired listeners tested in the present experiments were usually only insensitive to pitch-change direction when the stimuli were roved over a wide frequency range. Put another way, there was a three-way interaction between listener, task, and roving range. Neither the CVD model nor those suggested by Micheyl et al. (2008) are equipped to explain this interaction, since according to them I/D ratios should always be between 0.64 and 1. The CVIO model predicts increases in DDLFs and IDLFs with roving, but that DDLFs should increase *more* than IDLFs—an interaction between task and roving in the opposite direction to the one actually observed. Clearly the existing models need to be modified substantially if they are to account for this and the other findings from real listeners.

6.4 Modifying the existing models

6.4.1 Modelling stimulus uncertainty (UVD model)

A potential modification that could be implemented to account for the second of the three features above—the general influence of frequency roving in unimpaired listeners—is considered first. As mentioned in Section 3.4, reduced sensitivity has been modelled in previous studies as an increase in the deleterious effect of stimulus uncertainty (e.g., Durlach et al., 2005; Schlauch & Hafter, 1991; Lutfi, 1993). A similar characterisation could be achieved here by abandoning the assumption of constant internal sensory noise under all conditions, and instead allowing the noise to change relative to the degree of uncertainty. For a given run of trials, the parameter σ that limits the performance of Λ could be related by an appropriate function to the frequency-roving range of the tones within the run[5], and in all other respects Λ would behave exactly as under the standard CVD model. I refer to this modification as the uncertainty-based-variance differencing (UVD) model.

To speculate as to the shape of the function, the geometric mean DDLFs and IDLFs measured in the expert listeners in Experiment 3 were re-analysed. Only the DLFs from the conditions involving level roving between and within pairs were included in the analysis. The frequency-roving ranges used in those conditions were 0 cents (no roving), 31 cents (narrow roving), 310 cents (medium roving), and 3102 cents (wide

---

[5] Alternatively, γ could be related to roving range by the inverse of this function.

roving). According to SDT, an estimate of σ can be obtained from any measured DLF using the following rule:

$$\sigma = \frac{\gamma \cdot \mathrm{DLF}}{d'}. \tag{6.4.1}$$

If one assumes that listeners are using a differencing decision strategy, the values of d' reported earlier (2.10 when listeners perform the detection task, and 1.34 when they perform the identification task) can be plugged into Equation 6.4.1. More accurate estimates of d' might be achieved by assuming an independent-observations strategy, but these require a priori knowledge of the correlation coefficients between the sensory observations, which are not straightforward to calculate. Besides, the discussion in the previous section suggests that real listeners may not use the independent-observations strategy, so the simpler differencing strategy was assumed for the purposes of this analysis. With the value of γ again set to 1, estimates of each listener's σ_D and σ_I per roving range were obtained from their geometric mean DDLFs and IDLFs, respectively.

The results of the re-analysis are shown in Figure 6.5. Each panel contains the data from one expert listener. The abscissae and ordinates of the symbols represent the frequency-roving range of the condition in cents and a listener's estimated σ_D or σ_I, respectively. The roving-range values are plotted on a linear axis so that the values when the range was 0 cents (i.e., no roving) can be seen. In each panel, the following exponential-rise-to-maximum function has been fitted to listeners' σ_D estimates:

$$\sigma_\mathrm{D} = a + b\left(1 - c^{f_\mathrm{range}}\right), \tag{6.4.2}$$

where f_range represents the frequency-roving range over the condition, and a, b, and c are all free parameters. The parameter a controls the value of σ_D under minimal stimulus uncertainty (i.e., no roving), b controls the value of σ_D under maximal uncertainty (i.e., wide roving), and c controls the shape of the function. The best-fitting values of a, b, and c—along with the goodness-of-fit (adjusted R^2) of the function—were obtained using the Marquardt-Levenberg algorithm for non-linear regression implemented in SigmaPlot, and are shown in Table 6.1. These goodness-of-fits should not be taken to indicate the veracity of the model since only four data points per curve were predicted using three parameters.

A function has also been fitted to listeners' σ_I estimates in each panel of Figure 6.5. Recall that the values of d' used in the analysis predicted I/D ratios of 0.64 under the CVD model, and also that in Experiments 1 and 2 (and in the study by Micheyl et al., 2008), listeners who were not direction-impaired usually had I/D ratios that were somewhere between 0.64 and 1. This was also true for the expert listeners in Experiment 3, whose geometric mean I/D ratio across all conditions was 0.79, approximately 1.23 times larger than the 'lower bound' of 0.64. The use of these d' values led inevitably to larger σ_I estimates than σ_D estimates, and therefore a different

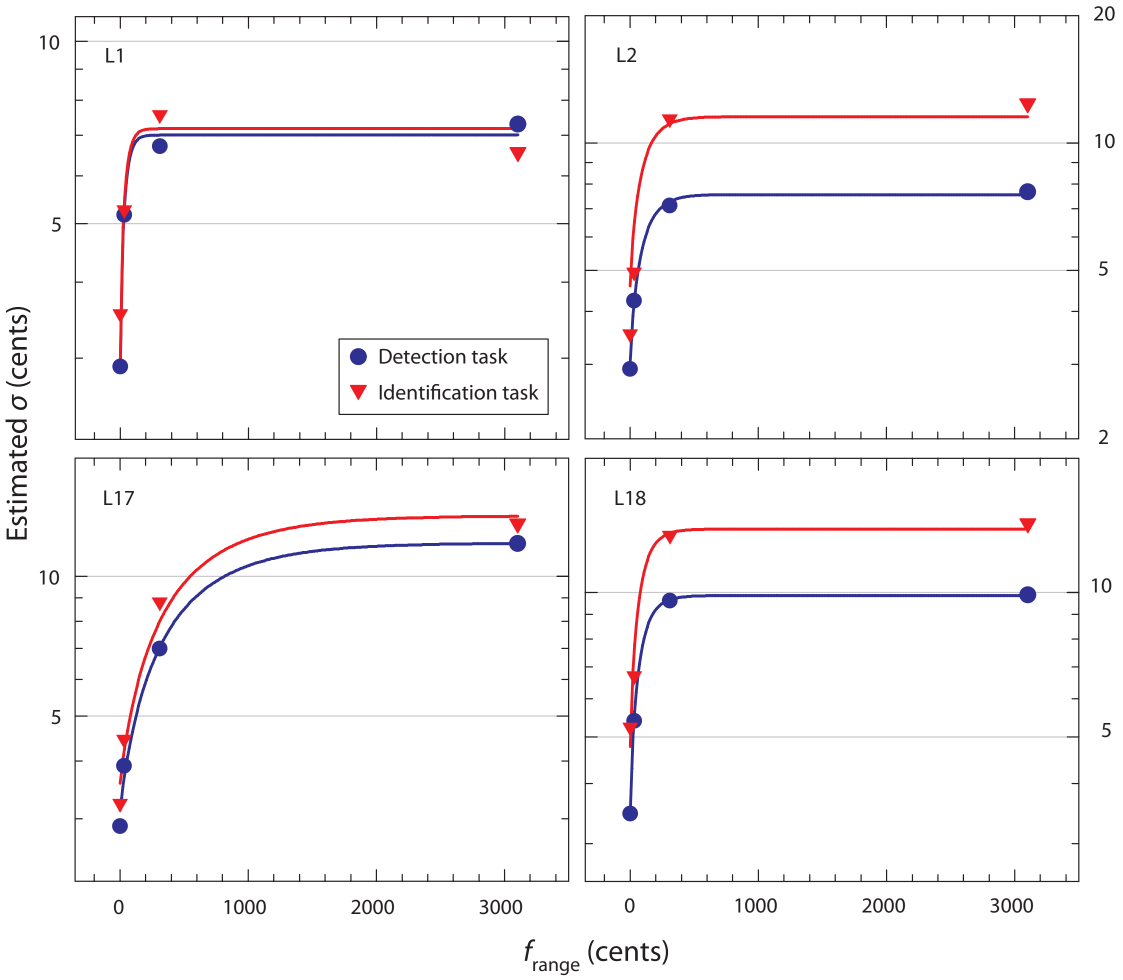


Figure 6.5. Estimated values of σ_D and σ_I (ordinates, scaled logarithmically) obtained from the four expert listeners' DLFs tested in Experiment 3. Only the conditions in which level was roved both between and within trials are included. The abscissa of each symbol represents the frequency-roving range (in cents), scaled on linear axes so that the values of σ when f_range equals 0 can be seen. The curves are described in the text.

Table 6.1. Estimates of function parameters and goodness-of-fit values derived from the UVD model.

	σ_D function				σ_I function	
Listener	a	b	c	Adjusted R^2	k	Adjusted R^2
L1	2.903	4.104	0.975	0.955	1.024	0.905
L2	3.012	4.547	0.991	0.988	1.527	0.938
L17	3.126	8.680	0.998	0.992	1.145	0.981
L18	3.473	6.366	0.989	1.000	1.376	0.985

function was fitted to the data:

$$\sigma_{\mathrm{I}} = k\left[a + b\left(1 - c^{f_{\mathrm{range}}}\right)\right], \tag{6.4.3}$$

where a, b, and c are no longer free parameters and are instead the values obtained for the curve fitted to the detection task data, and k represents a new free parameter. The best-fitting values of k and the adjusted R^2 of these curves are also shown in Table 6.1. In a sense, the function fitted to σ_{I} is simply a scaled version of the one fitted to σ_{D}, in which the parameter k compensates for the fact that listeners' sensory observations may be something other than perfectly continuous Gaussian variables: it is equivalent—plus or minus measurement variability—to a listener's true I/D ratio divided by 0.64. Therefore, if one was to implement one of the models suggested by Micheyl et al. (2008), k could be removed from the function entirely, and the same internal noise would limit both pitch-change detection and pitch-direction identification.

An alternative to the UVD model would be to take the approach used by Durlach and Braida (1969) in their model of intensity discrimination. Their model differs somewhat from the present one because it assumes several additive sources of internal noise limiting the performance of Λ rather than modelling the noise as a single distribution. When operating in sensory-trace mode, sensitivity is limited by noise in the formation of observations (sensation noise) and by the rate at which traces become diminished over time, either through decay or through interference from other sources (memory noise). When operating in context mode, Λ makes use of one or more stored referents, described as 'perceptual anchors' in later formulations of the model (e.g., Braida et al., 1984); sensitivity in the latter mode is not limited by memory noise but by noise in the anchor locations. The model assumes that Λ uses whichever mode has the least associated internal noise on any given trial, or that Λ is able to combine both modes without loss of sensitivity to achieve optimal performance. However, if one was to approximate Durlach and Braida's construct of multiple internal noises with a single variable, then the use of the two coding modes is effectively the same as saying that the ensemble variance of the internal noise changes with the frequency-roving range employed over a run of trials. Thus, Durlach and Braida's model and the UVD model would lead to very similar predictions.

6.4.2 Modelling sequential interference

The UVD model may be able to account for the general influence of frequency roving in listeners who do not experience difficulty identifying pitch-change direction. However, the model is less successful at accounting for the three-way interaction between listener, task, and frequency-roving range described earlier. Using the same method as before, σ_{D} and σ_{I} estimates were obtained from the novice listeners'

DDLFs and IDLFs measured in Experiment 3. The group geometric means of these values are shown in the right-hand panel of Figure 6.6, along with the expert group means in the left-hand panel for comparison. The functions described in Equations 6.4.2 and 6.4.3 were fitted again, and the parameter values and goodness-of-fits are shown in Table 6.2. The first function provided a very good fit to the novices' group σ_{D} estimates, but the second function provided a considerably poorer fit to their σ_{I} estimates. In other words, the UVD model cannot adequately account for the internal noise limiting novice listener performance in the identification task.

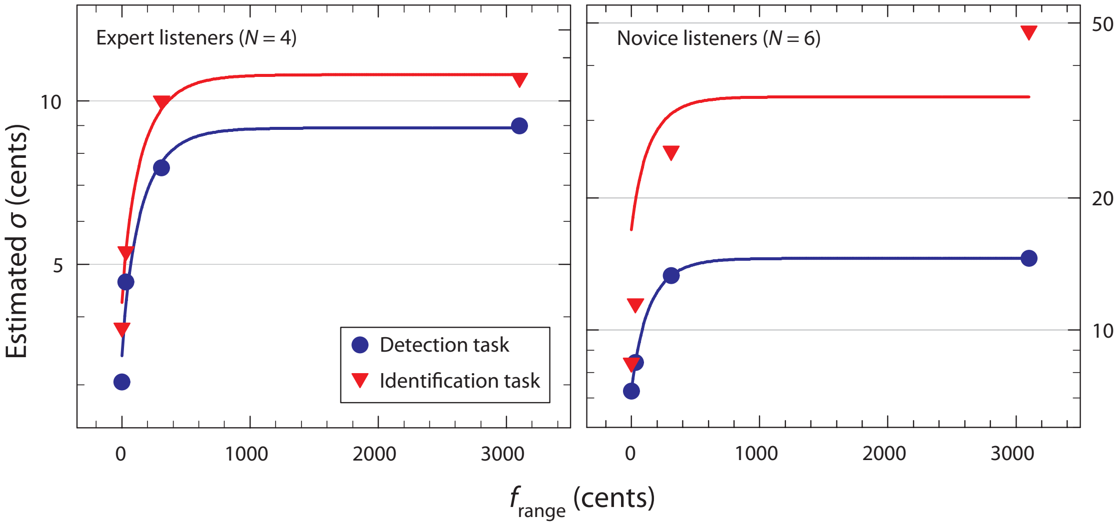


Figure 6.6. Same as Figure 6.5, except for the geometric group mean DLFs in Experiment 3.

Table 6.2. Estimates of function parameters and goodness-of-fit values derived from the modified model (group data).

	σ_{D} function				σ_{I} function	
Group	a	b	c	Adjusted R^2	k	Adjusted R^2
Experts (N = 4)	3.394	5.517	0.995	0.948	1.253	0.991
Novices (N = 6)	7.251	7.304	0.994	1.000	2.333	0.625

One way to account for this relationship without throwing away the CVD or UVD models would be to abandon the parsimonious assumption that sensitivity in the dual-pair detection and identification tasks is limited by the same internal noise, and assume instead that, for some reason, the internal noise is greater in the latter than the former. In the initial formulation of the CVD model, frequency discrimination in a dual-pair experiment was considered a two-step decision process: first Λ detects a change in frequency, and then assigns the correct direction to the change (Figure 6.1). The detection task requires only the first step in this process, whereas the identification task requires both steps. The first step involves comparing observations and so is limited by sensory noise. However, when Λ is required additionally to perform the second step, a novel source of internal noise could be introduced. This additional noise does not necessarily have to reflect sensory noise, and therefore would not have to be modelled as an increase in the variability of the observations. For those listeners with no difficulty identifying pitch-change direction, the additional noise would be minimal. Conversely, for direction-impaired listeners, the additional noise would be larger.

To justify such an extensive modification, the potential generators of additional, identification-specific noise need to be identified. One generator proposed earlier in the thesis is the effect of sequential interference from irrelevant pitch changes. When Λ assigns a direction to a perceived pitch change, the decision (up or down) could be influenced by other pitch changes occurring during the experiment using weighted linear combinations. For example, consider a dual-pair identification trial in which the relevant pitch change occurs in the second pair. Under the CVD or UVD models, Λ selects the correct pair if $|x_4 - x_3| > |x_2 - x_1|$, and then selects a direction based on the outcome of $x_4 - x_3$. The decision variable for the second step can be rewritten as ΔX; an 'up' decision is made if ΔX is positive, and a 'down' decision is made if it is negative. To incorporate sequential interference effects, the decision variable ΔX could be generated as follows:

$$\Delta X = \Delta X_{\text{rel}} + \sum \left(w_i \cdot \Delta X_{\text{irrel}_i} \right), \tag{6.4.4}$$

where ΔX_{rel} represents the difference between the two relevant sensory observations (e.g., $x_4 - x_3$), $\Delta X_{\text{irrel}_i}$ represents an irrelevant difference, and w_i represents a weight applied to an irrelevant difference. There is no constraint on the number of irrelevant differences that can be incorporated into the model, or where the irrelevant differences originate (e.g., between pairs within trials, or between trials). The modification can be thought of as incorporating a filter in the temporal domain, centred on the point in time when ΔX_{rel} occurs. An optimal filter would be sufficiently narrow so that weights of 0 are applied to all irrelevant differences; in this case no sequential interference would occur and the modified model would be equivalent to the CVD or UVD models. Alternatively, non-zero weights would result in irrelevant differences contaminating the decision variable, elevating IDLFs.

Theoretically, the incorporation of a temporal filter could explain both the influence of incongruent between-pair pitch changes described in Experiments 1 and 3 (Section 4.4.3), and the influence of the random-frequency first tone in Experiment 6 (Section 5.2). However, the validity of the approach cannot be tested effectively using the data from Experiments 1–6: in those experiments, the values of ΔX_{irrel} were always random due to roving, as were the values of ΔX_{rel} because ΔF was always manipulated adaptively. Future experiments could use a method of constant stimuli, in which the relevant frequency difference is fixed at an appropriate point on the psychometric function, and the irrelevant pitch changes are manipulated systematically.

6.5 Discussion

In this chapter, three important features of the data from Experiments 1–6 were identified and were used to discuss the validity of SDT-based models of frequency discrimination. Neither the standard model originally described by Semal and Demany (2006) nor other more complex models in the literature were able to account for all of these features. Two new modifications were suggested, but more data are needed to determine whether these can explain the range of individual differences observed in real listeners' DLFs.

The first modification suggested that in listeners who are not insensitive to pitch-change direction, the influence of frequency roving on sensory noise followed an exponential-rise-to-maximum function. The function is appealing from an intuitive point of view. One could imagine that it reflects the use of different listening strategies or different discrimination mechanisms, wherein the efficacy of the first strategy/mechanism becomes poorer with increases in the frequency-roving range, until at some point—when the function reaches asymptote—the first is overtaken by a second strategy/mechanism that is not influenced by roving. It remains to be seen whether the function also describes listeners' DLFs or d' for frequency discrimination in other experiments (e.g., Demany & Semal, 2005; Jesteadt & Bilger, 1974), and if so, whether the best-fitting parameter values are similar to those reported here.

The second modification attempted to provide a quantification of sequential interference effects from task-irrelevant pitch changes. Clearly much more data are required to assess the validity of this modification, although there is an indication from Experiment 2—in which two listeners exhibited larger IDLFs than DDLFs even in the absence of frequency roving—that at least one more source of internal noise influencing pitch-direction identification is required. At present, the origin of this additional noise is not clear.

Two paramount considerations when formulating future models of frequency discrimination are simplicity and ecological validity. The models outlined in this chapter are phenomenological and arguably the simplest that can be applied to the present data. They make few assumptions as to how pitch is determined in the auditory system, merely that listeners' decisions in a frequency-discrimination task are based on noisy internal observations. The default approach under SDT is to model observations as random Gaussian variables—as in all of the models described here—which may or may not be an ecologically valid assumption. It would be a relatively simple matter to replace the Gaussian probability distribution function in these models with another. For instance, Kaernbach (1991a) suggested that SDT-based models assuming Poisson-distributed observations account for the results of many experiments where traditional Gaussian models fail. A Poisson model may also have greater ecological validity in many cases, since Poisson processes are often used to predict the spontaneous firing rates of neurons (e.g., Rieke, Warland, Steveninck, & Bialek, 1997).

In their theoretical discussion, Micheyl et al. (2008) included a 'neurophysiologically inspired' Poisson model, incorporating Demany and Ramos' (2005) concept of frequency-shift detectors (FSDs). The model contains a neural increment detector, which compares the two tones in a pair, and whose mean firing rate increases linearly with the difference in frequency between the two, but only if this difference is positive (i.e., $\Delta F > 0$). For negative frequency differences (i.e., $\Delta F < 0$), the output of the detector is distributed following the same distribution as for physically identical stimuli (i.e., $\Delta F = 0$). The model also contains a decrement detector, whose mean firing rate increases linearly with the magnitude of negative frequency differences, and as if $\Delta F = 0$ with positive differences. The difference between the outputs of these increment and decrement detectors is used to decide whether a change occurred, and if a change did occur, to decide in which direction it was. Crucially, the output of each detector was modelled as a Poisson process with a baseline firing rate μ_0.

The Poisson model is able to predict I/D ratios between 0.64 and 1 by varying μ_0, making it a better fit than the CVD model to unimpaired listeners' DLFs. However, the Poisson model is similar to the CVD model in all other respects, meaning that it is not appropriate when one considers the influence of frequency roving or some listeners' insensitivity to pitch-change direction. Simple modifications to the Poisson model, such as introducing an asymmetry in the sensitivity or baseline firing rates of the detectors, did not improve its ability to explain these features of the present data. Moreover, as discussed in Section 2.2.4.5, more evidence is needed to support the existence of FSDs.

Chapter 7

Correlations with self-reported everyday hearing and musical ability

7.1 Introduction

Psychoacoustical experiments are usually quite unnatural listening experiences. Although some experiments are more realistic than others (cf. Kitterick, Bailey, & Summerfield, 2010), researchers often use synthesised or processed stimuli that bear little similarity to naturally occurring sounds, and test their listeners in exceptionally quiet environments. This is particularly true of studies of pitch perception, since most commonly used pitch stimuli—such as pure tones, complex tones with equal-amplitude harmonics, and IRN—are rarely (if ever) encountered outside a laboratory. It is not always obvious, therefore, if or how individual differences observed in such experiments relate to listeners' everyday hearing experiences or difficulties.

All of the listeners who participated in Experiments 1–6 (L1–36) were tested at some point under relatively comparable psychoacoustical conditions. Experiment 1, Experiment 3, and the two practice sessions prior to Experiment 6 all involved measuring 20 DDLFs and 20 IDLFs per listener using the basic dual-pair paradigm without maskers and with frequency roving over the range 400–2400.1 Hz. As mentioned previously, the listeners also completed two questionnaires prior to testing: the speech, spatial and qualities of hearing scale (SSQ), and a questionnaire about their musical education and expertise. DLFs and questionnaire data were also collected from an additional 23 listeners. The following chapter reports the results of analyses exploring the correlations between listeners' psychoacoustical measurements (DDLFs, IDLFs, and I/D ratios) and their self-reports. The primary aim of this work was to investigate whether the relative insensitivity to pitch-change direction observed in some individuals is related of aspects of their real-world hearing.

The SSQ is a 53-item questionnaire designed originally to measure a range of hearing disabilities across several domains, and to assess outcomes in clinical patients after the fitting of auditory prostheses (Gatehouse & Noble, 2004). The questionnaire has since proven a reasonably popular tool in audiology and in basic auditory research (e.g., Agus, Akeroyd, Noble, & Bhullar, 2010; Gatehouse & Akeroyd, 2006; Noble & Gatehouse, 2004; Noble, Tyler, Dunn, & Bhullar, 2009). Each item on the SSQ comprises a vignette of a normal circumstance involving hearing (e.g., 'You are talking with one other person and there is a TV on in the same room, can you follow what the person you're talking to says?'), and a visual analogue scale with which to make a response (e.g., going from 'not at all' to 'perfectly'). Items are split across three scales and cover many aspects of hearing, including understanding speech in a variety of competing contexts, perceiving distance, perceiving movement, segregating sounds, ease of listening, naturalness, and so on.

The broad scope of the SSQ made it an excellent choice for an exploratory investigation, but simply reporting the correlations between DDLFs, IDLFs, I/D ratios, and each item on the SSQ could lead to issues of interpretation due to the non-trivial family-wise error rate (Vul, Harris, Winkielman, & Pashler, 2009). It was also felt a priori that statistical corrections for multiple comparisons would not be appropriate since no specific hypothesis was being tested. Therefore, the SSQ data were first reduced by exploratory factor analysis, and the psychoacoustical measurements were correlated with individuals' factor scores rather than with each SSQ item separately. A previous study suggested that the original three scales of the SSQ could be further divided into 10 subscales (Gatehouse & Akeroyd, 2006). The subscales distinguish between understanding speech in the presence of different masking sounds, the localisation of sound sources and perceiving movement, sound quality and listening effort, and so on. To my knowledge, however, no previous attempt has been made to assess the validity of the subscales empirically. The similarities and differences between the theoretically and empirically derived structures of the SSQ are discussed briefly in this chapter.

It is conceivable that musical background would be related to performance in the experiments, particularly as insensitivity to pitch-change direction has been implicated as a symptom of amusia (Foxton et al., 2004) and previous research has found that musicians have at least initially smaller DLFs than non-musicians (Section 2.3.3.1). The SSQ contains a few items concerning listening to music, but does not assess an individual's musical education or expertise. The Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003) contains a comprehensive (but lengthy) music questionnaire, which was used as a basis for a short bespoke questionnaire completed by the listeners in the present experiments. Some of the items on the questionnaire were ordinal and these were entered into a second exploratory factor analysis; the remaining items were considered on an individual basis. Copies of the SSQ and music questionnaires are provided in the Appendix.

7.2 DLFs and I/D ratios

In total, DLFs and I/D ratios from 57 listeners were analysed. This figure includes 35 of the 36 listeners taking part in Experiments 1–6 (L18 was an outlier in terms of age and was excluded), and 22 of the further 23 listeners for whom DLFs and questionnaire data were also collected. The remaining listener was excluded on the basis of being an extreme outlier in terms of their IDLF, which was over an octave in size (1368 cents). Four of the listeners were experts (L1–3 and L17) who had experience in psychoacoustical experiments involving frequency and/or pitch discrimination prior to their DLFs being measured. The others were all novices selected using the pre-test (Section 3.2). All of the included listeners were in the age range 18–30 years, and with the exceptions of L4, L12, L16, and L19 (see earlier chapters), none had hearing levels greater than 20 dB HL at frequencies between 250 and 4000 Hz, inclusively.

Listeners' geometric mean DLFs are plotted in Figure 7.1. Overall, the DLFs are far larger than the ones typically reported in classic studies of frequency discrimination. Many of the listeners have very large IDLFs, or very large DDLFs and IDLFs. This finding is unsurprising because the majority of the group (red circles) were selected specifically because of their poor performance in the pre-test. The figure also highlights the considerable individual differences over the group.

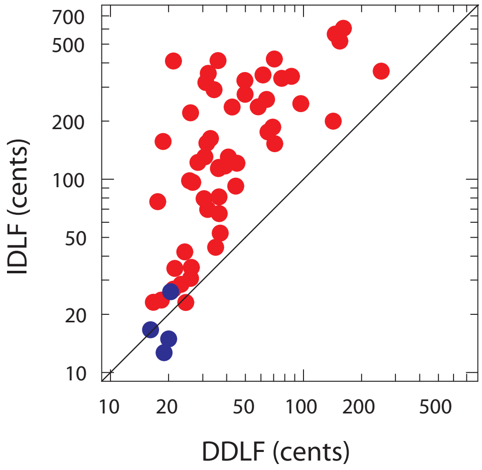


Figure 7.1. DLFs from 57 listeners entered into the analyses reported in this chapter. Each listener is represented by a symbol (blue are experts, red are novices). The abscissa and ordinate of each symbol represents that listener's geometric mean DDLF and IDLF, respectively. Error bars have been omitted. The diagonal line in each panel represents the locus of equivalent performance.

7.3 Exploratory factor analyses

7.3.1 The SSQ

7.3.1.1 Method

SSQ data from 118 individuals were entered into the factor analysis, including data from the 58 listeners mentioned above (all except L17), and data from 60 additional listeners from experiments conducted by other members of the research group at the University of York. All additional listeners were in the age range 18–30 years and had normal audiograms.

Since factor analysis is susceptible to the influences of deviations from normality, strong inter-item correlations, influential cases, and outliers (Field, 2005), the SSQ data were screened carefully. Histograms, Q-Q plots, and box plots revealed a strong negative skew in the data. Listeners tended to rate their hearing favourably, which is perhaps not surprising given that the group comprised young adults and that the SSQ was designed initially to be administered to older adults and individuals with hearing difficulties. A cubic transformation appeared to be most appropriate for the non-normally distributed items (selected using the Stata software package), and the same transformation was applied to all items because applying a transformation to only a subset of variables could artificially increase or decrease the correlations between variables (Field, 2005).

After transformation, outlier cases were identified and their values were replaced with the value of the closest non-outlying case from the same item (Field, 2005). A case was defined as an outlier if its value differed from the mean by more than plus or minus 3.29 standard deviations (i.e., exceeding the limits of bi-directional 99.9% confidence intervals). There were three outliers in total: one in item 3 and one in item 10 from the speech scale, and one in item 3 from the spatial scale. Methods to detect influential cases (bivariate outliers) could not be undertaken given the number of items in the analysis. Three items—item 5 from the speech scale, and 2 and 17 from the spatial scale—had one missing case each; these values were imputed with the mean of that item.

Some SSQ items were considered problematic and were excluded from the factor analysis. Item 14 from the spatial scale (‘Do the sounds of things you are able to hear seem to be inside your head or out there in the world?’) was removed because it contained more than one missing case and a number of individuals complained that the question was either confusing or difficult to answer. Three items from the spatial scale (15, 16, and 17) were all essentially slightly different wordings of the same question, and because strong correlations between specific items can influence the sensitivity of the overall analysis, items 15 and 16 were removed. Items 15 and 20–22 from the qualities scale were not completed by the group because they referred to auditory prostheses. Item 16 from the qualities scale (‘When you are the driver in a car, can you easily hear what someone is saying who is sitting alongside you?’) was removed because it was not completed by non-drivers.

Exploratory factor analysis was performed using SPSS in two stages. The first stage involved running the analysis with the remaining 45 SSQ items. Factors were extracted using the principle components method and only factors with eigenvalues greater than 1 were considered. An orthogonal rotation (varimax) was applied to the items’ factor-loading coefficients and items were considered to load significantly on a given factor if the coefficient was greater than .4. This analysis yielded nine factors, and item 13 from the speech scale (‘Can you easily have a conversation on the telephone?’) did not load on any factor significantly. This item is arguably more open-ended than most on the SSQ because issues not related to hearing per se—such as errors in conversational turn taking—could lead to difficulty talking on the telephone. For greater sensitivity, the factor analysis was repeated in a second stage with item 13 removed, leaving 44 items in total.

7.3.1.2 Results

The final analysis yielded nine factors with eigenvalues greater than 1, explaining approximately 72.25% of the total variance in the SSQ data. The rotated factor-loading matrix is shown in Table 7.1. The table also contains brief descriptions of each of the SSQ items for quick reference; the reader is advised to consult the Appendix or the original report by Gatehouse and Noble (2004) for the full vignette of each item. The solution suggests that the SSQ is factorially complex, supporting the assertion from a previous study that smaller, more specific subdivisions of the original three SSQ scales may be usefully made (Gatehouse & Akeroyd, 2006). The number of exploratory factors in the present solution was similar to the number of subscales (10) suggested by those authors, and in some cases the factors and subscales correspond more or less exactly.

Table 7.1. Summary of the SSQ items and the factor-loading matrix with varimax rotation. Loadings > 0.4 are in bold.

		Factors								
Item	Desciption	1	2	3	4	5	6	7	8	9
Speech 1	One person taking, TV on	−.152	**.414**	.306	.284	.198	.398	.169	−.084	−.091
Speech 2	One person taking, quiet and no echoes	−.095	.235	.210	**.521**	.360	.017	.278	−.221	.008
Speech 3	Five people talking, quiet	.129	**.534**	.335	.283	.025	.225	−.133	−.274	.173
Speech 4	Five people talking, busy restaurant	.215	**.747**	.161	.092	.019	.348	.055	−.078	.098
Speech 5	One person taking, continuous noise	.069	**.691**	.128	.196	.061	.042	.278	.108	.177
Speech 6	Same as 4, cannot see all talkers	.302	**.661**	.074	.120	.005	.384	.127	−.142	.183
Speech 7	One person taking, echoes	.160	**.723**	.201	.116	.068	.016	.136	.006	−.012
Speech 8	Ignore voice same pitch	.130	**.713**	.128	.056	.200	.203	.214	.301	−.005
Speech 9	Ignore voice different pitch	.170	**.726**	.097	.141	.256	.106	.185	.266	.055
Speech 10	Follow person talking and TV	.308	**.404**	.006	.068	.044	**.639**	.085	.015	.061
Speech 11	One person taking, other talkers	.338	**.723**	.044	.015	.145	.112	.094	.055	.043
Speech 12	Follow switching talkers	.253	**.737**	.234	.129	.143	.017	−.066	−.079	.002
Speech 14	Follow person talking and telephone	.267	.268	.088	−.016	.146	**.711**	.109	−.042	.089
Spatial 1	Locate lawnmower	**.488**	.269	.091	.211	**.548**	.236	.142	.153	.222
Spatial 2	Locate talker	**.511**	.333	.132	.195	**.548**	.097	−.060	.107	.091
Spatial 3	Talker left or right	.380	.107	.067	**.481**	**.412**	−.074	.275	.012	.212
Spatial 4	Locate door slam	**.540**	.210	.179	.064	**.589**	.216	−.024	.100	.044
Spatial 5	Locate sounds above/below	**.519**	.342	.346	.055	**.470**	.032	.099	−.054	−.099
Spatial 6	Locate dog barking	**.547**	.277	.240	.043	**.603**	.157	.079	.041	.081

Table 7.1. (Continued). Summary of the SSQ items and the factor-loading matrix with varimax rotation. Loadings > 0.4 are in bold.

Item	Desciption	Factors								
		1	2	3	4	5	6	7	8	9
Spatial 7	Locate moving vehicle	**.688**	.150	.064	.315	.211	.255	−.019	.085	.115
Spatial 8	Distance from voice/steps	**.729**	.191	.036	.139	.120	.315	.083	.104	−.013
Spatial 9	Distance from vehicle	**.649**	.050	.156	.231	.192	.348	.158	.074	.054
Spatial 10	Lateral movement vehicle	**.853**	.176	.156	.122	.012	.116	.109	.035	.146
Spatial 11	Lateral movement voice/steps	**.780**	.255	.283	.008	.146	.009	.178	.010	.180
Spatial 12	Voice/steps moving towards or away	**.767**	.228	.274	.123	.008	.020	.107	.073	.108
Spatial 13	Vehicle moving towards or away	**.781**	.248	.242	.030	.133	−.049	.124	.073	.128
Spatial 17	Sounds where expected	**.406**	.072	.107	.157	.201	.049	.098	.048	**.690**
Qualities 1	Sounds separate	.232	.301	.363	**.505**	.022	.061	.114	.159	.218
Qualities 2	Sounds not jumbled	.154	.187	**.485**	.078	−.039	.082	.008	.082	**.660**
Qualities 3	One person taking, radio	.195	.332	**.497**	**.424**	−.139	−.022	.135	.080	.241
Qualities 4	Identify person by voice	.241	.172	.251	**.697**	.023	.158	.037	.106	−.042
Qualities 5	Distinguish familiar music	.346	.104	.229	**.519**	.067	−.122	.099	**.410**	.151
Qualities 6	Distinguish sound sources	.220	.186	.388	**.499**	.162	.127	.026	.327	.131
Qualities 7	Identify musical instruments	.104	.012	.138	.129	.029	−.050	.009	**.852**	.015
Qualities 8	Music natural	.119	.108	**.652**	.203	.113	−.025	.192	**.445**	.142
Qualities 9	Sounds clear	.289	.181	**.692**	.314	.101	.020	.130	.216	.029
Qualities 10	Voices natural	.247	.173	**.765**	.174	.072	.042	.125	.126	.076
Qualities 11	Sounds natural	.193	.196	**.723**	.073	.154	.100	.044	−.046	.374

Table 7.1. (Continued). Summary of the SSQ items and the factor-loading matrix with varimax rotation. Loadings > 0.4 are in bold.

Item	Desciption	Factors 1	2	3	4	5	6	7	8	9
Qualities 12	Own voice natural	.094	.082	**.521**	.051	.287	.037	**.438**	−.116	.278
Qualities 13	Tell person's mood from voice	.237	.152	**.536**	.135	.069	.364	.222	−.064	−.183
Qualities 14	Concentrate when listen	.226	.219	.192	.180	.047	.200	**.761**	.115	−.037
Qualities 17	Understand speech when passenger	.290	.357	**.519**	.291	.230	.075	.108	.092	−.060
Qualities 18	Effort of conversation	.292	.379	.262	.203	.014	.028	**.666**	−.048	.050
Qualities 19	Ignore competing sounds	.120	.284	.243	−.088	.028	.210	**.451**	.337	.265

Factor 1: Movement perception. All except one of the items from the spatial scale of the SSQ loaded on the first factor. These items were concerned with determining the location of a sound source relative to the listener, or determining the direction of a travelling source, and covered a range of different types of environmental sound (e.g., vehicle noise, speech). That all these items loaded on the factor could suggest that it reflects general spatial awareness; however, it was most strongly loaded by those items involving judging the distance or direction of moving sounds (items 8–13). Since judging sound movement would necessarily require a listener to be aware of the sound's location at any given moment in time, it follows that judging movement would rely in part on the same mechanisms important for localising a stationary object. Therefore, it makes sense for the other items from the spatial scale to load also on a movement perception factor. This interpretation is entirely consistent with the 'distance and movement' subscale (items 8–13) suggested by Gatehouse and Akeroyd (2006). Item 3—the only item to involve a completely lateral spatial judgement (i.e., purely left or right)—did not load on this factor.

Factor 2: Speech intelligibility. All except two of the items from the speech scale loaded on the second factor. These items were concerned with understanding speech in everyday environments that are not optimal in some way, such as because of continuous environmental noise, irrelevant speech from other talkers, or reverberation. In their theoretical analysis, Gatehouse and Akeroyd (2006) divided the speech scale of the SSQ into four subscales: speech in quiet (items 2 and 3); speech in noise (items 1, 4, 5, and 6); speech in speech contexts (items 7, 8, 9, and 11); and multiple speech-stream processing and switching (items 10, 12 and 14). The speech factor observed here, however, was not loaded more strongly by any particular kind of situation or any particular type of masker. Item 2 did not load on the factor, possibly because it is the only item involving speech perception under essentially ideal conditions (i.e., no background noise whatsoever, and no reverberation). Item 14 also did not load on the factor.

Factor 3: Quality and naturalness. Only items from the qualities scale loaded on the third factor. Items from Gatehouse and Akeroyd's (2006) 'quality and naturalness' subscale (8–12) loaded most strongly, but the factor was also loaded by items concerned with perceiving someone's mood (13), sound segregation (2 and 3) and the relative ease or difficulty of understanding speech (17)—all of which are conceivably influenced by the quality of the sound's percept.

Factor 4: Segregation and identification. The fourth factor yielded by the analysis is more difficult to interpret. The factor was loaded by most but not all the items from two subscales suggested by Gatehouse and Akeroyd (2006): identification of sound and objects (4–7 and 13 from the qualities subscale) and segregation of sounds (1–3 from the qualities subscale). That item 13 did not load on the factor might be expected because in its case the source of the sound (a person's voice) has already been identified. Item 2 was worded slightly differently to items 1 and 3, and was slightly more ambiguous because it did not name explicitly the sources of the sounds to be segregated. Curiously, item 2 from the speech scale and item 3 from the spatial scale—both of which involve listening to speech in the *absence* of other sounds—loaded significantly on the factor. Why this is the case is not clear.

Factor 5: Localisation. The fifth factor is the counterpart of the first factor, loaded only by items from the spatial scale that involve localising sound sources relative to the listener (1–6). The factor matches exactly Gatehouse and Akeroyd's (2006) localisation subscale.

Factor 6: Following multiple conversations. Two items from the speech scale of the SSQ loaded on the sixth factor (10 and 14). These items are the only two that involved following conversations simultaneously, requiring greater demands on a listener's attention that the other items in the speech scale. Gatehouse and Akeroyd's (2006) 'multiple speech-stream processing and switching' subscale included item 12 in addition to 10 and 14; the former item did not involve attending to and understanding two streams of speech at the same time.

Factor 7: Listening effort. Items from Gatehouse and Akeroyd's (2006) 'listening effort' subscale loaded on the seventh factor (14, 18, and 19 from the qualities scale), along with the item concerning the clarity of one's own voice (12).

Factor 8: Music perception. Three items from the qualities scale concerning the perception of music per se loaded on the eighth factor (5, 7, and 8). These items assessed a listener's ability to identify familiar pieces of music, their ability to distinguish between instruments, and the perceived clarity or naturalness of music. These items were not considered to represent a separate subscale by Gatehouse and Akeroyd (2006). Incidentally, although no other items loaded on the factor above .4, the most strongly loading sub-threshold items all involved the segregation of other sounds from a background, two of which mentioned pitch explicitly (8 and 9 in the speech scale).

Factor 9. The ninth factor was loaded by item 17 in the spatial scale and item 2 on the qualities scale. It is not obvious how these items relate to each other, so the ninth factor is not considered in any more detail.

7.3.2 Music questionnaire

Seventeen of the items on the music questionnaire required listeners to make an ordinal response via a visual analogue scale. The responses from these items were entered into an exploratory factor analysis. Questionnaire data from 67 individuals were available in total. The data were not heavily skewed, so no transformation was performed. Item 19—concerned with humming along to music—contained a single outlier case ($|z| > 3.29$), the value of which was replaced with the closest non-outlying value from that item. There were no missing cases. Exploratory factor analysis was performed using all 17 items in a single stage. Factors were extracted using the principle components method and only factors with eigenvalues greater than 1 were considered. An orthogonal rotation (varimax) was applied to the items' factor-loading coefficients and items were considered to load significantly on a given factor if the coefficient was greater than .4. The analysis revealed three factors, explaining approximately 62.9% of the total variance. The rotated factor-loading matrix is shown in Table 7.2.

Table 7.2. Summary of the music questionnaire items and the factor-loading matrix with varimax rotation. Loadings > 0.4 are in bold.

		Factors		
Item	Desciption	1	2	3
Music 1	Music in childhood (never–all the time)	.208	**.523**	.264
Music 9	Sing in private (never–all the time)	.325	.283	**.581**
Music 10	Sing in public (never–all the time)	−.122	.259	**.721**
Music 11	Listen to music (never–all the time)	−.051	**.794**	.200
Music 12	Perceive seperate notes (impossible–effortless)	**.537**	.306	**.540**
Music 13	Perceive seperate instruments (impossible–effortless)	**.778**	.109	.161
Music 14	Sing (impossible–effortless)	.045	−.020	**.869**
Music 15	Memorise song (impossible–effortless)	**.407**	.299	**.561**
Music 16	Repeat tune recently sung (impossible–effortless)	**.446**	**.505**	**.533**
Music 17	Accuracy of Happy Brithday in head	**.514**	**.541**	.203
Music 18	Recognise song from first notes (never–all the time)	.398	**.620**	.280
Music 19	Hum along (impossible–effortless)	**.801**	.133	.191
Music 20	Sing a note to match piano (impossible–effortless)	.331	.156	**.684**
Music 21	Identify pitch-change direction (impossible–effortless)	**.866**	.174	.100
Music 22	Self out of tune (unable–able)	.137	**.626**	.103
Music 23	Tune stuck in head (never–all the time)	.372	**.714**	.036
Music 24	Clap/tap familiar melody (impossible–effortless)	**.760**	.392	.059

Factor 1: Analytical listening. Items that logically required a listener to perform some sort of on-line analysis when listening to music loaded most strongly on the first factor. The item concerning pitch-direction identification (21) obviously requires the pitch of sequential sounds to be determined; humming, clapping, and tapping along to music (19 and 24) require the listener to anticipate features of the piece, such as its melody, meter, and so on; and isolating different elements of a musical piece (12 and 13) requires segregation and grouping. The factor was also moderately loaded by three items concerned with learning or recalling a learnt melody (15–17).

Factor 2: Musical exposure and memory for music. The item concerning musical exposure during childhood (1) and the item concerning the extent to which a listener actively listens to music (11) loaded on the second factor, along with four items concerned with recalling music from memory (including the 'earworm' effect; 16–18 and 23). It is not clear why item 22—concerned with telling if oneself is out of tune—loaded on this factor and not the third (see below).

Factor 3: Singing. The final factor was most strongly loaded by four items that mentioned singing explicitly (9, 10, 14, and 20), and by one item concerning segregating notes when listening to music (12), and one item concerning memorising a musical piece (15).

7.4 Correlations between measures

For the 57 listeners in which DDLFs and IDLFs were measured, individual factor scores were calculated for each of the interpretable factors reported above using the regression method implemented in SPSS. This method derives scores by calculating the weighted sum of a listener's responses to all the items entered into the factor analysis, with weights determined by the factor-loading coefficients. Table 7.3 shows the correlations between listeners' log-transformed DLFs, log-transformed I/D ratios, SSQ factor scores, and music questionnaire factor scores. Although using factor scores rather than individual item scores greatly reduced the number of variables entered into this analysis, the number of statistical tests performed was still large, and the Bonferroni-corrected significance criterion (p_{bf}) was extremely conservative ($< .0008$). For this reason, correlations that would be considered significant at the uncorrected level are also highlighted in the table. Note that because an orthogonal rotation was applied to the factor-loading coefficients, individuals' scores on any one factor are uncorrelated with their scores on all of the other factors from that analysis. These coefficients were therefore omitted.

Table 7.3. Correlations (Pearson's r) between listeners' log-transformed psychoacoustical measurements and their factor scores ($N = 57$).

	Log-transformed			Music factor scores		
	DDLF	IDLF	I/D ratio	An. list.	Exp./mem.	Singing
Log-transformed IDLF	.709†					
Log-transformed I/D ratio	.142					
SSQ factor scores						
Movement perception	.079	.107	.082	.070	.345**	.128
Speech intelligibility	−.036	−.037	−.021	.280*	−.076	−.142
Quality and naturalness	−.174	−.104	.003	.111	−.008	−.051
Segregation and identification	−.069	−.267*	−.315*	.209	.161	−.167
Localisation	.027	.065	.068	.040	.007	−.180
Multiple conversations	−.123	−.057	.025	.024	.195	.113
Listening effort	−.256	−.147	.012	.179	−.042	−.066
Music perception	−.262*	−.241	−.115	.502†	.057	.160
Music factor scores						
Analytical listening	−.383**	−.370**	−.193			
Musical exposure/memory	−.282*	−.229	−.080			
Singing	−.087	.006	.084			

*$p < .05$, **$p < .01$, †$p_{\text{bf}} < .0008$

In their original study, Semal and Demany (2006) reported a strong positive correlation between listeners' log-transformed DDLFs and log-transformed I/D ratios (r = .90). In the present study, although there was strong positive correlation between listeners' log-transformed DDLFs and their log-transformed IDLFs, there was a very weak correlation between the DDLFs and I/D ratios. The lack of consistency between Semal and Demany's results and the present ones could stem from the fact that the latter were based on more data, or that the latter contained a much greater proportion of direction-impaired listeners.

The correlations between the log-transformed psychoacoustical measurements and scores on several of the SSQ factors—movement perception, speech intelligibility, quality and naturalness, localisation, and multiple conversations—and scores on the singing factor from the music questionnaire were all very small ($|r| < .2$). This suggests that neither a listener's basic sensitivity to frequency changes nor their sensitivity to pitch-change direction is related to these aspects of self-reported everyday hearing. Most of the correlations between SSQ and music questionnaire factor scores were also small, although there was a strong positive correlation—significant at the corrected level—between scores on the SSQ music factor and scores on the analytical listening factor. This relatioship is to be expected given the similarity between the interpretations of the two factors. Two more correlations were significant at the uncorrected level: the movement perception factor correlated postively with the musical exposure/memory factor, and the seech intelligibility factor correlated with the analytical listening. Why these factors in particular correlated with each other is not clear.

There were weak correlations between listeners' IDLFs and I/D ratios and their scores on the segregation and identification factor from the SSQ, which were statistically significant at the uncorrected level. The correlations were negative, indicating that larger IDLFs and I/D ratios were associated with rating one's own hearing more poorly. As discussed earler, the interpretation of this factor is made difficult because it is loaded by two items that are not from the qualities scale of the SSQ. To determine which items were driving the relationship, a second correlation analysis was run using listeners' log-transformed psychoacoustical measurements and their cubic-transformed scores on the SSQ items that loaded strongly on the factor. The results (Table 7.4) show that none of the correlations were significant at the Bonferroni-corrected level (< .002). IDLFs and I/D ratios correlated at the uncorrected level with item 1 on the qualities scale ('Think of when you hear two things at once ...Do you have the impression of these as sounding separate from each other?'), and DDLFs and IDLFs correlated item 4 from the qualities scale ('Do you find it easy to recognise different people you know by the sounds of each one's voice?').

There was a weak negative correlation between listeners' DDLFs and their scores on the music perception factor. A slightly weaker correlation, below the criterion for

Table 7.4. Correlations (Pearson's r) between listeners' log-transformed psychoacoustical measurements and their scores on the individual SSQ items that loading on the segregation and identification factor. These correlations were performed on the transformed item scores that were when entered into the factor analysis ($N = 57$).

	Log-transformed		
Item	DDLF	IDLF	I/D ratio
Speech 2	−.161	−.151	−.074
Spatial 3	.028	−.054	−.100
Qualities 1	−.130	−.270*	−.268*
Qualities 3	−.177	−.211	−.145
Qualities 4	−.268*	−.283*	−.168
Qualities 5	−.195	−.256	−.192
Qualities 6	−.021	−.063	−.070

*$p < .05$

uncorrected statistical significance, was observed between listeners' scores on the factor and their IDLFs. The same was true for the musical exposure/memory factor from the music questionnaire, and both DDLFs and IDLFs correlated negatively with the analytical listening factor. These results suggest that performance in the present experiments was related to music perception or musical experience. A final correlation analysis was performed to determine which items from these factors were driving the observed relationships. Table 7.5 shows that most of the items from all three factors correlated significantly with listeners' DLFs. These items are concerned with segregating notes and instruments (qualities 5 and 7, music 12 and 13), the naturalness of music (qualities 8), exposure to music during childhood (music 1), time spent listening to music (music 11), repeating a tune (music 16 and 22), humming, tapping, and clapping along to music (music 19 and 24), judging the direction of a pitch change between subsequent piano notes (music 21), and having a tune stuck in one's head (music 23). The non-significant correlations were with items concerning the accuracy of a memorised melody (15 and 17). None of the correlations with listeners' I/D ratios were significant. Again, the significant correlations were all negative.

7.4.1 Non-ordinal items

There were nine non-ordinal items on the music questionnaire. The items assessed whether the listener came from a musical family (2 and 3), whether they at any time practised a musical instrument (and if so, how accomplished they became; 4–8), whether they or anyone else thought they might be tone deaf (25), and whether they had any other difficulties or skills relevant to music (26).

The 57 listeners were grouped according to whether they responded yes or no on

Table 7.5. Same as Table 7.4, for the music, analytical listening, and musical exposure/memory factors ($N = 57$).

	Log-transformed		
Item	DDLF	IDLF	I/D ratio
Qualities 5	−.233	−.285*	−.206
Qualities 7	−.366**	−.301*	−.108
Qualities 8	−.370**	−.355**	−.182
Music 1	−.361**	−.259*	−.051
Music 11	−.263*	−.216	−.076
Music 12	−.532†	−.449†	−.153
Music 13	−.451†	−.403**	−.179
Music 15	−.072	−.147	−.149
Music 16	−.283*	−.140	.052
Music 17	−.301	−.195	.012
Music 18	−.392**	−.416**	−.252
Music 19	−.345*	−.316*	−.149
Music 21	−.435**	−.374**	−.152
Music 22	−.308*	−.214	−.033
Music 23	−.285*	−.290*	−.166
Music 24	−.474†	−.442**	−.215

*$p < 0.05$, **$p < 0.01$, †$p_{\mathrm{bf}} < 0.001$

item 2 (‘During your childhood, were any of your family members musicians?’ no: $N = 42$; yes: $N = 15$). Independent-samples t-tests were used to compare if the groups differed in their DDLFs, IDLFs, or I/D ratios. In all the t-tests reported in this section, equal variances across groups have not been assumed if the Levene’s test was significant. The tests revealed statistically significant differences between the groups’ log-transformed DDLFs [$t(41.16) = 2.72$, $p < .01$, $r = .39$], and the groups’ log-transformed IDLFs [$t(55) = 2.27$, $p < .05$, $r = .29$]; in both cases, responding yes to item 2 was associated with smaller DLFs. The difference between the groups’ log-transformed I/D ratios was not significant [$t(55) = 1.29$, $p = .21$].

Listeners were grouped according to whether they responded yes or no on item 3 (‘Were musical instruments played during your childhood?’ no: $N = 23$; yes: $N = 34$). Independent-samples t-tests revealed that log-transformed DDLFs were reliably smaller in the listeners who answered yes to this question [$t(33.24) = 2.14$, $p < .05$, $r = .35$], but neither log-transformed IDLFs nor I/D ratios differed reliably between the groups [$t(55) \leq .75$, $p \geq .12$].

The music questionnaire revealed than none of the 57 listeners was a highly experienced or regularly practising musician. Responses on items 4–8 were used to group the listeners according to whether they had received music lessons (other than compulsory school lessons) or had attempted to learn to play an instrument for more than one year at some point in the past (no: $N = 31$; yes: $N = 26$). Listeners who had received music lessons or practiced an instrument had reliably lower DDLFs [$t(55) = 2.24$, $p < .05$, $r = .29$], but neither log-transformed IDLFs nor I/D ratios differed reliably between the groups [$t(55) \leq .97$, $p \geq .34$]. The remaining items on the questionnaire were either not completed by all the listeners or not enough detail was given, which prevented running further analyses.

7.5 Discussion

This chapter aimed to investigate whether the relative insensitivity to pitch-change direction observed in some individuals was indicative of aspects of their everyday hearing, which were measured using self-report questionnaires. Although potentially limited by listeners' opinions or a lack of insight in a way that would not affect the outcomes of psychophysical tests, this approach provided a convenient way to assess a wide range of different aspects of hearing. Overall, the results suggest that in young adults, neither DLFs nor I/D ratios are strongly related to aspects of their self-reports. Listeners' scores on many of the latent factors underlying the SSQ, particularly those concerning speech intelligibility or auditory spatial awareness, were uncorrelated with their psychoacoustical measurements. Weak correlations were observed with scores on some factors. DDLFs and IDLFs appeared to be related to musical experience and music perception, and IDLFs and I/D ratios were weakly related to segregation and identification self-reports.

The results of the exploratory factor analysis of the SSQ—performed here for the purposes of data reduction—supported the suggestion from Gatehouse and Akeroyd (2006) that the SSQ is factorially complex, and there was a reasonable amount of agreement between the empirical factors and the theoretical subdivisions proposed by those authors. The results could be informative for future investigations. In practical terms, the factor loadings could guide decisions about which particular items to include or exclude if researchers wished to abridge the SSQ for use when the time available for testing is at a premium. The analysis included data only from healthy young adults, and a potentially rewarding comparison might be made between the present results and those of similar analyses performed using data from older adults or users of auditory prostheses, for whom the SSQ was originally intended. It would be interesting if the factor structure of the SSQ were different in these different groups of listeners. However, since exploratory factor analyses are data-driven techniques, their best-fitting solutions are not always the most parsimonious or the most appropriate in terms of theory; many authors argue that a solution obtained using exploratory methods must be verified using confirmatory factor analysis (e.g., J. Stevens, 1996). With an appropriate factor solution confirmed in young adults, structural equation modelling could be used to estimate the strength of the causal relations linking the latent factors to listeners' self-reports, which in turn could be compared between different populations. An attempt was made here to run confirmatory factor analysis, but there was not enough data per item on the SSQ to implement it.

The finding that DLFs did not correlate with self-reports of speech intelligibility is consistent with the results of the studies employing psychometric test batteries to explore individual differences in auditory perception (Section 2.3.1). In those studies, performance on tasks involving the discrimination of basic acoustic features, such as spectral content, intensity, or duration, loaded on factors that were distinct from those loaded by tests of speech perception. This stands to reason since the demands of speech intelligibility are likely to be quite different from those of non-categorical discrimination along a basic dimension. Moreover, as many functional imaging studies have shown, the neural substrates of pitch and lexical/phonetic discrimination are very different (e.g., Zatorre et al., 1992). The studies discussed in Section 2.3.1 did not include tests of auditory spatial awareness, such as measuring listeners' discrimination thresholds for interaural time or level differences, presumably because they would have been difficult to implement when listeners were tested groups using loudspeakers rather than individually using headphones. The present results suggest that listeners' performance on those kinds of tests would be uncorrelated with frequency discrimination.

The weak correlations between listeners' DLFs and their scores on the music factor of the SSQ were substantiated by the analysis of the music questionnaire data. Listeners' reports of their exposure to music during childhood and whether or not they had attempted to learn to play an instrument had weak-to-moderate associations with their DDLFs, and rating one's own musical abilities more favourably was generally associated with smaller DDLFs and IDLFs. These modest effects are in the same direction as those reported in several studies comparing DLFs in professional or highly skilled musicians and non-musicians (Section 2.3.3.1), suggesting that the correlations reported here might have been stronger if any of the listeners tested were experienced musicians. The 20 DDLFs and 20 IDLFs were measured within approximately 2 hr of testing per listener. One study suggests that the differences between musician and non-musician groups is diminished after considerable practice (Micheyl et al., 2006), so perhaps the present correlations would also have dissappeared if more DLFs per listener were measured .

That DLFs correlated more strongly with items on the music questionnaire than did I/D ratios suggests that it was listeners' basic sensitivity to frequency changes, rather than their abilitiy to identify the direction of those changes per se, that was related to musical experience. A study by Foxton et al. (2004) suggests that IDLFs are markedly more impaired than DDLFs in people with amusia. The results reported in this chapter are not necessarily inconsistent with those from Foxton et al.'s study because the direction-impaired listeners tested here were almost certainly not amusic. Nothing from the questionnaires suggested that any of the listeners suffered from a profoundly disturbed perception of music, which is symptomatic of the disorder (cf. Peretz et al., 2003).

Chapter 8

Experiment 7: Neural correlates of pitch-change detection and pitch-direction identification

8.1 Rationale

Section 2.4.3 described a recent study by Foxton et al. (2009) that used magnetoencephalography (MEG) to investigate the neural basis of poor pitch-direction identification. In their first experiment, the authors presented two groups of seven individuals—a group whose IDLFs were in the range 8–25 cents, and a group whose IDLFs were in the range 1.4–2.5 semitones—with stimuli containing linear frequency glides whilst their extra-cranial auditory evoked fields (AEFs) were recorded. The listeners performed a pitch-direction identification task during the MEG session, and the data were analysed using dipoles fitted to listeners' left and right temporal areas in order to delineate the relative contributions of the auditory cortices in the two hemispheres. The authors concluded from the results that when identifying the direction of the glides, their high-threshold listeners relied more than did their low-threshold listeners on non-optimal cortical mechanisms that were lateralised to their left hemispheres.

The aim of the Experiment 7 was to replicate and extend Foxton et al.'s (2009) basic finding that listeners with large IDLFs exhibit an increased contribution of left-hemisphere cortex when identifying the direction of pitch changes. Like their study, the present experiment used MEG to measure listeners' AEFs in response to stimuli containing frequency changes. As in the previous experiments in this thesis, Experiment 7 included a group of listeners who were direction-impaired, and a group who were unimpaired. The data were analysed using a dipole approach that was similar to the one employed by Foxton et al..

A major difference between Foxton et al.'s (2009) study and others that have identified listeners who are insensitive to pitch-change direction (Johnsrude et al., 2000; Semal & Demany, 2006; Tramo et al., 2002) is that the former used frequency-glide stimuli rather than pairs of temporally discrete pure tones (cf. Foxton et al., 2004). Experiment 4 in this thesis found that listeners' I/D ratios were very similar regardless of whether the target frequency change was discrete or continuous (see also Demany, Carlyon, & Semal, 2009), suggesting that similar mechanisms are involved in the perception of both. Based on that finding it seems reasonable to predict that Foxton et al. would have found the same pattern of results in their data if they had used tone pairs rather than glides. Experiment 7 tested the hypothesis that direction-impaired listeners exhibit a leftward asymmetry of processing (i.e., stronger AEFs in their left-hemisphere dipole than in their right-hemisphere dipole) when identifying the direction of frequency changes between temporally discrete pure tones.

Experiment 7 also investigated the influence of task on listeners' AEFs. It was pointed out in Section 2.4.3 that, since the listeners in Foxton et al.'s (2009) first experiment were always required to identify the direction of the glides they heard, the group difference they observed might not have been related to the mechanisms involved in pitch discrimination per se. Their second experiment, which was designed to evoke mismatch negativity (MMN), involved passive listening to the stimuli. The authors did not find a statistically significant difference between the groups' MMNs, suggesting that perhaps active pitch-direction identification was a critical feature of the first experiment. In Experiment 7, listeners performed two tasks in the MEG session. One of the tasks was analogous to the detection task used in Experiments 1–4, whilst the other was analogous to the identification task—only the latter required listeners to determine the direction of the pitch changes. If the group differences observed by Foxton et al. were influenced by the fact that listeners performed a pitch-direction identification task, these differences might be less pronounced when listeners hear the same stimuli but are not required to make pitch-direction judgements.

Another feature of Foxton et al.'s (2009) study was that their direction-impaired listeners did not have much experience with frequency or pitch discrimination before taking part in the first experiment. DDLFs and IDLFs were measured prior to the MEG session, but at most the listeners heard only 180 glide stimuli before their AEFs were recorded. DLFs measured with such a small number of trials might be quite unreliable, and with more practice the large performance difference between Foxton et al.'s groups might have diminished. The results of Experiment 3 in this thesis support this possibility: the starting frequencies of Foxton et al.'s glide stimuli were randomly chosen values in the range 513–595 Hz (256.7 cents); this frequency-roving range is slightly narrower than the one used in medium-roving conditions of in Experiment 3 (310 cents), in which on average novice listeners' IDLFs were only slightly larger than their DDLFs. All of the direction-impaired listeners taking part in Experiment 7 had their DDLFs and IDLFs measured during several hours of testing prior to their taking part in the main experiment.

8.2 Method

8.2.1 Listeners

Twenty-three listeners took part in Experiment 7, 11 of whom were considered to have no difficulty identifying the direction of small pitch changes (7 females; 10 right-handed; age range 19–62 years), and 12 of whom were considered to be direction-impaired (8 females; 10 right-handed; age range 18–30 years). Data were collected from another unimpaired listener, but due to a temporary fault with the MEG scanner, their data had an extremely low signal-to-noise ratio and were omitted from the analyses.

The unimpaired group included three expert listeners from the previous experiments (L2, L3, and L18) and six more listeners who were either research students or faculty members of the University of York (L25–30). All six had considerable prior experience in psychoacoustical experiments and completed the online version of the pre-test, their scores indicating no difficulty with pitch-direction identification. The remaining two individuals were undergraduates who also completed the pre-test (L31 and L32). These listeners' scores on the pre-test indicated initially that they might be direction-impaired, but in both cases the results turned out to show that they were not (Section 8.2.3).

The direction-impaired group included L7, L9, L12, and L19–22 from the previous experiments. The group also included five new listeners (L33–37) selected using the pre-test. None of the listeners was a professional or practising musician.

8.2.2 Stimuli

Each trial in the main part of Experiment 7 contained a pair of pure tones. The tones in each pair were shorter than in those in the previous experiments (150 ms rather than 250 ms), and were separated by a shorter silent ISI (150 ms). All tones had random starting phases and 20-ms cosine-squared on/off ramps. The frequency of the first tone in a pair was always roved over the range 400–2400.1 Hz. The key manipulation was the frequency of the second tone, which was either the same as the first, upward or downward in frequency by 50 cents, or upward or downward in frequency by 350 cents.

Unlike in the previous experiments, stimulus generation in Experiment 7 was done offline before the experimental sessions. Three hundred and sixty-six tone pairs were generated in total: 122 contained no frequency change, 122 contained a 50-cents change (61 upward and 61 downward), and 122 contained a 350-cents change (61 upward and 61 downward). These sounds were delivered diotically via silicone Etymotic ER30 tubes connected to ear inserts. Since the frequency response of the ER30s declines at approximately 10 dB/octave above 1000 Hz, the stimuli were pre-emphasised to approximate a flat response when presented to the listener in a manner consistent with previous studies using the same audio-delivery system (Kitterick, 2008). This involved making recordings of a complex tone with a fundamental frequency of 100 Hz and 59 harmonics presented through the ER30s using a Brüel & Kjær ear canal coupler and sound level meter. The fast Fourier transform of the recorded tone was used to calculate an attenuation value for each of the harmonics relative to the frequency with the maximum output level. These values were then inverted and used as coefficients in a digital filter that pre-emphasised the tones. The pre-generated, pre-emphasised tones were roved in level over a 7-dB range, centred at 65 dB SPL.

8.2.3 Procedure

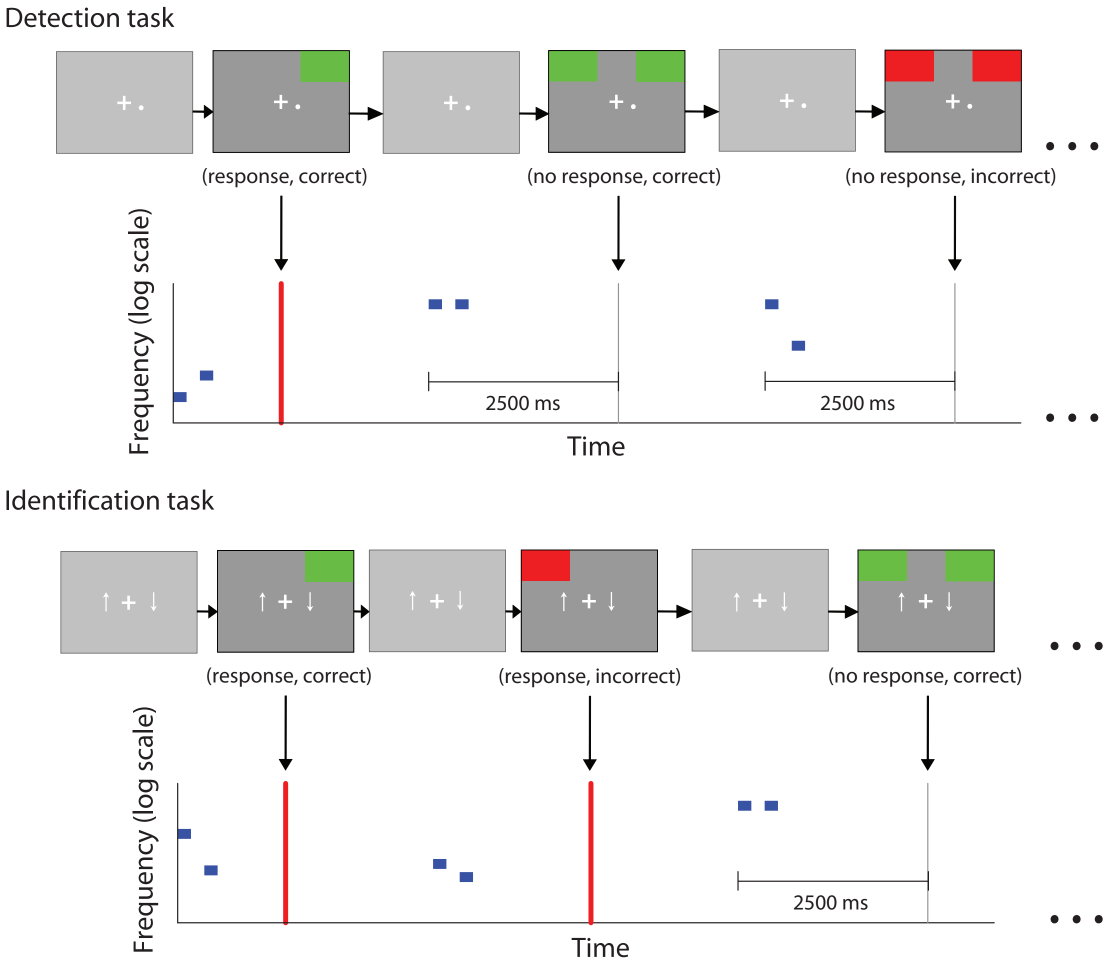
Before the main part of Experiment 7, the listeners who had not taken part in any previous experiments underwent preliminary assessments. First, pure-tone audiometry confirmed that none of them had hearing levels greater than 20 dB HL at octave frequencies between 250 and 4000 Hz, inclusive. As mentioned above, six of the new listeners (L31–37) were selected using the pre-test, and to ensure that they were reliably insensitive to pitch-change direction, 20 DDLFs and 20 IDLFs were measured in each using the standard dual-pair paradigm (method described in Section 5.2). As in Experiments 1 and 2, Mann–Whitney tests with a Bonferroni-corrected significance criterion ($\alpha = .00833$) were used to compare DDLFs and IDLFs on a listener-by-listener basis. The results indicated that the IDLFs measured in L33–37 were significantly larger than their DDLFs at the corrected level ($U \leq 86.00$, $z \leq -3.08$, $p < .001$, $r \leq -.49$; I/D ratios in the range 2.12–10.16). DDLFs and IDLFs did not differ significantly at this level in L31 and L33 ($U \geq 178.00$, $z \geq -0.60$, $p \geq .57$, $r \geq -.09$), whose I/D ratios were 1.33 and 0.94, respectively. These two listeners were therefore included in the unimpaired group rather than the direction-impaired group.

The main part of the experiment consisted of two sessions, completed on different days. The first session was carried out in a sound-attenuating booth (IAC) in the Psychology department at the University of York. The whole session took around 40 min, during which time the listeners heard each of the 366 tone pairs twice, in two randomly-ordered runs of trials (732 trials in the session in total). In one of the runs, listeners performed the 'detection task', and in the other they performed the 'identification task'. It is important to point out that these tasks are analogous to but not the same as those used in the dual-pair paradigm of Experiments 1–4. The order in which the tasks were performed was counterbalanced across the group.

During the detection task, listeners were instructed to listen to the tones and respond using a PC keyboard whenever they heard a pair of tones containing a frequency change. Listeners were told that not all pairs would contain a change, and that their response times were limited: they had a maximum of 2.5 s to respond after hearing a pair, after which time their lack of a response would be counted as them not perceiving a change on that trial. After a response was made or 2.5 s had elapsed (whichever came first), visual feedback was provided on a PC monitor. The next trial began after a jittered silent interval between 1 and 1.5 s. During the identification task, listeners again only responded when they heard a pair of tones containing a frequency change, but indicated the direction of the change with their response. If listeners heard an upward change, they pressed a key on the left-hand side of the keyboard (e.g., a, s, or d), and if they heard a downward change, they pressed a key on the right-hand side of the keyboard (e.g., :, @, or ~). Again, listeners were told that not all pairs would contain a change, and that they had a maximum of 2.5 s to respond.

The visual display contained a central cross (+) at all times. During the detection task, the display also contained a small dot, lateralised to either the left- or right-hand side of the cross. Although a correct response could be made with any key, listeners were instructed to respond using keys on the side of the keyboard indicated by the dot. The position of the dot switched from left to right or vice versa after listeners completed approximately 25%, 50%, and 75% of the trials within the run. This feature of the experiment ensured that listeners responded using their left and right hands a roughly equal number of times during the detection task. During the identification task, the display contained instead two arrows—an upward-pointing arrow on the left-hand side of the cross and a downward-pointing arrow on the right-hand side of the cross—as a reminder of the appropriate responses. Feedback took the form of green and red rectangles appearing in the top left-hand and/or right-hand corners of the display for approximately 100 ms (Figure 8.1).

The general design of the second session of Experiment 7 was identical to the first. Listeners heard the same stimuli (in two new randomly ordered runs), performed the same tasks (in the same order), and were presented with the same visual display. The session was carried out in a magnetically shielded room at the York Neuroimaging Centre, and extra-cranial magnetic fields were recorded using MEG for the entire 40-min session. Responses were made with two Photon Control Lumitouch response pads rather than a PC keyboard. One response pad was placed underneath each hand, and listeners were instructed to use any button on the left-hand pad for a left-key press, and any button on the right-hand pad for a right-key press. Listeners were



Detection task

Identification task

Figure 8.1. Example trials illustrating the design of Experiment 7. The top panels represent the stimuli and the corresponding visual display during trials in the detection task, and the bottom trials represent the same in the identification task. Blue horizontal lines represent pure tones, the vertical red lines represent responses made by the listener, and the vertical grey lines represent the response-time limits. In the visual display, a trial would be followed by a red or green rectangle (for approximately 100 ms) lateralised to the top corner corresponding to the response; when no response was made on a given trial, rectangles appeared in both corners.

instructed additionally to fixate on the central cross throughout the experiment, and were supine during the session; the visual display was presented overhead using a projector and a mirror.

8.2.4 MEG data analysis

8.2.4.1 Data acquisition

MEG data were acquired continuously using a 248-sensor, whole-head, magnetometer system (Magnes 3600 WH, 4-D Neuroimaging) at a sampling rate of 678.17 Hz (16-bit). The data were DC coupled and low-pass filtered online at 200 Hz. Bipolar vertical and horizontal electro-oculograms (EOGs) were also recorded from four electrodes (SynAmps system, NeuroScan) attached to the left and right outer canthus and above and below the right eye, and an electro-cardiogram (ECG) was recorded via two electrodes, one on each forearm. The apparatus used to deliver the stimuli (MOTU) sent triggers marking the onset of each tone pair. EOGs, ECGs, triggers, and listener responses made with the Lumitouch pads were all encoded within the MEG data sets.

8.2.4.2 Processing and averaging

After acquisition, the data sets were imported into the BESA software package (Brain Electrical Source Analysis, MEGIS software, Germany). Three sensors that were known to perform unreliably (including one over the right temporal area) were marked as bad in each set and were excluded from all further processing and analysis steps. In each data set, a characteristic ocular artifact was created by averaging the MEG signal occurring within a period over peaks in the vertical EOG (around 100 exemplars per data set). Principle components analysis was used to extract a single topographic component from this average, which usually explained more than 90% of the variance. A characteristic cardiac artifact was identified in the same way, with a time window over the Q peak in the ECG (around 2000 exemplars per data set). Principle components analysis was used to extract two topographic components, usually explaining more than 90% of the variance. These average artifact topographies were saved for later incorporation into the source solutions, and were used to correct the MEG data temporarily using the surrogate function in BESA.

The stimulus triggers were used to define 732 epochs extending from 100 ms before to 1400 ms after the onset of the tone pairs. Any epoch within the artifact-corrected data in which the recorded magnetic field strength at any sensor exceeded 3.5 pT, or which contained a difference between two adjacent samples greater than 2.5 pT at any sensor, was considered to be contaminated by a non-ocular and non-cardiac artifact, and was excluded from all further analyses. There were usually less than 20 such epochs in each set.

Artifact correction was turned off before averaging. Ten average epochs were created per MEG data set. The first was the average of all the uncontaminated epochs within the set (maximum 732). Three separate averages of each of the three stimulus types (no frequency change, 50-cents change, 350-cents change) were created, which included the stimuli heard during both tasks (maximum 244 each). Three corresponding averages including only those trials completed during the detection task, and three corresponding averages including only trials completed during the identification task (maximum 122 each) were created. Like those in Foxton et al.'s study (2009), these averages all included epochs containing both correct and incorrect responses, and tone pairs containing upward and downward frequency changes. The average epochs were baseline-corrected to the 100-ms pre-stimulus interval and were low-pass filtered at 20 Hz (zero-phase Butterworth, 12 dB per octave).

8.2.4.3 Source analysis

Spatio-temporal dipoles were fitted to the average epochs in BESA. The method used here was very similar to the method used in many previous MEG studies of frequency and pitch perception (e.g., Gutschalk et al., 2004; Krumbholz et al., 2003; Seither-Preisler et al., 2006a). The analysis involved fitting two dipoles simultaneously to a time window encapsulating an AEF component within an average epoch (see Section 8.3.2 for which AEFs and which averages were used), and determining their positions and orientations using a spherical-conductor model of the head and an iterative least-squares algorithm with no constraints regarding symmetry of the two dipoles (Sherg, 1990). Fitting was done for each listener's data set separately, using their scalp information determined by the Polhemus system to position the sphere, and using their three average artifact topographies as additional sources (Berg & Sherg, 1994).

Once the positions and orientations of the dipoles were established for a given AEF and listener, these parameters were held constant, and left- and right-hemisphere dipole-moment waveforms were reconstructed for each of the 10 average epochs per listener. The amplitudes and latencies of the peaks within the source waveforms were used as dependent variables in later analyses of the data.

The source analysis used in Experiment 7 was different in two ways from the one used by Foxton et al. (2009). The first difference was that those authors used a regional-source montage; that approach is less common in the literature, and simply involves averaging together two dipoles with the same spatial location and orthogonal orientations. The source-space data from Experiment 7 appeared to be almost identical when the dipoles were converted to regional sources, so the traditional dipole approach was adopted here instead. The second difference was that Foxton et al. kept the regional sources in the same positions in each of their listener's data, rather than creating a specific source solution for each individual. I saw no reason to adopt this less sensitive method.

8.2.4.4 Anatomical images

T1-weighted, whole-head, magnetic resonance images were recorded from each listener using a 3-Tesla Signa Excite HDx scanner (GE Healthcare; flip angle = 20°, TE = 3.07 ms, TR = 8.03 ms, FOV = 290 × 290 × 176, matrix size = 256 × 256 × 176, voxel size = 1.13 × 1.13 × 1.0 mm). These images were recorded for later anatomical co-registration with the MEG data. However, the main of objective of Experiment 7 was to compare the relative contributions of the two hemispheres to listeners AEFs rather than to localise their generators precisely. The results (see below) were such that, for the present purposes, considerably greater insight was unlikely to have been gained by co-registration.

8.3 Results

8.3.1 Behvioural performance and reaction times

Listeners' behavioural data were analysed using signal detection theory. In the detection task, if a tone pair containing a 50-cents or a 350-cents change was followed by a response, it was labelled as a hit. If a response followed a tone pair without a change, it was labelled as a false alarm. Tone pairs without subsequent responses were labelled as misses and correct rejections, and values of d' were calculated seperately for 50-cents and 350-cents changes using the yes/no equation [$d' = z(H) - z(F)$]. In the identification task, tone pairs without a change and tone pairs after which listeners did not respond were excluded from the analysis. Upward changes were treated as targets (i.e., an 'up' response following an upward change was labelled as a hit, etc.) and d' was calculated using the 2AFC equation [$d' = 1/\sqrt{2} \cdot \{z(H) - z(F)\}$]. Thus, d' in the detection task provides a measure of a listener's ability to detect a frequency change, whereas d' in the identification task provides a measure of their ability to identify pitch-change direction. This analysis was performed seperately for the data from the first session (without MEG) and for the data from second session (with MEG).

Figure 8.2 shows the results of the analysis. The data were entered into a mixed-measures ANOVA with session (first, second), task (detection, identification) and stimulus type (50-cents change, 350-cents change) as within-subjects factors, and with listener group (unimpaired, direction-impaired) as a between-subjects factor. The main effect of session was not significant [$F(1,21) = 2.80$, $p = .11$, $\eta^2 = .12$], there was a significant interaction between session and task [$F(1,21) = 7.58$, $p < .05$, $\eta^2 = .27$], and session did not interact significantly with any other factor. These results indicate that overall performance during the detection task was slightly better and performance during the identification task was slightly worse in the first session than the second, but overall performance was extremely similar in the two sessions (there was also a very high test-retest reliability between the sessions; $r = .93$). There were significant main effects of task [$F(1,21) = 36.24,\ p < .001,\ \eta^2 = .63$], stimulus [$F(1,21) = 81.10,\ p < .001,\ \eta^2 = .79$], and group [$F(1,21) = 43.89,\ p < .001,\ \eta^2 = .63$]; and significant interactions between task and group [$F(1,21) = 5.14,\ p < .05,\ \eta^2 = .20$], and stimulus and group [$F(1,21) = 5.22,\ p < .05,\ \eta^2 = .20$]. The interactions indicate that although the direction-impaired listeners performed more poorly than the unimpaired listeners overall, the group difference was more pronounced in the identification task than the detection task, and more pronounced for smaller frequency changes than for larger changes. The remaining interactions were not significant.

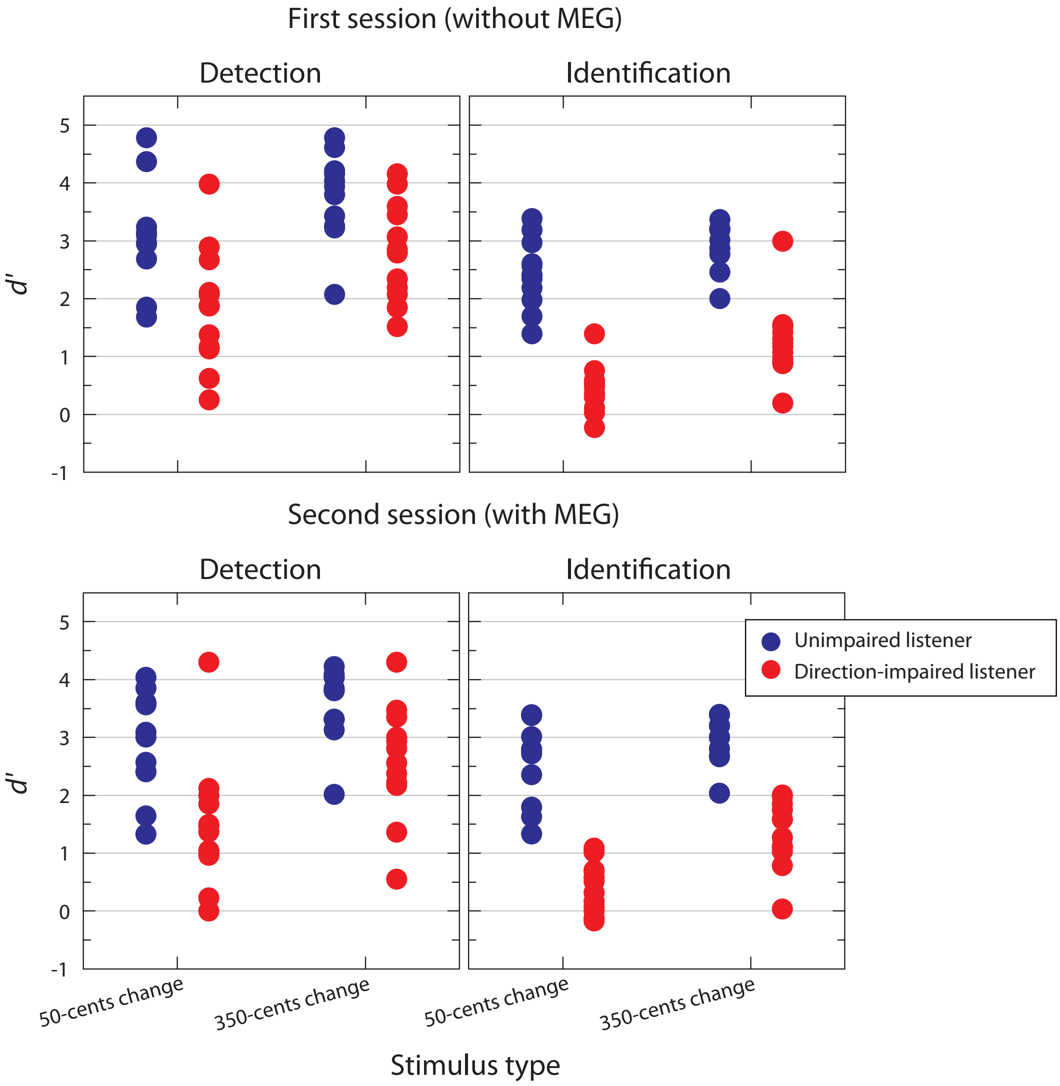


Figure 8.2. Behavioural performance in Experiment 7. In each panel, listeners' d' scores (see text) are represented by symbol ordinates. The upper panels show the data from the first session and the lower panels show the data from the second session.

Listeners' reaction times (RTs) were not recorded during the first session of Experiment 7. However, RTs from the second session were determined from the MEG data sets with a high degree of accuracy (Figure 8.3. Listeners' mean RTs—measured from the onset of the first tone in a trial—were entered into a mixed-measures ANOVA with stimulus type (50-cents change, 350-cents change) and task (detection, identification) as within-subjects factors, and group (unimpaired, direction-impaired) as a between-subjects factor. All of the main effects were statistically significant ($F \geq 5.47$, $p < .05$, $\eta^2 \geq .21$), indicating that the listeners were faster to respond to trials containing a larger frequency change, faster at detection than identification, and that the unimpaired listeners were faster than the direction-impaired listeners. None of the interactions was significant.

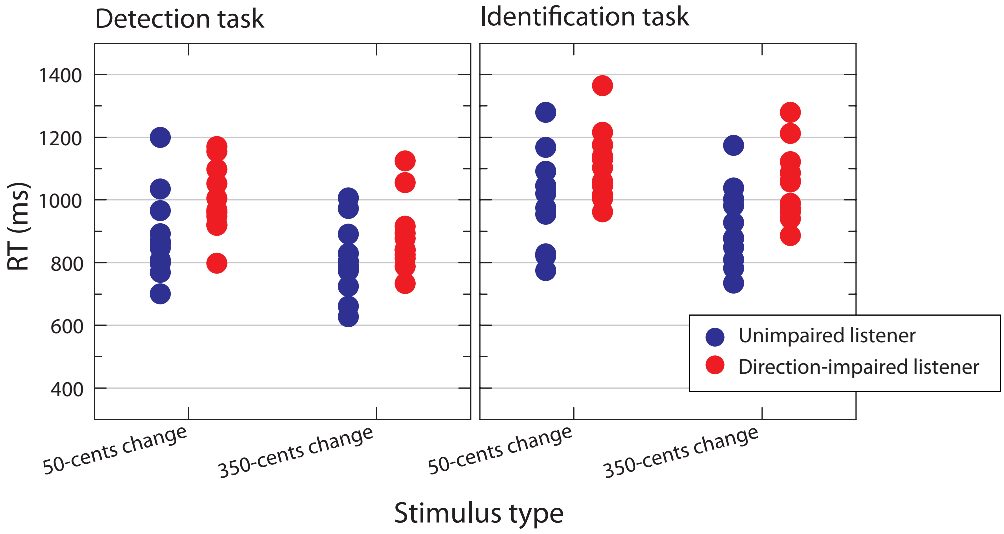


Figure 8.3. Same as Figure 8.2, but for the RT data from the second session of Experiment 7. For the identification task, these means contain both correct and incorrect responses.

8.3.2 AEFs

8.3.2.1 Overview

Figure 8.4 summarises the key features of the sensor-level MEG data from Experiment 7. The lower part of the figure shows the timings of three important events within the epochs: the onsets of the first and second tones—which always occurred at 0 and 300 ms, respectively—and listeners' overall mean RT when the pairs contained a frequency change (971.3 ms). The main part of the figure is the grand-average MEG signal. The trace represents the time course of the root-mean-square (RMS) magnetic field strength of all 245 sensors, collapsed across all stimulus types, both tasks, and all 23 listeners. This waveform contained five clear peaks, three of which occurred after the onset of the first tone but before the onset of the second tone, and two of which occurred after the onset of the second tone but before the mean RT. The most prominent of the peaks reached its maximum amplitude at 103.2 ms, and the scalp topography at this point suggests that the source was bilateral and emanated from the temporal lobes. The component is a typical M100 AEF, evoked by the first tone in each pair. It is referred to hereafter as the first M100.

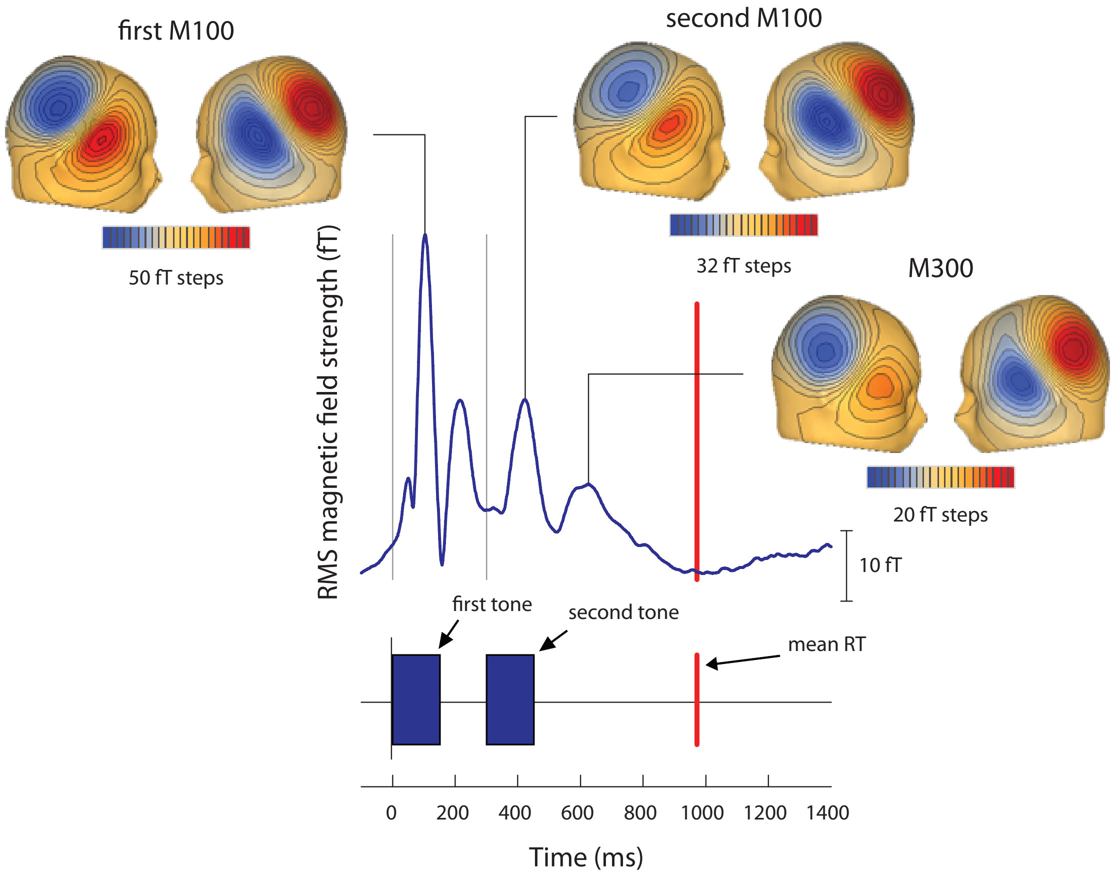


Figure 8.4. The grand-average MEG signal from Experiment 7. The trace shows the time course of the RMS magnetic field strength, below which is a schematic indicating the timings of the important events within the epoch. The diagrams surrounding the trace are the topographies of the magnetic field transposed onto a standard scalp shape at time points corresponding to the first M100, the second M100, and the M300.

In their study, Foxton et al. (2009) also observed a typical M100 after the onset of their frequency-glide stimuli. All of their stimuli began with a 100-ms plateau, so their listeners' M100s should not have been influenced by the perception of the task-relevant frequency glide within each stimulus, or by the neural mechanisms associated with identifying the direction of the glide. The authors used the M100 as a control AEF, and reported that its amplitude did not differ significantly between their two listener groups or between the source hemispheres. Similarly, the first M100 in Experiment 7 occurred before the onset of the second tone and therefore before the onset of the task-relevant frequency change[1]. The M100 was analysed as a control AEF here also.

---

[1] Since Experiment 7 used frequency roving, the frequency of the first tone was always unpredictable and different to the frequency of the last tone listeners' heard. Thus, the first M100 might contain some contribution from the neural generators involved in perceiving those frequency changes, but

The waveform in Figure 8.4 contained a peak that reached its maximum amplitude at 123.2 ms after the onset of the second tone. Again, the scalp topography at this point suggests that the source was bilateral and emanated from the temporal lobes. The AEF probably shared some of its generators with the first M100, but may also have included contributions from frequency- or pitch-change detection mechanisms when the tone pairs contained either a 50- or a 350-cents frequency change. For convenience, this AEF is referred to as the second M100.

The final peak in the waveform reached its maximum at approximately 625.3 ms (325.3 ms after the onset of the second tone). The AEF was broader and less prominent than those occurring earlier in the waveform, and strongly resembled a peak with approximately the same latency and scalp topography in the waveforms reported by Foxton et al. (2009; reproduced in Figure 2.26). The authors interpreted the activity around this time as reflecting the neural processes involved in perceiving and determining the direction of frequency glides. Based on their results, one would expect the amplitude of this AEF to be larger in the left-hemisphere dipoles of the direction-impaired listeners. It is referred to as the M300.

It might be expected that an asymmetry in the contributions of the two hemispheres to the MEG signals would be visible at the sensor level, before the application of complex source models. As a preliminary step, two grand averages were created that were split by listener group but contained all of the uncontaminated epochs from the experiment (i.e., collapsed across all stimulus types and both tasks). For both averages, the time course of the RMS magnetic field strength of 50 sensors over the left temporal area and the RMS of 49 sensors over the right temporal area were calculated. These traces are shown in Figure 8.5. Inspection of the figure suggests that the MEG signals were generally stronger in the right-hemisphere sensors than in the left-hemisphere sensors, and that this difference was more pronounced in the direction-impaired listeners. This result is surprising because it is the opposite to what was reported by Foxton et al. (2009). However, since they do not take into account the position of a listener's head inside the scanner, sensor-space data of this kind can only provide a very rough indication of where the neural activity originated.

The three specific AEFs mentioned above—the first M100, the second M100, and the M300—were identified in each listeners' averages, and for each AEF the contributions of the left and right hemispheres were delineated using the dipole approach described in Setion 8.2.4.3. The effects of stimulus type, task, and listener group on the amplitudes and latencies of these AEFs (in listeners' source waveforms) are investigated in the remainder of this section. Note that there are two more AEFs visible in Figure 8.5: an M50 and an M200. These AEFs are not investigated further because I did not have any specific predictions about their amplitudes and latencies.

---

this contribution did not differ systematically between the different stimulus types or tasks. The same was also true of Foxton et al.'s (2009) M100.

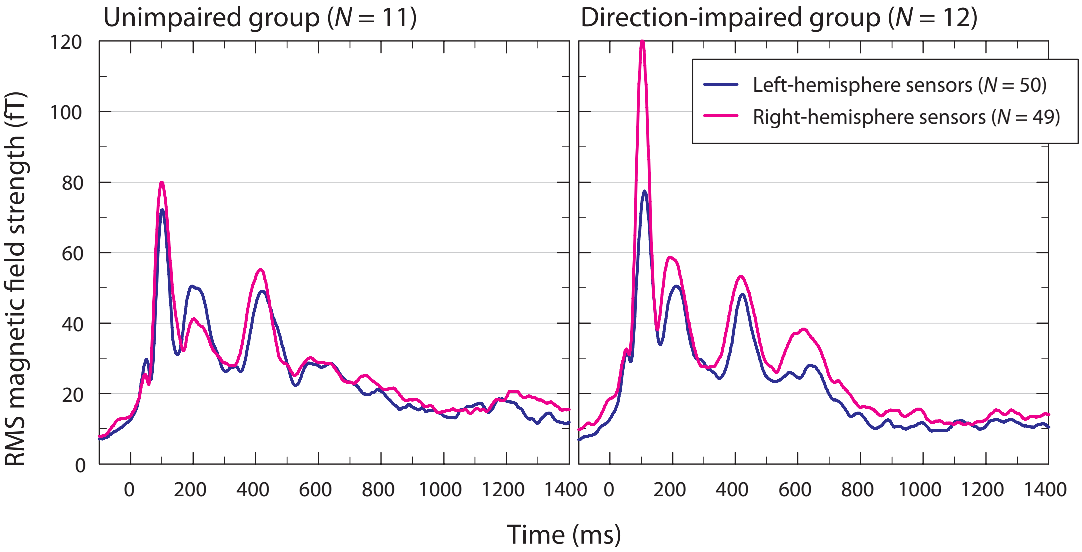


Figure 8.5. Sensor-space averages from Experiment 7. The traces in both panels show the time courses of the group mean RMS magnetic field strengths of the left- and right-hemisphere sensors.

8.3.2.2 The first M100

Dipoles were fitted to a time window of approximately 20 ms encapsulating the first M100 in each listeners' data. Fitting was done using the average AEF that included all uncontaminated epochs since the first M100 had the highest signal-to-noise ratio in this average. Because the source models were not co-registered with listeners' specific anatomical brain images, it is not possible to pinpoint the precise anatomical locations of the dipoles, although both were always situated within the temporal lobes in standard Talairach space.

Figure 8.6 illustrates the influences of hemisphere, stimulus type, task, and listener group on the mean dipole waveforms over the range 0–200 ms. The waveforms of the different stimuli and tasks were all very similar over this period; this was expected because prior to 300 ms there were no systematic differences between any of the conditions in the experiment. For each individual, the dipole moments corresponding to the peak of the M100 were identified in their specific source waveforms (defined as the minimum value within the range 60–150 ms, each checked by visual inspection) and the data were entered into a mixed-measures ANOVA with dipole (left hemisphere, right hemisphere), stimulus type (no change, 50-cents change, 350-cents change), and task (detection, identification) as within-subjects factors, and with listener group (unimpaired group, direction-impaired group) as a between-subjects factor. The ANOVA revealed that the interaction between dipole and group was close to the level of two-tailed statistical significance [$F(1,21) = 3.54$, $p = .07$, $\eta^2 = .14$]; this feature of the data is relevant to the discussion later on in the chapter. None of the main effects or the remaining interactions was significant ($F \leq 2.30$, $p \geq .12$, $\eta \leq .10$).

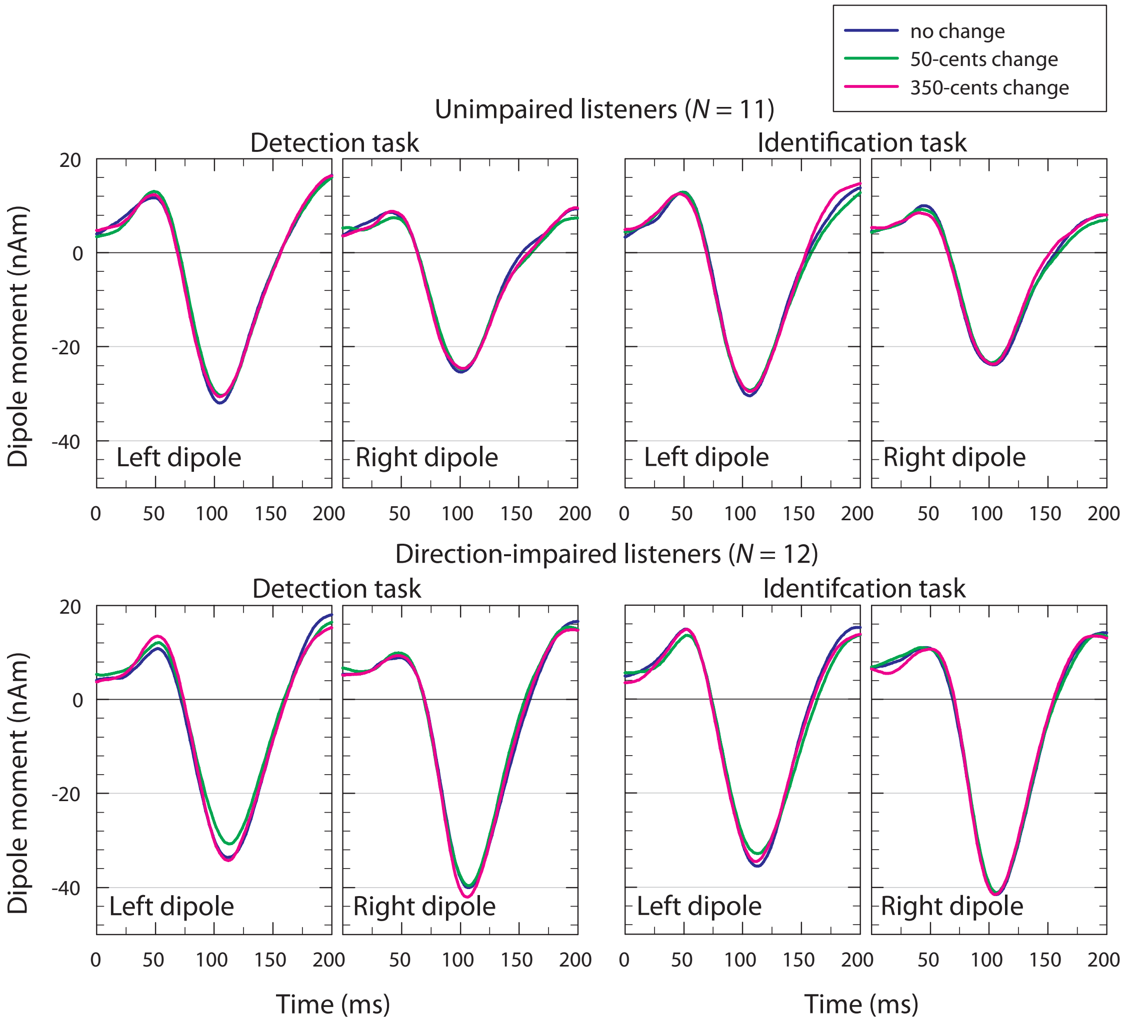


Figure 8.6. Mean dipole waveforms over 0–200 ms.

The latencies of the M100 peaks (defined as the sample number at which the minimum value was observed) were entered into an ANOVA with the same design as the previous one. The ANOVA revealed a significant main effect of dipole [$F(1,21) = 12.10$, $p < .01$, $\eta^2 = .37$], indicating that the M100 tended peak reliably earlier (approximately 4 ms earlier on average) in the right-hemisphere dipole than in the left-hemisphere dipole. None of the other main effects or interactions was statistically significant ($F \leq 2.16$, $p \geq .13$, $\eta \leq .93$).

8.3.2.3 The second M100

Dipoles were re-fitted to a time window of approximately 20 ms encapsulating the second M100 in each listeners' data. Fitting was done using the average AEF that included the epochs containing a 350-cents frequency change (both tasks), since the AEF had the highest signal-to-noise ratio in this average[2]. Mean dipole waveforms over the range 300–500 ms are shown in Figure 8.7.

The dipole moments corresponding to the peak of the second M100 were identified in the listeners' waveforms (minimum value within the range 300–480 ms, each checked by visual inspection) and were entered into an ANOVA with the same design as the two previous ANOVAs. There was a significant main effect of stimulus type [degrees of freedom Greenhouse-Geisser corrected; $F(1.12,23.52) = 31.61$, $p < .001$, $\eta^2 = .60$]. The main effect indicates that the amplitude of the second M100 was influenced by the frequency relationship between the tones within each pair. Planned comparisons revealed that the M100 was larger when there was a 350-cents change than when there was no change [$F(1,21) = 28.96$, $p < .001$, $\eta^2 = .58$]. This result is consistent with many previous studies that have found that the amplitude of the M100 [and its equivalent from electroencephalography (EEG), the N1] is smaller in response to a tone preceded by another of the same frequency than in response to a tone preceded by another of a different frequency (e.g., Budd, Barry, Gordon, Rennie, & Michie, 1998; Rosburg, 2004), and that frequency changes can evoke new M100s/N1s (e.g., Lavikainen, Huotilainen, Ilmoniemi, Simola, & Näätänen, 1995; Yamashiro, Inui, Otsuru, & Kakigi, 2010). However, the M100 was not significantly larger when there was a 50-cents change than when there was no change [$F(1,21) = 0.27$, $p = .61$, $\eta^2 = .01$], perhaps because the present experiment lacked the sensitivity to detect this weaker modulation of M100 amplitude. Stimulus type did not interact significantly with any other factor in the ANVOA ($F \leq 2.33$, $p \geq .11$, $\eta^2 \leq .10$).

The ANOVA also revealed a significant main effect of task [$F(1,21) = 11.32$, $p < .01$, $\eta^2 = 0.35$], indicating that the second M100 was generally larger during the detection task than the identification task. Although the interaction between task and stimulus

---

[2] Using the average with the highest signal-to-noise ratio is standard practice (e.g., Krumbholz et al., 2003). To determine whether this had any effect on the results, the dipoles were re-fitted using the average of all the uncontaminated epochs, and the ANOVA of the second M100 amplitude was re-run. The second ANOVA yielded very similar results to the first.

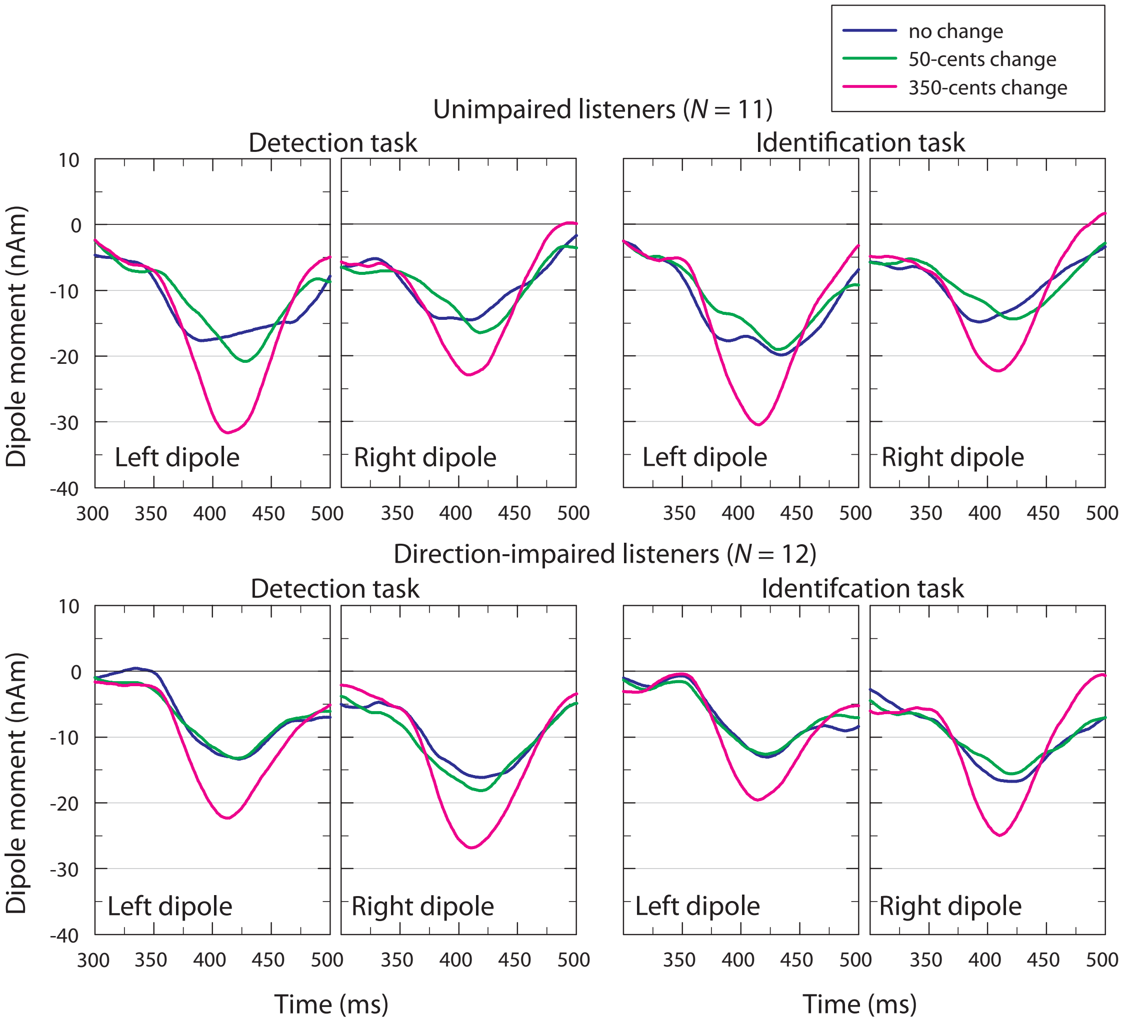


Figure 8.7. Mean dipole waveforms over 300–500 ms.

type was not statistically significant [$F(2,42) = 2.33$, $p = .11$, $\eta^2 = .10$], post-hoc paired-samples t-tests were used nevertheless to compare the amplitudes of the second M100 in the detection task to those in the identification task separately for each dipole and for each stimulus type. The tests revealed that the AEFs evoked by 350-cents changes were reliably larger in the detection task than in the identification task, in both dipoles [left: $t(22) = -3.18$, $p < .008$, $r = -.56$; right: $t(22) = -3.22$, $p < .008$, $r = -.57$]. The difference was also significant for the 50-cents changes, but only in the right-hemisphere dipole [left: $t(22) = -1.64$, $p = .114$, $r = -.33$; right: $t(22) = -3.30$, $p < .008$, $r = -.58$]. The AEFs evoked by tone pairs without frequency changes were not significantly different between the tasks in either dipole [left: $t(22) = -0.61$, $p = .55$, $r = -.13$; right: $t(22) = -0.39$, $p = .70$, $r = -.08$].

The task-related modulation of the second M100 could reflect differences in listeners' state of attention during the detection and identification tasks. Previous studies have found that the M100/N1 to pure tones is enhanced by selective attention, such that its amplitude is increased when listeners perform a task that requires them to attend to the stimuli compared to passive listening to the same stimuli (e.g., Hillyard, Hink, Schwent, & Picton, 1973), although in many circumstances this effect may be the product of an enhancement of other components overlapping with the M100 in time (see Alho, 1992). In an EEG study, Mulert et al. (2007) presented listeners with pure tones with frequencies of 800, 1000, or 1200 Hz, in six runs of trials. The listeners identified the frequencies of the tones with button presses, but the stimuli presented and the precise task performed differed from run to run. The N1 differed in amplitude between the runs, which was interpreted by the authors as indicating that the N1 is enhanced by increasing task difficulty and mental demands[3]. Given that the identification task probably required greater attention and effort than the detection task, one might predict that the M100 would be larger during the former than the latter. However, the M100 was actually smaller when listeners performed the identification task in Experiment 7. The interaction between task and dipole—which would have implicated a greater involvement of either the left- or right-hemisphere auditory cortex in the identification task—was not significant [$F(1,21) = 0.00$, $p = .998$, $\eta^2 = .00$].

There was a statistically significant interaction between dipole and listener group on the amplitude of the second M100 [$F(1,21) = 5.57$, $p < .05$, $\eta^2 = .21$]. The bottom left-hand panel of Figure 8.8 illustrates the interaction: the second M100 tended to be larger on average in the left-hemisphere dipole in the unimpaired group, but larger on average in the right-hemisphere dipole in the direction-impaired group. This interaction is indicative of a *rightward* asymmetry of processing (i.e., a stronger AEF in the right hemisphere than in the left hemisphere) in the brains of direction-impaired

---

[3] The authors' interpretation of their results is complicated by the fact that the stimuli presented were different in the different runs, and because the pattern of the N1 amplitudes in the different runs did not match precisely their ranking of task difficulty.

listeners. The bottom-right-hand panel of the figure shows listeners' individual 'asymmetry scores'—calculated by subtracting the peak amplitude of the AEF in their right dipole from the peak amplitude in their left dipole—for each stimulus type and each task. Although the interaction was significant, the figure shows that the individual differences in these scores were considerable, and that there was an extensive overlap between the groups in the distribution of the scores. The three-way interaction between dipole, stimulus type, and group was not significant [bottom-right-hand panel of Figure 8.8; $F(2,42) = 0.48$, $p = .65$, $\eta^2 = .06$]. In other words, the rightward asymmetry in the direction-impaired group was just as apparent when there was no frequency change as it was when the stimuli contained a frequency change. For comparison, the top panels of the figure show the interaction and the asymmetry scores for the first M100. The general pattern is very similar, in that the amplitude of the first M100 also tended to be larger in the right-hemisphere dipole in the direction-impaired listeners.

The interaction reported above is inconsistent with the results of Foxton et al.'s (2009) study. Those authors conducted a similar analysis of the M100 evoked by the onset of their glide stimuli, and found—as was found here—that the main effects of dipole and listener group were not statistically significant. However, the authors did not report the interaction term of the ANOVA, and visual inspection of their mean dipole waveforms (Figure 3, p. 1308; reproduced in Figure 2.26) suggests that even if it was not significant, there was clearly a trend for a leftward asymmetry in the amplitude of the M100 in the direction-impaired listeners. This interaction was in the same direction as the significant interaction they observed later on in listeners' AEFs, but was in the opposite direction to the interaction observed here. This issue is discussed in detail later.

The latencies of the second M100 peaks were entered into an ANOVA with the same design as above, which revealed a significant main effect of dipole [$F(1,21) = 8.00$, $p < .05$, $\eta^2 = .28$]. As with the first M100, the peak of the second M100 occurred earlier on average in the right-hemisphere dipole. There was also a significant main effect of stimulus type [$F(2,42) = 4.39$, $p < .05$, $\eta^2 = .17$]. Planned comparisons revealed that the second M100 occurred later when there was a 50-cents frequency change than when there was no change [$F(1,21) = 10.41$, $p < .01$, $\eta^2 = .33$], but did not occur later when there was a 350-cents frequency change than when there was no change [$F(1,21) = 0.30$, $p < .86$, $\eta^2 = .00$]. Why this effect occurred is not clear. None of the other main effects or interactions was significant ($F \leq 3.20$, $p \geq .09$, $\eta^2 \leq .13$).

8.3.2.4 The M300

Dipoles were re-fitted to a time window of approximately 40 ms encapsulating the M300 in each listeners' data. Fitting was done using the average that included the epochs containing a 350-cents frequency change (both tasks), since the response

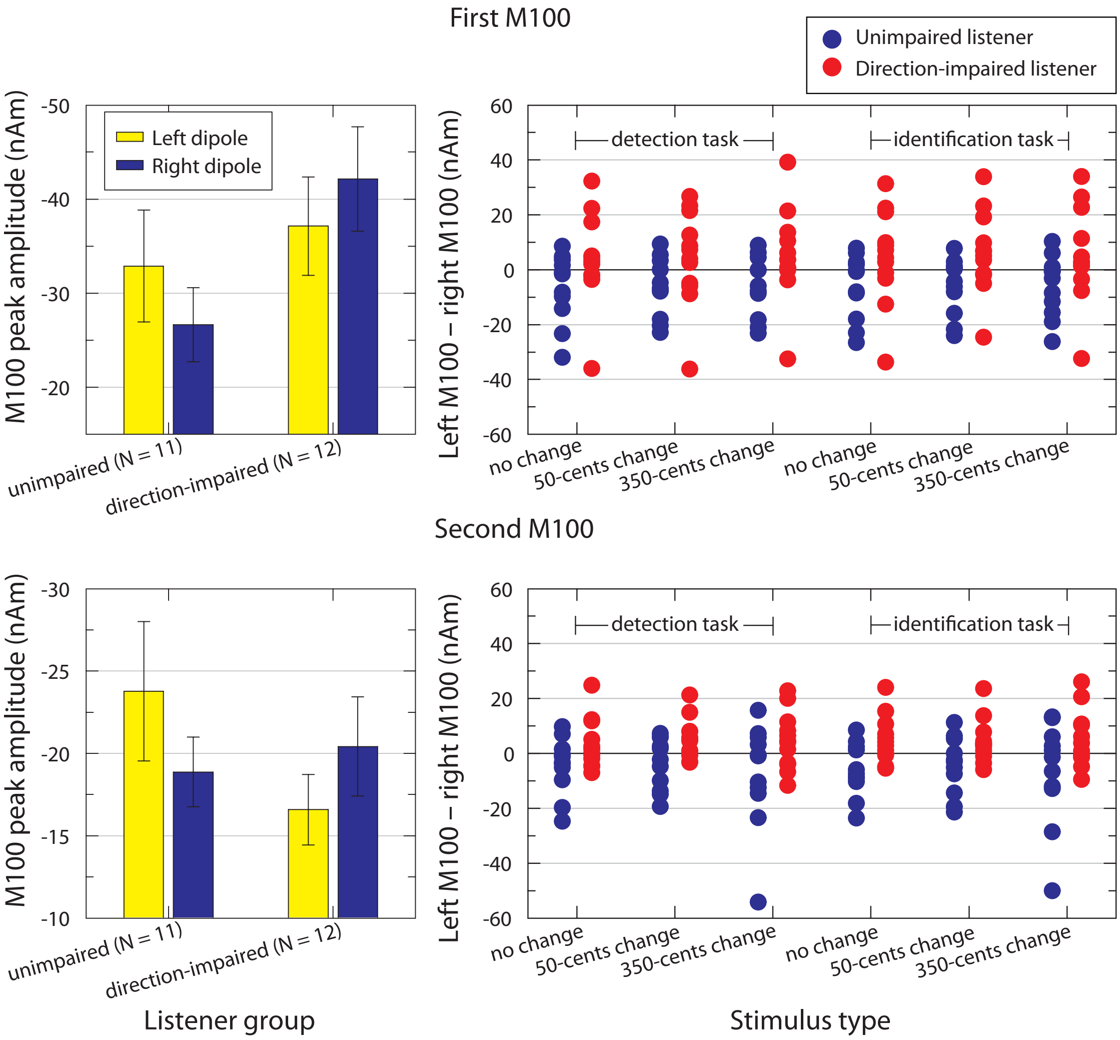


Figure 8.8. Interaction between dipole and group on M100 amplitude. The top- and bottom-left-hand panels show the group mean M100 amplitudes collapsed across stimulus type and task (with standard errors) on inverted axes. The right-hand panels show listeners' individual 'asymmetry scores', calculated by subtracting their right M100 peak amplitude from their left M100 peak amplitude. A positive asymmetry score indicates a larger right-hemisphere M100.

had the highest signal-to-noise ratio in this average. Mean source waveforms over the range 520–720 ms are shown in Figure 8.9. Visual inspection of the individual waveforms revealed that the M300 was much broader than the previous AEFs, and sometimes contained multiple peaks; it was felt that treating the minima of the M300 as a dependent variable might produce erratic results. Therefore, for each listener and for each dipole waveform, the mean dipole moment over the time range 550–700 ms was recorded and entered into an ANOVA with the same design as those reported above.

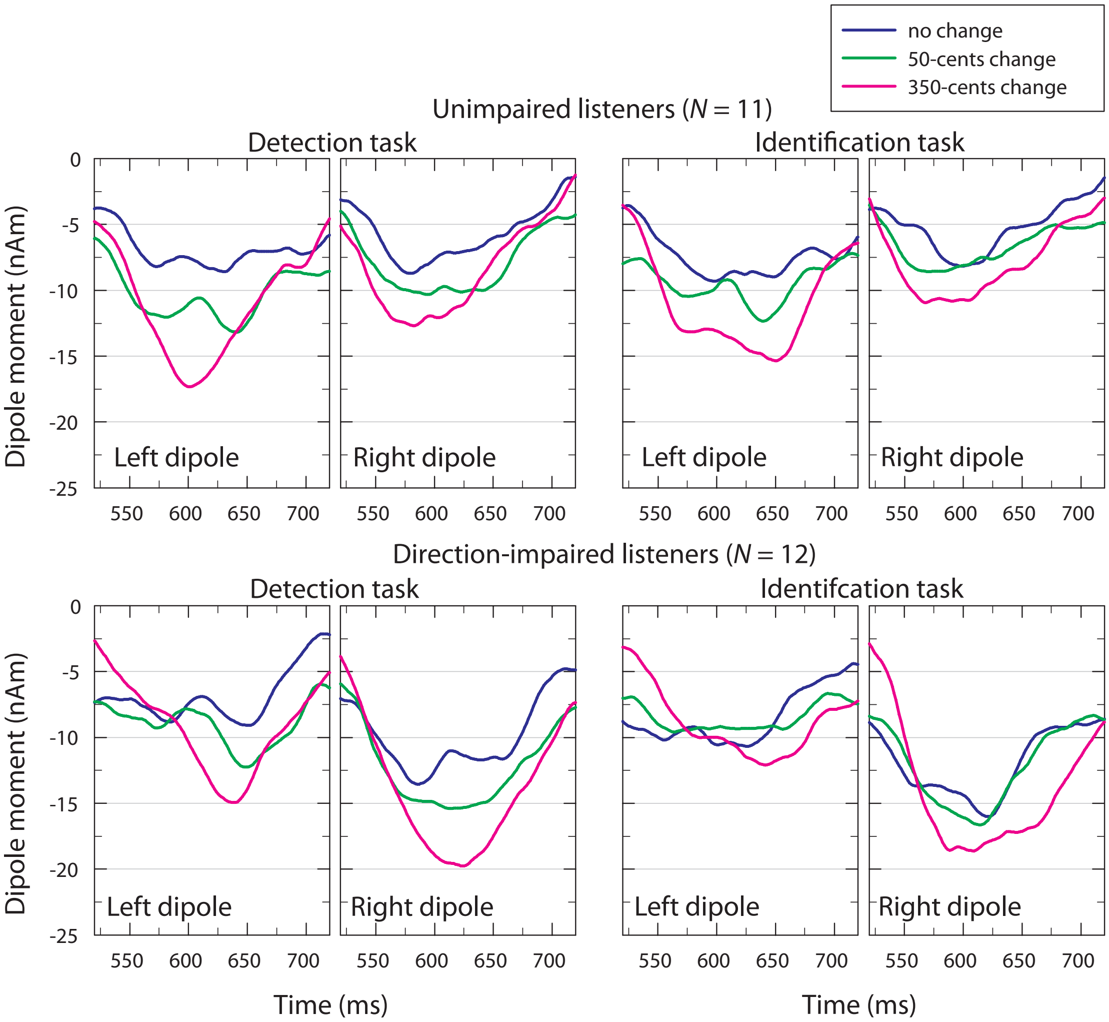


Figure 8.9. Mean dipole waveforms over 520–720 ms.

There was a significant main effect of stimulus type on the amplitude of the M300 [degrees of freedom Greenhouse-Geisser corrected; $F(1.19, 24.90) = 15.04$, $p < .001$, $\eta^2 = .42$]. Planned comparisons revealed that the M300 was larger when the stimuli contained either a 50-cents frequency change or a 350-cents change than when there was no change ($F \geq 15.36$, $p < .001$, $\eta^2 \geq .44$). The main effect of task was not significant [$F(1,21) = 0.21$, $p = .66$, $\eta^2 = .01$], but the interaction between stimulus type and task was close to significance [$F(2,42) = 2.83$, $p = .07$, $\eta^2 = .12$]. The interaction suggests a trend for the differences between the M300s evoked by the different stimulus types to be less pronounced during the identification task.

As mentioned in Section 8.3.2.1, the M300 was very similar to a response in the AEFs measured by Foxton et al. (2009). There was a significant interaction between dipole and group, and a significant three-way interaction between stimulus type, dipole, and group on the mean amplitude of the response in their experiment. In the present experiment, there was a significant interaction between dipole and listener group on the amplitude of the M300 [$F(1,21) = 6.42$, $p < .05$, $\eta^2 = .23$] and a significant three-way interaction between stimulus type, dipole, and group [$F(2,42) = 6.40$, $p < .01$, $\eta^2 = .23$]. However, as illustrated in Figure 8.10, both terms indicated that the M300 in the direction-impaired listeners was larger in the right-hemisphere dipole than in the left-hemisphere dipole, the reverse of what was reported by Foxton et al.. The four-way interaction between these factors and task was not significant $F(2,42) = .12$, $p = .90$, $\eta^2 = .05$].

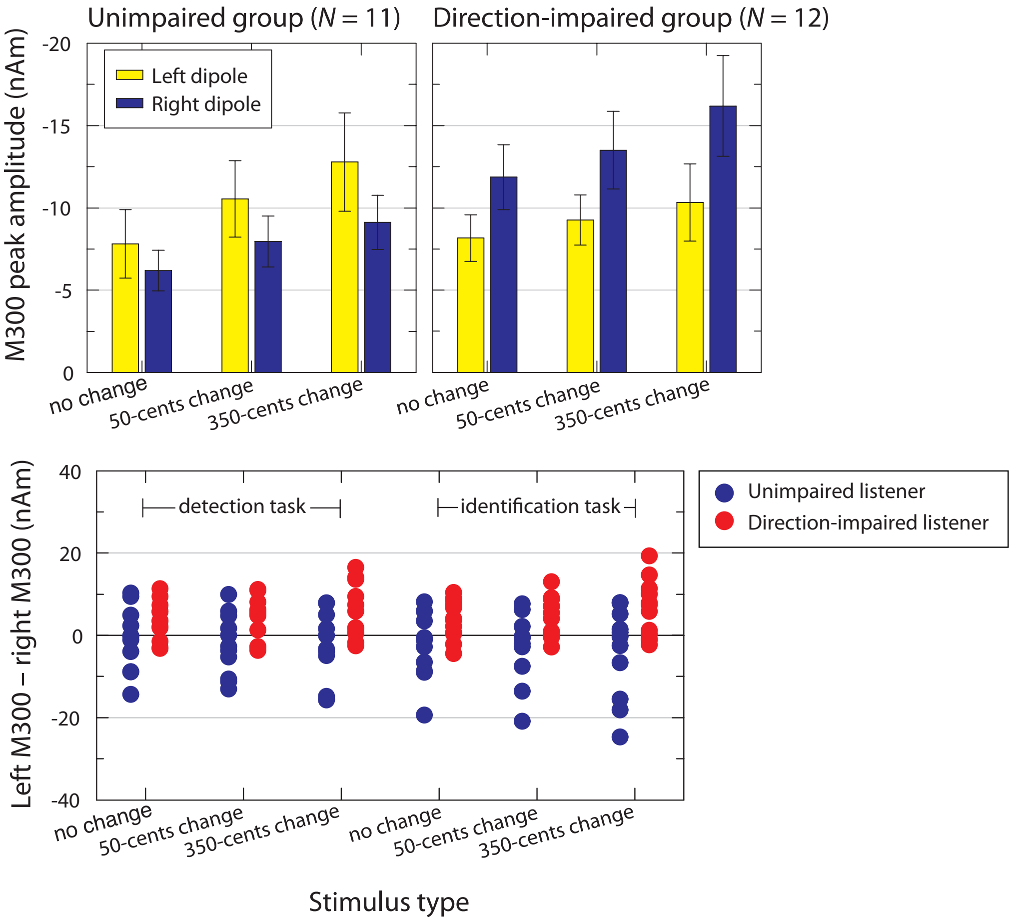


Figure 8.10. Interactions between dipole and group, and between stimulus type, dipole, and group on M300 amplitude. The top panels show the group mean M300 amplitudes collapsed across task (with standard errors) on inverted axes. The bottom panel shows listeners' individual asymmetry scores (see Figure 8.8). A positive asymmetry score indicates a larger right-hemisphere M300.

8.3.2.5 Correlations with performance

The significant interactions in the ANOVAs reported above imply that the extent to which a listener's AEFs are larger in one hemisphere is related to their ability to identify the direction of a pitch change. If this interpretation is correct, listeners' asymmetry scores should correlate with their performance in the experiment. Listeners' d' scores over all identification trials in Experiment 7 were calculated. This calculation included tone pairs containing both 50-cents and 350-cents frequency changes, and trials from both sessions of the experiment. Correlations were used to determine whether listeners' overall identification d' scores were related to their asymmetry scores (from the grand-average MEG signals, as above). Listeners' identification d' scores for 50-cents changes only and for 350-cents changes only were also calculated, and these were correlated with their asymmetry scores derived from their MEG averages for 50-cents and 350-cents identification trials, respectively. The results are shown in Table 8.1. All of the correlations were negative, indicating that rightward asymmetries were associated with poorer pitch-direction identification, and most were statistically significant at least at the uncorrected level. The correlations were stronger for 50-cents pitch changes than for 350-cents changes, and strongest for the M300 AEF. These results are consistent with the ANOVAs reported in the previous sections.

Table 8.1. Correlations (Pearson's r) between between listeners' identification d' and their asymmetry scores ($N = 23$). See text for how the scores were calculated in each case.

Asymmetry scores	Identification d'		
	Overall	50 cents	350 cents
Grand average			
First M100	−.402		
Second M100	−.450*		
M300	−.536**		
Ident. trials only (50 cents)			
First M100		−.432*	
Second M100		−.500*	
M300		−.637†	
Ident. trials only (350 cents)			
First M100			−.461*
Second M100			−.344
M300			−.414*

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{\dagger}p_{\mathrm{bf}} < 0.006$

8.4 Discussion

In Experiment 7, AEFs in response to pairs of temporally discrete pure tones were measured with MEG in 11 unimpaired and 12 direction-impaired listeners. The tones within the pairs were either different in frequency by 50 or 350 cents, or had the same frequency. In one run of trials, listeners responded whenever they heard a frequency change (the detection task), and in the other run of trials, listeners responded with its direction whenever a change was heard (the identification task). Average source waveforms were extracted from the MEG signals using dipoles fitted in listeners' left and right hemispheres, and three prominent responses were identified and analysed further: an M100 evoked by the first tone in each pair, an M100 evoked by the second tone, and an M300 occurring after the second tone. Of particular interest was the relationship between the amplitudes of these AEFs in both hemispheres. The main findings of the experiment are as follows:

1) The amplitude of the first M100 did not differ significantly between the different stimuli or the two tasks in the experiment. The difference in M100 amplitude between the two groups of listeners was also not significant. However, it may be relevant that there was a non-significant trend for the right-hemisphere M100 to be larger than the left-hemisphere M100 in the direction-impaired group, and vice versa in the unimpaired group.

2) The amplitude of the second M100 was largest for pairs containing a 350-cents frequency change, and was significantly larger on average during the detection task than during the identification task. The trend for the response to be larger in the right-hemisphere dipole than the left-hemisphere dipole in the direction-impaired listeners was observed again, and this time the interaction between dipole and group was significant.

3) The mean amplitude of the M300 was smallest for pairs without a frequency change, larger for pairs containing a 50-cents change, and largest for pairs containing a 350-cents change. Again, there was a significant interaction between dipole and group, and there was a three-way interaction between dipole, stimulus type, and group.

4) The relative amplitude differences of the AEFs between the left and right dipoles correlated with listener performance during the identification task, with smaller d' being associated with more rightward laterality.

The main findings of Experiment 7 and those reported by Foxton et al. (2009) are precisely contradictory. What could account for this discrepancy? It cannot be the case that Experiment 7 lacked sufficient power to replicate their results, because it was actually the stronger of the two studies. Experiment 7 included more listeners (23 versus 14), and more stimuli were presented during the MEG sessions (732 versus 320). One major difference between the two studies is that the present one used pairs of tones as stimuli rather than frequency glides. Although it is difficult to imagine that this difference lead to the discrepancy in the results, it cannot be ruled out without further investigation.

Another more substantial difference between the studies is that in the present investigation the direction-impaired listeners were tested extensively before their AEFs were recorded. As mentioned above, Foxton et al.'s (2009) direction-impaired listeners did not have much experience with frequency or pitch discrimination before taking part in the first experiment, and with more practice the large performance difference between their groups might have diminished. Thus, it can be claimed with more confidence that the two groups in Experiment 7 truly differed in terms of their underlying sensitivity to pitch-change direction.

An alternative explanation for the contradictory results is that—rather than either experiment revealing a genuine group difference in the specific neural mechanisms involved in the active determination of pitch-change direction—listeners simply vary enormously in the inter-hemispheric differences in the amplitudes of their MEG signals. Previous studies have commented on the considerable inter-individual variability in listeners' AEFs (e.g., Lütkenhöner, 2003; Lütkenhöner, Krumbholz, & Seither-Preisler, 2003)[4]. In particular, the report by Lütkenhöner, Krumbholz, and Seither-Preisler (2003) contains a figure that shows the M100 response to pure tones of different frequencies (250, 500, 1000, and 2000 Hz) in the left and right hemispheres of nine individuals (Figure 1, p.938). In some listeners the M100 was larger in the right hemisphere, in others it was larger in the left hemisphere, and in many cases the hemispheric difference exceeded 50 fT. The listeners in that study listened passively to the tones, so it is not known if their asymmetry correlated with their pitch-discrimination abilities. Nevertheless, simply considering the extent of the individual differences in those listeners and the ones here suggests that the 23 listeners tested in Experiment 7 and the 14 tested by Foxton et al. (2009) might be too few to determine convincingly with MEG whether poor pitch-direction identification is associated with a greater dependence on either left- or right-hemisphere auditory cortex.

Clearer results might have been obtained if Experiment 7 had employed a transition paradigm, or had used another kind of pitch-evoking stimulus such as iterated rippled noise (IRN), like the studies described in Section 2.4.2.3. It could be that direction-impaired and unimpaired listeners differ in terms of their transition AEFs, which have been suggested to reflect the neural correlates of pitch perception (e.g., Gutschalk et al., 2004; Krumbholz et al., 2003; Ritter, Dosch, Specht, & Rupp, 2005; Ritter, Dosch, Specht, Schneider, & Rupp, 2007; Seither-Preisler et al., 2006a; Seither-Preisler, Patterson, Krumbholz, Seither, & Lütkenhöner, 2006b). However, it

---

[4] Both of those studies stress that, in contrast to the high inter-individual variability, the *intra*-individual variability of listeners' AEFs is rather low. In other words, although the MEG signals measured in any particular individual can be quite consistent across experiments or MEG sessions, those measured in different individuals are likely to vary substantially.

might prove difficult to interpret the results of such an experiment because there is some debate over whether transition AEFs are related to pitch perception or to the perception of other stimulus features instead (e.g., Chait et al., 2007). An issue with using IRN is that it cannot produce pitches higher than 500 Hz, which obviously imposes a limitation on the size of the roving range that can be used in an experiment.

Experiments 1–6 of this thesis revealed a critical role of wide frequency roving in some listeners' insensitivity to pitch-change direction, and suggested that this might be because roving introduces random, irrelevant pitch changes. It is quite possible, therefore, that the neural mechanisms by which the pitch of a given sound is determined are completely normal in direction-impaired listeners, and that their difficulties lie instead at some more central stage of the decision process, such as when trying to ignore the irrelevant features of the stimulus ensemble. This suggestion is consistent with the sequential interference model described in Section 6.4.2, which would not necessarily require the standard deviation of the sensory noise to increase to account for the pitch-direction impairment. Further analysis of the data from Experiment 7 could focus on these processes. For example, tone pairs that were preceded by an irrelevant frequency change that was incongruent with the direction of the target change could be averaged separately from trials that were a preceded by a congruent change, and the group differences between the AEFs evoked by congruent and incongruent changes could be examined.

As mentioned in Section 2.4.3, Foxton et al.'s (2009) findings were broadly consistent with a popular hypothesis about hemispheric specialisations in the processing of sounds (e.g., Zatorre, 2003; Zatorre & Gandour, 2008). The hypothesis states that due to differences in their degree of myelination, the left-hemisphere auditory cortex is specialised to represent acoustic features that change rapidly over time, whereas the right-hemisphere cortex is specialised to represent acoustic features leading to the perception of pitch. A number of problems with this basic hypothesis and the application of the hypothesis to Foxton et al.'s data were described earlier in the thesis. Experiment 7 adds a further complication to their interpretation. The increased myelination of the left hemisphere should lead to the more rapid transmission of neural signals; however, in the present experiment the first and second M100s were consistently earlier in the right hemisphere, suggesting slower transmission in the left hemisphere.

Chapter 9

General discussion

9.1 Summary of the thesis

The aim of the empirical work reported in this thesis was to investigate further the origins and consequences of individual differences in pitch perception reported in adult listeners from the normal population. The work followed on from a study by Semal and Demany (2006), who found that some normally hearing listeners are insensitive to the direction of small but detectable frequency changes between pure tones (Section 2.3.3.2). The thesis investigated this phenomenon in four ways: by extending Semal and Demany's findings through psychoacoustical experimentation; by considering the data in terms of models of frequency discrimination; by investigating whether or not the individual differences were correlated with aspects of everyday hearing; and by investigating the brain activity evoked during frequency discrimination using magnetoencephalography (MEG).

Experiment 1 (Chapter 3) replicated Semal and Demany's (2006) main finding. As in their study, listeners were tested using a procedure that measured their DLFs for unsigned pitch-change detection (DDLFs) and for pitch-direction identification (IDLFs) independently whilst maintaining equivalent stimulus arrangements. In the detection task, listeners heard two pairs of pure tones on every trial, and responded by indicating which of the pairs contained tones of a different frequency. In the identification task, listeners heard the same stimuli, but instead responded with the direction of the change occurring in the pair with the different tones. Roughly half of the listeners tested in the experiment had IDLFs that were considerably larger than their DDLFs—in other words, they were relatively insensitive to pitch-change direction. Their insensitivity endured despite several hours of practice with feedback. The remaining listeners had no such difficulty, and consistent with previous results their IDLFs were roughly equal to or smaller than their DDLFs.

An important feature of the stimuli in Experiment 1 and in Semal and Demany's (2009) first experiment was that the standard frequency of the tones was roved over a wide range (400–2400.1 Hz, or 3102 cents) between trials, and between pairs within a trial. Most of the listeners from the first experiment were re-tested in Experiment 2 (Chapter 3), which had essentially the same design as Experiment 1, except that the standard frequency was always 979.8 Hz (the geometric centre of the previous roving range). Consistent with earlier studies, listeners' DLFs were smaller in Experiment 2 than in Experiment 1. The key result, however, was that the large discrepancies between IDLFs and DDLFs (reflected in large I/D ratios) were reduced, and many of those identified as direction-impaired in Experiment 1 could no longer be described as such in Experiment 2.

Experiment 3 (Chapter 4) substantiated and clarified the role of frequency roving in some listeners' insensitivity to pitch-change direction. DDLFs and IDLFs were measured in a group of expert listeners and in a new group of pre-selected novices with no prior experience in psychoacoustical experiments. The experiment measured DLFs at four different frequency-roving ranges (no roving, as in Experiment 2; and 1%, 10%, and 100% of the range used in Experiement 1). Those conditions also included level roving to control for possible cues other than pitch, and conditions were completed in random order to control for order effects. The DLFs in all listeners tended to increase monotonically with increasing frequency-roving range. However, in the novices, I/D ratios also increased with increasing frequency-roving range. The novices appeared to be direction-impaired only when a medium (310 cents) or a wide (3102 cents) range was used, and not when level was roved but standard frequency was fixed.

Experiment 4 (Chapter 4) demonstrated that frequency roving also affected listeners' abilities to identify the direction of continuous frequency modulations. The listeners heard two pure tones on each trial, one of which contained a half-cycle sinusoidal sweep in frequency, and DDLFs and IDLFs were measured as in the previous experiments. Consistent with the previous results, novice listeners' I/D ratios were larger when the starting frequency of the tones was roved over a wide range, and were smaller when the starting frequency was fixed at 979.8 Hz.

In Experiment 5 (Chapter 5), DLFs were measured again in individuals identified as direction-impaired in Experiment 1. The listeners heard two pure tones on each trial, and responded with the direction of the frequency change. In most of the conditions in the experiment, the standard frequency of the tones was fixed throughout a run of trials, and listeners completed the different conditions and runs in a prescribed order. The first condition measured DLFs with a standard that was fixed at 979.8 Hz. In the second and third conditions, the standard frequency was switched to 400 and 2400.1 Hz, respectively, and feedback was not provided after trials. The standard frequencies were 400 and 2400.1 Hz in fourth and fifth conditions, respectively, which included feedback. The main finding of Experiment 5 was that the listeners' DLFs were not reliably different in those five conditions, suggesting that they were not relying on feedback to perform pitch-direction identification in the absence of frequency roving in the earlier experiments (Section 4.4.2).

The findings of Experiments 1–5 were consistent with an explanation of the insensitivity to pitch-change direction based on sequential interference (Section 4.4.3). According to this explanation, the direction-impaired listeners were more easily confused or distracted by the irrelevant pitch changes that were introduced to the stimulus ensemble by frequency roving. These irrelevant pitch changes affected pitch-direction identification, but not pitch-change detection. Experiment 6 (Chapter 5) aimed to test the explanation further by introducing additional irrelevant pitch changes. Experts and a new group of novices heard three tones on each trial, and identified the direction of frequency change occurring between the second and third tones. The standard frequency was manipulated in the same way as in Experiment 3, and in separate conditions, the first tone either was the same frequency as the second, or had a random frequency. As in Experiment 3, DLFs increased monotonically with increasing frequency-roving range, and this effect was much more pronounced in the novices. The novices' DLFs were also much larger when the first tone was random than when it was the same as the second; this finding supported the sequential interference hypothesis. Moreover, there was a strong positive correlation between the degree to which a listener's DLF was influenced by roving, and the degree to which their DLF was influenced by the random-frequency tone. In keeping with the hypothesis, the correlation suggests that the two effects have similar origins.

An alternative explanation for the findings of Experiment 6 was that the results stemmed from a beneficial cuing effect when the first tone was the same as the second, rather than from deleterious interference when it was random. A supplementary experiment (Chapter 5) aimed to test this possibility, and the results suggested that the first tone could have either a beneficial or a deleterious effect, depending on its relationship with the standard frequency. Although this finding did not contradict the sequential interference hypothesis, it raised further questions. Does the repetition of the standard frequency provide listeners with an opportunity for multiple looks, provide a frequency cue for selective attention, or simply suppress interference effects from earlier stimulus changes? Are the DLFs of direction-impaired listeners more or less influenced by any of these factors? Thus, in summary, the results of the psychoacoustical experiments reported in this thesis mostly support the sequential interference hypothesis, but also suggest that the hypothesis requires further refinement.

The results from the experiments were compared to the predictions of several models based on signal detection theory (Chapter 6). The DLFs measured in some of the listeners were approximately consistent with the standard model proposed originally by Semal and Demany (2006), which—counter-intuitively— predicted smaller IDLFs than DDLFs. However, the model could not account for three key features of the empirical data: the considerable individual differences in listeners' I/D ratios, the general deleterious effects of frequency roving, and the interaction between roving range and task in direction-impaired listeners. Two modifications to the model were suggested. In the first (Section 6.4.1), the sensory noise limiting both DDLFs and IDLFs, which was held constant under all conditions in the original model, was varied proportionally with the degree of uncertainty in the stimuli (i.e., the size of the frequency-roving range used for a run of trials). The modification provided a good fit to the data from four experts who did not experience difficulty identifying pitch-change direction, but did not fit well to the data from direction-impaired listeners.

The second modification (Section 6.4.2) involved considering frequency or pitch discrimination as a two-stage decision process. In the first stage, listeners detect the presence of a pitch change; their sensitivity in this stage is limited by internal sensory noise. After detecting the change, listeners next identify its direction as up or down. The modified model involved adding an additional source of noise at the second stage, which affected listeners' DLFs in Experiments 5 and 6 (and the supplementary experiment), and listeners' IDLFs but not their DDLFs in Experiments 1–4. If the standard deviation of the additional noise were small, listeners' IDLFs would be smaller than or similar to their DDLFs. If the standard deviation of the noise were large, IDLFs would be elevated relative to DDLFs, as found in the direction-impaired listeners. The additional noise could be used to quantify the effects of sequential interference if one adopts the idea of a temporal filter: irrelevant pitch changes, occurring before or after the target pitch change, could be incorporated into the decision variable using weighted linear combinations. Developing the model further was beyond the scope of the present thesis.

All of the listeners taking part in the psychoacoustical experiments completed two questionnaires before their DLFs were measured, one of which (the SSQ) assessed experiences and difficulties in everyday hearing, and the other assessed the extent of their musical experience and expertise. For the purposes of data reduction, listeners' SSQ responses were collated with the responses of listeners from other experiments, and a factor analysis was performed (Chapter 7). The analysis revealed a clear multi-factor structure, which turned out to be very similar to a theoretical structure proposed in a previous study (Gatehouse & Akeroyd, 2006). The DDLFs, IDLFs, and I/D ratios measured in 57 listeners were then correlated with their scores on the SSQ factors. The psychoacoustical measurements were not strongly correlated with any of the factors, although weak relationships were observed with some factors, including those related to sound segregation and music perception. These relationships were negative: larger thresholds were associated with rating one's own hearing less favourably.

The music questionnaire investigated further the relationship between the psychoacoustical measurements and music perception (Chapter 7). Although none of the listeners was a professional or practising musician, DDLFs and IDLFs were correlated with several aspects of musical experience and music perception. Again, the significant relationships were all negative. DDLFs were correlated more strongly than IDLFs with items on the music questionnaire, suggesting that it was listeners' basic sensitivity to frequency or pitch changes, rather than the ability to identify the direction of those changes, that was related to musical experience.

A final experiment (Chapter 8) measured listeners' extra-cranial magnetic activity associated with detecting and identifying the direction of small frequency differences between pure tones. The contributions of the left and right hemispheres to these signals were delineated using a dipole-source approach, and the hypothesis tested was that direction-impaired listeners would experience a leftward asymmetry of processing during the identification task. Contrary to the hypothesis—and the results of a previous experiment (Foxton et al., 2009)—the opposite result was found: on average, the direction-impaired listeners had stronger right-hemisphere dipoles than the unimpaired listeners.

9.2 Future research

Based on the findings of Experiments 1–6, the sequential interference hypothesis provided the most convincing explanation of the pitch-direction impairment. However, as the supplementary experiment demonstrates, there are still a number of outstanding issues. As mentioned above and discussed in Section 5.3, the hypothesis does not take into account the potential beneficial effects of a preceding tone (e.g., cuing and multiple looks) on pitch-direction identification, or how the extent of such effects might differ between direction-impaired and unimpaired listeners. Another issue is whether direction-impaired listeners are susceptible to backward as well as forward interference. Experiment 6 investigated the role of irrelevant pitch changes occurring before the onset of the relevant pitch change (i.e., forward interference). However, in the dual-pair procedures of Experiments 1–4, when the first tone pair contained the relevant pitch change, the irrelevant between-pair change occurred after the relevant change and could have caused backward interference. The trial-by-trial analysis of the data from Experiments 1 and 3 (Section 4.4.3) did not distinguish between backward and forward effects. A re-analysis of those data could reveal whether interference occurs in both directions, and if so, which has the greater impact on IDLFs in direction-impaired listeners.

Another issue for future work is how best to characterise the relationship between the susceptibility of a direction-impaired listener to sequential interference and the size of irrelevant pitch changes they hear. A logical first step is to assume that the relationship is systematic and that larger irrelevant changes cause more interference than smaller changes. Preliminary support for this idea comes from the results of Experiment 3, in which there was a monotonic relationship between frequency-roving range and I/D ratio in the novice listeners (Section 4.2.2.2). However, the experiment was not designed to test this idea specifically, and so it did not control the magnitude of the irrelevant pitch changes in the different conditions. Future experiments could use the method of constant stimuli. As in Experiment 6, listeners could be presented with three tones per trial and could judge the direction of the frequency difference between the second and third tones. The relevant pitch change could be fixed at a value corresponding to some nominal point on the psychometric function, and d' could be measured as a function of the irrelevant pitch change between the first and second tones. An experiment of this kind could be used firstly to test the validity of the temporal filter model described in Section 6.4.2, and secondly—if the model's validity is confirmed—to estimate the weights applied to the irrelevant pitch changes. Note that the irrelevant pitch changes in such an experiment could be expressed either in physical terms (e.g., cents) or relative to a listener's DLF. It would be of interest if there were differences between unimpaired and direction-impaired listeners in the relationship between d' and irrelevant-change magnitude, and of particular interest if the differences were dependent on whether the irrelevant changes are expressed in cents, or expressed relative to listeners' DLFs.

Even if the sequential interference hypothesis turns out to be correct, it still does not explain the main source of the individual differences: why are some listeners more susceptible to interference than others? Experiment 7 (Chapter 8) found that direction-impaired listeners had stronger right-hemispheric dipoles when listening to frequency changes, although it is not clear how this effect influences or is influenced by sequential interference. Future work could use functional imaging techniques to examine the neural correlates of interference, and investigate how these correlates differ between individuals.

Another possible source of the individual differences is pitch memory. It could be the case that direction-impaired listeners have a poorer memory for pitch than do unimpaired listeners. For instance, pitch traces in direction-impaired listeners might become diffuse over time more rapidly than normal, or might be retained less effectively as new pitches are perceived and memorised. It might be informative to test direction-impaired listeners in traditional pitch-memory experiments, as has recently been done with individuals with congenital amusia (Williamson, McDonald, Deutsch, Griffiths, & Stewart, 2010; Williamson & Stewart, 2010). direction-impaired listeners' memories for pitch could then be compared to their memories for other attributes of auditory sensation, their memories for visual sensory attributes, and/or their memories for categorical and semantic information. The results of such experiments could even provide evidence for or against the idea that the human auditory system has a unique memory store for pitch (Section 2.2.5).

Appendix A

The speech, spatial, and qualities of hearing scales (SSQ) questionnaire

The matter in square brackets following each vignette represents the extrema of the visual analogue scale for that item.

Speech

1. You are talking with one other person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you're talking to says? [not at all–perfectly].
2. You are talking with one other person in a quiet, carpeted lounge-room. Can you follow what the other person says? [not at all–perfectly].
3. You are in a group of about five people, sitting round a table. It is an otherwise quiet place. You can see everyone else in the group. Can you follow the conversation? [not at all–perfectly].
4. You are in a group of about five people in a busy restaurant. You can see everyone else in the group. Can you follow the conversation? [not at all–perfectly].
5. You are talking with one other person. There is continuous background noise, such as a fan or running water. Can you follow what the person says? [not at all–perfectly].
6. You are in a group of about five people in a busy restaurant. You cannot see everyone else in the group. Can you follow the conversation? [not at all–perfectly].

7. You are talking to someone in a place where there are a lot of echoes, such as a church or railway terminus building. Can you follow what the other person says? [not at all–perfectly].
8. Can you have a conversation with someone when another person is speaking whose voice is the same pitch as the person you're talking to? [not at all–perfectly].
9. Can you have a conversation with someone when another person is speaking whose voice is different in pitch from the person you're talking to? [not at all–perfectly].
10. You are listening to someone talking to you, while at the same time trying to follow the news on TV. Can you follow what both people are saying? [not at all–perfectly].
11. You are in conversation with one person in a room where there are many other people talking. Can you follow what the person you are talking to is saying? [not at all–perfectly].
12. You are with a group and the conversation switches from one person to another. Can you easily follow the conversation without missing the start of what each new speaker is saying? [not at all–perfectly].
13. Can you easily have a conversation on the telephone? [not at all–perfectly].
14. You are listening to someone on the telephone and someone next to you starts talking. Can you follow what's being said by both speakers? [not at all–perfectly].

Spatial

1. You are outdoors in an unfamiliar place. You hear someone using a lawnmower. You can't see where they are. Can you tell right away where the sound is coming from? [not at all–perfectly].
2. You are sitting around a table or at a meeting with several people. You can't see everyone. Can you tell where any person is as soon as they start speaking? [not at all–perfectly].
3. You are sitting in between two people. One of them starts to speak. Can you tell right away whether it is the person on your left or your right, without having to look? [not at all–perfectly].

4. You are in an unfamiliar house. It is quiet. You hear a door slam. Can you tell right away where that sound came from? [not at all–perfectly].
5. You are in the stairwell of a building with floors above and below you. You can hear sounds from another floor. Can you readily tell where the sound is coming from? [not at all–perfectly].
6. You are outside. A dog barks loudly. Can you tell immediately where it is, without having to look? [not at all–perfectly].
7. You are standing on the footpath of a busy street. Can you hear right away which direction a bus or truck is coming from before you see it? [not at all–perfectly].
8. In the street, can you tell how far away someone is, from the sound of their voice or footsteps? [not at all–perfectly].
9. Can tell how far away a bus or a truck is, from the sound? [not at all–perfectly].
10. Can you tell from the sound which direction a bus or truck is moving, for example, from your left to your right or right to left? [not at all–perfectly].
11. Can you tell from the sound of their voice or footsteps which direction a person is moving, for example, from your left to your right or right to left? [not at all–perfectly].
12. Can you tell from their voice or footsteps whether the person is coming towards you or going away? [not at all–perfectly].
13. Can you tell from the sound whether a bus or truck is coming towards you or going away?
14. Do the sounds of things you are able to hear seem to be inside your head rather than out there in the world? [inside my head–out there].
15. Do the sounds of people or things you hear, but cannot see at first, turn out to be closer than expected when you do see them? [much closer–not closer].
16. Do the sounds of people or things you hear, but cannot see at first, turn out to be further away than expected when you do see them? [much further–not further].
17. Do you have the impression of sounds being exactly where you would expect them to be? [not at all–perfectly].

Qualities

1. Think of when you hear two things at once, for example, water running into a basin (a power-tool being used)(a plane flying past) and, at the same time, a radio playing (the sound of hammering)(a truck driving past). Do you have the impression of these as sounding separate from each other? [not at all–perfectly].
2. When you hear more than one sound at a time, do you have the impression that it seems like a single jumbled sound? [jumbled–not jumbled].
3. You are in a room and there is music on the radio. Someone else in the room is talking. Can you hear the voice as something separate from the music? [not at all–perfectly].
4. Do you find it easy to recognise different people you know by the sound of each one's voice? [not at all–perfectly].
5. Do you find it easy to distinguish different pieces of music that you are familiar with? [not at all–perfectly].
6. Can you tell the difference between different sounds, for example, a car versus a bus; water boiling in a pot versus food cooking in a frying pan? [not at all–perfectly].
7. When you listen to music, can you make out which instruments are playing? [not at all–perfectly].
8. When you listen to music, does it sound clear and natural? [not at all–perfectly].
9. Do everyday sounds that you can hear easily seem clear to you (not blurred)? [not at all–perfectly].
10. Do other people's voices sound clear and natural? [not at all–perfectly].
11. Do everyday sounds that you hear seem to have an artificial or unnatural quality? [unnatural–natural].
12. Does your own voice sound natural to you? [not at all–perfectly].
13. Can you easily judge another person's mood from the sound of their voice? [not at all–perfectly].
14. Do you have to concentrate very much when listening to someone or something? [concentrate hard–no need to concentrate].
15. Item excluded.

16. When you are the driver in a car can you easily hear what someone is saying who is sitting alongside you? [not at all–perfectly].
17. When you are a passenger can you easily hear what the driver is saying sitting alongside you? [not at all–perfectly].
18. Do you have to put in a lot of effort to hear what is being said in conversation with others? [lot of effort–no effort].
19. Can you easily ignore other sounds when trying to listen to something? [not easily ignore–Easily ignore].
20. Item excluded.
21. Item excluded.
22. Item excluded.

Appendix B

Music questionnaire

Items 1 and 8–23 required listeners to respond using a visual analogue scale. The matter in square brackets following the vignettes of these items represents the extrema of the scale. Items 2, 3, and 24 required yes/no responses, and blank spaces were left for responses after all of the other items.

1. How often did you hear music in your childhood? [never–all the time].
2. During your childhood, were any of your family members musicians? [yes/no].
3. Were musical instruments played during your childhood? [yes/no].
4. Please list the forms of musical education you have had, if any. This could include private or group tuition, self-tuition, concert recitals etc.
5. Please list the musical instruments you play, if any, indicating your primary instrument (this may include singing).
6. If you play an instrument, how many years have you played your (primary) instrument?
7. If you play an instrument, at the peak of your interest, roughly how many hours per week did you play/practice your instrument?
8. I sing in private (e.g., in my car, in the shower, etc.) [never–all the time].
9. I sing in public (as part of a group or solo: e.g., a choir, carols, a sing-a-long, etc.) [never–all the time].
10. How often do you choose to listen to music (as opposed to hearing music on TV, in a store, etc.)? [never–all the time].
11. When you listen to music, how difficult do you find it to hear the difference between the notes? [impossible–effortless].

12. When you listen to music, how difficult do you find it to hear the difference between the instruments being played? [impossible–effortless].
13. How difficult do you find singing in general? [impossible–effortless].
14. Rate your ability to memorize a short song [completely unable–perfectly able].
15. How difficult would you find it to repeat a tune someone else has recently sung to you? [impossible–effortless].
16. If you imagine the tune *Happy Birthday*, how accurately can you hear the melody in your head? [not at all–completely]
17. When you hear music (on the radio, in a store, on TV, etc.), how often can you recognise familiar songs by the first two or three notes? [never–all the time].
18. How difficult do you find it to hum along with your favourite music? [impossible–effortless].
19. Singing a note to match one played on a piano is a task you would find ... [impossible–effortless].
20. If someone played two notes on a piano, separately, and asked you to say which one was higher in pitch, how difficult would you find it? [impossible–effortless].
21. When you sing, how able are you to tell if you are out of tune? [completely unable–perfectly able].
22. How often do you get a tune stuck in your head? [never–all the time].
23. How difficult do you find it to clap or tap along with a familiar melody? [impossible–effortless].
24. Have you ever thought (or has anyone told you) that you might be tone deaf? [yes/no].
25. Finally, if you think you may have any other difficulties or skills (such as "perfect pitch") to do with music or hearing in general that have not been addressed by this questionnaire, please give details below.

Abbreviations and symbols

Λ	synthetic listener
d'	index of sensitivity (pronounced ‘dee-prime’)
F	frequency
F_0	Fundamental frequency
PC	proportion correct
2AFC	two-alternative forced choice
2I-2AFC	two-interval, two-alternative forced choice
ABRM	auditory backwards recognition masking
AEF	auditory evoked field
AM	amplitude modulation
ANSI	American National Standards Institute
ASA	American Standards Association
BOLD	blood-oxygen-level dependent
CoRE	component-relative entropy
CVD	constant-variance differencing
CVIO	constant-variance independent-observations
DDLF	detection DLF
DLF	difference limen for frequency
ECG	electro-cardiogram
EEG	electroencephalography
EOG	electro-oculogram
ERP	event-related potential
fMRI	functional magnetic resonance imaging
FSD	frequency-shift detector
HG	Heschl’s gyrus
HL	hearing level
I/D	identification/detection (as in I/D ratio)
IDLF	identification DLF
IM	informational masking
IRN	iterated rippled noise
ISI	interstimulus interval
MEG	magnetoencephalography

MMN	mismatch negativity
MRI	magnetic resonance imaging
PET	positron emission tomography
POR	pitch-onset response
PPAM	pre-perceptual auditory memory
RMS	root-mean-square
RT	reaction time
SDT	signal detection theory
SPL	sound pressure level
SQUID	super-conducting quantum interface device
SSQ	the speech, spatial, and qualities of hearing scale
STAM	short-term auditory memory
TBAC	tests of basic auditory capabilities
UVD	uncertainty-based-variance differencing

Abbreviations and symbols for statistics, for units of measurement under the International System of Units, those appearing in references, and others used very commonly in publications from the American Psychological Association are ommitted from the list above.

References

Agus, T. R., Akeroyd, M. A., Noble, W., & Bhullar, N. (2010). An analysis of the masking of speech by competing speech using self-report data. *Journal of the Acoustical Society of America, 125*, 23–26.

Alexander, J. M., & Lutfi, R. A. (2004). Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *Journal of the Acoustical Society of America, 116*, 2234–2247.

Alho, K. (1992). Selective attention in auditory processing as reflected by event-related brain potentials. *Psychophysiology, 29*, 247–263.

Amitay, S., Hawkey, D. J. C., & Moore, D. R. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception and Psychophysics, 67*, 691–698.

Anderson, B., Southern, B. D., & Powers, R. E. (1999). Anatomic asymmetries of the posterior superior temporal lobes: A postmortem study. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology., 12*, 247–254.

Anderson, D. J., Rose, J. E., Hind, J. E., & Brugge, J. F. (1971). Temporal position of discharges in single auditory nerve fibres within the cycle of a sine-wave stimulus: Frequency and intensity effects. *Journal of the Acoustical Society of America, 49*, 1131–1139.

ANSI. (1973/1994). *American National Standards acoustical terminology.* New York, NY: American Nation Standards Institute.

ASA. (1960). *Acoustical terminology.* New York, NY: American Standards Association.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology, 84*, 147–166.

Bachem, A. (1948). Chroma fixation at the ends of the musical frequency scale. *Journal of the Acoustical Society of America, 20*, 704–705.

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine, 18*, 14–30.

Bella, S. D., & Peretz, I. (2003). Congenital amusia interferes with the ability to synchronize with music. *Annals of the New York Academy of Sciences, 999*, 166-169.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature, 436*, 1161–1165.

Berg, P., & Sherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*, 229–241.

Besson, M., Faita, F., & Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience Letters*, *168*, 101–105.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100–130.

Bland, D., & Perrott, D. R. (1978). Backward masking: Detection versus recognition. *Journal of the Acoustical Society of America*, *63*, 1215–1217.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*, 389–395.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., & Purks, S. R. (1984). Intensity perception: XIII. Perceptual anchor model of context-coding. *Journal of the Acoustical Society of America*, *76*, 722–731.

Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex*, *15*, 578–587.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Brown, M., Irvine, D. R. F., & Park, V. N. (2004). Perceptual learning on an auditory frequency discrimination task by cats: Association with changes in primary auditory cortex. *Cerebral Cortex*, *14*, 952–65.

BSA. (2004). *Recommended procedure: Pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels*. British Society of Audiology.

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology*, *31*, 56–68.

Bull, A. R., & Cuddy, L. L. (1972). Recognition memory for pitch of fixed and roving stimulus tones. *Perception and Psychophysics*, *11*, 105–109.

Burns, E. M., & Sampat, K. S. (1980). A note on possible culture-bound effects in frequency discrimination. *Journal of the Acoustical Society of America*, *68*, 1886–1888.

Cansino, S., & Williamson, S. J. (1997). Neuromagnetic fields reveal cortical plasticity when learning an auditory discrimination task. *Brain Research*, *764*, 53–66.

Cansino, S., Williamson, S. J., & Karron, D. (1994). Tonotopic organization of human auditory association cortex. *Brain Research*, *663*, 38–50.

Carcagno, S., & Plack, C. J. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. *Journal of the Association for Research in Otolaryngology*, *12*, 89–100.

Carr, C. E. (1993). Processing of temporal information in the brain. *Annual Review of Neuroscience, 16,* 223–243.

Cary, H. (1922). Are you a musician? Professor Seashore's specific psychological tests for specific musical abilities. *Scientific American, 12,* 326–327.

Cattell, R. B. (1949). *Culture free intelligence test, scale 1, handbook.* Champaign, IL: Institute of Personality and Ability.

Chait, M., Poeppel, D., de Cheveigné, A., & Simon, J. Z. (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. *Journal of Neuroscience, 27,* 5207–5214.

Chait, M., Poeppel, D., & Simon, J. Z. (2006). Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cerebral Cortex, 16,* 835–848.

Chait, M., Simon, J. Z., & Poeppel, D. (2004). Auditory M50 and M100 responses to broadband noise: Functional implications. *NeuroReport, 15,* 2455–2458.

Christman, R. J. (1954). Shifts in pitch as a function of prolonged stimulation with pure tones. *American Journal of Psychology, 67,* 484–491.

Christman, R. J., & Williams, W. E. (1963). Influence of the time interval on experimentally induced shifts of pitch. *Journal of the Acoustical Society of America, 35,* 1030–1033.

Christopherson, L. A., & Humes, L. E. (1992). Some psychometric properties of the tests of basic auditory capabilities (TBAC). *Journal of Speech and Hearing Research, 35,* 929–935.

Clément, S., Demany, L., & Semal, C. (1999). Memory for pitch versus memory for loudness. *Journal of the Acoustical Society of America, 106,* 2805–2811.

Cohen, A. (1961). Further investigation of the effects of intensity upon the pitch of pure tones. *Journal of the Acoustical Society of America, 33,* 1363–1376.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of the mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Human Perception and Performance, 19,* 909–921.

Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., & Frisina, R. D. (1994). Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *Journal of the Acoustical Society of America, 95,* 2720–2727.

Cuddy, L. L., Balkwill, L. L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: A study of "tone deafness" among university students. *Annals of the New York Academy of Sciences, 1060,* 311–324.

Dai, H., Nguyen, Q. T., & Green, D. M. (1995). A two-filter model for frequency discrimination. *Hearing Research, 85,* 109–114.

Davis, A. (1995). *Hearing in adults.* London: Whurr.

de Cheveigné, A. (2005). Pitch perception models. In C. J. Plack, A. J. Oxenham,

R. R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 147–168). New York, NY: Springer.

Demany, L. (1985). Perceptual learning in frequency discrimination. *Journal of the Acoustical Society of America, 78*, 1118–1120.

Demany, L., Carlyon, R. P., & Semal, C. (2009). Continuous versus discrete frequency changes: Different detection mechanisms? *Journal of the Acoustical Society of America, 125*, 1082–1090.

Demany, L., Clément, S., & Semal, C. (2001). Does auditory memory depend on attention? In D. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, & R. Schoonhoven (Eds.), *Physiological and psychophysical bases of auditory function* (pp. 461–467). Maastricht: Shaker.

Demany, L., Montandon, G., & Semal, C. (2004). Pitch perception and retention: Two cumulative benefits of selective attention. *Perception and Psychophysics, 66*, 609–617.

Demany, L., Montandon, G., & Semal, C. (2005). Internal noise and memory for pitch. In D. Pressnitzer, A. de Cheveigné, S. McAdams, & L. Collet (Eds.), *Auditory signal processing: Physiology, psychoacoustics, and models* (pp. 230–236). New York, NY: Springer.

Demany, L., Pressnitzer, D., & Semal, C. (2009). Tuning properties of the auditory frequency-shift detectors. *Journal of the Acoustical Society of America, 126*, 1342–1348.

Demany, L., & Ramos, C. (2005). On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *Journal of the Acoustical Society of America, 117*, 833–841.

Demany, L., & Semal, C. (2002). Learning to perceive pitch differences. *Journal of the Acoustical Society of America, 111*, 1377–1388.

Demany, L., & Semal, C. (2005). The slow formation of a pitch percept beyond the ending time of a short tone burst. *Perception and Psychophysics, 67*, 1376–1383.

Demany, L., & Semal, C. (2008). The role of memory in auditory perception. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 77–114). New York, NY: Springer.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science, 168*, 1604–1605.

Deutsch, D. (1999). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (pp. 349–411). Academic Press.

Deutsch, D., & Feroe, J. (1975). Disinhibition in pitch memory. *Perception and Psychophysics, 17*, 320–324.

Digman, J. M. (1990). Personality structure — emergence of the 5-factor model. *Annual Review of Psychology, 41*, 417-440.

Doughty, J. M., & Garner, W. R. (1947). Pitch characteristics of short tones I: Two kinds of pitch threshold. *Journal of Experimental Psychology*, *37*, 351–365.

Doughty, J. M., & Garner, W. R. (1948). Pitch characteristics of short tones II: Pitch as a function of duration. *Journal of Experimental Psychology*, *38*, 478–494.

Drake, R. M. (1954). *Manual for D musical aptitude tests.* Chicago, IL: University of Chicago Press.

Durlach, N. I. (2006). Need for improved conceptual structure (L). *Journal of the Acoustical Society of America*, *120*, 1787–1790.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, *46*, 372–383.

Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., & Kidd, G. Jr. (2005). Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. *Journal of the Acoustical Society of America*, *118*, 2482–2497.

Durlach, N. I., Mason, C. R., Kidd, G. Jr., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *Journal of the Acoustical Society of America*, *113*, 2985–2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G. Jr. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *Journal of the Acoustical Society of America*, *114*, 368–379.

Dynes, S., & Delgutte, B. (1992). Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea. *Hearing Research*, *58*, 79–90.

Elliott, D. N., Riach, W. D., Sheposh, J. P., & Trahiotis, C. (1966). Discrimination performance of high school sophomores on a battery of auditory tests. *Acta Oto-Laryngologica Supplement*, *216*, 1–59.

Emmerich, D. S., Ellermeier, W., & Butensky, B. (1989). A reexamination of the frequency discrimination of random-amplitude tones, and a test of Henning's modified energy-detector model. *Journal of the Acoustical Society of America*, *85*, 1653–1659.

Erné, S., & Hoke, M. (1990). Short-latency evoked magnetic fields from the human auditory brainstem. *Advances in Neurology*, *54*, 167–176.

Erné, S., Scheer, H., M., H., Pantev, C., & Lütkenhöner, B. (1987). Brainstem auditory evoked magnetic fields in response to stimulation with brief tone pulses. *International Journal of Neuroscience*, *37*, 115–125.

Feth, L. L., & O'Malley, H. (1977). Two-tone auditory spectral resolution. *Journal of the Acoustical Society of America*, *62*, 940–947.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.

Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, *12*, 47–65.

Formisano, E., Kim, D., Di Salle, F., Moortele, P. van de, Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*, 859–869.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain, 127*, 801–810.

Foxton, J. M., Weisz, N., Bauchet-Lecaignard, F., Delpuech, C., & Bertrand, O. (2009). The neural bases underlying pitch processing difficulties. *Neuroimage, 45*, 1305–1313.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience, 6*, 1216–1223.

Frodl-Bauch, T., Kathmann, N., Möller, H.-J., & Hegerl, U. (1997). Dipole localization and test-retest reliability of frequency and duration mismatch negativity generator processes. *Brain Topography, 10*, 3–8.

Gaab, N., & Schlaug, G. (2003). The effect of musicianship on pitch memory in performance matched groups. *NeuroReport, 14*, 2291–2295.

Galuske, R. A., Schlote, W., Bratzke, H., & Singer, W. (2000). Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science, 289*, 1946–1949.

Gatehouse, S., & Akeroyd, M. (2006). Two-eared listening in dynamic situations. *International Journal of Audiology, 45*, 120–124.

Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology, 43*, 85–99.

Gelfand, S. A. (2010). *Hearing: An introduction to psychological and physiological acoustics* (5th ed.). Colchester: Informa Healthcare.

Giard, M. H., Lavikahen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., et al. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience, 7*, 133–143.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., et al. (2000). Representation of the temporal envelope of sounds in the human brain. *Neurophysiology, 84*, 1588–1598.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*, 408–414.

Gleich, O., & Narins, P. (1988). The phase response of primary auditory afferents in a songbird (sturnus vulgaris L.). *Hearing Research, 32*, 81–92.

Gold, J. M., Murray, R. F., Sekuler, A. B., Bennett, P. J., & Sekuler, R. (2005). Visual memory decay is deterministic. *Psychological Science, 16*, 769–774.

Gosselin, N., Jolicoeur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences, 1169*, 270–272.

Green, D. M. (1971). Temporal auditory acuity. *Psychological Review, 78*, 540–551.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New

York, NY: John Wiley.

Green, T. J., & McKeown, J. D. (2001). Capture of attention in selective frequency listening. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 1197–1210.

Greenberg, G. S., & Larkin, W. D. (1968). Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe–signal method. *Journal of the Acoustical Society of America, 44*, 1513–1523.

Griffiths, P. (2006). *A concise history of western music.* Cambridge: Cambridge University Press.

Griffiths, T. D. (2008). Tone deafness: A model complex cortical phenotype. *Clinical Medicine, 8*, 592–595.

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience, 1*, 422–427.

Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., & Patterson, R. D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nature Neuroscience, 4*, 633–637.

Gussenhoven, C. (2004). *The phonology of tone and intonation.* Cambridge: Cambridge University Press.

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *Neuroimage, 15*, 207–216.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *Neuroimage, 22*, 755–66.

Guttman, N., & Pruzansky, S. (1962). Lower limits of pitch and musical pitch. *Journal of Speech and Hearing Research, 5*, 207–214.

Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (2002). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 12*, 140–14.

Hall, D. A., & Plack, C. J. (2007). The human 'pitch center' responds differently to iterated noise and huggins pitch. *NeuroReport, 18*, 323–327.

Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerbral Cortex, 19*, 576–585.

Hall, J. W., Buss, E., & Grose, J. H. (2005). Informational masking release in children and adults. *Journal of the Acoustical Society of America, 118*, 1605–1613.

Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography — theory, instrumentation, and applications to noninvasive studies of signal processing in the human brain. *Reviews of Modern Physics, 65*, 413–497.

Hari, R., Hämäläinen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., et al. (1984). Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: Neuromagnetic recordings in man. *Neuroscience Letters, 50*, 127–132.

Harms, M. P., & Melcher, J. R. (2002). Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. *Journal of Neurophysiology, 88*, 1433–1450.

Harris, J. D. (1952a). The decline of pitch discrimination with time. *Journal of Experimental Psychology, 43*, 96–99.

Harris, J. D. (1952b). Pitch discrimination. *Journal of the Acoustical Society of America, 24*, 750–755.

Hartmann, R., & Klinke, R. (1987). Impulse pattern in auditory-nerve fibres to extra- and intra-cochlear electrical stimulation. In *Cochlear implants: Current situation.* Düren, Germany: International Cochlear Implant Symposium.

Hartmann, W. M. (1978). The effect of amplitude envelope on the pitch of sinewave tones. *Journal of the Acoustical Society of America, 63*, 1105–1113.

Hartmann, W. M., Rakerd, B., & Packard, T. N. (1985). On measuring the frequency difference limen for short tones. *Perception and Psychophysics, 38*, 199–207.

Hashimoto, I., Mashiko, T., Yoshikawa, K., Mizuta, T., Imada, T., & Hayashi, M. (1995). Neuromagnetic measurements of the human primary auditory response. *Electroencephalography and Clinical Neurophysiology, 96*, 348–356.

Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Strome in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche. *Annalen Der Physik Und Chemie, 165*, 211–233.

Henning, G. B. (1966). Frequency discrimination of random-amplitude tones. *Journal of the Acoustical Society of America, 39*, 336–339.

Henning, G. B. (1970). A comparison of the effects of signal duration on frequency and amplitude discrimination. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing.* Leiden: A. W. Sijthoff.

Hill, K., Stange, G., & Mo, J. (1989). Temporal synchronization in the primary auditory response in the pigeon. *Hearing Research, 39*, 63–74.

Hillebrand, A., & Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head meg to activity in the adult human cortex. *Neuroimage, 16*, 638–650.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in human brain. *Science, 182*, 171–180.

Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31*, 759–767.

Houtsma, A. J. M. (1997). Pitch and timbre: Definition, meaning and use. *Journal of New Music Research, 26*, 104–115.

Howard III, M. A., Volkov, I. O., Abbas, P. J., Damasio, H., Ollendiek, M. C., & Granner, M. (1996). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Research, 724*, 260–264.

Humes, L. E., & Christopherson, L. (1991). Speech identification difficulties of hearing-impaired elderly persons — the contributions of auditory processing deficits. *Journal of Speech and Hearing Research, 34*, 686–693.

Hutsler, J., & Gazzaniga, M. (1996). Acetylcholinesterase staining in human auditory and language cortices—regional variation of structural features. *Cerebral Cortex, 6*, 260–270.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more. *Journal of Neuroscience, 27*, 13028–13032.

Hyde, K. L., & Peretz, I. (2003). "Out-of-pitch" but still "in-time". An auditory psychophysical study in congenital amusic adults. *Annals of the New York Academy of Sciences, 999*, 173–176.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*, 356–360.

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2010). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex*, advanced online publication.

Iramina, K., & Ueno, S. (1995). Measurement of brain-stem auditory-evoked magnetic fields using a highly sensitive squid magnetometer with a variable base-line. *IEEE Transactions on Magnetics, 31*, 4271–4273.

Irvine, D. R. F., Martin, R., Klimkeit, E., & Smith, R. (2000). Specificity of perceptual learning in a frequency discrimination task. *Journal of the Acoustical Society of America, 108*, 2964–2968.

Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology, 38*, 723–727.

Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex, 16*, 1266–1275.

Jeffress, L. A., & McFadden, D. (1971). Differences of interaural phase and level in detection and lateralization. *Journal of the Acoustical Society of America, 49*, 1169–1179.

Jesteadt, W., & Bilger, R. C. (1974). Intensity and frequency discrimination in one- and two-interval paradigms. *Journal of the Acoustical Society of America, 55*, 1266–1276.

Jezzard, P., Matthews, P. M., & Smith, S. M. (2003). *Functional magnetic resonance imaging: An introduction to methods.* Oxford: Oxford University Press.

Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *Journal of the Acoustical Society of America, 68,* 1115–1122.

Johnson, D. M., & Hafter, E. R. (1980). Uncertain-frequency detection: Cueing and condition of observation. *Perception and Psychophysics, 28,* 143–149.

Johnson, D. M., Watson, C. S., & Jensen, J. K. (1987). Individual differences in auditory capabilities I. *Journal of the Acoustical Society of America, 81,* 427–438.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain, 123,* 155–163.

Jones, J. L., Zalewski, C., Brewer, C., Lucker, J., & Drayna, D. (2009). Widespread auditory deficits in tune deafness. *Ear and Hearing, 30,* 63–72.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America, 97,* 11793–11799.

Kaernbach, C. (1991a). Poisson signal-detection theory: Link between threshold models and the Gaussian assumption. *Perception and Psychophysics, 50,* 498–506.

Kaernbach, C. (1991b). Simple adaptive testing with the weighted up-down method. *Perception and Psychophysics, 49,* 227–229.

Kaernbach, C., & Schlemmer, K. (2008). The decay of pitch memory during rehearsal. *Journal of the Acoustical Society of America, 123,* 1846–1849.

Karlin, J. E. (1942). A factorial study of auditory function. *Psychometrika, 7,* 251–279.

Keller, T. A., Cowan, N., & Saults, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, memory, and Cognition, 21,* 635–645.

Kettner, R., Feng, J., & Brugge, J. (1985). Postnatal development of the phase-locked response to low frequency tones of auditory nerve fibers in the cat. *Journal of Neuroscience, 5,* 275–283.

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *Journal of the Acoustical Society of America, 122,* 418–435.

Kidd, G. Jr., Arbogast, T. L., Mason, C. R., & Walsh, M. (2001). Informational masking in listeners with sensorineural hearing loss. *Journal of the Association for Research in Otolaryngology, 3,* 107–119.

Kidd, G. Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *Journal of the Acoustical Society of America, 95,* 3475–3480.

Kidd, G. Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 143–189). New York, NY: Springer.

Kinchla, R., & Smyzer, F. (1967). A diffusion model of perceptual memory. *Perception and Psychophysics, 2,* 219–229.

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology*, *12*, 125S–143S.

Kitterick, P. (2008). *The bases of difficulties in spatial hearing for speech: Investigations using psychoacoustic techniques and magneto-encephalography.* Unpublished doctoral dissertation, Department of Psychology, University of York.

Kitterick, P., Bailey, P., & Summerfield, A. Q. (2010). Benefits of knowing who, where, and when in multi-talker listening. *Journal of the Acoustical Society of America*, *127*, 2498–2508.

Koelsch, S., Schroger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport*, *10*, 1309–1313.

Köppl, C. (1997). Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba. *Journal of Neuroscience*, *17*, 3312–3321.

Krumbholz, K., Patterson, R. D., & Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *Journal of the Acoustical Society of America*, *108*, 1170–1180.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's Gyrus. *Cerebral Cortex*, *13*, 765–72.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch* (Vol. 17). Oxford: Oxford University Press.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 739–751.

Kuriki, S., & Murase, M. (1989). Neuromagnetic study of the auditory responses in right and left hemispheres of the human brain evoked by pure tones and speech sounds. *Experimental Brain Research*, *77*, 127–134.

Kuriki, S., Nogai, T., & Hirata, Y. (1995). Cortical sources of middle latency responses of auditory evoked magnetic field. *Hearing Research*, *92*, 47–51.

Langner, G., Sams, M., Heil, P., & Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: Evidence from magnetoencephalography. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, *181*, 665–676.

Larkin, W. D. (1978). Pitch shifts following tone adaptation. *Acustica*, *41*, 110–116.

Lauter, J. L. (1982). Dichotic identification of complex sounds: Absolute and relative ear advantages. *Journal of the Acoustical Society of America*, *71*, 701–707.

Lauter, J. L. (1983). Stimulus characteristics and relative ear advantages: A new look at old data. *Journal of the Acoustical Society of America*, *74*, 1–17.

Lavikainen, J., Huotilainen, M., Ilmoniemi, R., Simola, J. T., & Näätänen, R. (1995).

Pitch change of a continuous tone activates two distinct processes in human auditory cortex: A study with whole-head magnetometer. *Electroencephalography and Clinical Neurophysiology*, *96*, 93–96.

Levänen, S., Hari, R., McEvoy, L., & Sams, M. (1993). Responses of the human auditory cortex to changes in one versus two stimulus features. *Experimental Brain Research*, *97*, 177–83.

Levänen, S., & Sams, M. (1997). Disrupting human auditory change detection: Chopin is superior to white noise. *Psychophysiology*, *34*, 258–265.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.

Liégeois-Chauvel, C., Giraud, K., Badier, J. M., Marquis, P., & Chauvel, P. (2003). Intracerebral evoked potentials in pitch perception reveal a functional asymmetry of human auditory cortex. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 152–167). Oxford: Oxford University Press.

Liégeois-Chauvel, C., Graaf, J. de, Laguitton, V., & Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral Cortex*, *9*, 484–496.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology*, *18*, R331–R332.

Lounasmaa, O., Hari, R., Joutsiniemi, S., & Hämäläinen, M. (1989). Multi-SQUID recordings of human cerebral magnetic-fields may give information about memory processes. *Europhysics Letters*, *9*, 603–608.

Lutfi, A. (1993). A model of auditory pattern analysis based on component-relative-entropy. *Journal of the Acoustical Society of America*, *94*, 748–758.

Lütkenhöner, B. (2003). Single-dipole analyses of the N100m are not suitable for characterizing the cortical representation of pitch. *Audiology and Neurootology*, *8*, 222–233.

Lütkenhöner, B., Krumbholz, K., Lammertmann, C., , Seither-Preisler, A., Steinsträter, O., et al. (2003). Localization of primary auditory cortex in humans by magnetoencephalography. *Neuroimage*, *18*, 58–66.

Lütkenhöner, B., Krumbholz, K., & Seither-Preisler, A. (2003). Studies of tonotopy based on wave N100 of the auditory evoked field are problematic. *Neuroimage*, *19*, 935–949.

Lütkenhöner, B., Lammertmann, C., & Knecht, S. (2001). Latency of auditory evoked field deflection N100m ruled by pitch or spectrum? *Audiology and Neurootology*, *6*, 263–27.

Lütkenhöner, B., Lammertmann, C., Ross, B., & Pantev, C. (2000). Brain stem auditory evoked fields in response to clicks. *NeuroReport*, *11*, 913–918.

Lütkenhöner, B., & Steinsträter, O. (1998). High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiology and Neuro-*

otology, 3, 191–213.

Lyzenga, J., Carlyon, R. P., & Moore, B. C. J. (2004). The effects of real and illusory glides on pure-tone frequency discrimination. *Journal of the Acoustical Society of America, 116*, 491–501.

Macmillan, N. A., & Creelman, C. D. (2005). *Dection theory: A user's guide.* Mahwah, NJ: Lawrence Erlbaum Associates.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. (1977). Psychophysics of categorical perception. *Psychological Review, 84*, 452–471.

Macmillan, N. A., & Schwartz, M. (1975). A probe–signal investigation of uncertain-frequency detection. *Journal of the Acoustical Society of America, 58*, 1051–1058.

Maiwald, D. (1967). Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas. *Acustica, 18*, 193–207.

Mäkelä, J., Hämäläinen, M., Hari, R., & McEvoy, L. (1994). Whole-head mapping of middle-latency auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology, 92*, 414–421.

Manley, G., Köppl, C., & Yates, G. (1997). Activity of primary auditory neurones in the cochlear ganglion of the emu Dromaius novaehollandiae I: Spontaneous discharge, frequency tuning and phase locking. *Journal of the Acoustical Society of America, 101*, 1560–1573.

Massaro, D. W. (1970a). Preperceptual auditory images. *Journal of Experimental Psychology, 85*, 411–417.

Massaro, D. W. (1970b). Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology, 83*, 32–39.

Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review, 69*, 124–145.

Massaro, D. W., & Idson, W. L. (1977). Backward recognition masking in relative pitch judgments. *Perceptual and Motor Skills, 45*, 87–97.

Massaro, D. W., & Loftus, G. R. (1996). Sensory and perceptual storage. In E. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 67–99). San Diego, CA: Academic Press.

Mauermann, M., Long, G. R., & Kollmeier, B. (2004). Fine structure of hearing threshold and loudness perception. *Journal of the Acoustical Society of America, 116*, 1066–1080.

McEvoy, L., Mäkelä, J., Hari, R., Hämäläinen, M., & Hari, R. (1994). Effect of interaural time differences on middle-latency and late auditory evoked magnetic fields. *Hearing Research, 78*, 249–257.

McFadden, D., Jeffress, L. A., & Russell, W. E. (1973). Individual differences in sensitivity to interaural differences in time and level. *Perceptual and Motor Skills, 37*, 755–761.

McKenna, T. M., Weinberger, N. M., & Diamond, D. M. (1989). Responses of single auditory cortical neurons to tone sequences. *Brain Research, 481*, 142–153.

McLeish, J. (1950). The validation of Seashore's measures of musical talent by factorial methods. *British Journal of Psychology*, *3*, 129–140.

Menning, H., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport*, *11*, 817–822.

Merzenich, M. M., & Brugge, J. F. (1973). Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Research*, *50*, 275–296.

Micheyl, C., Arthaud, P., Reinhart, C., & Collet, L. (2000). Informational masking in normal-hearing and hearing-impaired listeners. *Acta Oto-Laryngologica*, *120*, 242–246.

Micheyl, C., & Carlyon, R. P. (1998). Effect of temporal fringes on fundamental-frequency discrimination. *Journal of the Acoustical Society of America*, *104*, 3006–3018.

Micheyl, C., & Dai, H. (2009). Likelihood ratio, optimal decision rules and relationship between proportion correct and d' in the dual-pair AB-versus-BA identification paradigm. *Attention, Perception, and Psychophysics*, *71*, 1426–1433.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*, 36–47.

Micheyl, C., Kaernbach, C., & Demany, L. (2008). An evaluation of psychophysical models of auditory change perception. *Psychological Review*, *115*, 1069–1083.

Micheyl, C., & Messing, D. P. (2006). Likelihood ratio, optimal decision rules and correct response probabilities in a signal detection theoretic, equal-variance Gaussian model of the observer in the 4IAX paradigm. *Perception and Psychophysics*, *68*, 725–735.

Molholm, S., Martinez, A., Ritter, W., Javitt, D. C., & Foxe, J. J. (2005). The neural circuitry of pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex*, *15*, 545–551.

Mondor, T. A., & Breau, L. M. (1999). Facilitative and inhibitory effects of location and frequency cues: Evidence of a modulation in perceptual sensitivity. *Perception and Psychophysics*, *61*, 438–444.

Mondor, T. A., Breau, L. M., & Milliken, B. (1998). Inhibitory processes in auditory selective attention: Evidence of location-based and frequency-based inhibition of return. *Perception and Psychophysics*, *60*, 296–302.

Mondor, T. A., & Bregman, A. S. (1994). Allocating attention to frequency regions. *Perception and Psychophysics*, *56*, 268–276.

Moore, B. C. J. (1973a). Frequency-difference limens for short duration tones. *Journal of the Acoustical Society of America*, *54*, 610–619.

Moore, B. C. J. (1973b). Some experiments relating to the perception of complex tones. *Quarterly Journal of Experimental Psychology*, *25*, 451–475.

Moore, B. C. J. (1974). Relation between critical bandwidth and frequency-difference limen. *Journal of the Acoustical Society of America*, *55*, 359–359.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). San Diego, CA: Academic Press.

Moore, B. C. J., & Glasberg, B. R. (1986). The role of frequency selectivity in the perception of loudness, pitch and time. In B. C. J. Moore (Ed.), *Frequency selectivity in hearing* (pp. 251–308). London: Academic Press.

Moore, B. C. J., & Glasberg, B. R. (1989). Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation. *Journal of the Acoustical Society of America*, *86*, 1722–1732.

Moore, B. C. J., Glasberg, B. R., Low, K. E., Cope, T., & Cope, W. (2006). Effects of level and frequency on the audibility of partials in inharmonic complex tones. *Journal of the Acoustical Society of America*, *120*, 934–944.

Moreau, P., Jolicoeur, P., & Peretz, I. (2009). Automatic brain responses to pitch changes in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 191–194.

Morel, A., Garraghty, P. E., & Kaas, J. H. (1993). Tonotopic organzation, architectonic fields, and connections of auditory cortex in macaque monkeys. *Journal of Comparative Neurology*, *335*, 437–459.

Morgan, C. T., Garner, W. R., & Galambos, R. (1951). Pitch and intensity. *Journal of the Acoustical Society of America*, *23*, 658–663.

Mosher, J. C., Leahy, R. M., & Lewis, P. S. (1999). EEG and MEG: Forward solutions for inverse methods. *IEEE Transactions on Bio-Medical Engineering*, *46*, 245–259.

Mulert, C., Leicht, G., Pogarell, O., Mergl, R., Karch, S., Juckel, G., et al. (2007). Auditory cortex and anterior cingulate cortex sources of the early evoked gamma-band response: Relationship to task difficulty and mental effort. *Neuropsychologia*, *45*, 2294–2306.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*, 2544–2590.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *Journal of the Acoustical Society of America*, *98*, 1909–1920.

Neff, D. L., & Callaghan, B. P. (1988). Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. *Journal of the Acoustical Society of America*, *83*, 1833–1838.

Neff, D. L., & Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *Journal of the Acoustical Society of America, 98,* 125–134.

Neff, D. L., Dethlefs, T. M., & Jesteadt, W. (1993). Individual differences in simultaneous masking with random frequency, multicomponent maskers. *Journal of the Acoustical Society of America, 94,* 3112–3126.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception and Psychophysics, 41,* 409–415.

Neff, D. L., Kessler, C. J., & Dethlefs, T. M. (1996). Sex differences in simultaneous masking with random-frequency maskers. *Journal of the Acoustical Society of America, 100,* 2547–2550.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Nelson, D. A., Stanton, M. E., & Freyman, R. L. (1983). A general equation describing frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America, 73,* 2117–2123.

Noble, W., & Gatehouse, S. (2004). Interaural asymmetry of hearing loss, speech, spatial and qualities of hearing scale (SSQ) disabilities, and handicap. *International Journal of Audiology, 43,* 100–114.

Noble, W., Tyler, R., Dunn, C. C., & Bhullar, N. (2009). Younger- and older-age adults with unilateral and bilateral cochlear implants: Speech and spatial hearing self-ratings and performance. *Otology and Neurotology, 30,* 921–929.

Nordmark, J. O. (1968). Mechanisms of frequency discrimination. *Journal of the Acoustical Society of America, 44,* 1533–1540.

Noreen, D. (1981). Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology: Proceedings of the symposium in applied mathematics of the american mathematical society and the society for industrial and applied mathematics* (Vol. 13, pp. 237–279). Providence, RI: American Mathematical Society.

Oh, E. L., & Lutfi, R. A. (1998). Nonmonotonicity of informational masking. *Journal of the Acoustical Society of America, 104,* 3489–3499.

Oh, E. L., & Lutfi, R. A. (2000). Effect of masker harmonicity on informational masking. *Journal of the Acoustical Society of America, 108,* 706–709.

Ohgushi, K., & Hatoh, T. (1989). On the perception of the musical pitch of high frequency tones. *Proceedings of the 13th International Congress on Acoustics, Belgrade, Yugoslavia, 3,* 27–30.

Okada, Y. (1983). Neurogenesis of evoked magnetic fields. In S. H. Williamson, G. L. Romani, L. Kaufman, & I. Modena (Eds.), *Biomagnetism: An interdisciplinary approach* (pp. 399–408). New York, NY: Plenum Press.

Okada, Y. (2003). Toward understanding the physiological origins of neuromagnetic signals. In Z. L. Lu & L. Kaufman (Eds.), *Magnetic source imaging of the human brain* (pp. 43–76). Mahwah, NJ: Lawrence Erlbaum Associates.

Okamoto, H., Stracke, H., Draganova, R., & Pantev, C. (2009). Hemispheric asymmetry of auditory evoked fields elicited by spectral versus temporal stimulus change. *Cerebral Cortex, 19*, 2290–2297.

Oxenham, A. J., Bernstein, J. G. W., & Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences, 101*, 1421–1425.

Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G. Jr. (2003). Informational masking and musical training. *Journal of the Acoustical Society of America, 114*, 1543–1549.

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1991). Right-hemisphere dominance of different mismatch negativities. *Electroencephalography and Clinical Neurophysiology, 78*, 466–479.

Paavilainen, P., Jaramillo, M., Näätänen, R., & Winkler, I. (1999). Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neuroscience Letters, 265*, 179–182.

Paavilainen, P., Karlsson, M., Reinikainen, K., & Näätänen, R. (1989). Mismatch negativity to change in spatial location of an auditory stimulus. *Electroencephalography and Clinical Neurophysiology, 73*, 129–141.

Paavilainen, P., Saarinen, J., Tervaniemi, M., & Näätänen, R. (1995). Mismatch negativity to changes in abstract sound features during dichotic-listening. *Journal of Psychophysiology, 9*, 243–249.

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

Pantev, C., Bertrand, C., O.and Eulitz, Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology, 94*, 26–40.

Pantev, C., Engelien, A., Candia, V., & Elbert, T. (2001). Representational cortex in musicians. plastic alterations in response to musical practice. *Annals of the New York Academy of Sciences, 930*, 300–314.

Pantev, C., Eulitz, C., Elbert, T., & Hoke, M. (1994). The auditory evoked sustained field: Origin and frequency dependence. *Electroencephalography and Clinical Neurophysiology, 90*, 82–90.

Pantev, C., Hoke, M., Lehnertz, K., & Lütkenhöner, B. (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science, 246*, 486–488.

Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G., & Wittkowski,

W. (1988). Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology, 69*, 160–170.

Pantev, C., Hoke, M., Lütkenhöner, B., Fahrendorf, G., & Stöber, U. (1990). Identification of sources of brain neuronal activity with high spatiotemporal resolution through combination of neuromagnetic source localization (NMSL) and magnetic resonance imaging (MRI). *Electroencephalography and Clinical Neurophysiology, 75*, 173–184.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*, 811–814.

Pardo, P. J., & Sams, M. (1993). Human auditory-cortex responses to rising versus falling glides. *Neuroscience Letters, 159*, 43–45.

Parkkonen, L., Fujiki, N., & Mäkelä, J. (2009). Sources of auditory brainstem responses revisited: Contribution by magnetoencephalography. *Human Brain Mapping, 30*, 1772–1782.

Patterson, R., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36*, 767–776.

Pelizzone, M., Hari, R., Mäkelä, J., Huttunen, J., Ahlfors, S., & Hämäläinen, M. (1987). Cortical origin of middle-latency auditory evoked responses in man. *Neuroscience Letters, 82*, 303–307.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *J. Opt. Soc. Am. A., 2*, 1508–1532.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functionalmagnetic resonance imaging. *Journal of Neuroscience, 24*, 6810–6815.

Penhune, V. R., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex, 6*, 661–672.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33*, 185–191.

Peretz, I., Brattico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Annals of Neurology, 58*, 478–482.

Peretz, I., Champod, S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999*, 58–75.

Peretz, I., Cummings, S., & Dube, M. P. (2007). The genetics of congenital amusia (tone deafness): A family-aggregation study. *American Journal of Human Genetics, 81*, 582–588.

Pfeuty, M., & Peretz, I. (2010). Abnormal pitch–time interference in congenital amusia: Evidence from an implicit test. *Attention, Perception, and Psychophysics, 72*, 763–774.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics, 16*, 283–290.

Plack, C. J., & Oxenham, A. J. (2005a). *Pitch: Neural coding and perception* (C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper, Eds.). New York, NY: Springer.

Plack, C. J., & Oxenham, A. J. (2005b). The psychophysics of pitch. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 7–55). New York, NY: Springer.

Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P., Rowley, H. A., Wexler, K., et al. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Brain Research. Cognitive Brain Besearch, 4*, 321–242.

Pollack, I. (1975). Auditory informal masking. *Journal of the Acoustical Society of America, 57*, 85.

Pressnitzer, D., & Patterson, R. D. (2001). Distortion products and the pitch of harmonic complex tones. In D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, P. V. F., & R. Schoonhoven (Eds.), *Physiological and psychophysical bases of auditory function.* Maastricht: Shaker.

Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *Neuroimage, 49*, 1641–1649.

Rakowski, A., & Hirsh, I. J. (1980). Post-stimulatory pitch shifts for pure tone. *Journal of the Acoustical Society of America, 68*, 467–474.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience, 13*, 87–103.

Rieke, F., Warland, D., Steveninck, R. de Ruyter van, & Bialek, W. (1997). *Spikes: Exploring the neural code.* Cambridge, MA: MIT Press.

Ritter, S., Dosch, H. G., Specht, H.-J., & Rupp, A. (2005). Neuromagnetic responses reflect the pitch changes of regular interval sounds. *Neuroimage, 27*, 533–543.

Ritter, S., Dosch, H. G., Specht, H.-J., Schneider, P., & Rupp, A. (2007). Latency effect of the pitch response due to variations of frequency and spectral envelope. *Clinical Neurophysiology, 118*, 2276–2281.

Roberts, T. P., Ferrari, P., & Poeppel, D. (1998). Latency of evoked neuromagnetic M100 reflects perceptual and acoustic stimulus attributes. *NeuroReport, 14*, 3265–3269.

Roberts, T. P., & Poeppel, D. (1996). Latency of auditory evoked M100 as a function of tone frequency. *NeuroReport*, *7*, 1138–1140.

Robinson, K., & Summerfield, A. Q. (1996). Adult auditory learning and training. *Ear and Hearing*, *17*, 51S–65S.

Romani, G., Williamson, S. J., & Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, *216*, 1339–1340.

Rosburg, T. (2003). Left hemispheric dipole locations of the neuromagnetic mismatch negativity to frequency, intensity and duration deviants. *Cognitive Brain Research*, *16*, 83–90.

Rosburg, T. (2004). Effects of tone repetition on auditory evoked neuromagnetic fields. *Clinical Neurophysiology*, *115*, 898–905.

Rosburg, T., Haueisen, J., & Sauer, H. (2002). Stimulus duration influences the dipole location shift within the auditory evoked field component N100m. *Brain Topography*, *15*, 37–41.

Rosenblith, W. A., & Stevens, K. N. (1953). On the DL for frequency. *Journal of the Acoustical Society of America*, *25*, 980–985.

Rossing, T. D., & Houtsma, A. J. M. (1986). Effects of signal envelope on the pitch ofshort sinusoidal tones. *Journal of the Acoustical Society of America*, *76*, 1926–1933.

Ruusuvirta, T. (2000). Proactive interference of a sequence of tones in a two-tone pitch comparison task. *Psychonomic Bulletin and Review*, *7*, 327–331.

Ruusuvirta, T., Wikgren, J., & Astikainen, P. (2008). Proactive interference in a two-tone pitch-comparison task without additional interfering tones. *Psychological Research*, *72*, 74–78.

Saarinen, J., Paavilainen, P., Schröger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory-stimuli in the human brain. *NeuroReport*, *2*, 1149–1151.

Sachs, M. B., Woolf, N. K., & Sinnott, J. M. (1980). Response properties of neurons in the avian auditory system: comparisons with mammalian homologues and consideration of the neural encoding of complex stimuli. In A. N. Popper & R. R. Fay (Eds.), *Comparative studies of hearing in vertebrates.* Berlin: Springer.

Salajegheh, A., Link, A., Elster, C., Burghoff, M., Sander, T., Trahms, L., et al. (2004). Systematic latency variation of the auditory evoked M100: From average to single-trial data. *Neuroimage*, *23*, 288–95.

Salvi, R. J., Saunders, S. S., Powers, N. L., & Boettcher, F. A. (1992). Discharge patterns of cochlear ganglion neurons in the chicken. *Journal of Comparative Physiology*, *170*, 227–241.

Sams, M., Hämäläinen, M., Antervo, A., Kaukoranta, E., Reinikainen, K., & Hari, R. (1985). Cerebral neuromagnetic responses evoked by short auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, *61*, 254–256.

Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine and Biology, 32*, 11–22.

Scherg, M., Hari, R., & Hämäläinen, M. (1989). Frequency-specific sources of the auditory N19–P30–P50 response detected by a multiple source analysis of evoked magnetic fields and potentials. In S. Williamson (Ed.), *Advances in biomagnetism* (pp. 97–100). New York, NY: Plenum Press.

Schlauch, R. S., & Hafter, E. R. (1991). Listening bandwidths and frequency uncertainty in pure-tone signal-detection. *Journal of the Acoustical Society of America, 90*, 1332–1339.

Schlaug, G. (2001). The brain of musicians. A model for functional and structural adaptation. *Annals of the New York Academy of Sciences, 930*, 281–299.

Schönwiesner, M., Rubsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience, 22*, 1521–1528.

Schönwiesner, M., & Zatorre, R. J. (2008). Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. *Experimental Brain Research, 187*, 97–105.

Schröger, E. (1995). Processing of auditory deviants with changes in one versus two stimuli dimensions. *Psychophysiology, 32*, 55–65.

Scott, S. K., Blank, S. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*, 2400–240.

Scott, S. K., & Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition, 92*, 13–45.

Seashore, C. E. (1919a). *Manual of instructions and interpretations for measures of musical talent.* New York, NY: Columbia Graphophone Co.

Seashore, C. E. (1919b). *The psychology of musical talent.* Boston, NY: Silver, Burdett, and Company.

Seashore, C. E. (1938). *Psychology of music.* New York, NY; London: McGraw-Hill.

Seashore, C. E., Lewis, D., & Saetveit, J. C. (1960). *Manual of instructions and interpretations for the seashore measures of musical talent.* New York, NY: The Psychological Corporation.

Seither-Preisler, A., Krumbholz, K., & Lütkenhöner, B. (2003). Sensitivity of the neuromagnetic N100m deflection to spectral bandwidth: A function of the auditory periphery? *Audiology and Neuro-otology, 8*, 322–337.

Seither-Preisler, A., Krumbholz, K., Patterson, R. D., Seither, S., & Lütkenhöner, B. (2004). Interaction between the neuromagnetic responses to sound energy onset and pitch onset suggests common generators. *European Journal of Neuroscience, 19*, 3073–3080.

Seither-Preisler, A., Patterson, R. D., Krumbholz, K., Seither, S., & Lütkenhöner, B. (2006a). Evidence of pitch processing in the n100m component of the auditory evoked field. *Hearing Research, 213*, 88–98.

Seither-Preisler, A., Patterson, R. D., Krumbholz, K., Seither, S., & Lütkenhöner, B. (2006b). From noise to pitch: Transient and sustained responses of the auditory evoked field. *Hearing Research*, *218*, 50–63.

Sek, A., & Moore, B. C. J. (1995). Frequency discrimination as a function of frequency, measured in several ways. *Journal of the Acoustical Society of America*, *97*, 2479–2486.

Sek, A., & Moore, B. C. J. (1999). Discrimination of frequency steps linked by glides of various durations. *Journal of the Acoustical Society of America*, *106*, 351–359.

Semal, C., & Demany, L. (1990). The upper limit of "musical" pitch. *Music Perception*, *8*, 165–176.

Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America*, *89*, 2404–2410.

Semal, C., & Demany, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *Journal of the Acoustical Society of America*, *94*, 1315–1322.

Semal, C., & Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *Journal of the Acoustical Society of America*, *120*, 3907–3915.

Sheldon, H. (1981). Structure of human auditory cortex. II: Axon distributions and morphological correlates of speech perception. *Brain Research*, *229*, 295–310.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 343–390). New York, NY: Academic Press.

Sherg, M. (1990). Fundamentals of dipole source potential analysis. In F. Grandori, M. Hoke, & G. L. Romani (Eds.), *Advances in audiology* (Vol. 6, pp. 40–69). Karger, Basel.

Shower, E. G., & Biddulph, R. (1931). Differential pitch sensitivity of the ear. *Journal of the Acoustical Society of America*, *3*, 275–287.

Siegel, R. J. (1965). A replication of the mel scale of pitch. *American Journal of Psychology*, *78*, 615–620.

Sloboda, J. A., Wise, K. J., & Peretz, I. (2005). Quantifying tone deafness in the general population. *Annals of the New York Academy of Sciences*, *1060*, 255–261.

Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America*, *118*, 3177–3186.

Smolders, J. W. T., & Klinke, R. (1986). Synchronized responses of primary auditory fiber-populations in caiman crocodilus (L.) to single tones and clicks. *Hearing Research*, *24*, 89–103.

Snow, W. B. (1936). Change of pitch with loudness at low frequencies. *Journal of the Acoustical Society of America*, *8*, 14–19.

Sparks, D. W. (1976). Temporal recognition masking—or interference? *Journal of the Acoustical Society of America*, *60*, 1347–1353.

Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology*, *15*, 201–292.

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and non-musicians. *Journal of the Acoustical Society of America, 76*, 1690–1695.

Stankov, L., & Horn, J. L. (1980). Human abilities revealed through auditory tests. *Journal of Educational Psychology*, *72*, 21–44.

Starr, G. E., & Pitt, M. A. (1997). Interference effects in short-term memory for timbre. *Journal of the Acoustical Society of America*, *102*, 486–494.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Stevens, S. S. (1935). The relation of pitch to intensity. *Journal of the Acoustical Society of America*, *5*, 150–154.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude of pitch. *Journal of the Acoustical Society of America*, *8*, 185–190.

Stufflebeam, S. M., Poeppel, D., Rowley, H. A., & Roberts, T. P. (1998). Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. *NeuroReport*, *9*, 91–94.

Suzuki, Y., & Takeshima, H. (2004). Equal-loudness-level contours for pure tones. *Journal of the Acoustical Society of America*, *116*, 918–933.

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (2001). Preattentive processing of spectral, temporal and structural characteristics of acoustic regularities: A mismatch negativity study. *Psychophysiology*, *38*, 92–98.

Tanner, W. P., & Rivette, G. L. (1964). Experimental study of 'tone deafness'. *Journal of the Acoustical Society of America*, *36*, 1465–1467.

Tanner, W. P. Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*, 401–409.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, *161*, 1–10.

Tervaniemi, M., Maury, S., & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport*, *5*, 844–846.

Tervaniemi, M., Rytkonen, M., Schroger, E., Ilmoniemi, R. J., & Naatanen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learning and Memory*, *8*, 295–300.

Tervaniemi, M., Winkler, I., & Näätänen, R. (1997). Pre-attentive categorization of sounds by timbre as revealed by event-related potentials. *NeuroReport*, *8*, 2571–2574.

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago, IL: University of Chicago Press.

Tillmann, B., Jolicoeur, P., Ishihara, M., Gosselin, N., Bertrand, O., Rossetti, Y., et al. (2010). The amusic brain: Lost in music, but not in space. *PloS one*, *5*, e10173.

Tillmann, B., Peretz, I., Bigand, E., & Gosselin, N. (2007). Harmonic priming in an amusic patient: the power of implicit tasks. *Cognitive neuropsychology*, *24*, 603–622.

Tramo, M. J., Shah, G. D., & Braida, L. D. (2002). Functional role of auditory cortex in frequency processing and pitch perception. *Journal of Neurophysiology*, *87*, 122–139.

Verschuure, J., & van Meeteren, A. A. (1975). The effect of intensity on pitch. *Acustica*, *32*, 33–44.

von Békésy, G. (1960). *Experiments in hearing.* New York, NY: McGraw-Hill.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*, 274–290.

Ward, L. M., & Mori, S. (1996). Attention cueing aids auditory intensity resolution. *Journal of the Acoustical Society of America*, *100*, 1722–1727.

Ward, W. D. (1954). Subjective musical pitch. *Journal of the Acoustical Society of America*, *26*, 369–380.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 10038–10042.

Watson, B. U. (1991). Some relationships between intelligence and auditory discrimination. *Journal of Speech and Hearing Research*, *34*, 621–627.

Watson, B. U., & Miller, T. K. (1993). Auditory perception, phonological processing, and reading ability/disability. *Journal of Speech and Hearing Research*, *36*, 850–863.

Watson, C. S. (2004). Temporal acuity and the identification of temporal order: Related, but distinct, auditory abilities. *Seminars in Hearing*, *25*, 219–227.

Watson, C. S., Johnson, D. M., Lehman, J. R., Kelly, W. J., & Jensen, J. K. (1982). An auditory discrimination test battery. *Journal of the Acoustical Society of America*, *71*, S73.

Watson, C. S., Kelly, W., & Wroton, H. W. (1976). Factors in the discrimination of tonal patterns II: Selective attention and learning under various levels of stimulus uncertainty. *Journal of the Acoustical Society of America*, *60*, 1176–1186.

Wechsler, D. (1981). *Manual for the wechsler adult intelligence scale — revised.* New York, NY: The Psychological Corporation.

Weinberger, N. M., & McKenna, T. M. (1988). Sensitivity of single neurons in auditory cortex to contour: Toward a neurophysiology of music perception. *Music*

Perception, 5, 355–390.

Weiss, T. F., & Rose, C. (1988). A comparison of synchronization filters in different auditory receptor organs. *Hearing Research, 33*, 175–180.

Whitfield, I. C. (1967). *The auditory pathway.* Baltimore, MD: Williams and Wilkins.

Whitfield, I. C., & Evans, E. F. (1965). Responses of auditory cortical neurons to stimuli of changing frequency. *Journal of Neurophysiology, 28*, 655–672.

Wier, C. C., Jesteadt, W., & Green, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America, 61*, 178–184.

Williamson, V., McDonald, C., Deutsch, D., Griffiths, T. D., & Stewart, L. (2010). Faster decline of pitch memory over time in congenital amusia. *Advances in Cognitive Psychology, 6*, 15-22.

Williamson, V., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch perception problem. *Memory, 18*, 657-669.

Wing, H. D. (1962). A revision of the Wing musical aptitude test. *Journal of Research in Musical Education, 10*, 743–791.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*, 420–422.

Woods, D. L. (1995). The component structure of the n1 wave of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology Supplement, 44*, 102–109.

Wright, B. A. (1998). Generalization of auditory-discrimination learning. *Abs. for Association for Research in Otolaryngology, 413*, 104.

Wright, B. A., & Saberi, K. (1999). Strategies used to detect auditory signals in small sets of random maskers. *Journal of the Acoustical Society of America, 105*, 1765–1775.

Yamashiro, K., Inui, K., Otsuru, N., & Kakigi, R. (2010). Change-related responses in the human auditory cortex: An MEG study. *Psychophysiology*, advanced online publication.

Yip, M. (2002). *Tone.* Cambridge: Cambridge University Press.

Yoshiura, T., Ueno, S., Iramina, K., & Masuda, K. (1995). Source localization of middle latency auditory evoked magnetic fields. *Brain Research, 703*, 139–144.

Yost, W. A. (1996a). Pitch of iterated rippled noise. *Journal of the Acoustical Society of America, 100*, 511–518.

Yost, W. A. (1996b). Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America, 100*, 3329–3335.

Yost, W. A., Berg, K., & Thomas, G. B. (1976). Frequency recognition in temporal interference tasks: A comparison among four psychophysical procedures. *Perception and Psychophysics, 20*, 353–359.

Yvert, B., Crouzeix, A., Bertrand, O., Seither-Preisler, A., & Pantev, C. (2001). Multiple supratemporal sources of magnetic and electric auditory evoked middle latency components in humans. *Cerebral Cortex, 11*, 411–423.

Zatorre, R. J. (2003). Neural specializations for tonal processing. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 231–246). Oxford: Oxford University Press.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946–953.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience, 14*, 1908–1919.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch processing in speech perception. *Science, 256*, 846–849.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*, 1087–1104.

Zhang, L. I., Tan, A. Y. Y., Schreiner, C. E., & Merzenich, M. M. (2003). Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature, 424*, 201–205.

Zwicker, E. (1970). Masking and psychological excitation as consequences of the ear's frequency analysis. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing.* Leiden: Sijthoff.

Zwicker, E., & Fastl, H. (1990). *Psychoacoustics: facts and models.* Berlin; New York, NY: Springer.