UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
INSTITUTE OF SOUND AND VIBRATION RESEARCH

LEARNING TO DISCRIMINATE INTERAURAL TIME DIFFERENCES

BY

DANIEL ROWAN

THESIS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

2006

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
INSTITUTE OF SOUND AND VIBRATION RESEARCH

Doctor of Philosophy

LEARNING TO DISCRIMINATE INTERAURAL TIME DIFFERENCES

by Daniel Rowan

Normal–hearing adults can learn to discriminate sensory stimuli that were initially indiscriminable. For example, the smallest discriminable change in acoustic frequency may reduce by a factor of 2 or more across thousands of trials. It has been suggested that this may partly reflect changes in auditory processing of frequency. Few studies have considered such learning in the context of spatial perception. One study trained listeners on discrimination with the primary cues to horizontal–plane localisation: interaural time difference (ITD) at low frequency and interaural level difference (ILD) at high frequency. The authors suggested that these conditions were associated with fundamentally different characteristics of learning, such as the time–course of learning appearing to be longer for ILD than ITD. These results may have important implications for clinical populations. For example, complex electronic prostheses are increasingly used to alleviate bilateral hearing loss but may distort the normal perceptual representation of binaural cues. While users may learn to utilise disrupted ILDs, it is less clear if this would be the case with ITDs. The aim of this study was to investigate further the apparent disparity in learning between these conditions, leading to a better understanding of the conditions required for learning in normal–hearing listeners as background to future studies with hearing–impaired populations.

Three experiments were undertaken on learning to discriminate 'ongoing' ITD at low and high frequencies, the latter using amplitude–modulated tones. In Experiment 1, ability improved substantially with training using high–frequency stimuli associated with fused or unfused percepts. This was in apparent contrast with the findings of the previous study at low frequencies. In Experiment 2, indirect evidence was found for differential learning with ITD discrimination at low and high frequencies through a comparison of inexperienced and experienced listeners using stimuli associated with comparable asymptotic performance. However, this was not confirmed by Experiment 3 which measured the time–courses of learning directly. This latter experiment also found that the time–courses were broadly comparable to those reported by the previous study with ILD. Learning was also found to generalise across frequency, unlike the findings of the previous study with ILD. A detailed examination of the data from the previous study indicated that their data on ITD was difficult to interpret and that the authors' interpretation of different time–courses of learning between conditions may not be justified. It is concluded that training influences discrimination of ITD and ILD in a broadly comparable manner. Nonetheless, subtle, but potentially important, differences may exist. Future research is required to explore further the specific conditions required for learning with localisation cues, differences in learning between cues and the implications of this learning for hearing–impaired populations.

Contents

List of figures

List of tables

Acknowledgements

I would like to acknowledge and thank my PhD supervisor Prof. Mark Lutman. His academic and technical support, suggestions during the study and comments on previous drafts of the dissertation were invaluable. I particularly appreciated his wisdom, faith that I could and would complete the PhD and apparent limitless supply of patience. I also thank Drs. Sarosh Kapadia and Kevin Munro for their inspiration and encouragement to undertake the PhD, and my review panel, Drs. Kapadia and Ian Flindell, for their time and patience.

I thank Dr. Beverly Wright, Northwestern University, for kindly and generously sharing her data and her post–graduate student Yuxuan Zhang for answering copious questions. I also thank Dr. Les Bernstein, University of Connecticut Health Centre, for sharing data from one of his 'transposed' experiments, for general advice and for hosting a visit to his laboratory (and for the jazz). I acknowledge Mr. Antony Wood and Dr. Ollie Bauman, University of Southampton, for technical assistance and help with programming, respectively. Countless others have provided, one way or another, opinions, advice, guidance, insight or encouragement that has been much appreciated. Jane Burgneay, Roberta Buhagiar and Mona Mahmoud deserve special mention for their calming influence during the PhD. I would also like to thank those individuals who participated in my experiments.

I would like to dedicate this thesis to my family, particularly my parents. Thanks Mum and Dad for nurturing my tenacity and confidence, and for your unfaltering encouragement and support; thanks my two girls for your love, patience and raspberries; thanks Rebecca for your love, support and cake.

List of abbreviations

AFC	Alternative forced choice
AM	Amplitude modulated or modulation
ANOVA	Analysis of variance
CF	Characteristic frequency
CN	Cochlear nucleus
IC	Inferior colliculus
IFD	Interaural carrier frequency difference
ILD	Interaural level difference
IPD	Interaural phase difference
ITD	Interaural time difference
LSO	Lateral superior olive
MLD	Masking level difference
MSO	Medial superior olive
SAM	Sinusoidal amplitude modulation
SC	Superior colliculus
SD	Standard deviation
SE	Standard error
SOC	Superior olivary complex
SPL	Sound pressure level

Chapter 1. Introduction

Normal–hearing adults can learn to discriminate sensory stimuli that were initially indiscriminable. A classic example of this is the improvement in the ability of wine–tasters to distinguish between similar wines with experience. However, this also occurs with relatively simple stimuli, such as the discrimination of the alignment of short lines or the frequency of pure tones, and relatively simple tasks, such as detecting which of three consecutive stimuli is the 'odd one out'. The past two decades have witnessed a proliferation of evidence that the structure and function of the mammalian adult brain is plastic; that is, influenced by experience (Calford, 2002; Buonomano and Merzenich, 1998; Gilbert, 1998). It is widely held that learning on simple discrimination tasks with simple stimuli may provide insight into plasticity in sensory processing. However, the specific conditions under which learning occurs, the associated plasticity and the implications for clinical populations remain unclear.

There is a substantial body of evidence on learning with discrimination tasks (hereafter referred to as 'discrimination learning') with visual perception (Fahle and Poggio, 2002). Recent years have also witnessed a growth in research with auditory discrimination learning (Wright, 2001). This stems in part from a report that this learning may be correlated with changes in the neural representation of stimuli (Recanzone et al., 1993) and from observations that the auditory capabilities of users of auditory prostheses often improve markedly with experience (Clarke, 2002; Tyler and Summerfield, 1996). Research is also motivated by the possibility that training on simple discrimination tasks may be used to treat auditory and language–based disorders and to maximise the use of auditory prostheses (Wright, 2001).

Of the studies of auditory discrimination learning, few have considered it in the context of spatial perception. This is despite considerable evidence of learning and plasticity with other aspects of spatial performance (Moore and King, 2004). One study trained normal–hearing listeners to discriminate two acoustic cues to spatial position with pure tones: interaural time difference (ITD) at low frequency and interaural level difference (ILD) at high frequency (Wright and Fitzgerald, 2001). These cues arise naturally from having two ears either side of the head and when a sound source is closer to one ear than the other. The advantages of having two ears over one arise from the perception of these 'binaural' cues. However, ITD and ILD can also be presented over earphones, as was done by Wright and Fitzgerald, allowing each cue to be varied independently. Wright and Fitzgerald compared learning in the trained groups to that in untrained, control groups. With ITD, the authors concluded that the trained group did not learn more than the untrained group (i.e. 'training–induced learning' did not occur) and the learning that was observed occurred over a short time–course. With ILD, they concluded that training–induced learning was apparent and thus learning occurred over a longer time–course than with ITD. Further, they concluded that less learning on ILD was apparent with untrained stimuli at a different frequency (i.e. learning 'generalised' incompletely). They suggested that only the learning with ILD was potentially related to auditory plasticity.

This research may have important implications for clinical populations. Hearing impairment is usually associated with impaired binaural hearing and it has been suggested that the use of electronic prostheses, such as hearing aids or cochlear implants, may restore binaural hearing. Studies of discrimination learning in normal–hearing people may provide insight into the potential for hearing–impaired users of auditory prostheses to learn to use the information provided by the protheses and whether clinical auditory training may facilitate this process. While Wright and Fitzgerald's results suggest this may be the case with ILD, it is less clear with ITD.

The present study sought to investigate further discrimination learning with binaural cues and the potential implications for hearing–impaired populations. A review of the literature was conducted and is described in Chapter 2. This review indicated a number of gaps in knowledge. For example, it is not clear if the difference in time–courses of learning reported by Wright and Fitzgerald (2001) reflects the different frequencies or cues. Further, learning on ITD discrimination across training sessions was apparent with a number of Wright and Fitzgerald's listeners, and the authors did not compared directly learning between trained and untrained groups. There is therefore uncertainty as to whether Wright and Fitzgerald's conclusions are justified. Nevertheless, if the time–course of binaural discrimination learning depends on the cue and/or frequency, all else being equal, this may have implications for binaural theory. Further research with ITD discrimination learning in particular may also have implications for clinical populations. Whilst, ITD is thought to be the dominant binaural cue for localisation in normal–hearing people (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002) these cues may be distorted by auditory prostheses (Wilson et al., 2003). Acoustic studies of ITD discrimination learning may provide greater insight into the potential for learning and plasticity under these conditions. However, the stimuli used by Wright and Fitzgerald (2001) may not be relevant to the complex stimuli presented by prostheses. During the literature review, it was also found that few studies of ITD sensitivity with hearing–impaired people report providing training. In contrast, studies using normal–hearing listeners usually provide extensive training. Uncertainty regarding the influence of training complicates comparison across groups.

Based on the literature review summarised above, the broad aims of the experimental part of the study were determined as follows:

(1) Develop suitable methodology for studies of discrimination learning in normal–hearing and hearing–impaired populations.

(2) Gain insight into ITD discrimination learning under conditions that may be relevant to auditory prostheses.

(3) Evaluate the possibility that characteristics of ITD discrimination learning depend on frequency.

Chapter 3 describes the general methodology for investigations of learning. Chapter 4 describes Experiment 1, the primary aims of which were to evaluate this methodology and explore learning under conditions that may be relevant to cochlear implants. Specifically, this experiment addressed issues related to the time–course and generalisation of learning and the effect of IFD in a larger number of listeners than hitherto reported. Sixteen normal–hearing young adults were trained on high–frequency ITD discrimination using amplitude-modulated (AM) stimuli. One stimulus contained an interaural difference in spectral frequency (IFD). Measurements with both stimuli were obtained before and after training. Chapter 5 describes two associated experiments, both of which were aimed partly at evaluating additional refinements in the methodology. Experiment 2a obtained data from five trained listeners using various stimuli. Experiment 2b obtained data with a subset of these stimuli using seven inexperienced listeners. The results from these two experiments were compared to explore indirectly an effect of frequency on ITD discrimination learning. Chapter 6 describes Experiment 3, which was designed to identify direct evidence for an effect of frequency on learning with ITD. Specifically, this experiment addressed issues related to training–induced learning and the time–course of learning at low and high frequencies, generalisation of ITD learning across frequency and inter–individual differences in learning. Chapter 7 describes a detailed comparison of data from Experiment 3 with Wright and Fitzgerald's (2001) data, generously provided by the authors.

The primary contribution to knowledge made by this thesis is a clearer understanding of learning with ITD discrimination. This study shows that learning over many hundreds to thousands of trials is apparent for ITD discrimination with a variety of stimuli and the magnitude and time–course of learning is not inherently strongly dependent on frequency. While these results initially seem inconsistent with Wright and Fitzgerald's conclusions, a re–analysis of their data indicates there may be less disagreement on this issue. The present study also shows that learning on ITD generalises across frequency. This study also contributes to knowledge on ITD sensitivity at low and high frequencies and on the effect of IFD. These findings may have implications for binaural theory and clinical populations. For example, the results indicate that training is important in measuring ITD discrimination in hearing–impaired populations, at least if the aim is to determine potential capability. However, valid measurements of the effect of IFD may not require training to asymptotic levels. This study also raises the possibility that studies that do not provide extensive training may underestimate potential performance and overestimate underlying 'impairment' in ITD discrimination. The finding of training–induced learning that generalises across frequency is also encouraging with regard to the possibility of developing therapeutic training tools based on binaural discrimination.

Aspects of this study have been reported at a number of international auditory research meetings:

Rowan D, Lutman ME. High–frequency ITD discrimination with 'transposed' stimuli in trained and untrained normal–hearing adults. Poster presented at the 27th Annual Mid Winter Research Meeting of the Association for Research in Otolaryngology; 2004 Feb 22–26; Daytona Beach, Florida, USA.

Rowan D, Lutman ME. Effects of training on ITD discrimination using "transposed stimuli": preliminary analysis. Poster presented at the British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness; 2004 Sept 16–17; London, UK.

Rowan D, Lutman ME. Generalisation of learning with ITD discrimination across frequency and type of cue. Poster presented at the 28th Annual Mid Winter Research Meeting of the Association for Research in Otolaryngology; 2005 Feb 19–24; New Orleans, Louisiana, USA.

Rowan D, Lutman ME. Learning on binaural discrimination tasks in humans. Oral presentation at the British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness; 2005 Sept 12–13; Cardiff, Wales.

Chapter 2. Background

2.1 Introduction

This chapter begins with a review of psychoacoustical and physiological experiments on and theories of the binaural perception of unmodulated and modulated tones[1]. Before doing so, terminology used in the field and adopted in this thesis and the basic acoustics of spatial hearing are described.

2.1.1 Terminology

A stimulus can be presented to a single ear (i.e. 'monotic' or 'unilateral' stimulation) or to both ears simultaneously (i.e. 'bilateral' stimulation). The latter can involve identical stimulation of the ears (i.e. 'diotic' stimulation) or stimulation with interaural temporal or spectral disparities (i.e. 'dichotic' stimulation). These terms refer to the manner in which the auditory system is stimulated and do not imply the manner in which the stimulus is then processed. In contrast, monaural hearing or processing refers to auditory capabilities that require the use of only a single ear, and binaural hearing or processing refers 'in a broad sense to the functions underlying any human capabilities that are rendered possible or superior by the use of two ears rather than one' (Grantham, 1995). The term 'binaural sensitivity' is used specifically to refer to the perception of dichotic stimuli, usually over earphones.

The imaginary line that connects the two ears is referred to as the interaural axis. Three–dimensional space is divided into upper and lower halves by the horizontal plane, which passes through the interaural axis, and into right and left halves by the median plane, which is perpendicular to the interaural axis and passes through the centre of the head. The angles *between* a sound source and the median and horizontal planes are referred to as azimuth and elevation, respectively. When presented with a sound via a loudspeaker at zero azimuth and elevation, a normal–hearing listener will generally report hearing a single percept (i.e. a 'fused image') outside of the head (i.e. is 'externalised') and located directly ahead. In contrast, when presented with a diotic

[1]Broader and more detailed reviews of binaural hearing are provided elsewhere. Binaural psychoacoustics: Blauert (1997), Hafter and Trahiotis (1997), Grantham (1995), Durlach and Colburn (1978); binaural neurophysiology: Yin (2002), Tollin (2003), Irvine (1991), Casseday et al. (2002); binaural models: Colburn (1995), Colburn and Durlach (1978), Stern and Trahiotis (1995).

stimulus over earphones, the listener will generally report hearing a fused image *within* the head (i.e. is 'internalised') and midway along the interaural axis. The terms 'localisation' and 'lateralisation' are used when images are externalised and internalised, respectively, and therefore generally with experiments using loudspeakers and earphones, respectively[2]. Perceived location along the interaural axis with lateralisation is referred to as the lateral position or extent of laterality.

Perceptual learning refers to a change in performance on a psychophysical task linked with experience (i.e. interaction with the environment) and presumably reflects a change in structure or function within the brain (i.e. 'plasticity'). Perceptual learning is thought to comprise different forms of learning, reflecting plasticity in different parts of the brain. For example, authors have differentiated between perceptual and cognitive learning (Goldstone, 1998), perceptual and association learning (Hall, 1991) and stimulus, task and procedural learning (Robinson and Summerfield, 1996). Consequently, perceptual, or stimulus, learning is often used specifically to refer to learning assumed to reflect plasticity within the sensory system. However, the precise meaning of these terms and how they interrelate is unclear.

2.1.2 Acoustics of spatial hearing

The interactions of a sound with the torso, head and pinnae result in location–dependent temporal and spectral alterations to the stimulus. For example, when a sound source is on the median plane (i.e. has zero azimuth) the sound arrives at both tympanic membranes simultaneously and with the same sound pressure level (SPL). When the sound source is displaced from the median plane, and is therefore closer to one ear, the sound arrives at this ear momentarily before the other ear. This delay is referred to as the interaural time difference (ITD). The SPL may also be higher at the closer ear due the acoustic shadowing effect of the head, and is referred to as an interaural level difference (ILD). These 'interaural cues' are illustrated in Figure 2.1, and can only be determined through binaural hearing. These cues are systematically related to azimuth, and are therefore useful to horizontal–plane localisation, but not to elevation. Elevation–dependent spectral cues are introduced by the pinnae with complex stimuli above 3000–4000 Hz. Interaural differences in these spectral cues are also dependent on azimuth, albeit less robustly than ITD and ILD (Duda, 1996).

[2] Under special conditions, images associated with stimuli presented over loudspeakers or earphones may be internalised or externalised, respectively (Plenge, 1974; Hartmann and Wittenberg, 1996).

Both ITD and ILD are maximal at an azimuth of ±90°. In adults, the maximum ITD is 600–700 μs, depending on the dimensions of the head but not strongly on frequency. In contrast, ILD varies strongly with frequency, being as much as 25 dB at 4000 Hz and less than 5 dB below 500 Hz (Kuhn, 1987). An ITD may be apparent in the initial or final pressure wave of a stimulus with an abrupt onset or offset, or in the remaining, 'ongoing' portion of a sustained stimulus. This ongoing ITD has a corresponding interaural phase difference (IPD) when the stimulus is periodic, such as with pure tones and amplitude–modulated (AM) tones. An example of an AM tone is provided in Figure 2.2. Here, the peak amplitude of a high–frequency 'carrier' tone varies over time in a manner and at a rate determined by a modulation function. In Figure 2.2, this function (or 'modulator') is a sinusoid, and is known as a sinusoidal amplitude modulation (SAM). The SAM tone has a spectral locus at the frequency of the carrier but a bandwidth of twice the modulation rate (see Section 3.4.2). The shape of the modulator superimposed on the carrier, as in Figure 2.2, is referred to as the waveform envelope. The two panels represent recordings at either ear as in Figure 2.1. The ongoing ITD is apparent in both the carrier (i.e. the waveform 'fine–structure') and the envelope, and it is possible to manipulate these independently over earphones. Ongoing ITD in the fine–structure and envelope are similarly dependent on azimuth in the free–field (Middlebrooks and Green, 1990).

Acoustical considerations led Lord Rayleigh (1907) to propose the duplex theory of horizontal–plane localisation for tones, which states that ongoing ITD determines ability at low frequencies and ILD determines ability at high frequencies. Ongoing ITD–based localisation is predicted to be difficult above 700–800 Hz because of phase ambiguities that occur when half the period of a tone is equal to and greater than the maximum ITD afforded by the head; ILD–based localisation is predicted to be difficult below 500 Hz because of the minimal acoustic shadowing effect of the head. The role of envelope–based ongoing ITD in localisation is not explicit in this theory. However, phase ambiguities would occur with the envelope rather than the fine–structure, and so there is not an obvious high–frequency spectral limit on this cue based on acoustical considerations.

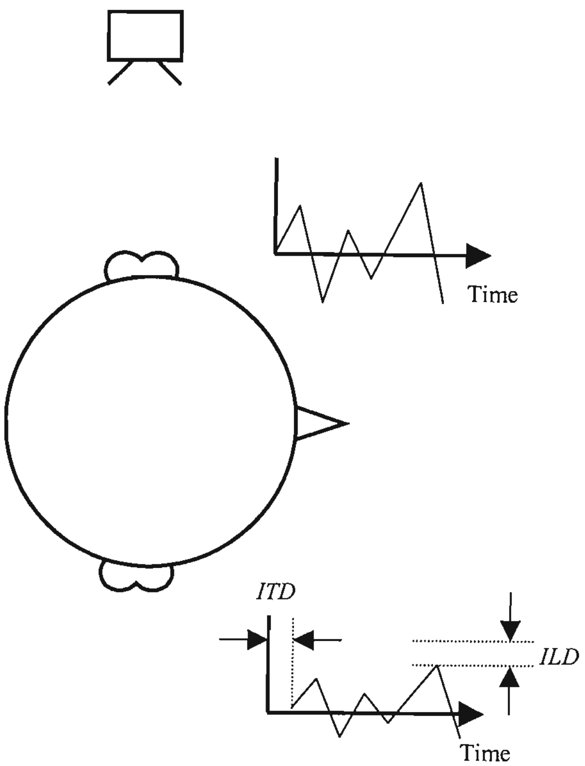


Figure 2.1. Schematic illustration of the interaural cues in the waveforms measured at the two ears: interaural time difference (ITD) and interaural level difference (ILD).

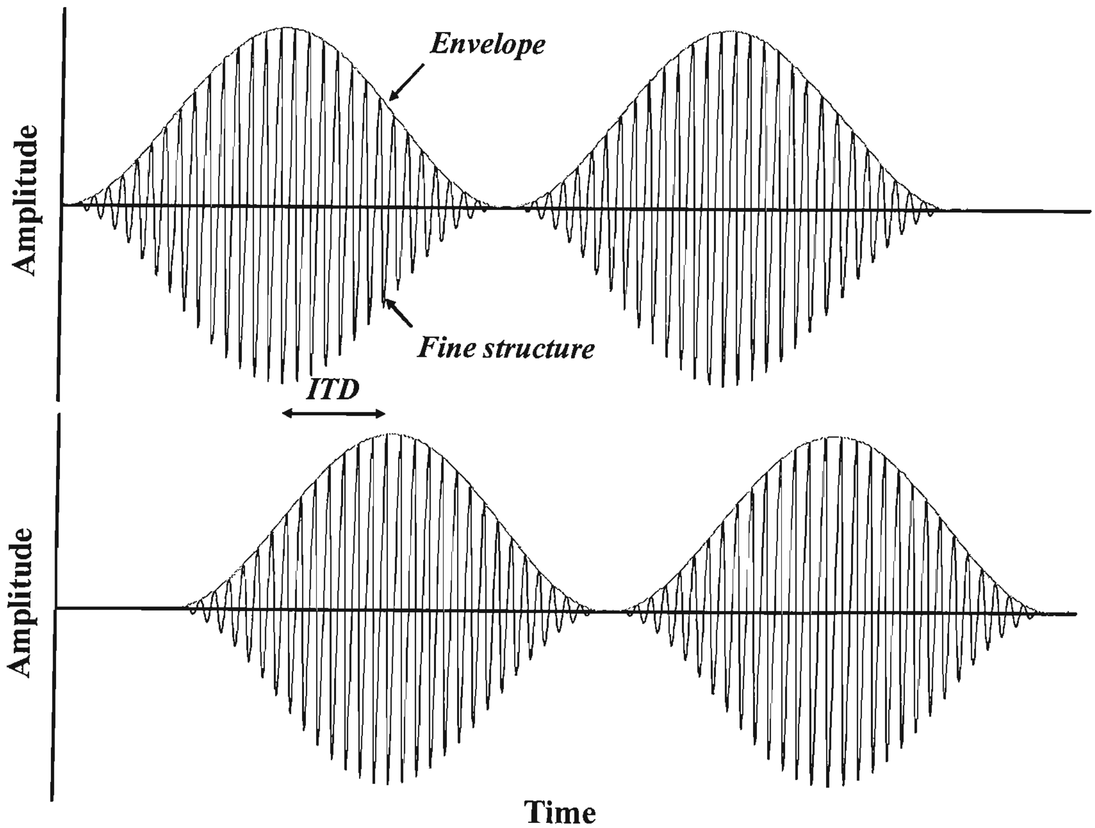


Figure 2.2. Waveform of an AM tone (sinusoidal amplitude modulation, SAM) recorded at each ear, as in Figure 2.1. The envelope of stimulus is superimposed in grey.

2.2 Binaural psychoacoustics

Three general aspects of most psychoacoustical studies described in this section are important. Firstly, the experiments were conducted over earphones, permitting the independent control of the waveforms at the ears. Secondly, the experiments generally manipulated ongoing ITD (i.e. IPD) and guarded against providing a salient onset ITD by turning on the stimuli relatively slowly and simultaneously to both ears. Thirdly, listeners were generally given extensive training and may also have been selected on the basis of having particularly acute sensitivity. This section also concentrates on results from unmodulated and modulated tones, although data are available from a variety of stimuli.

2.2.1 Unmodulated tones

A common measure of binaural sensitivity is ‘interaural discrimination’. Typically, a listener is presented with one diotic and one dichotic stimulus (e.g. containing an ITD or ILD) in random order and is required to indicate which was dichotic. This is referred to as a two–alternative forced choice task (2AFC). The smallest interaural disparity required for the listener to select the dichotic stimulus a criterion proportion of the time is identified, and referred to as the ‘threshold’. In recent years, this is usually achieved using an adaptive procedure whereby the interaural disparity is varied in a step–wise manner over a number of trials in order to target a particular percent correct on the psychometric function. Figure 2.3(a) plots average ITD and ILD thresholds from Klump and Eady (1956) and Yost and Dye (1988), respectively. The ITD threshold reduces (i.e. sensitivity improves) gradually with increasing frequency until sharply worsening above 1000 Hz, such that no ITD sensitivity is found for tones above 2000 Hz[3]. In contrast, the ILD threshold is broadly independent of frequency. One difficulty with interaural discrimination is that it may not be clear how listeners detect the dichotic stimulus. While ITD and ILD influence lateral position, they may also influence the width (i.e. diffuseness) and number of images (Durlach and Colburn, 1978), and ILD theoretically introduces non–spatial (e.g. loudness) cues (Bernstein, 2004).

[3] While Klump and Eady did not gate the stimuli to the ears simultaneously, they used relatively long onset and offset ramps (300 ms). Had they used substantially shorter ramps, as is more common, ITD sensitivity would probably have been apparent above 1500 Hz by virtue of an onset ITD cue.

The influence of ITD or ILD on lateral position may be measured directly using some form of visual scale or by manipulating a mechanical or acoustic 'pointer'. For example, Schiano et al. (1986) required listeners to adjust the ILD of a 200–Hz–wide band of noise centred on 500 Hz (the pointer) to coincide with the lateral position of a tone with an ITD of 150 μs. The data in Figure 2.3(b) is from one representative listener. The extent of laterality associated with this ITD (i.e. the ILD of the pointer) gradually increases with frequency until sharply reducing above 1000 Hz, consistent with ITD discrimination. For frequencies below 1500 Hz, extent of laterality increases in a roughly linear manner with IPD up to about 60–70° (Yost, 1981). Above 90°, laterality increases at a slower rate and the image becomes more diffuse and eventually breaking into multiple images by 180°. This process is reversed from 180–360°. Lateral position also increases with ILD up to approximately 30 dB but is not strongly dependent on frequency (Yost, 1981).

Alternatively, the lowest SPL of a tonal 'signal' required for the signal to be detected over a criterion proportion of trials, the 'absolute threshold', may be measured in the presence of a noise 'masker', when the signal and masker are both diotic (denoted 'N_0S_0') and, typically, when the signal has an IPD of 180° (denoted '$N_0S\pi$'). The difference (in dB) between thresholds in these conditions is referred to as the *masking level difference* (MLD). Figure 2.3(c) displays MLDs as a function of signal frequency, based on data from Webster (1951). The bandwidth of the noise masker was 80–6600 Hz. The MLD is maximal for tones below 500 Hz and gradually reduces to an asymptotic value of 3–4 dB above 1500 Hz. The MLD does not reduce to zero above 1500 Hz, as is the case with ITD thresholds, due to the interaction of the signal with the envelope of the noise. If an ILD rather than an IPD is applied to the signal, the MLD can be as much as 7 dB (Durlach and Colburn, 1978).

A further type of experiment aims to measure the ILD (or ITD) required to re–centre an image displaced by an ITD (or ILD), referred to as the 'time–intensity trading ratio' and expressed in μs/dB. Values of 20–50 μs/dB are typical found for low–frequency tones (Durlach and Colburn, 1978). However, centred images associated with opposing ITD and ILD may be readily discriminated from images produced by diotic stimulation (Hafter and Carrier, 1972), apparently because highly unnatural combinations of ITD and ILD lead to diffuse and/or multiple images.

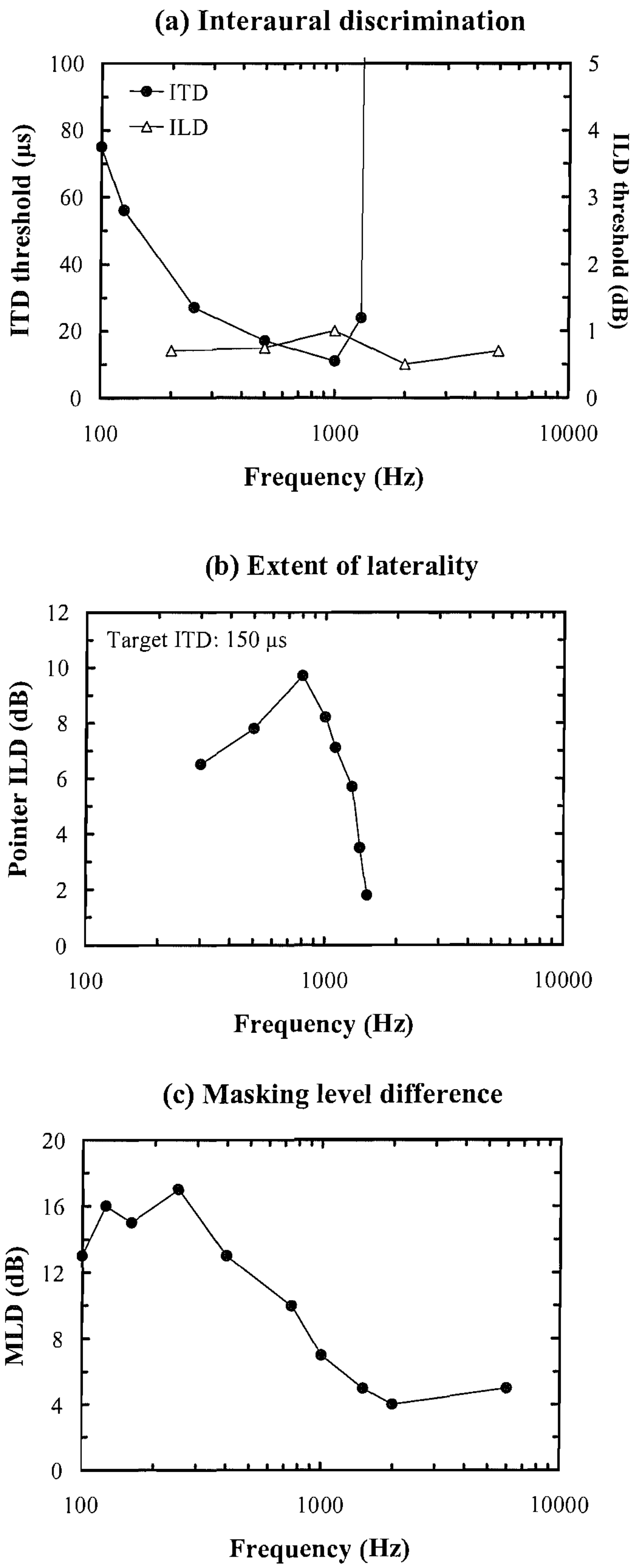


Figure 2.3. Experimental data with three measures of binaural sensitivity.

2.2.2 High–frequency complex stimuli

Although Klump and Eady (1956) found that high–frequency ITD sensitivity was absent for tones, ITD thresholds remained below 100 μs for bands of noise above 3000 Hz. Subsequent work in the 1970s using a variety of high–frequency complex stimuli confirmed this and showed that sensitivity was not related to the use of low–frequency spectral information (SAM: Henning, 1974 1980, Nuetzel and Hafter, 1976, 1981; bands of noise: McFadden and Pasanen, 1976, Bernstein and Trahiotis, 1982; beating tones: McFadden and Pasanen, 1976; band–pass clicks: Yost et al., 1971). For example, the use of intermodulation distortion was avoided by presenting low–pass masking noise during testing. Some studies have reported comparable thresholds with low– and high–frequency stimuli (e.g. Henning, 1974). However, a review of these and other studies concluded that, when the use of low–frequency spectral information and an onset ITD had been avoided, ITD thresholds were generally 2–10 times poorer with these high–frequency stimuli (Bernstein, 2001).

A recent study by Bernstein and Trahiotis (2002) compared ITD sensitivity with low–frequency unmodulated tones, SAM tones and novel AM tones referred to as 'transposed' stimuli, described shortly. They used 2AFC but added a diotic stimulus before and after the two primary stimuli to act as 'cues'. An adaptive procedure was used to measure ITD thresholds. The carrier frequency of the AM tones was 4000 Hz, and all stimuli were presented at 70 dB SPL. Gaussian noise, low–passed at 1300 Hz, was presented continuously and diotically[4], and at approximately 60 dB SPL, during testing. Thresholds were measured as a function of tonal frequency or modulation rate. Figure 2.4 illustrates how the transposed stimulus was generated. The modulator was a half–wave rectified (i.e. all negative values of the waveform are set to zero) and low–pass filtered at 2000 Hz. The theoretical rationale for this is described in Section 2.3.5. This stimulus is similar to a SAM tone, except that the peaks in the envelope have a more rapid onset and offset and are separated by a brief silent gap, and the bandwidth is twice the cut–off frequency of the low–pass filter. The differences in the waveforms can be seen in the legend of Figure 2.5.

[4] Alternative arrangements (e.g. gated or interaurally uncorrelated noise) may *impair* performance (Bernstein and Trahiotis, 1995), a phenomena often referred to as *binaural interference*.

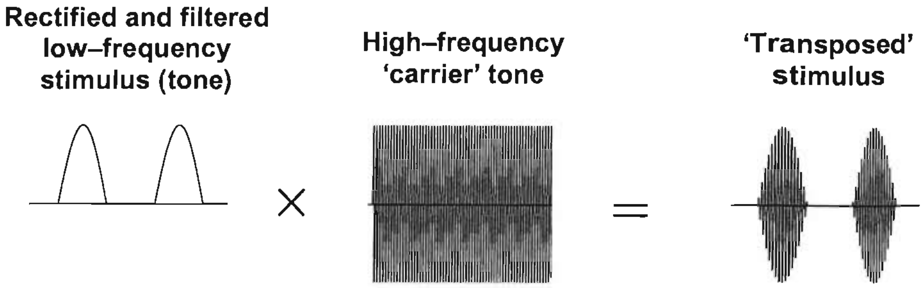


Figure 2.4. Schematic illustration of the generation of a 'transposed stimulus'. A low–frequency stimulus (e.g. a pure tone) is half–wave rectified and low–pass filtered (e.g. at 2000 Hz) to produce the modulator, and is then multiplied with a high–frequency tone.

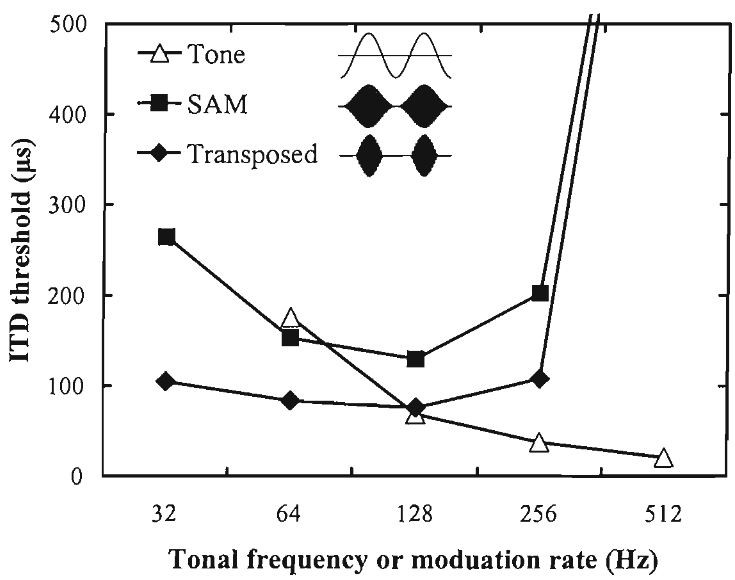


Figure 2.5. Mean ITD threshold across four listeners as a function of tonal frequency or modulation rate for three stimuli. The waveforms of the stimuli are illustrated in the legend. Redrawn from Bernstein and Trahiotis (2002). Data kindly supplied by Dr. Bernstein and used with permission.

Figure 2.5 presents ITD thresholds with the various stimulus conditions, averaged across four listeners. Four main trends are apparent. Firstly, tonal ITD thresholds gradually improved with increasing frequency, as expected from Figure 2.3(a). Secondly, the thresholds with the SAM tones were poorer than with the unmodulated tones, except at 64 Hz. Thirdly, performance with SAM tones improved with modulation rate between 32–128 Hz but deteriorated markedly above 128 Hz in most listeners, consistent with previous research (e.g. Nuetzel and Hafter, 1981). Fourthly, thresholds with the transposed tones were better than with the SAM tones although displayed a similar effect of modulation rate above 128 Hz. Consequently, thresholds at and below a modulation rate of 128 Hz were comparable to or better than with the unmodulated tone at 128 Hz. This has been confirmed by Oxenham et al. (2004).

Extents of laterality with ITD have also been measured with high–frequency stimuli (e.g. Leakey et al., 1957; Bernstein and Trahiotis, 1985, 2003; Trahiotis and Bernstein, 1986). While ITD influences laterality for these stimuli in a broadly similar way as with low–frequency stimuli, the extent of laterality for a given ITD is generally smaller. Interestingly, Trahiotis and Bernstein (1986) found that ITD had a stronger effect on laterality with bands of noise compared to SAM tones, despite ITD thresholds being smaller with SAM tones (e.g. Bernstein and Trahiotis, 1994). This illustrates the general difficulty in predicting ITD thresholds from extents of laterality. Transposed bands of noise have been found to produce greater extents of laterality than high–frequency bands of noise (Bernstein and Trahiotis, 2003) and SAM tones (Bernstein, 2001), and are, under some conditions, comparable to low–frequency stimuli. Van de Par and Kohlrausch (1997) showed that MLDs are also enhanced with transposed stimuli compared to conventional high–frequency stimuli. Overall, studies with transposed stimuli indicate that binaural sensitivity can be comparable at low and high frequencies, under certain conditions.

2.2.3 Salience of cues to spatial position

Two studies have investigated the relative salience of the various spatial cues to perceived location with stimuli presented over earphones but leading to externalised images (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002). The stimuli were based on recordings in the listeners' own ear canals during sound–field presentation (often referred to as the 'virtual auditory space' technique) and so contained the natural combination of localisation cues. The authors measured perceived localisation whilst manipulating one (spurious) cue and permitting the other cues to vary 'naturally' with spatial position. The basic stimulus used by both studies was broadband Gaussian noise, which was also low– and high–pass filtered at approximately 2000 Hz. Spurious ITD and ILD cues were imposed by applying a whole–waveform time shift and attenuation, respectively, to the stimulus presented to one ear. In addition, Macpherson and Middlebrooks varied the onset and offset duration and amplitude–modulated the high–pass noise (using a 100–Hz square–wave modulator) to investigate the relative saliency of the high–frequency ITD cues. Taken together, the findings of these studies can be summarised as follows:

- Ongoing fine–structure–based ITD is the most salient cue for horizontal–plane localisation for stimuli containing low–frequency information. Both studies found that even substantial spurious ILDs had surprisingly little effect on perceived location with the low–pass noise.
- The ILD cue is the most salient cue for horizontal–plane localisation with high–frequency stimuli[5].
- The onset ITD cue played a role for stimuli with onset durations of 1 ms or less. Envelope–based ongoing ITD was a far less salient cue compared to ILD in most listeners, even when enhanced with AM. However, the evidence with transposed stimuli (previous section) suggests this conclusion may not apply to all stimuli and to the processing of ongoing ITD at high–frequencies in general.

2.2.4 The influence of experience

Studies have examined the influence of experience on binaural hearing in adults, and generally fall into two categories: (1) studies of modified localisation cues; (2) studies of training with unmodified localisation cues. The number of studies in the first category far exceeds that in the second, and Moore and King (2004) and Byrne and Dirks (1996) reviewed these in detail. These studies have used a number of approaches to modify localisation cues, for example, using pinna or ear–canal plugs/inserts, signal processing in 'hearing aids' and the virtual auditory space

[5] Changes in the relative distribution of ILD across frequency and the interaural pinna cue played little, if any, role (Macpherson and Middlebrooks, 2002).

technique. Most of these studies have altered ILD or spectral cues. For example, while a unilateral ear–canal plug, which distorts the ILD cue, initially displaces perceived spatial position away from the plugged ear with an otherwise diotic stimulus, experience with the plug over a number of weeks generally results in perceived position returning to the centre (e.g. Florentine, 1976). Javer and Schwarz (1995) fitted normal–hearing listeners with bilateral 'hearing aids' used to produce a fixed interaural delay of 171 or 684 μs, and therefore distort the ITD cue. Listeners correctly localised a sound–source at an azimuth and elevation of 0° prior to introducing the delay, but responses were biased towards one ear immediately after introducing the delay. This bias reduced after several days of continuously wearing the devices. Interestingly, Hofman et al. (2002) fitted listeners with devices that reversed the polarity of the binaural cues, whist maintaining the integrity of the spectral cues. Responses on a horizontal–plane localisation task generally corresponded, or were close, to the actual location prior to the cue reversal but reversed after the cue reversal, consistent with the intended manipulation. Curiously though, performance did not change following 3 days, with one listener, and 20 days, with another listener, of wearing the devices. The reasons for the apparent absence of learning in this experiment, compared to the others are not clear. However, overall, these studies have demonstrated that perceived spatial position depends on experience. These studies do not indicate, though, if the basic sensitivity to localisation cues is also dependent on experience. Such information may be obtained from studies whereby listeners are trained with unmodified cues.

Three studies were identified that trained listeners on various horizontal–plane localisation tasks. Butler (1987) examined the ability to localise sound sources in the one hemifield (i.e. from straight ahead to behind on one side) in listeners with a contralateral ear–canal plug. The accuracy of localisation ability improved with modest training (repeated testing with trial–by–trial feedback). This mainly reflected a reduction in bias towards the side when stimuli were presented from straight ahead or behind, presumably reflecting learning of subtle monaural cues. Interestingly, a similar improvement was not apparent for an untrained stimulus with a different centre frequency. Terhune (1985) measured the ability to discriminate an azimuth angle of 8° close to the midline, using both ears. For one listener and with an 8000–Hz tone, the percentage of correct responses improved from near chance to nearly

90% over 2500 trials of practice. In contrast, Recanzone et al. (1998) measured discrimination thresholds for azimuth with a baseline position of 0° and the average error of localisation identification. They found 'essentially no learning effect over the course of the experiment for either task.' The reasons for the difference between studies are unclear, but could reflect inter–individual differences or differences in methodology or stimuli. One general difficulty with studies of training with localisation is that, as sound–source position is manipulated, it may not be clear which localisation cue or combination of cues listeners use to perform the task. This can be resolved using binaural earphone studies.

A number of experimenters have commented on the potential for improvements in performance on binaural earphone tasks with training (e.g. Hafter, 1984; Trahiotis et al., 1990). However, few studies have examined this in detail and different studies often reach disparate conclusions. For example, Hafter and Carrier (1970) indicated that MLDs may improve substantially over any months, whereas Trahiotis et al. (1990) and Bernstein et al. (1998) concluded that training has little if any effect of MLDs. Several studies have also commented on observations of training effects with ITD discrimination when using high–frequency SAM tones (e.g. Nuetzel and Hafter, 1976, 1981) and pairs of clicks (Saberi and Perrott, 1990; Litovsky et al., 2000; Saberi and Antonio, 2003). Prior to starting this project, only one study had systematically investigated learning on interaural discrimination (Wright and Fitzgerald, 2001). This study appeared to show different characteristics of learning with the discrimination of low–frequency ongoing ITD and high–frequency ILD. The authors speculated that this might reflect fundamental differences in the way the stimuli are processed by the auditory system. Before describing these studies in greater detail, current knowledge of the processing of the different binaural cues in the mammalian brain and the effects of training on non–binaural discrimination tasks is reviewed.

2.3 Binaural neurophysiology

2.3.1 Overview

Numerous pathways and nuclei are thought to be involved in binaural processing in the mammalian brain. The complexity of the neural organisation of the brainstem, important in the extraction of the interaural cues, is illustrated in Figure 2.6.

Processing (monaural) in the auditory periphery results in different acoustic waveforms being represented by differences in the temporal characteristics of action potentials in auditory–nerve fibres and how this activity is distributed across the array of fibres (e.g. Palmer, 1995). For example, the rate of action potentials may increase as stimulus level increases. Action potentials are also synchronised to one phase of waveform (i.e. are 'phase–locked') for unmodulated tones below approximately 4000 Hz and to the envelope of high–frequency AM stimuli (Joris et al., 2004). However, fibres are not equally sensitive to the audible frequencies, but rather are 'tuned' to a narrow range of frequencies. The frequency to which a fibre is most sensitive is called the 'characteristic frequency' (CF). The CF is correlated with the position that the fibre innervates the cochlea, becoming progressively higher towards the base, called 'cochleotopicity'. Phase locking and cochleotopicity are found within most auditory pathways and nuclei in the brain, although the upper frequency limit of phase–locking reduces with ascent towards the auditory cortex.

All auditory–nerve fibres project to the homolateral cochlear nucleus (CN) in the brainstem. From this, two parallel pathways arise (although this is a simplification). The 'binaural pathway' arises from cells in the anteroventral portion of the CN and conveys information to a collection of nuclei called the superior olivary complex (SOC) on either side of the brainstem. The SOC therefore receives bilateral ascending input, and is involved in the extraction of ITD and ILD. Cells in the SOC largely project to the inferior colliculus (IC), via the lateral lemniscus. The 'monaural pathway' bypasses the SOC and projects, again via the lateral lemniscus, predominately to the contralateral IC. In the IC, the binaural and monaural pathways converge, and further binaural interactions occur. Information from the IC regarding binaural hearing is sent to the superior colliculus (SC), which receives multi–sensory input and interacts with the motor system, and to the auditory cortex via the medial geniculate body.

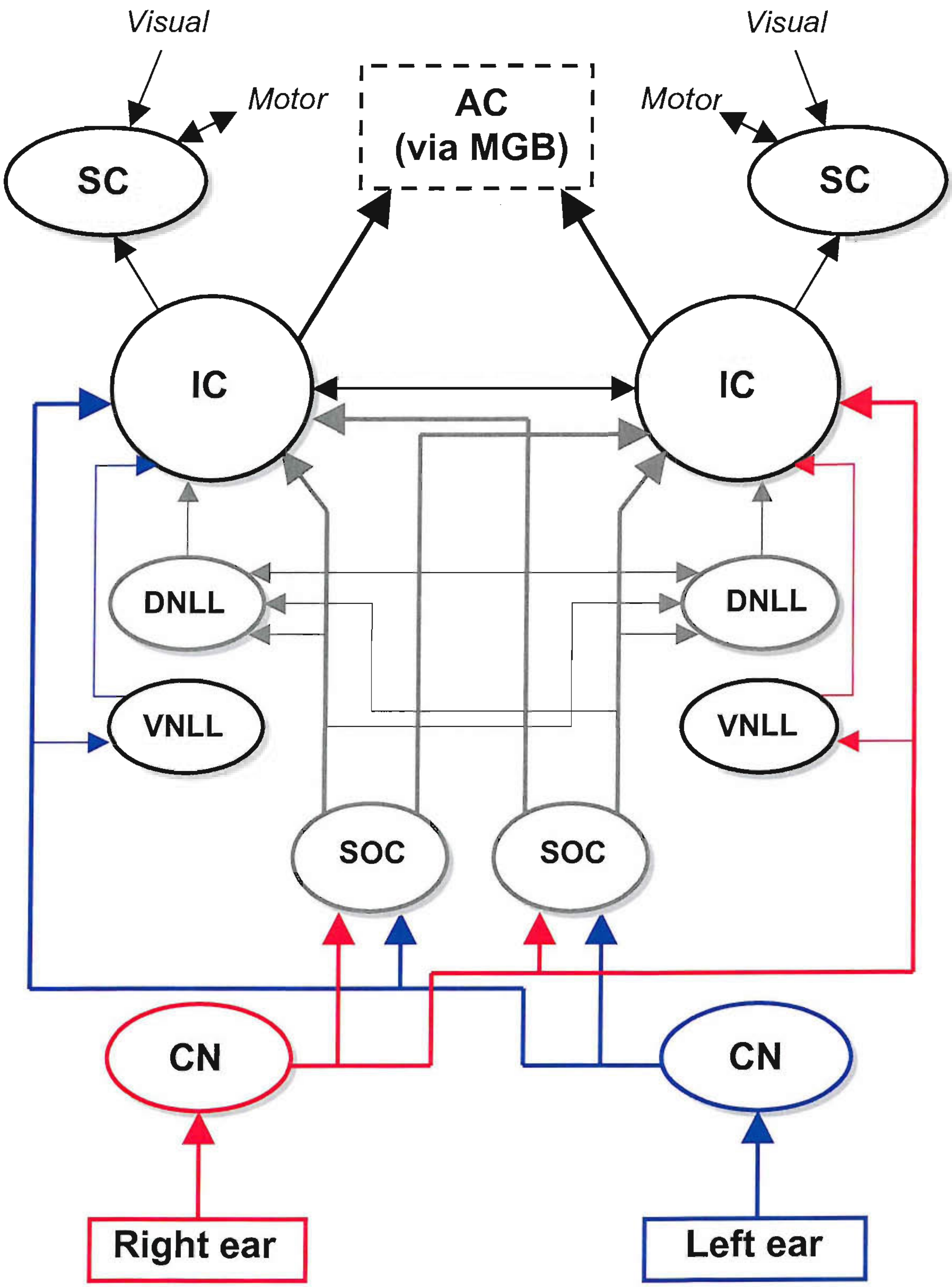


Figure 2.6. Schematic illustration of the complexity of the mammalian brainstem pathways involved in binaural hearing. (Cochlear nucleus: CN; superior olivary complex: SOC; ventral nucleus of the lateral lemniscus: VNLL; dorsal nucleus of the lateral lemniscus: DNLL; inferior colliculus: IC; superior colliculus: SC; medial geniculate body: MGB; auditory cortex: AC).

2.3.2 Unmodulated tones

The traditional scheme for the processing of interaural cues relevant to unmodulated tones in the mammalian brainstem is illustrated in Figure 2.7. The medial superior olive (MSO) and lateral superior olive (LSO) receive bilateral input directly or indirectly from both cochlear nuclei. Although the MSO and LSO are cochleotopically organised, the MSO receives predominately low–frequency, and therefore phase–locked, information whereas the LSO receives predominately high–frequency, and therefore not phase–locked, information. Consequently, the MSO is generally considered to extract ITD (Yin, 2002) and the LSO is generally considered to extract ILD (Tollin, 2003). Most low–CF MSO cells receive bilateral excitatory input, and are referred to as 'EE–type' cells. Most high–CF LSO cells receive ipsilateral excitatory input and contralateral inhibitory input, via the ipsilateral medial nucleus of the trapezoid body, and are referred to as 'EI–type' cells. Cells in the MSO and LSO project mainly to the ipsilateral and contralateral central nucleus of the IC, respectively. The firing rate of EE–type MSO cells to a CF–tone is usually strongly modulated by ITD, as illustrated in the upper right panel of Figure 2.7. Composite graphs based on the responses to tones of various frequencies (or broadband noise) show that these cells are maximally responsive to a particular, 'characteristic' ITD, referred to as 'peak–type' responses. These cells are thought to act as coincidence detectors, responding only when bilateral action potentials arrive simultaneously. The lower right panel of Figure 2.7 illustrates the typical effect of ILD on the firing rate of EI–type LSO cells using a CF–tone. Firing rate is greatest when the tone is presented to the ipsilateral ear only and reduces as the level of the tone presented to the contralateral ear is increased.

The scheme illustrated in Figure 2.7 is complicated by inhibitory input to the MSO and contralateral excitatory input to the LSO (Yin, 2002). Also, some ITD–sensitive cells within the SOC display inverted peak–type (i.e. 'trough–type') and intermediate–type responses, consistent with inhibitory influences, and some LSO cells have non–monotonic ILD functions (Fitzgerald et al., 2002). Further, the ILD functions of LSO cells are often dependent on overall SPL (Park et al., 2004), whereas the ILD arising from a sound source at a fixed azimuth is not.

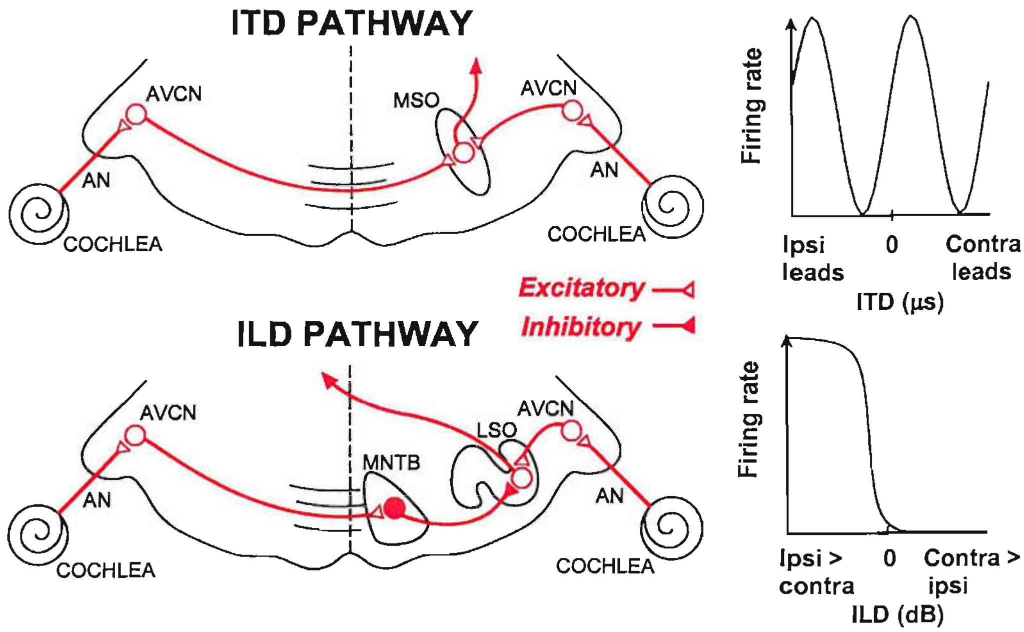


Figure 2.7. Illustration of the traditional scheme for the processing of interaural cues in the brainstem relevant to tones, and idealised cell responses. Reproduced and modified with permission from Prof. Yin[6]. (Auditory nerve: AN; anteroventral CN: AVCN; medial superior olive: MSO; lateral superior olive: LSO; medial nucleus of the trapezoid body: MNTB.)

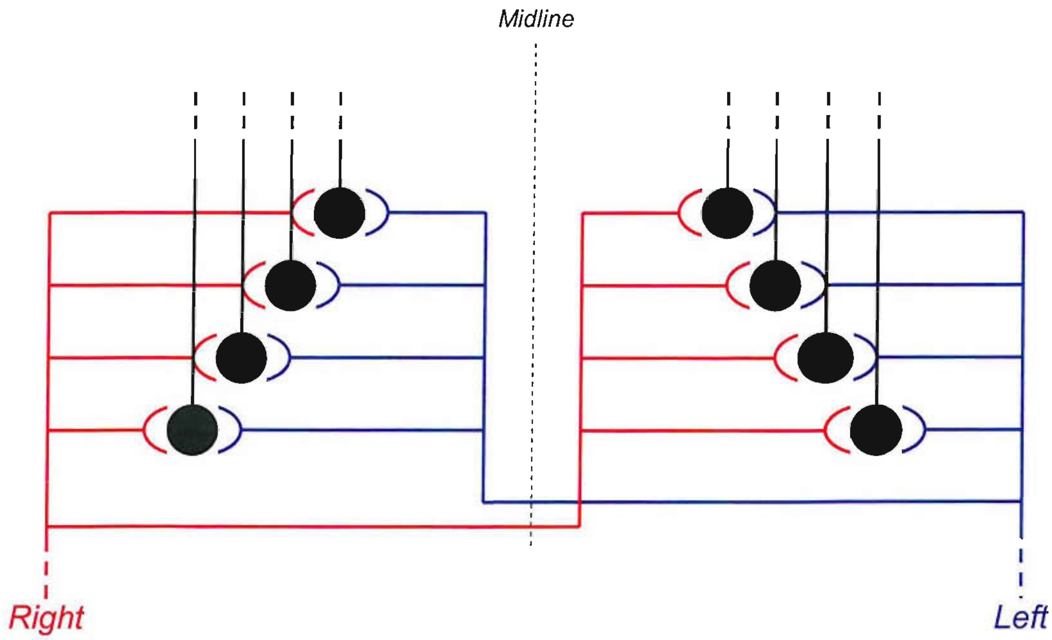


Figure 2.8. Schematic illustration of the Jeffress coincidence detection model of ITD processing. Based on Figure 1 from Jeffress (1948).

[6] Obtained from http://www.physiology.wisc.edu/yin/labproj.html (July 2004)

Despite the extensive research on the responses of binaural cells in the SOC and IC, precisely how ITD and ILD are encoded has not been settled. Figure 2.8 illustrates the traditional view of ITD coding, as proposed by Jeffress (1948). He assumed that bilateral phase–locked information was conveyed to an array of coincidence detectors on each side of the brainstem. It was thought that the continuum of ITDs was coded by virtue of a series of internal interaural time delays, sensitising different cells to different ITDs. These internal delays were thought to arise from interaural asymmetries in the length of axons innervating each coincidence detector, referred to as delay lines. The delay lines were assumed to be organised topographically resulting in the coding of ITD by the distribution of activity across the array. There is a general consensus that low–CF EE–type MSO cells receive phase–locked information, the synchrony of which may be even greater than in the auditory nerve (Joris et al., 1994), and behave like coincidence detectors. Some researchers have further argued that the MSO appears to conform broadly to other aspects of Jeffress's scheme (e.g. Joris et al., 1998). For example, studies of the cat appear to indicate the existence of distributed ITD tuning within the animal's natural range (Yin and Chan, 1990) and delay lines (Smith et al., 1993), albeit more complex than envisaged by Jeffress. However, the little evidence there is for a topographical code of ITD is indirect (Yin and Chan, 1990).

Others (e.g. McAlpine and Grothe, 2003) have suggested that Jeffress's scheme may not apply to mammals in general. For example, studies of the guinea pig and Mongolian gerbil suggest that the characteristic ITD is generally *outside* the animals' natural range except with high–CF cells (McAlpine et al., 2001; Brand et al., 2002). Further, this appears dependent on inhibitory input to the MSO. These authors have proposed an alternative code of ITD based on the comparison of the rate of activity across the two sides of the brain.

2.3.3 High–frequency complex stimuli

The previous section indicated that the initial stage of binaural processing can be thought of as essentially two parallel pathways, one conveying low–frequency information to the MSO and the other conveying high–frequency information to the LSO. This suggests that the LSO may also be primarily responsible for processing high–frequency envelope–based ongoing ITD. On the other hand, it is not immediately obvious that high–CF LSO cells would respond to ITD, given that most are EI–type, unlike most cells in the MSO, and respond to ILD. One might also imagine that the additional synapse in the pathway to the LSO from the contralateral, but not ipsilateral, CN may adversely affect temporal coding. However, high–CF cells throughout the brainstem do phase–lock to the envelope of AM stimuli (Joris et al., 2004). Further, the pathway to the LSO appears to have important anatomical and physiological specialisations for temporal coding (e.g. Oertel, 1999; Trussel, 1999). For example, the synapse in the medial nucleus of the trapezoid body is the largest in the mammalian brain. Numerous experiments have also shown that high–CF EI–type LSO cells are sensitive to envelope–based ITD, although their response troughs, rather than peaks, at the characteristic ITD ('trough–type responses') (e.g. Joris and Yin, 1995; Batra et al., 1997; Fitzgerald et al., 2002)[7].

A complication for these considerations is that the human LSO is far less developed than in other mammals (Irvine, 1991), such as those used in some of the above experiments. The MSO of experimental mammals also contains some high–CF cells that are sensitive to ongoing envelope–based ITD, showing peak–type responses (Batra et al., 1997; Yin and Chan, 1990). Thus, the processing of this cue may be less divided between the MSO and LSO, and EE–type and EI–type cells, than is apparent for ITD and ILD with tones. Further, projections from the MSO and LSO with similar frequency tuning converge in the IC, and high–CF IC cells are sensitive to envelope–based ITD, showing a range of peak–, trough– and intermediate–type responses (Batra et al., 1993; Fitzgerald et al., 2002).

[7] An ongoing envelope–based ITD can also be thought of as a dynamically varying ILD.

2.3.4 Computation of spatial position

The neural substrate for the processing of perceived spatial position is unclear. Many experiments have sought to identify a topographical representation of auditory space, as exists with visual space. However, a simple topographical map of auditory space has not been identified in the mammalian IC or auditory cortex (Yin, 2002; Middlebrooks et al., 2002). For example, the IC and primary auditory cortex are cochleotopically organised, whereas the computation of spatial location for broadband sounds, such as speech, presumably requires substantial convergence across frequency. However, spatially–sensitive cortical cells in cats and mammals respond preferentially to sound–sources located in the contralateral hemifield and lesions to one side of the cortex in humans often impair localisation performance to the contralateral hemifield. Human imaging studies also suggest that spatial processing may have a distinct location in the auditory cortex (e.g. Alain et al., 2001).

Evidence for a topographical code of auditory space has been identified in the deep layers of the SC of the ferret (King et al., 2001) and the cat (Middlebrooks, 1988). Cells in the SC tend not to be highly frequency–tuned and so are not cochleotopically organised. Instead, they tend to be tuned to spatial location. The SC receives input from the IC via the nucleus of the brachium of the IC, which may be involved in the formation of the space code. It is also thought that the space map in the SC is aligned with visual and somatosensory space maps. The response properties of cells in the SC are known to be highly plastic and neurophysiological correlates with behavioural measures of learning following ear–canal plugging have been identified (e.g. Moore and King, 2004).

2.3.5 Models of ITD discrimination

The specific neural mechanisms underlying ITD discrimination are unclear, although performance is presumably based largely on the output of the SOC. Despite the limited evidence for Jeffress–type ITD processing, and the recent evidence against it, most psychoacoustical models of ITD sensitivity are based on coincidence–detection (i.e. EE–type interactions) following internal time delays (Colburn and Durlach, 1978; Colburn, 1995; Stern and Trahiotis, 1995). [Breebaart et al. (2001) have reported a binaural model based on EI–type rather than EE–type interactions.] Typically, the output of the coincidence–detectors is represented as an interaural cross–correlation function (i.e. the correlation of the signals from the two ears as a function of internal time delay and CF). Overall, these models, coupled with realistic models of peripheral processing, can account for ITD discrimination, and other measures of ITD sensitivity, in a range of conditions.

Models of ITD discrimination indicate that performance is limited by the variance in the internal representation of the ITD. This 'noise' may have its origins in mainly peripheral (i.e. monaural) processing. Models also show that it is not necessary to explicitly predict lateral position in order to account for discrimination ability. For example, Bernstein and Trahiotis (2002) have reported a model that assumes that ITD discrimination can be accomplished by detecting interaural decorrelation of the signals. Many models assume ITD discrimination requires the pooling of information across many cells and across CFs. Physiological studies in the guinea pig IC, however, indicate that sufficient information is contained within individual EE–type cells to account for human ITD thresholds with unmodulated tones (Skottun et al., 2001; Shackleton et al., 2003). This also indicates that there may not be a simple relationship between the number of neurons sensitive to ITD and the associated psychoacoustical thresholds, at least for simple stimuli. On the other hand, Hancock and Delgutte (2004) suggest it is necessary to pool information across cells within and across CF in order to predict performance under more complex conditions.

Psychoacoustical models of ITD discrimination have also been used with high–frequency complex stimuli. For example, Colburn and Esquissaud (1976) suggested that the poorer ITD thresholds with SAM compared to unmodulated tones (see Section 2.2.2) reflected differences in the salience of the temporal information at the inputs to binaural processing, due to peripheral processing, rather than inherent differences in binaural sensitivity across frequency. Figure 2.9 illustrates the influence of specific aspects of peripheral processing on the waveforms of various stimuli. This processing consists of a simple model of cochlear inner hair cells: bandpass filtering, half–wave rectification and low–pass filtering at 500 Hz (Delgutte, 1995). The output of the hair cell in Figure 2.9 drives the generation of action potentials in the auditory nerve, and essentially represents the probability of an action potential as a function of the phase of the input. The top input waveform is a 128–Hz–unmodulated tone and the middle waveform is a 128–Hz modulated SAM tone at

4000 Hz. The shape of the output to the latter is less well defined in time compared to the former. This can be thought of as representing greater variance in the distribution of action potentials with the SAM tone, and was suggested by Colburn and Esquissaud to account for the trends in ITD discrimination. The transposed stimulus described in Section 2.2.2 was designed to attempt to mimic the output of this processing to low–frequency unmodulated tones but in high–frequency auditory channels, and thus provide comparable input to the binaural system, all else being equal. The bottom waveform in Figure 2.9 represents a 128–Hz–modulated transposed tone at 4000 Hz. It can be seen that the output of the inner hair cell model is essentially equivalent to that with the 128–Hz tone. Assuming binaural processing is at least functionally comparable at 128 and 4000 Hz, ITD sensitivity with the transposed stimulus would be expected to be enhanced compared to that with SAM tones, and to be comparable to ITD sensitivity with unmodulated tones, again all else being equal. Bernstein and Trahiotis (2002) were able to account for their data, reproduced in Figure 2.5, using a standard model of monaural processing and comparable binaural processing at low and high frequencies. This was supplemented with a further stage of monaural processing with AM stimuli only, to account for the rapid loss of sensitivity with modulation rates above about 200 Hz. Recent measurements from the guinea pig IC also indicate that ITD sensitivity in high–CF cells is enhanced with transposed tones compared to SAM tones, consistent with the psychoacoustical data (Griffin et al., 2005). It is less clear if sensitivity is comparable to that found with unmodulated tones and in the SOC.

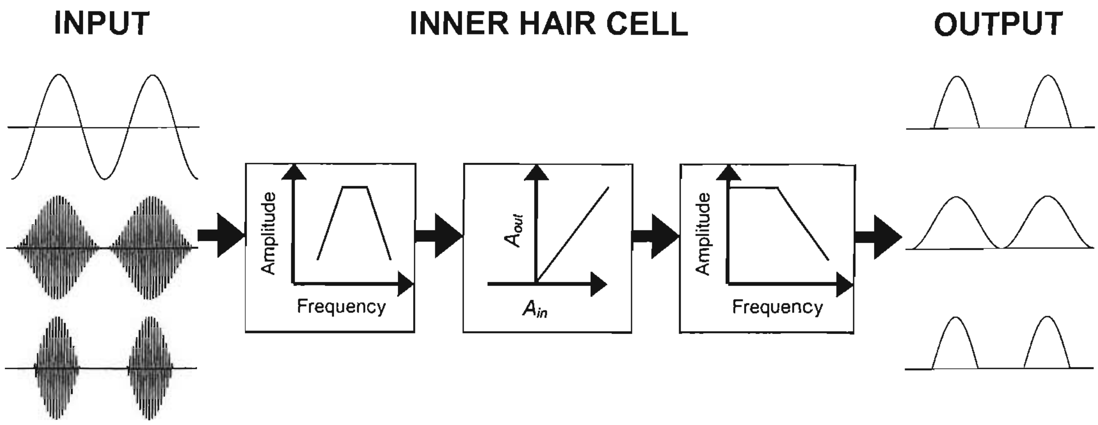


Figure 2.9. Simulations of the effect of peripheral processing on waveforms of an unmodulated tone, top, and SAM and transposed tones, middle and bottom, respectively.

2.4 Training–induced perceptual learning

2.4.1 Overview

Substantial evidence has accrued over many decades that discrimination thresholds often improve with practice, referred to as 'training–induced discrimination learning', the bulk of which is from visual research (Fahle and Poggio, 2002; Fahle, 2005; Fine and Jacobs, 2002; Goldstone, 1998). In contrast with other forms of learning, discrimination learning is not related to the development of associations between events and 'does not lead to conscious insights that can be (easily) communicated' (Fahle, 2004). Overall, visual research has shown that discrimination learning occurs in a variety of conditions, displays a variety of characteristics and may be associated with a variety of physiological changes in a variety of cortical areas.

Table 2.1 presents a survey of the methods used in systematic studies of auditory discrimination learning using simple stimulus conditions[8]; most of these have studied frequency discrimination and used unmodulated tones. All studies recruited listeners inexperienced with psychoacoustical testing. The general experimental paradigms used are illustrated in Figure 2.10. The approach referred to as 'train' involves measuring repeatedly discrimination ability with the trained stimulus, and can be used to investigate the time–course of learning and to compare time–courses between different stimuli. A more powerful, and more common, paradigm is 'pre–post'. Here,

[8]Watson (1980, 1991) reviews of studies of learning with more complex stimulus conditions.

a listener undergoes baseline testing, usually with multiple stimuli, before and after formal training, usually with one stimulus. Unlike the previous paradigm, pre–post permits examination of the influence of training with one stimulus on performance with another, untrained stimulus. The effects of alternative training regimes on performance with a common stimulus can also be evaluated. Some experimenters have extended this paradigm using a mid–training test. Others have placed one or two tests after training, referred to as 'train–test'. Experiments based on the pre–post paradigm have often conducted brief testing prior to pre–test, referred to as 'familiarisation'. The aim of this is usually to provide listeners with experience of the procedures (e.g. trial format) or orientate listeners to the perceptual dimension of interest (e.g. lateral position) and so minimise the influences of learning related to these during pre–testing and training.

Table 2.1. Studies of auditory discrimination learning (familiarisation: famil.). Continued overleaf.

Study	Parameter discriminated	Paradigm	N	Groups	Untrained groups	Famil.	General		
							Task	Feedback	Procedure
Amitay et al. 2005	Frequency	Train–test	39	3	–	–	2AFC	?	Adaptive
Ari–Even Roth et al. 2003	Frequency	Pre–post	10	2	1	N	2AFC	Throughout	Adaptive
Ari–Even Roth et al. 2004	Frequency	Train–test	20	2	–	–	2AFC	Throughout	Adaptive
Delhommeau et al. 2002	Frequency	Pre–post	10	1	0	Y	1 cue 2AFC	Throughout	Adaptive
Delhommeau et al. 2005	Frequency	Pre–post	32	4	0	N	1 cue 2AFC	?	Adaptive
Demany 1985	Frequency	Pre–post	70	4	0	N	2 cue 2AFC	No	Adaptive
Demany and Semal 2002	Frequency	Complex	8	2	0	Y	1 cue 2AFC	Throughout	Adaptive
	Frequency	Complex	16	2	0	N	1 cue 2AFC	Throughout	Adaptive
Grimault et al. 2002	Frequency	Pre–post	12	2	1	Y	1 cue 2AFC	Training	Adaptive
Grimault et al. 2003	Frequency and modulation rate	Pre–mid–post	15	5	0	Y	1 cue 2AFC	Training	Adaptive
Hawkey et al. 2004	Frequency	Train–test	80	4	–	–	2AFC	Throughout	Adaptive
Irvine et al. 2000	Frequency	Pre–post	16	2	0	Y	3AFC	No	Adaptive
Wright and Fitzgerald 2003	Frequency	Pre–post	16	2	1	?	2AFC	Throughout	Adaptive
	Level	Pre–post	12	2	1	?	2AFC	Throughout	Adaptive
	Duration	Pre–post	16–30	2	1	?	2AFC	Throughout	Adaptive
Wright et al. 1997	Temporal interval	Pre–post	14	1	0	N	2AFC	Throughout	Adaptive
Karmarkar and Buonomano 2003	Temporal interval	Pre–post	45	>1	0	N	2AFC	Throughout	Adaptive
Drennan and Watson 2001	Spectral envelope	Train	11	1	–	N	1 cue 2AFC	Throughout	Adaptive
Constantinides 2004	ILD	Train–test	32	4	–	–	1 of 4 based on 2AFC	Throughout	Adaptive
	ILD	Train–test	16	2	–	–	2AFC	Throughout	Adaptive or fixed
	ILD or ITD	Complex	51	8	2	Y	2AFC	Throughout	Adaptive
Wright and Fitzgerald 2001	ITD	Pre–post	13	2	1	Y	2AFC	Throughout	Adaptive
	ILD	Pre–post	19	2	1	Y	2AFC	Throughout	Adaptive

Table 2.1 continued.

Study	Test phase			Training phase		
	'Tests'	Conditions	Total trials	Blocks	Conditions	Total trials
Amitay et al. 2005	2	>1	1300–1400	500 trials	1	3500
Ari–Even Roth et al. 2003	2	3 or 6	c. 210 or 420	>8 revs	1	c. 2800
Ari–Even Roth et al. 2004	2	1	210	>8 revs	1	c. 700
Delhommeau et al. 2002	2	4	c. 600	16 revs	1	c. 4500
Delhommeau et al. 2005	2	8	c. 1200	12 revs	1	c. 4500
Demany 1985	2	1	50	70 trials	1	700
Demany and Semal 2002	–	–	–	80 trials	>1	>12000
	–	–	–	80 trials	1	>12000
Grimault et al. 2002	1	6	c. 1800	16 revs	1	c. 21600
Grimault et al. 2003	3	6	c. 2520	16 revs	1	c. 37800
Hawkey et al. 2004	2	1	500	500 trials	1	500
Irvine et al. 2000	2	2	c. 300	11 revs	1	c. 3600
Wright and Fitzgerald 2003	2	3	900	60 trials	1	2880–9000
	2	5	1500	60 trials	1	2880–9000
	2	4	1200	60 trials	1	2880–9000
Wright et al. 1997	2	3 or 5	900–1800	60 trials	1	≥9000
Karmarkar and Buonomano 2003	2	6	720	60 trials	1 or 2	>7000
Drennan and Watson 2001	–	–	–	80 trials	>1	2000
Constantinides 2004	1	4	400	50 trials	1	400
	1	4	400	50 trials	1	400
	4	1	240	60 trials	1	360
Wright and Fitzgerald 2001	2	5	1500	60 trials	1	6480–7200
	2	5	1500	60 trials	1	6480–7200

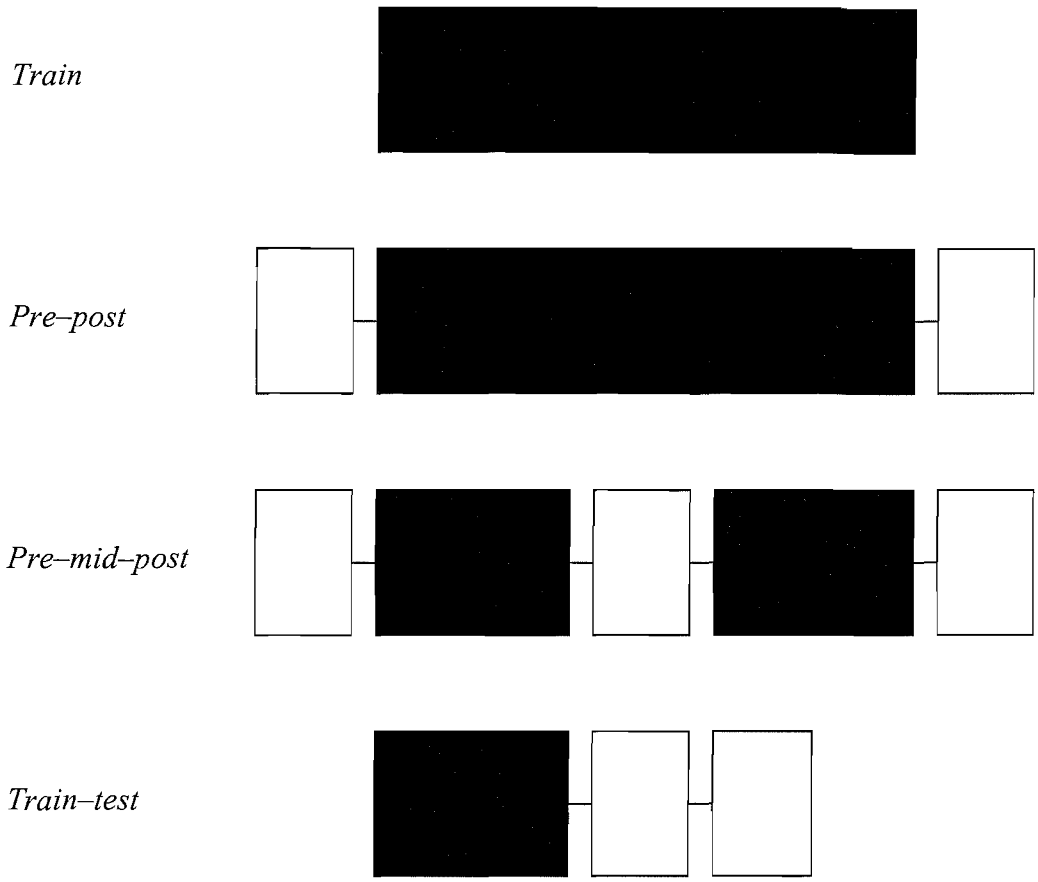


Figure 2.10. Illustration of paradigms used in auditory discrimination research. Open blocks: tests; filled blocks: training.

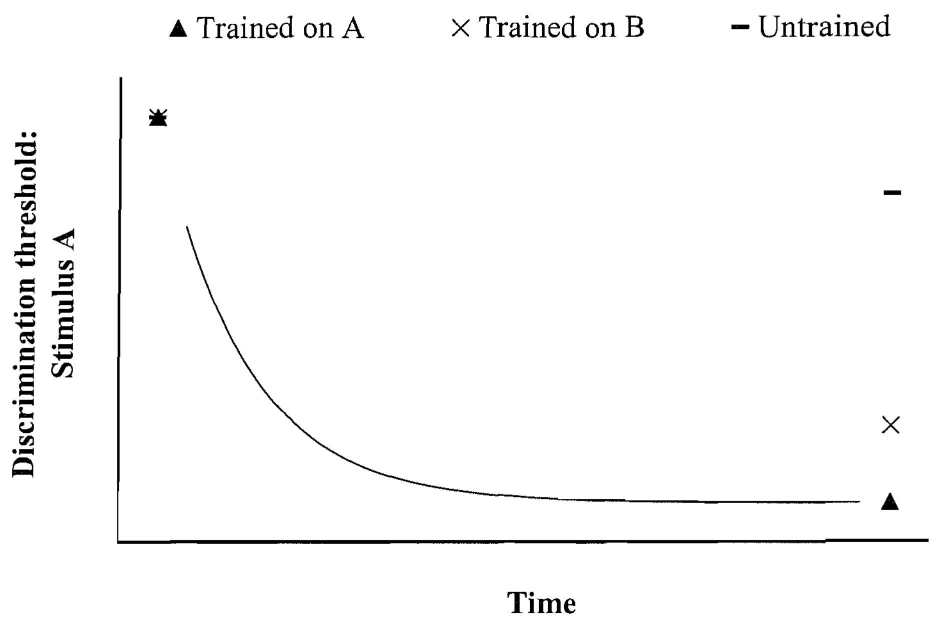


Figure 2.11. Hypothetical results on discrimination learning with the pre–post paradigm.

Figure 2.11 plots hypothetical results with the pre–post paradigm. Group A was trained with Stimulus A, Group B was trained with Stimulus B and Group C did not participate in training (i.e. were untrained controls). However, all groups were tested with both stimuli during pre– and post–tests. Thresholds were measured using a forced–choice task and an adaptive procedure, as is usually the case. The forced–choice task minimises certain response biases, and the adaptive procedure holds the percentage of correct responses constant and presents stimuli close to threshold without requiring prior knowledge of its value (Macmillan and Creelman, 1991). Figure 2.11 presents thresholds with Stimulus A. Post–test thresholds are lower than those at pre–test in all groups, even the untrained group (e.g. Wright and Fitzgerald, 2001, 2003; Ari–Even Roth et al., 2003). Learning in the untrained group presumably reflects experience of pre– and post–tests. Wright and Fitzgerald (2001) also suggested that learning in the untrained group might reflect the 'consolidation' of experience gained from pre–test (i.e. that learning occurred between rather than within test sessions). However, this is difficult to judge since this study and others employing untrained groups do not report analyses of learning within test sessions.

Considering the data from Group A (i.e. trained with Stimulus A), the amount of learning is greater than in the untrained group, indicating that learning was induced by training. A best fit to the thresholds during training (i.e. the 'learning curve') is also plotted in Figure 2.11. This shows the characteristic negative exponential–type function with most improvement occurring early during training. It is commonly assumed that this early rapid phase of learning primarily reflects so–called 'procedural learning' (i.e. learning to execute the task). The later slower phase of learning is often assumed to primarily reflect perceptual learning *per se* (i.e. learning to hear differences between the stimuli). However, the time–course of learning has been found to vary widely between studies and listeners.

Consider now data from Group B. (Assume that the results from Group B with Stimulus B are comparable to those from Group A with Stimulus A.) This group also learned with Stimulus A, and by more than the untrained group. This indicates that training on Stimulus B also induced learning on Stimulus A, or, put differently, that the learning with Stimulus B 'generalised' to Stimulus A. However, the learning with Stimulus A in Group B was less than found in Group A. Training on Stimulus B was therefore less effective than training on Stimulus A in inducing learning with Stimulus A. Put differently, the training–induced learning with Stimulus B generalised incompletely to Stimulus A; it showed a degree of 'stimulus specificity'. Stimulus specificity has been reported widely in the vision literature. For example, discrimination of the alignment of lines may be specific to the trained orientation, position and eye (Fahle, 2004). It is often assumed that learning that generalises widely across stimuli occurs relatively early during training whereas stimulus–specific learning occurs only after more protracted training. Further, learning that generalises is also commonly assumed to reflect mostly procedural learning. However, there is little empirical evidence for these assumptions. Further, while learning may be specific to one stimulus parameter, it may not be to another.

The data in Figure 2.11 are idealised in a number of ways. In particular, pre–test thresholds are comparable across groups, although this is rarely the case. Material differences in pre–test thresholds between groups may arise by chance during randomisation, and is confounded by the small samples in most experiments. This is a problem because learning is generally related to pre–test ability; listeners with higher pre–test thresholds generally learn more than listeners with lower pre–test thresholds. Listeners with higher pre–test thresholds may also be more likely to display stimulus–specificity (Amitay et al., 2005). If the groups are not comparable in terms of pre–test thresholds on a common stimulus, or in terms of other factors that may influence learning, it may be difficult to determine if differences in learning between groups were due to effects of training.

2.4.2 Trends in studies of auditory discrimination learning

Overall, there is much variation in the results across studies, probably at least partly due to different methodologies and confounded by small samples. Many studies have reported evidence for stimulus–specific learning. For example, in a study of tonal frequency discrimination, Irvine et al. (2000) trained one group at 5000 Hz and another group at 8000 Hz, with pre– and post–testing at both frequencies. Learning was apparent in both groups on the trained stimulus. Much of this appeared to generalise to the untrained frequency, but a small statistically significant portion was specific to the trained stimulus. Other studies have reported stimulus–specificity in terms of the trained frequency (Demany and Semal, 2002; Delhommeau et al., 2005), duration and ear (Delhommeau et al., 2002). However, others have not found learning to be specific to frequency (Demany, 1985; Ari–Even Roth et al., 2003) or ear (Delhommeau et al., 2005; Ari–Even Roth et al., 2003; Demany and Semal, 2002). Further, while Demany and Semal (2002) found the frequency–specificity to resolve after only substantial training with the previously untrained stimulus, Delhommeau et al. (2005) found that relatively few trials were required.

Differences in the time–course or magnitude of learning may be apparent between stimuli, despite the use of identical procedures. For example, Grimault et al. (2002) found differences on fundamental–frequency discrimination between stimuli comprised of widely spaced harmonics and narrowly spaced harmonics. They suggested that the differences in learning between stimuli indicated that they were at least partially processed by distinct mechanisms and that the effects of training operated on each mechanism differently. However, these stimuli also differed in their associated asymptotic thresholds, and it is possible that the differences in learning were related to this. For example, modelling research apparently suggests that these findings could be accounted for by differences in inherent sensitivity to the stimuli rather than differences in processing (Carlyon, 2004, personal communication). Visual research also indicates that learning is related to the difficulty of the stimuli and the task (Ahissar and Hochstein, 1997).

Recent studies have questioned the assumption that the early rapid phase of learning reflects procedural learning. For example, Hawkey et al. (2004) used the train–test paradigm to investigate whether early frequency discrimination learning actually required training on frequency discrimination (i.e. was perceptual learning) or could occur with training on discrimination of a different stimulus parameter (i.e. was procedural learning). Training was provided over 500 trials (whereas training to asymptote often requires many thousands of trials): two groups received training on frequency discrimination at 1000 Hz, using slightly different trial formats/tasks; one group received training on intensity discrimination also at 1000 Hz; one group received training on a visual discrimination task. The two tests consisted of frequency discrimination, and the amount of learning between these tests was compared across groups. The groups trained on frequency discrimination displayed little improvement between tests, presumably because most learning had already taken place during training. However, the groups trained on auditory intensity or visual discrimination showed a significantly larger improvement in frequency discrimination between tests. That is, training with the procedures but not the discrimination of frequency was insufficient to induce early learning on frequency discrimination. This indicates that learning over the first 500 trials of training with tonal frequency discrimination reflects largely perceptual learning.

The specific conditions required for learning and, related to this, what is learned remain unclear. For example, while Hawkey et al. (2004) indicate that experience of changes in the specific stimulus parameter is important, it is unclear if listeners need to attend to these changes or whether learning would occur if listeners were passively exposed to the stimuli, perhaps whilst performing an unrelated task. Some visual research indicates that attention is not always necessary for perceptual learning. For example, learning has been observed even when the stimuli were not consciously perceptible or relevant to the particular task (Watanabe et al., 2001). On the other hand, learning to discriminate the alignment of horizontal lines does not seem to generalise to vertical lines, and vice versa, even when using an identical stimulus in both cases (i.e. based on a cross) (Fahle, 2004). It is also unclear from Hawkey et al. (2004) if the experience of subtle changes in frequency is necessary for learning or whether coarser changes would produce similar results. However, Constantinides (2004) found that training with changes in ILD that were close to threshold, using an adaptive procedure, was far more effective in inducing learning than training with a fixed change of 15 dB. However, listeners trained in the latter condition may also have had lower levels of attention given the ease of detecting a 15–dB ILD, and this may have interacted with the experimental manipulation. Trahiotis et al. (1990) also suggest that the potency of training (in binaural conditions at least) may be enhanced by increasing the proportion of trials that expose the listener to near–threshold changes, although did not provide direct evidence. Other cognitive factors may also influence learning, such as IQ and age (Moore et al., 2003; van der Elst et al., 2005).

A number of other methodological factors may influence learning. Colburn and Trahiotis (1991) suggested that trial formats whereby listeners can perform the task by detecting the 'odd–stimulus–out' might be associated with less learning than the standard 2AFC. This appears to be due to the added complexity with the standard

2AFC of listeners having to detect a difference between stimuli *and* to order them. Koehnke et al. (1995) also selected an oddity task for a study of binaural sensitivity following experience of larger training effects with other tasks. The number of trials during training and the organisation of training may also be important. However, Constantinides (2004) found that completing training in one day was as effective as distributing training over many days. The provision of feedback during training (e.g. telling the listener if they were correct or not) may also influence learning, although visual research is somewhat divided on this (e.g. Fahle, 2004; Gilbert, 1998).

The clinical implications and applications of discrimination learning are only beginning to be appreciated. Moore et al. (2003) have stressed the importance of across–stimulus generalisation if training can be used therapeutically. Emerging evidence suggests that learning with auditory training using relative simple stimuli may generalise to more complex tasks and stimuli, and may be effective in the treatment of language–based disorders (e.g. Merzenich et al., 1996; Moore et al., 2005), in the improvement of reading skills (Kujala et al., 2001) and in reduction of the severity of tinnitus (Flor et al., 2004).

2.4.3 Theories and physiological correlates of discrimination learning

According to the 'reverse hierarchy theory' of visual perceptual learning (Ahissar and Hochstein, 1997, 2004), learning proceeds from general to more specific aspects of the task and stimuli, and is correlated with changes in higher level and then lower level brain processing. Stimulus–specificity is therefore assumed to reflect relatively low–level changes in sensory processing. This is based on the additional, common but contentious assumption (cf. Mollon and Danilova, 1996) that stimulus–specific learning requires neurons that respond selectively along the stimulus dimension discriminated and empirical findings that low–level neurons are more often tuned to such basic stimulus parameters, whereas higher level neurons display more complex response properties. More detailed models of stimulus–specificity (Goldstone, 1998) have suggested that such low–level neurons that are useful in the discrimination become more influential or have their response properties modified (e.g. enhanced tuning) with training, or that the response properties of other neurons are modified (e.g. to increase the representation of the trained stimulus).

A number of studies have provided evidence for plasticity in early visual processing (e.g. primary visual cortex[9]) following discrimination training in experimental animals. However, in a review of these studies, Ghose (2004) suggests that the changes found are generally insufficient to account for the learning observed. Further, Ghose notes that other studies have been unable to identify clear evidence for plasticity in primary visual cortex, and challenges an assumption of the reverse hierarchy theory by providing evidence that neurons in a variety of cortical areas respond selectively to basic stimulus parameters.

One interpretation of frequency–specific learning in auditory frequency discrimination is that the frequency tuning of neurons with CFs at and close to the trained frequency are modified, but those tuned to the untrained frequency are not. Recanzone et al. (1993) appear to provide physiological evidence for this. Five adult owl monkeys were given extensive training of frequency discrimination and then physiological measurements were made from neurons in the primary auditory cortex. Physiological data were compared to measurements from three monkeys who had not participated in training and two monkeys who had been presented with the stimuli but were engaged in a non–auditory task. The authors reported that neurons with CFs at the trained frequency had sharper frequency tuning in the trained compared to control monkeys. They also reported that the area of the brain responsive to the trained frequency was much greater in the trained monkeys. They argued that these apparent differences between monkeys were the result of training. Others though have been unable to identify changes in the frequency response of neurons in the primary auditory cortex. For example, Brown et al. (2004) found no difference in the cortical representation of the trained frequency, using objective mapping techniques, between cats trained extensively on frequency discrimination and untrained cats. Only subtle, and not always statistically reliable, differences in response properties of individual neurons were found. The reasons for the apparent disparity between studies are unclear, although differences in species, training or analyses may be important.

Schulze et al. (2002) have reported evidence suggesting that changes occur in the cell and population responses in the guinea pig primary auditory cortex following extensive training on monaural modulation–rate discrimination with AM stimuli.

[9] Note that the visual pathway involves relatively little processing prior to the primary visual cortex compared to the auditory system.

There are also indications that cortical plasticity is facilitated by the nucleus basalis, which is thought to be involved in indicating the salience of sensory information. For example, the representation of the frequency of a tone in the primary auditory cortex was enhanced when the nucleus basalis was stimulated simultaneously with presentation of the tone (Bakin and Weinberger, 1996). Evidence for cortical plasticity in humans following auditory training measured using electrophysiological (e.g. Tremblay et al., 1997; Gottselig et al., 2004) or imaging techniques (e.g. Golestani and Zatorre, 2004) has been reported. However, these studies have used complex stimuli, such as speech sounds or complex tonal sequences. In animals, auditory plasticity following peripheral lesions has been reported throughout the brainstem (e.g. Iling, 2001; Friauf, 2004), and changes in the SC have been identified following altered spatial cues (Moore and King, 2004). A recent study has reported, apparently for the first time, evidence of plasticity in the human brainstem following training (Russo et al., 2005). Electrophysiological measurements of the auditory brainstem response to speech sounds were collected in nine children (8–12 years old) with 'learning disabilities' before and after auditory training consisting of a number of speech perception tasks. Measurements were compared to control children (both with and without learning disabilities). A measure of neural temporal processing was reported to change in the trained children only. Although this experiment was conducted in children (who may display greater potential for plasticity than adults) and with speech sounds, it suggests that training–induced plasticity may occur at lower levels of processing than previously thought.

2.5 Training–induced learning in binaural hearing

Section 2.2.4 described evidence that perceived spatial position depends on experience but noted that there is paucity of evidence on the influence of experience on the basic sensitivity to the interaural cues. A few studies have investigated the influence of training on discrimination of sound–source position, but such studies are inherently difficult to interpret as the localisation cue used by listeners may be unclear. An alternative approach is to measure sensitivity to individual cues over earphones. Section 2.4 described evidence that training influences the discrimination of basic stimulus parameters, such as frequency, with simple stimuli, such as unmodulated tones. Although the neural substrate of these changes remains unclear, studies of training–induced discrimination learning may contribute to the understanding of stimulus processing and may lead to effective therapies for auditory and language disorders. In this section, studies on the influence of experience on binaural sensitivity and potential clinical applications are described.

2.5.1 Indications of learning

There are many anecdotal reports of improvements in performance over many weeks or months of training on various binaural tasks. For example, Trahiotis et al. (1990) reported that, in their experience, many 'thousands and thousands' of trials of training may be necessary for inexperienced listeners to be reach asymptotic performance. Similarly, Hafter (1984) notes:

'[A] common finding in studies of lateralization [is] that is takes an extraordinarily long time to train the listeners. Performance, especially with tones, can improve over a period of months by more than an order of magnitude.'

Nuetzel and Hafter (1976) studied ITD discrimination with SAM tones and note:

'The training period for listeners JN and BP extended over months and weeks, respectively, while BK was given only a few days of training. In addition, JN and BP had extensive experience in other lateralization tasks. Supplemental practice trials were permitted as frequencies were changed, with asymptotic performance typically being reached in 500 or fewer trials.'

These authors used a variety of stimulus manipulations and it is not clear which conditions these findings apply to. However, it would appear that even experienced listeners might require extensive training to reach asymptote in novel conditions. Further, there is a suggestion that learning may be stimulus specific. Other studies have provided data on learning and these are summarised in Table 2.2. Hafter and Carrier (1970) reported that absolute thresholds in the N_0S_π condition (250–Hz tonal signal) from one listener improved by over 10 dB over the course of 3 months, and implied that similar results were observed with two other listeners. In contrast, Trahiotis et al. (1990) and Bernstein et al. (1998) tested 10 and 19 inexperienced listeners, respectively, over as many as 25 sessions with similar stimulus conditions used by Hafter and Carrier (1970). Bernstein et al. also tested with high–frequency stimuli and again found no evidence of learning. One difficulty in interpreting data from Bernstein et al. (1998) is that training with the various stimuli was provided in an order designed to minimise learning in the subjectively more difficult high–frequency conditions. Nuetzel and Hafter (1976) also reported that, in their experience, discrimination requires longer training periods than signal detection.

A number of studies have discussed learning with ITD discrimination under conditions of the so–called precedence effect. This paradigm differs from simple ITD discrimination in that *each* stimulus interval contains a pair of bilateral stimuli, usually broadband clicks, separated by a short silent interval (i.e. a total of four stimuli with standard 2AFC). The first, 'lead' clicks of each pair are diotic and ITD discrimination is conducted with the second, 'lag' clicks. When the silent interval is sufficiently small that the images of the two clicks blur, ITD discrimination often worsens markedly; this is often considered to reflect the perceptual dominance of the first click in each pair (hence the 'precedence effect'). Saberi and Antonio (2003) reported that the thresholds of one listener, which were relatively poor initially, approximately halved over 66 hours of training, and had not obviously reached asymptote. Saberi and Perrott (1990) reported similar findings. In contrast, Litovsky et al. (2000) found far less learning, although only provided approximately 20 hours of training. Saberi and Antonio also showed that it was difficult to determine if learning had occurred with their listener over this time span.

These studies indicate that training–induced improvements in ITD sensitivity may be apparent, at least under some conditions. However, interpretation of and comparison across these studies is difficult because listeners may have had a variety of prior psychoacoustical experience and because training was often providing on an *ad hoc* basis using different approaches for different listeners. Further, it is not clear if the training effects observed only occur with listeners with relatively poor initial performance. Wright and Fitzgerald (2001) appeared to have conducted the first systematic study of learning with ITD and ILD discrimination, which was designed to address these limitations. Constantinides (2004) has also recently completed a MD thesis including several experiments of the effects of training on various measures of binaural hearing.

Table 2.2. Studies related to training–induced learning with binaural sensitivity.

Study	Measure	Comments
Hafter and Carrier 1970	Binaural masked absolute threshold	N_0S_π at 250 Hz – learning in one listener over many months. Implied also observed in two others.
Trahiotis et al. 1990	Binaural masked absolute threshold	No learning with low–frequency N_0S_π.
Bernstein et al. 1998	Binaural masked absolute threshold	No learning with low– or high–frequency N_0S_π.
Saberi and Perrott 1990	ITD discrimination (lead–lag clicks)	Learning over many months.
Saberi and Antonio 2003	ITD discrimination (lead–lag clicks)	Learning over many months.
Litovsky et al. 2000	ITD discrimination (lead–lag clicks)	No learning but fewer trials than Saberi studies.
Wright and Fitzgerald 2001	ITD discrimination	500 Hz – untrained and trained listeners learned but authors concluded trained listeners did not learn more the untrained, although learning in groups not directly compared.
	ILD discrimination	4000 Hz – untrained and trained listeners learned and trained learned more than untrained; indirect evidence of frequency–specific learning.
Constantinides 2004	ILD discrimination	Early learning generalises across trial formats based on 2AFC.
	ILD discrimination	Early trained–induced learning in group trained with adaptive procedure but not in group trained with fixed ILD of 15 dB.
	ILD and ITD discrimination	On ITD at 500 Hz – trained listeners learned and untrained listeners did not. On ILD at 4000 Hz – both groups learned but trained learned more than trained. untrained. No obvious differences in learning curves between ITD and ILD. However, fewer training trials than were provided by Wright and Fitzgerald during pre–test.

2.5.2 Wright and Fitzgerald (2001)

Wright and Fitzgerald (2001) investigated the effects of training on ongoing ITD and ILD discrimination in inexperienced listeners using the pre–post paradigm (see Table 2.1). Listeners in the ITD and ILD experiments were trained at 500 Hz and 4000 Hz, respectively, and each experiment included an untrained group. Pre– and post–testing consisted of the trained stimulus and four untrained stimuli, which included the trained stimulus from the other experiment. Training consisted of many thousands of trials over 9–10 days. Discrimination thresholds were measured using the standard 2AFC and an adaptive procedure. Figure 2.12 plots group mean ITD and ILD thresholds across all sessions from both groups. Consider first the ITD experiment (right panel). Wright and Fitzgerald reported that both groups improved between pre– and post–tests and, while there was a trend for the trained group to learn more than the untrained group, this was not borne out by the statistical analyses. Consider now the ILD experiment (left panel). Again both groups improved between pre– and post–tests, but the analysis this time indicated that the trained group did learn more than the untrained group. Figure 2.13 presents the learning curves of trained listeners. Only two of the ITD–trained listeners (right panel) met the authors' statistical criterion for learning across training sessions (L9 and L10), whereas only two ILD trained listeners (left panel) did *not* meet this criterion (L7 and L8).

On the basis of these results, Wright and Fitzgerald argued that low–frequency ITD and high–frequency ILD discrimination learning are associated with different time–courses. However, there are a number of difficulties with this interpretation. Firstly, Wright and Fitzgerald did not compare directly learning between trained and untrained groups. Rather, they compared the groups' pre–test and then post–test thresholds. For the ITD experiment, neither was significant at the 5% level although marginal at pre–test (p=0.078). This was taken as indicating that the trained group did not learn more than the untrained group. In contrast, these comparisons were statistically significant in the ILD experiment. However, this analysis may have been insensitive to differential learning between the groups in the ITD experiment. Indeed, the trained group did have slighter *higher* thresholds at pre–test but slightly *lower* thresholds at post–test. This may reflect a material difference in learning between groups that may not be detected through comparisons of pre– and post–test alone.

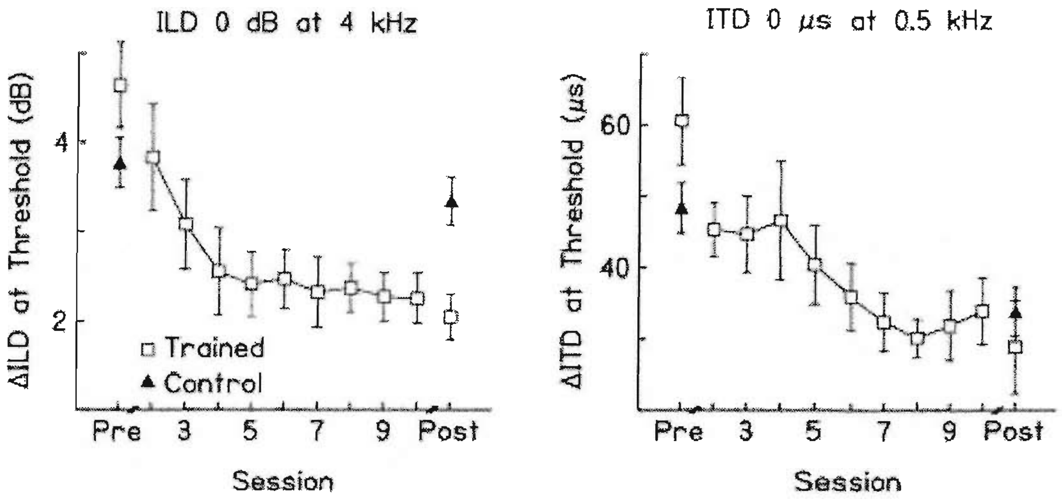


Figure 2.12. Group mean ITD (right panel) and ILD (left panel) threshold as a function of session number with the trained stimuli. Open and filled symbols indicate data from the trained and untrained groups, respectively; error bars indicate ±1 standard error. Reproduced from Wright and Fitzgerald (2001), with permission from Dr. Wright and publisher. Copyright (2001) National Academy of Sciences, U.S.A.

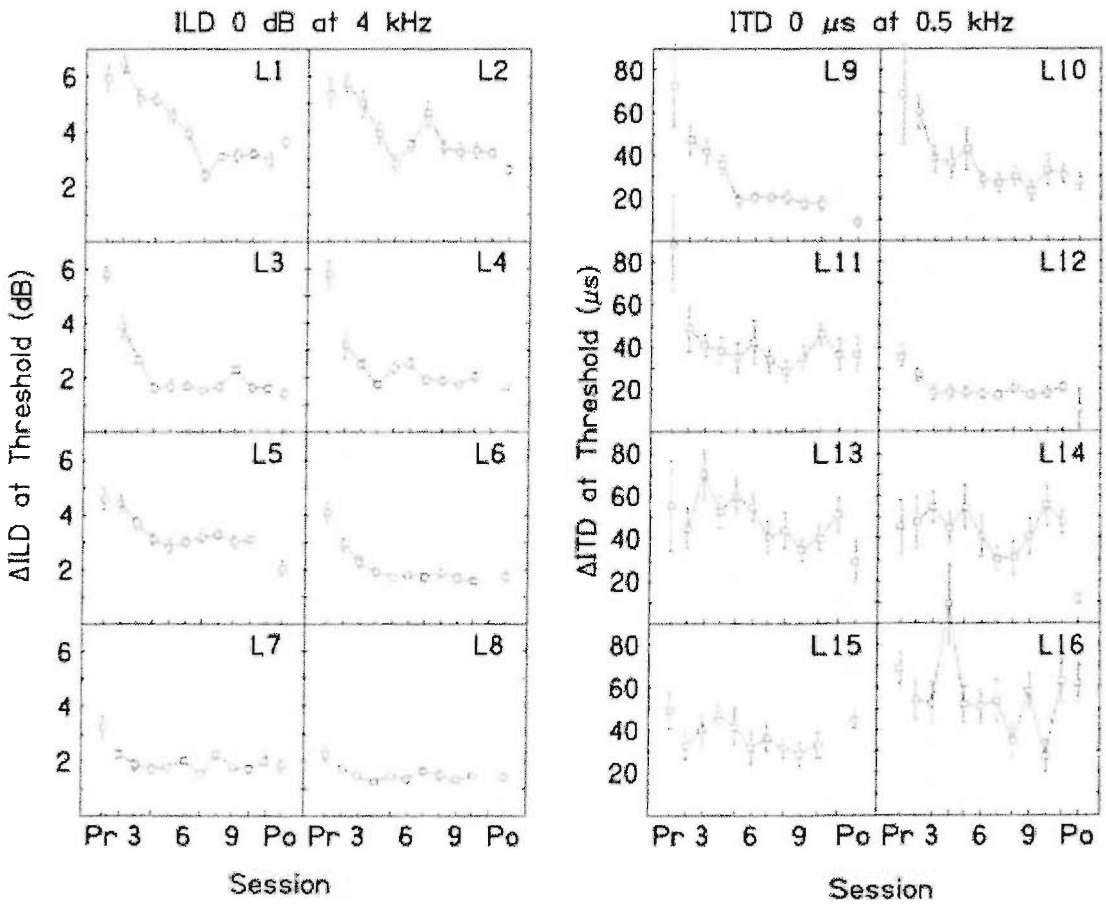


Figure 2.13. ITD (right panel) or ILD (left panel) discrimination learning curves for each listener separately. Reproduced from Wright and Fitzgerald (2001), with permission from Dr. Wright and publisher. Copyright (2001) National Academy of Sciences, U.S.A.

Another complication with Wright and Fitzgerald's data is that the mean ITD threshold of listeners in the ITD experiment was relatively good whereas the mean ILD threshold of listeners in the ILD experiment was relatively poor compared to previous research (e.g. Bernstein et al., 1998; also see Hinton et al., 2004), as acknowledged by the authors. Consequently, an apparent difference in the time–courses of learning may reflect at least partially the differences in general binaural sensitivity among listeners in each experiment. However, this is difficult to judge. While Wright and Fitzgerald measured pre–test thresholds in both conditions in both experiments, they did not report these data in a manner that can be compared across experiments. A third difficulty relates to the learning curves in Figure 2.13. Visual inspection of the curves from the ITD experiment indicates that there is a trend for other listeners (e.g. L11 and L12) to improve across training sessions, in addition to those who met the authors' criteria. Further, the data from other listeners are highly irregular and may be difficult to interpret. The statistical analysis of the curves may also have been insensitive to learning as it rested on linear regression of thresholds during training, whereas the learning curves are clearly non–linear.

Wright and Fitzgerald also argued that learning on ILD discrimination was specific to frequency but not baseline ILD. This was based on a comparison of pre–test and then post–test thresholds between groups with the untrained stimuli. This analysis indicated that the trained group learned more than the untrained group with a dichotic 4000–Hz tone containing a baseline ILD of 6 dB. This was interpreted as complete generalisation of learning across baseline ILD. The analysis was also taken to indicate that the trained group did not learn more than the untrained group with otherwise diotic stimuli at 500 and 6000 Hz. This was interpreted as incomplete generalisation of learning across frequency. However, there are a number of difficulties with this interpretation. Firstly, the possibility that the trained group learned more than the untrained group at 500 and 6000 Hz cannot be excluded because learning was not compared directly between the groups, as discussed above for the trained stimulus in the ITD experiment. Secondly, the finding that the trained group learned more than the untrained group with the untrained dichotic stimulus does not exclude the possibility that the generalisation was incomplete, as was illustrated in Figure 2.11. A third difficulty relates to the absence of data on the magnitude of learning with the untrained stimuli. Wright and Fitzgerald's argument that ILD discrimination learning generalised incompletely to 500 and 6000 Hz rests on the finding that the trained group did not appear to learn more than the untrained group at these untrained frequencies. However, this assumes that the potential magnitude of learning at these untrained frequencies was comparable to that at the trained frequency. While this seems likely at 6000 Hz, given its proximity to the trained frequency, the validity of this assumption at 500 Hz is less clear. Although previous studies indicate that asymptotic ILD thresholds at 500 and 4000 Hz are comparable (e.g. Grantham, 1984; Yost and Dye, 1988), little is known about ILD discrimination in untrained listeners, and Wright and Fitzgerald do not report pre–test thresholds in a manner that can be compared across stimuli.

Moore et al. (2003) have suggested that Wright and Fitzgerald's data also indicate that learning on ILD discrimination generalised incompletely to ITD discrimination. This argument appears to be based on Wright and Fitzgerald's conclusion from the ILD experiment that the trained group did not appear to learn more than the untrained group on ITD discrimination at 500 Hz. However, the situation is complex. For example, Wright and Fitzgerald argued that training–induced learning did not occur with ITD discrimination. The specific interpretation that learning generalises between stimuli requires performance with both stimuli to be influenced by training. Thus, if formal training on ITD discrimination was ineffective in inducing learning, the finding that training on ILD discrimination was also ineffective cannot be viewed as incomplete generalisation. However, Wright and Fitzgerald's interpretation on the effect of training with ITD discrimination may reflect insensitive analyses, as noted above; this also applies to ITD thresholds from the ILD experiment. Nevertheless, the hypothetical absence of generalisation from high–frequency ILD discrimination to low–frequency ITD discrimination may actually reflect incomplete generalisation across frequency, rather than across cue.

2.5.3 Constantinides (2004)

Constantinides's (2004) studies involving adult listeners are summarised in Table 2.2 (also see Table 2.1). Of particular interest, one experiment compared the effect of training in identical stimulus conditions to Wright and Fitzgerald (2001). The number of training trials was comparable to the number of trials Wright and Fitzgerald provided on the trained stimulus during pre–test. Training–induced learning and broadly comparable time–courses of learning were reported with both trained groups. This is in apparent contrast with Wright and Fitzgerald, who suggested that ILD discrimination learning occurred over a longer time–course. The reasons for this apparent disparity are unclear.

2.5.4 Theories of learning

Section 2.5.1 noted that tasks involving MLD and discrimination might be associated with different time–courses of learning, although there is no direct evidence of this. Similar observations have also been made with monaural tasks (Watson, 1980). The reasons for this apparent difference between tasks are unclear. Nevertheless, overall, studies of MLD with N_0S_0 vs. N_0S_π (Trahiotis et al., 1990; Bernstein et al., 1998) and ITD discrimination with tones (Wright and Fitzgerald, 2001) appear to indicate that low–frequency ITD sensitivity is associated with a relatively short time–course of learning. However, results from studies of ITD discrimination with lead–lag clicks (Saberi and Perrott, 1990; Saberi and Antonio, 2003) may indicate that the time–course of learning with ITD sensitivity under these conditions is considerably longer. This may reflect differences in the stimulus conditions. For example, some authors have suggested that the auditory cortex plays a greater role in the processing of ITD with lead–lag clicks under the conditions in these studies than with tones (Litovsky et al., 1999), which may have implications for plasticity and learning. However, modelling work by Hartung and Trahiotis (2001) showed that the poorer ITD sensitivity with lead–lag clicks could arise from temporal interactions in the cochlea. It is also possible that the results of Saberi and colleagues [and Hafter and Carrier (1976) with MLD] may indicate that substantial learning effects can be observed whenever training normal–hearing listeners who have particularly poor thresholds initially.

Although there are a number of uncertainties with Wright and Fitzgerald's data on ITD and ILD discrimination learning, the reported differences in time–course are intriguing. One might initially imagine that listeners learn to hear smaller changes in lateral position with training, and that this would be similar for both cues. However, different time–courses might be reflected in listeners learning to detect ITD and ILD with less obvious cues, and theoretically non–spatial cues with ILD. Introducing an ILD requires the experimenter to change the level at one or both ears and listeners may learn to take advantage of the associated monaural cues. However, Bernstein (2004) has recently reported evidence that ILD discrimination in highly trained normal–hearing listeners reflects the use of binaural rather than monaural cues, at least with otherwise diotic tones. Wright and Fitzgerald (2003) have also suggested that the time–course of learning on monaural level discrimination mirrors their results with ITD discrimination, although similar limitations are apparent with the analyses of both sets of data. Listeners may learn to detect small changes in ITD from the decorrelation of the signals, which would not be possible for ILD. However, it is unclear how this would lead to a longer time–course of learning with ILD.

Wright and Fitzgerald (2001) suggested that the seemingly shorter time–course of learning with ITD compared to ILD discrimination reflected procedural compared to perceptual learning. They also suggested that the learning with ITD discrimination occurred between test sessions, rather than within sessions. They have made a similar argument for other stimulus parameters (Wright and Fitzgerald, 2003). However, these suggestions were not supported by data, and the results of Hawkey et al. (2004) indicate that it may not be possible to account for early rapid learning in these terms.

Wright and Fitzgerald speculated that the apparent difference in time–courses of learning might be associated with differences in daily experience with the cues. Specifically, they suggested that listeners may have had less experience with high–frequency ILD prior to the experiment, and that the greater experience with low–frequency ITD had essentially saturated any available plasticity with this cue. However, many commonly occurring sounds contain information above 2000 Hz, such as speech, and these may produce substantial ILDs. Wright and Fitzgerald also speculated that the apparent difference in time–courses of learning may reflect differences in the inherent potential for training–induced plasticity in the processing of the stimulus conditions, and this may be associated with differences in brainstem processing. As described in Section 2.3.2, it is thought that low–frequency ITD and high–frequency ILD are processed primarily by distinct brainstem pathways (the MSO and LSO) and by cells with distinct binaural interactions (EE–type and EI–type cells). Wright and Fitzgerald suggested that the greater potential for plasticity in the high–frequency ILD pathway was related to the fundamental role of inhibition with EI–type cells. The general notion that discrimination learning reflects plasticity in the brainstem is controversial, at least partly because it has proven difficult to account for frequency discrimination learning even at the level of the primary auditory cortex (see Section 2.4.3). However, brainstem plasticity has been reported following profound hearing loss (e.g. Iling, 2001) and the recent study by Russo et al. (2005, Section 2.4.3) may indicate that training–induced plasticity occurs at lower levels than hitherto expected. The specific notion that the ILD pathway is associated with more potential for plasticity due to the fundamental role of inhibition is complicated by evidence described in Section 2.3.2. This suggests that EE–type cells also receive inhibitory input (e.g. Fitzpatrick et al., 2002) and that this may play a central role in shaping the response of the cell to ITD (Brand et al. 2002). A recent study has also reported that the ITD responses of MSO cells of the Mongolian gerbil are dependent on acoustic experience, at least prior to maturity (Seidl and Grothe, 2005).

It is unclear from Wright and Fitzgerald's study if the apparent difference in time–courses of learning reflects the different cues or frequencies. For example, if this reflects the different frequencies, different time–courses may be apparent for ITD discrimination with low–frequency unmodulated tones and high–frequency AM tones. Section 2.3.3 discussed the theory that low– and high–frequency cues are processed in distinct pathways. It was noted that this is complicated by a number of findings, such as that cells in both MSO and LSO, and with EE–type and EI–type responses, are sensitive to high–frequency ITD (e.g. Fitzpatrick et al., 2002). Further, ITD thresholds with narrowband stimuli may not be related to the number of ITD–sensitive cells tuned to it (e.g. Shackleton et al., 2003). Given these uncertainties, and the uncertainties with aspects of Wright and Fitzgerald's (2001) conclusions (discussed in Section 2.5.2), further research into learning on ITD and ILD discrimination is required before clear conclusions can be drawn.

Again, while there are uncertainties with Wright and Fitzgerald's conclusions regarding the apparent specificity of learning, the possibility that this is specific to frequency but not ILD is also intriguing. The apparent frequency–specificity parallels that found in some studies of frequency discrimination (Section 2.4.2), and may reflect similar mechanisms (e.g. related to cochleotopicity). In contrast to frequency, cells from the SOC to the cortex are not generally tuned to ILD, although tend to respond to ILDs favouring a single hemifield (Yin, 2002). These results may therefore reflect general differences in how frequency and ILD are represented.

Further research on learning with ITD at low and high frequencies may also have implications for psychoacoustical models of ITD processing. For example, Section 2.3.5 described models that assume that ITD discrimination is accomplished using the same basic mechanism independent of frequency (e.g. Bernstein and Trahiotis, 2002). While physiological data indicate that ITD processing in the mammalian brain is probably more complex than this, these models can account for a variety of psychoacoustical data with unmodulated and AM tones. Further, a recent physiological study suggests that ITD processing at low and high frequencies is at least functionally comparable (Griffin et al., 2005). However, if the time–course of ITD discrimination learning depends on frequency, this would seem difficult to account for with these simple psychoacoustical models. An important issue here, though, is potential differences in asymptotic thresholds between stimuli. For example, 'conventional' high–frequency AM stimuli are generally associated with poorer asymptotic thresholds than low–frequency unmodulated tones (see Section 2.2.2). This difference may lead to differences in the time–courses of learning, as discussed in Section 2.4.2, and so confound the effect of frequency. It may be possible to address this issue, and so isolate the effect of frequency, using transposed stimuli which, under some conditions, yield asymptotic thresholds that are comparable to those with low–frequency unmodulated tones.

2.5.5 Applied and clinical applications

Various types of hearing impairment are known to potentially influence binaural hearing. For example, studies have considered the effect of bilateral sensorineural hearing impairment (SNHL) on interaural discrimination (e.g. Hawkins and Wightman, 1980; Häusler et al., 1983; Smoski and Trahiotis, 1986; Gabriel et al., 1992; Koehnke et al., 1995; Koehnke and Besing, 1996; Smith–Olinde et al., 1998). It is generally concluded that SNHL is often associated with impaired ITD and ILD thresholds. While this is probably true in some cases, there are a number of difficulties when comparing data with normal–hearing listeners. For example, a plethora of other factors may influence performance on binaural tests, and these may also vary between hearing–impaired and normal–hearing listeners. These factors include experience with the test procedures and stimuli. Unlike studies with normal–hearing listeners, few studies with hearing–impaired listeners report providing extensive training prior to formal measurements. Hawkins and Wightman (1980) did provide training and, although the results from this were not reported, improvements in performance may have occurred over several hours judging from the total test time reported. Similar difficulties are mirrored in the literature on bilateral cochlear implantation. A number of studies have examined interaural thresholds with direct electric stimulation of the cochlea, bypassing the implant processor (e.g. van Hoesel and Clark, 1997; van Hoesel and Tyler, 2003; van Hoesel, 2004; Wilson et al., 2003), although these have provided little if any training. Interpretation of and comparison across studies not using training is difficult because the precise impact of training on performance is unclear.

New bilateral cochlear implant (BICI) users may have added difficulties related to the absence of binaural input for many years. Animals studies have shown that profound hearing loss arising from peripheral lesions is associated with plasticity in the CN, SOC and IC, the functional consequences of which are unclear (Illing, 2001; Syka, 2002; Friauf, 2004). While these changes may at least partially reverse following electrical auditory stimulation, there may be some added benefit in providing training on binaural tasks, perhaps ITD or ILD discrimination, following bilateral implantation. This may help maximize use of the available binaural cues or quicken the learning process. Some long–term users of unilateral hearing aids with bilateral hearing loss may develop difficulties in using information provided to the unaided ear, and possibly in binaural processing (Moore and King, 2004). These patients may also benefit from binaural training if being fitted with bilateral devices.

There are further complex issues relating to signal processing in auditory prostheses and the interaction of the prosthesis with the auditory system. For example, current cochlear implant electrodes are inserted only partially into the cochlea (e.g. Skinner et al., 2002) and BICIs may therefore represent low–frequency information in normally high–frequency channels. Further, current BICI processing may not preserve low–frequency, fine–structure–based ITD (Wilson et al., 2003). However, the view that this is the dominant horizontal–plane localisation cue in normal–hearing people (Section 2.2.3) has motivated reconsideration of BICI processing. It is therefore of interest to consider the potential for plasticity associated with the processing of ITD, but data on learning using low–frequency acoustic stimuli may not apply to BICIs.

Bilateral implants may also stimulate asymmetrical populations of auditory–nerve fibres, possibly because of asymmetrical electrode insertions, current spread or nerve survival. It has been suggested that this is analogous to presenting AM tones acoustically to normal–hearing listeners with an interaural carrier frequency difference (Long, 2000; Wilson et al., 2003). This manipulation can produce asymmetrical topographical distributions of activity within the two auditory nerves. A number of studies have shown that, if sufficiently large, ITD discrimination is also adversely affected (Henning, 1974; Nuetzel and Hafter, 1976, 1981; Long, 2000). These conditions are highly unnatural and data on the effects of training with otherwise diotic stimuli may not apply. There is also interest in the possibility of using binaural measures, such as ITD discrimination, to assist with the selection of the parameters of the bilateral implants after implantation in clinical contexts, perhaps given asymmetrical electrode insertions. An important issue here is the amount of training that is required for useful measurements to be obtained.

2.6 Summary and aims

The perception of perceived spatial position, such as relating to ITD and ILD cues, is dependent on experience. Far less is known about the influence of experience on basic binaural sensitivity, such as measured with discrimination tasks. This is in contrast to studies of monaural discrimination (e.g. of stimulus frequency) that have found that performance often improves substantially over many hundreds to thousands of trials. Anecdotal reports suggest this also occurs with interaural discrimination. The first systematic study of multi–hour training on interaural discrimination by Wright and Fitzgerald (2001) concluded that discrimination learning with low–frequency ongoing ITD and high–frequency ILD are associated with fundamentally different time–courses. However, there are several limitations with this study that may have influenced this conclusion. It is also unclear if this apparent difference reflects the difference cues or frequencies. Further research on ITD discrimination learning is indicated to clarify these issues.

Further research in this area may have implications for our understanding of binaural processing. Wright and Fitzgerald (2001) suggested that the apparent differences in time–course of learning observed might reflect differences in how the stimulus conditions are processed in the brainstem. If Wright and Fitzgerald's results reflect the different frequencies, differential learning effects may be apparent with low– and high–frequency ITD. Some physiological evidence suggests that low– and high–frequency ITD may be processed in distinct brainstem pathways, the latter sharing the 'high–frequency ILD' pathway, although the situation is probably more complex. However, there is currently no evidence that discrimination training is associated with plasticity in the brainstem, or that binaural processing is more plastic at high than at low frequencies. In fact, studies of plasticity in the auditory cortex associated with frequency discrimination learning have produced variable results. Nevertheless, if the time–course of ITD discrimination learning is dependent on frequency, all else being equal, this may have implications for models of ITD processing that assume a common mechanism independent of frequency.

Further research in this area may have relevance to clinical populations. For example, few studies of ITD sensitivity using unmodulated and AM stimuli with hearing–impaired people and users of BICI report providing training. In contrast, studies using normal–hearing listeners usually provide extensive training. Uncertainty regarding the influence of training complicates comparison across groups. However, data from low–frequency unmodulated tones may not apply to some clinical populations. For example, BICIs may present low–frequency information in normally high–frequency channels, and may produce asymmetrical peripheral excitation. Acoustic studies using high–frequency AM stimuli may provide insight into the potential for learning and plasticity under these conditions.

In summary, further research is required to address the question: does the time–course of ITD discrimination learning in normal–hearing listeners depend on the stimulus? Specifically, does the time–course of ITD discrimination learning differ under conditions that may be relevant to BICIs to that reported by Wright and Fitzgerald (2001) for low–frequency tones? These conditions include stimuli that convey ongoing ITD in high–frequency auditory channels (e.g. high–frequency transposed tones) and stimuli that produce an interaural asymmetry in the topographic excitation of the auditory nerves (i.e. an IFD). Closely related to this is the additional question: does the apparent difference in time–courses of learning reported by Wright and Fitzgerald depend on frequency or binaural cue? It is also of interest to consider whether the apparent absence of generalisation from high to low frequencies reported for ILD by Wright and Fitzgerald is also apparent with ITD. However, given the uncertainties as to whether Wright and Fitzgerald's conclusions were justified, it is important to revisit ITD discrimination learning at low frequencies and to make detailed comparisons with their results.

The aims of this thesis are therefore to:

(1) Design suitable methodology for studies of discrimination learning in normal–hearing and hearing–impaired populations.

(2) Gain insight into ITD discrimination learning under conditions that may be relevant to BICI, as background to future studies with BICIs.

(3) Evaluate the hypothesis that the time–course of ITD discrimination learning depends on frequency, and the implications of this for our understanding of binaural processing and for clinical populations.

(4) Gain insight into the extent to which training–induced learning with ITD discrimination generalises to alternative stimulus conditions, such as across frequency, and the implications of this for our understanding of binaural processing and for clinical populations.

(5) Compare these findings to those reported by Wright and Fitzgerald (2001) and address outstanding uncertainties.

Chapter 3. General methods

This chapter presents the methodology common to all experiments described in subsequent chapters. Additional methodology specific to individual experiments is given in the relevant chapter. Approval of the Institute of Sound & Vibration Research Human Experimentation Safety and Ethics Committee was obtained before commencing each experiment.

3.1 Learning experiments

Experiments 1 and 3 studied directly the effect of repeated performance of an ITD discrimination task (i.e. 'training') on ITD thresholds. The 'pre–post' design was used (Section 2.4.1). Here, the primary measure of learning is ITD discrimination measured at post–test relative to performance measured at pre–test. The specific approach was similar to previous studies of auditory discrimination learning (see Table 2.1), such as Wright and Fitzgerald (2001): listeners underwent familiarisation, pre–testing, formal training and finally post–testing. Many inexperienced listeners find adjusting to listening to changes in lateral position quite difficult and initial familiarisation was provided to assist with this. This was found necessary to avoid highly unstable performance on ITD discrimination during pre–test, which was undesirable because the added variance made detecting learning between pre– and post–test more difficult.

Pre– and post–test sessions were identical for each listener, were conducted on separate days to the training sessions and were nominally separated by 10 days (although the actual separation depended on listener availability). These sessions were restricted to 12 blocks of test trials. Thus, in Experiment 1 pre– and post–testing was conducted with two stimuli and so six blocks of test trials were completed for each stimulus; in Experiment 3 testing was conducted with three stimuli and so four blocks of test trials were completed for each stimulus. Twelve blocks per session, lasting about 80 minutes, appeared to reflect the limits of listeners' attention and motivation whilst permitting the enhancement of accuracy through the averaging of repeated measurements; testing beyond 12 blocks tended to lead to gradually worsening performance. Within each session, half of the total blocks were completed with each stimulus then, following a 5–minute break, repeated in reverse order; blocks with one stimulus were completed before moving to the next stimulus. The order of testing was varied across listeners. Formal training consisted of six sessions, each conducted on a different day, and six blocks were completed in each session. This amounted to approximately 2000 training trials. While this was less than used in some previous experiments, inspection of individual listeners' training data from Wright and Fitzgerald (2001) indicated that, in most cases, ITD and ILD thresholds approximated asymptotic performance within this time.

In Experiment 2, learning was estimated indirectly by comparing ITD thresholds in highly experienced listeners (Experiment 2a) to those in inexperienced listeners (Experiment 2b). This permitted an initial evaluation of hypotheses prior to a full, and more time intensive, learning experiment.

3.2 Listeners

Listeners were normal–hearing 18–35 year–olds recruited through advertisements around the University of Southampton, and thus were largely under– or post–graduate students. A significant history of ear or hearing problems, signs or symptoms of current ear or hearing problems or excessive wax were identified using a questionnaire (Appendix A), otoscopy and pure tone audiometry; listeners with such problems were excluded. Listeners had audiometric thresholds (British Society of Audiology, 1981), including at 3000 and 6000 Hz, of better than 20 dB HL; the thresholds at any one frequency did not differ by more than 15 dB between the two ears. Audiometry was not repeated during an experiment, although listeners informally reported having had no significant exposure to noise within 48 hours of each session and being free of colds and hearing problems. For Experiments 1, 2b and 3, listeners reported having had no significant experience with psychophysical testing. Listeners in Experiment 2a were recruited on the basis of having extensive experience of psychoacoustical, and in particular binaural, testing.

3.3 Psychophysical procedure

This section describes the procedure used to estimate ITD thresholds during both testing and training. The same general approach was also adopted for signal detection tasks in Experiments 2 and 3. In summary, measurements were made using a four–interval one–cue three–alternative forced choice (4I 1C 3AFC) task combined with an adaptive staircase procedure. The written instructions for ITD discrimination provided to listeners are given in Appendix B.

3.3.1 Task and trial format

The task is illustrated in Figure 3.1. Each trial consisted of four observation intervals containing a bilateral stimulus. The first interval, referred to as the 'cue', always contained the diotic 'standard' stimulus. Two of the following, 'primary' intervals also contained the standard. The remaining interval, selected at random, contained the dichotic 'target' stimulus containing the ITD. The listener was required to identify the target. In Figure 3.1, the target is presented in the second primary interval.

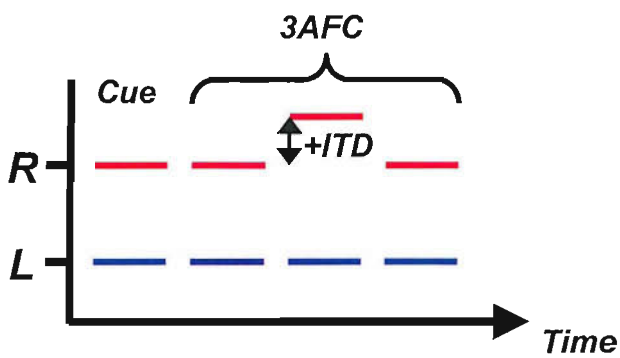


Figure 3.1. Schematic illustration of the four–interval one–cue three–alternative forced choice task. Red and blue markers indicate the four successive stimuli presented in the right and left channels, respectively.

Forced–choice paradigms are commonly used because they do not require the listener to form a subjective criterion regarding how to respond (Macmillan and Creelman, 1991). This is important since a systematic change in ITD threshold is therefore more likely to reflect a change in sensitivity rather than in this criterion. Three–alternative, as opposed to perhaps the more common 2AFC task was used for two reasons. Firstly, the probability of a correct response through guessing is lower with 3AFC. The adaptive procedure is therefore less likely to continue to reduce the ITD below the listener's threshold due to guessing, which may otherwise result numerous ineffective trials and a loss of the listener's attention. Secondly, 3AFC may reduce measurement variance over 2AFC (Leek, 2001). An initial cue was included because preliminary data from inexperienced listeners indicated a strong selection bias away the first primary interval. This bias was generally not apparent in subsequent work with the initial cue, described in Appendix C. A final cue, as in the paradigm used by Bernstein and Trahiotis (2002), was not included as there was no obvious selection bias relating to the third primary interval and it would have substantially lengthened the duration of the procedure.

The 3AFC task essentially involved the listener identifying the stimulus that was the 'odd–one–out', an example of what is often termed an 'oddity' paradigm. Many other ITD discrimination experiments have employed oddity paradigms, such as four–interval two–cue 2AFC (e.g. Bernstein and Trahiotis, 1982, 2002; Koehnke et al., 1995) and 3AFC (Oxenham et al., 2004). It has been suggested that learning may be reduced with these compared to two–interval 2AFC (Colburn and Trahiotis, 1991). This appears to be due to the added complexity with two–interval 2AFC of a listener having to detect a difference between stimuli *and* order the stimuli to identify the target rather than a difference in sensory processing *per se*. The use of an oddity paradigm may therefore provide a clearer indication of changes due to sensory processing. The use of an oddity paradigm may therefore be seen as an improvement over Wright and Fitzgerald (2001) who used two–interval 2AFC.

The temporal structure of each trial is illustrated in Figure 3.2. A light indicated the start of the trial 400 ms prior to the first interval. Each observation interval was 400 ms in duration, separated by a 400–ms silent gap. A light then indicated the response interval; the listener pressed one of three buttons on a keypad and was given unlimited time to respond. In some experiments, visual feedback was then provided in the form of a 400–ms–long display indicating the interval the target actually appeared in. Trials were separated by a pause of about 800 ms. A trial including feedback typically lasted 6–7 seconds.

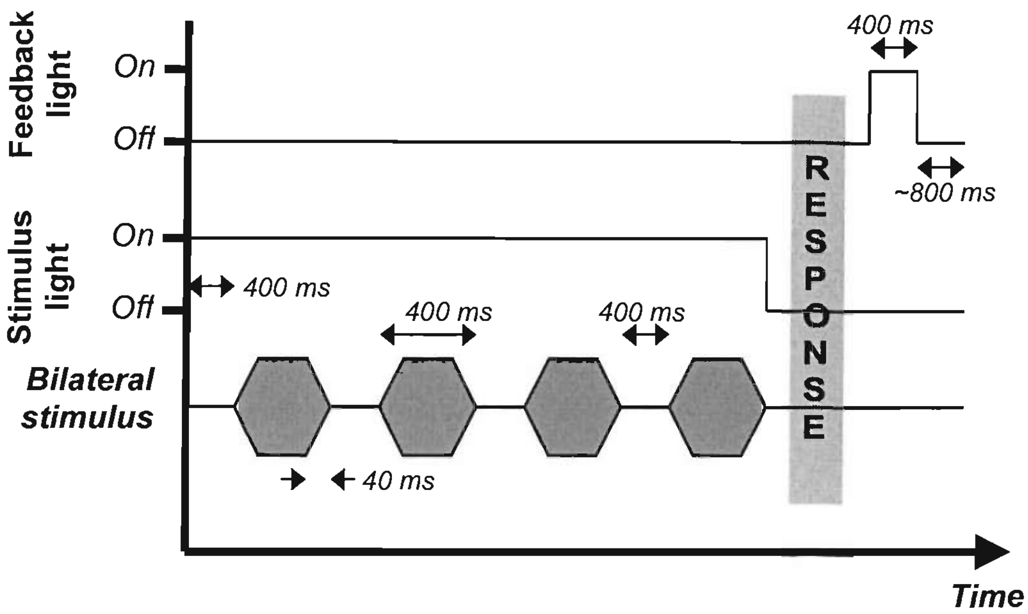


Figure 3.2. Illustration of temporal structure of trial used to measure ITD thresholds.

3.3.2 Adaptive procedure

An adaptive procedure was used in combination with the forced–choice task to measure ITD thresholds. Adaptive procedures adjust the ITD across consecutive trials in a stepwise manner depending on the responses of the listener and have been successfully used to measure ITD thresholds in experienced listeners (e.g. Bernstein and Trahiotis, 2002; Bernstein et al., 1998; Saberi, 1995) and inexperienced listeners (e.g. Wright and Fitzgerald, 2001; Saberi and Antonio, 2003; Saberi et al., 2004). Adaptive procedures are perhaps the most commonly used method of measuring discrimination thresholds and providing training in auditory discrimination learning experiments (Table 2.1). Their popularity is largely due to the ability to target a predetermined response level on the underlying psychometric function regardless of overall sensitivity, provide data in physical units and avoid 'floor' or 'ceiling' effects that may occur if a fixed set of values was used (Leek, 2001). The particular type of procedure used in these experiments was the popular 'staircase' or 'transformed up–down' procedure. This procedure does not require strong assumptions regarding the psychometric function and is robust to changes in sensitivity during testing (Levitt, 1971). Constantine (2004) found greater improvements in binaural discrimination thresholds with training when using an adaptive procedure compared a constant–stimulus procedure.

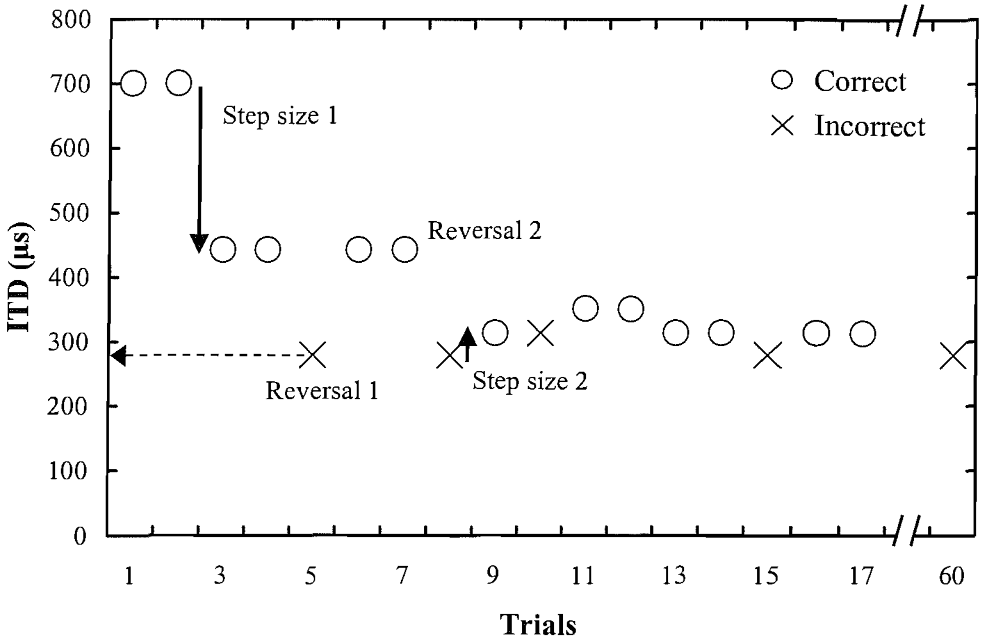


Figure 3.3. Portion of a hypothetical track of a staircase procedure used to measure ITD thresholds. The terms 'step size' and 'reversal' are illustrated.

A portion of a hypothetical track of a staircase procedure is illustrated in Figure 3.3, which plots the ITD of the target across trials. Circles indicate 'correct' responses and crosses indicate 'incorrect' responses. Although the details of the adaptive procedure were modified across experiments, the primary characteristics were:

- *Response scoring.* A 'correct' response was where the listener selected the interval containing the target stimulus. Conversely, an 'incorrect' response was where the listener did not select the interval containing the target stimulus.

- *Decision rule.* The ITD was decreased following a correct response across k trials and increased following an incorrect response on a single trial. In Experiment 1, $k = 3$ and the trials were *not* required to be consecutive; in Experiments 2 and 3, $k = 2$ and the trials were required to be consecutive. These rules theoretically target the 75% (Zwislocki and Relkin, 2001) or 71% (Levitt, 1971) correct response level, respectively. Further, an incorrect response on the first trial was ignored by the decision rule and the same ITD presented on the second trial; this was to avoid unnecessarily increasing the ITD due to initial inattention. The ITD at which the direction of the staircase changes is referred to as a 'reversal', as illustrated in Figure 3.3.

- *Initial ITD.* The ITD on the first trial was generally 700 μs, as in Figure 3.3, or two large step sizes above preliminary estimates of threshold.
- *Step size.* The step size refers to the magnitude by which the ITD changed on consecutive trials, as illustrated in Figure 3.3.
- *Termination rule.* Staircases were either terminated after eight 'scored' reversals (see below) or a given number of trials.
- *Permissible ITDs.* The resolution of ITD and minimum ITD (identical) differed across experiments. The maximum ITD was 2000 μs, corresponding to an IPD of approximately 90° at 128 Hz, in order to avoid the breakdown of binaural fusion and reversal of the psychometric function above this (Yost, 1981).
- *Calculation of threshold.* Data from the reversals with the largest step size and the first with the smallest step size were discarded. The next reversal was also discarded if required to leave an even number; reversals that arose from limiting the ITD to 2000 μs were also discarded. If there were less than four remaining reversals, threshold was scored as '≥2000'; otherwise, threshold was estimated as the average of the remaining ('scored') reversals. Computer simulations prior to each experiment determined the upper limit of ITD threshold that could produce a precise result.

3.4 Stimuli

3.4.1 General parameters

The primary stimuli were an unmodulated 128–Hz tone and transposed and SAM tones modulated at 128 Hz. This value was within the range of modulation rates that produce optimum ITD sensitivity with 'conventional' AM stimuli (Henning, 1974; Nuetzel and Hafter, 1976, 1981; McFadden and Pasanen, 1976), the greatest enhancement in ITD sensitivity using transposed stimuli (Bernstein and Trahiotis, 2002, 2003) and comparable sensitivity in general between transposed and unmodulated tones (Bernstein and Trahiotis, 2002; Oxenham et al., 2004).

Stimulus duration was 400 ms, including 40–ms $\cos^2$ onset and offset ramps, this being sufficiently long that further increases were unlikely to influence performance whilst minimising the length of each trial for inexperienced listeners. Ramps were applied after imposition of a phase delay (to produce the ITD) and ramp duration was selected to avoid audible transient distortion and to reduce the salience of the onset ITD cue. The starting phase, prior to imposition of a phase delay, was randomised across presentations, although identical in right and left channels.

An ITD was imposed by manipulating the phase in one channel and in the envelope only for AM stimuli. Listeners were randomised to receive either a right– or left–leading ITD although this was consistent across an experiment for each listener. Preliminary studies indicated that ITD thresholds for any one listener were not dependent on the side to which the ITD was presented and that a diotic stimulus was lateralised on the midline.

3.4.2 Amplitude–modulated stimuli

A general expression for an AM tone is:

$$x(t) = m(t)\sin(\omega_c t + \phi_c) \qquad (1)$$

where $m(t)$ is the modulator, and ω_c is the angular frequency and ϕ_c is the starting phase of the carrier. The modulator for a SAM tone, $m_{SAM}(t)$, is a raised cosine function:

$$m_{SAM}(t) = [0.5 + 0.5\cos(\omega_m t + \phi_m)], \qquad (2)$$

where ω_m is the angular frequency/rate and ϕ_m is the starting phase of the modulator. The parameter ϕ_m was manipulated to produce an envelope–based ongoing ITD. The waveform and amplitude spectrum of SAM given a carrier frequency of 4000 Hz are illustrated in Figure 3.4.

The modulator with the transposed tone consists of a half–wave rectified tone. Previous studies have generated the transposed tone and imposed the ITD computationally. However, an analytical expression for this stimulus can be obtained as follows. The Fourier series of the half–wave rectified tone, and thus the modulator of the transposed tone, $m_{trans}(t)$, is given by (Hartmann, 1998, p. 106):

$$m_{trans}(t)=\frac{1}{\pi}+\frac{1}{2}\cos(\omega_m t+\phi_\omega)-\frac{2}{\pi}\sum_{n=2,4,6...}^{\infty}\left(-1^{n/2}\right)\frac{1}{n^2-1}\cos[n(\omega_m t+\phi_\omega)] \qquad (3)$$

Previous studies also low–pass filtered the rectified tone in order to prevent spectral components of the transposed stimulus falling within the frequency range where these may contribute to ITD sensitivity. This can be achieved in Equation 3 by selecting an appropriate value of n. In most experiments reported here, n=2 to restrict the rectified tone to the DC component and the first two harmonics; Equation 3 becomes:

$$m_{trans}(t)=\frac{1}{\pi}+\frac{1}{2}\cos(\omega_m t+\phi_\omega)+\frac{2}{3\pi}\cos[2(\omega_m t+\phi_\omega)]$$

Substituting this into Equation (1) and expanding gives an expression for the transposed tone, $x_{trans}(t)$:

$$x_{trans}(t)=\frac{1}{\pi}\sin(\omega_c t+\phi_c)+\frac{1}{2}\cos(\omega_m t+\phi_m)\sin(\omega_c t+\phi_c)+\frac{2}{3\pi}\cos[2(\omega_m t+\phi_m)]\sin(\omega_c t+\phi_c)$$

The Fourier series of the transposed tone can be obtained using the following trigonometric identity:

$$\sin(A)\cos(B)=\frac{1}{2}[\sin(A+B)+\sin(A-B)],$$

giving:

$$\begin{aligned} x_{trans}(t) = {} & \frac{1}{3\pi}\sin[(\omega_c-2\omega_m)t+\phi_c-2\phi_m]+\frac{1}{4}\sin[(\omega_c-\omega_m)t+\phi_c-\phi_m] \\ & +\frac{1}{\pi}\sin(\omega_c t+\phi_c) \\ & +\frac{1}{4}\sin[(\omega_c+\omega_m)t+\phi_c+\phi_m]+\frac{1}{3\pi}\sin[(\omega_c+2\omega_m)t+\phi_c+2\phi_m] \end{aligned} \qquad (4)$$

The five terms correspond to the five central components of the transposed tone used by Bernstein and Trahiotis (2002), which was based on low–pass filtering the rectified tone at 2000 Hz. This can be seen in Figure 3.5, which plots the waveform and the amplitude spectrum of the transposed tone used by Bernstein and Trahiotis (2002), and also Experiment 2a, given a carrier frequency of 4000 Hz. The analytical approach was used because it permitted straightforward control over the number and amplitudes of the spectral components and reduced processing time in generating the stimulus. As with the SAM tone, the parameter ϕ_m was manipulated to produce an envelope–based ongoing ITD.

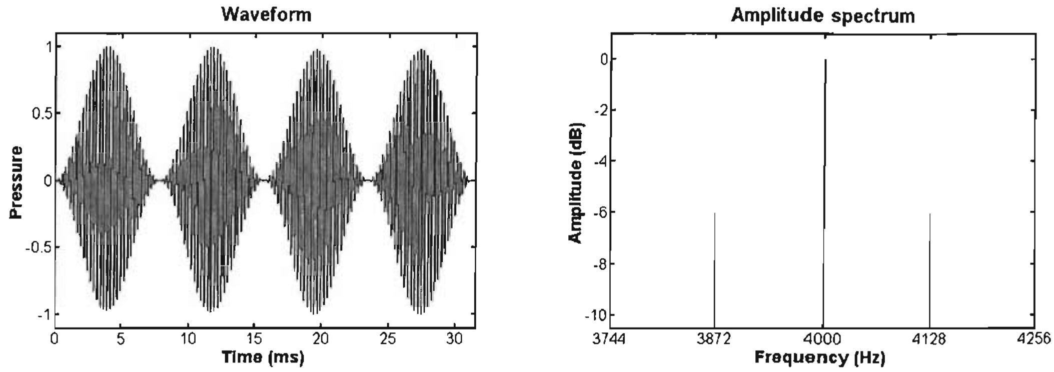


Figure 3.4. A portion of the waveform and the amplitude spectrum of a SAM tone. The carrier frequency was 4000 Hz and the modulation rate was 128 Hz.

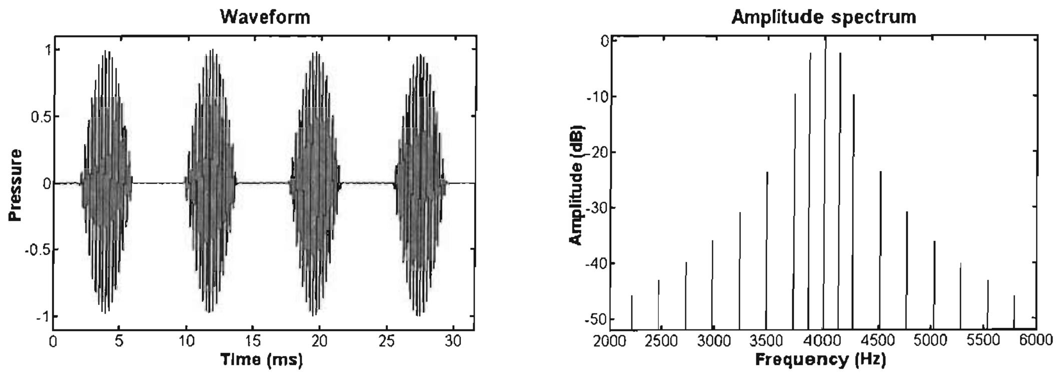


Figure 3.5. A portion of the waveform and the amplitude spectrum of a transposed tone as used by Bernstein and Trahiotis (2002) and in Experiment 2a. Note the different scales on the axes of the amplitude spectrum compared to Figure 3.4. The transposed tone used in other experiments consisted only of the central five spectral components. The carrier frequency and modulation rate was as in Figure 3.4.

The nominal carrier frequency used in these experiments was 4000 Hz. This is within the range of carrier frequencies which, when coupled with a modulation rate of 128 Hz, are associated with optimum ITD sensitivity for a diotic standard

(Nuetzel and Hafter, 1976, 1981; Henning, 1974) and as used by previous studies with transposed stimuli (Bernstein and Trahiotis, 2002; Oxenham et al., 2004).

3.5 Equipment

Stimuli were generated, procedures controlled and the responses recorded using MATLAB (The Math Works Inc., r.12)[10]. The digital waveforms were played out using a stereo soundcard (Experiment 1, Creative, SoundBlaster PCI; Experiments 2 and 3, Creative, Extigy). The output of the soundcard was passed through a custom–built passive stereo analogue mixer and was subject to pre–amplification (Experiment 1, via Kamplex, KC50; Experiments 2 and 3, Rega, Ear) before being routed to insert earphones (Experiment 1, Etymotic Research Inc., ER–5A; Experiments 2 and 3, Etymotic Research Inc., ER–2). Listeners were sat upright in a double–walled sound–treated booth, with an ambient noise level of rarely exceeding 30 dB(A). Responses were collected using a three–button keypad. Lights used to demarcate stimulus and response intervals and provide feedback were presented over a monitor directly in front of the listener.

Digital stimuli were produced with a sampling rate of 22050 Hz and 16–bit amplitude resolution. The soundcard used in Experiment 1 was initially thought to resolve inter–channel time delays down to only 20 μs, because resolution was (incorrectly) thought to be limited by the amplitude resolution. However, measurements conducted between Experiments 1 and 2 indicated that inter–channel delays of as small as 1 μs could be resolved, confirmed by reconsideration of theory[11]. (This had impact on the step sizes used in the adaptive procedure in these experiments.) Stimuli were calibrated in a 'Zwislocki' coupler taken from KEMAR (Knowles Electronics) in Experiment 1 or an occluded–ear simulator (IEC 711, 1981) otherwise.

Measurements of ITD discrimination with AM stimuli were generally conducted in the presence of continuous, diotic filtered Gaussian 'masking' noise. The noise was generated within MATLAB, filtered between 50–1500 Hz using a rectangular filter,

[10] MATLAB code was mostly written by the author, with the help of Dr. Bauman. Code from the Psychophysics Toolbox (http://psychtoolbox.org/index.html) was used to record responses and Dr. Bernstein provided code used to introduce a time delay with noise stimuli in Experiment 3.
[11] Dr. Bernstein is acknowledged for help in guiding this analysis.

recorded onto a CD and played out using a CD player (Technics SL–PG490) before being mixed with the primary stimuli. This stimulus was to used to guard against listeners using low–frequency intermodulation distortion products to detect the ITD (Nuetzel and Hafter, 1976; 1981; Bernstein and Trahiotis, 2002).

3.6 Statistics

Sample size calculations were conducted using formulae in Howell (1997 p. 221–5). Statistical analyses were conducted using SPSS (SPSS Inc., v. 12.0). The principal planned analysis for Experiments 1 and 3 (pre–post experiments) was an analysis of variance (ANOVA) on ITD threshold with related factors of Session (pre– vs. post–test) and Stimulus (trained vs. untrained) and unrelated factor of Training Group (e.g. trained vs. untrained). Learning was apparent as a statistically significant main effect of Session. Differential learning, such as between training groups, would be apparent as an additional statistically significant interaction between Session and one of the other factors. The initial criterion significance level was 0.05, although this was adjusted using a Bonferroni correction during sample size calculations according to the specific planned analyses. The degrees of freedom for related factors were adjusted using the Greenhouse–Geisser correction where Mauchly's test indicated a statistically significant departure from sphericity. Homogeneity of variance for unrelated factors was assessed using Levene's test.

There was uncertainty prior to conducting the experiments as to whether the statistical analyses should be based on the original ITD thresholds or subsequent to a non–linear transformation of the data. Such a transformation is indicated when the distribution of thresholds is not normally distributed and where the variance is related to the mean. Studies of frequency discrimination have found often reported requiring either a logarithmic (e.g. Irvine et al., 2000; Grimault et al., 2002) or reciprocal (Hawkey et al., 2004) transformation prior to parametric analyses. Few large–scale studies of ITD discrimination have been reported. Saberi and Antonio (2003) measured onset ITD thresholds for clicks in 89 inexperienced listeners at various SPLs and found the distributions to be positively skewed. Similarly, Hinton et al. (2004) measured ongoing ITD thresholds for 500–Hz tones in 24 modestly trained listeners using experimental conditions similar to those described here; a logarithmic transformation was required to correct for positive skew and variance that increased with mean threshold. Theoretical considerations of the shape of the psychometric function for ITD discrimination (Saberi, 1995) also indicate that variance should increase strongly with mean threshold unless a logarithmic transformation is applied. However, Wright and Fitzgerald (2001) appear to have found that ITD thresholds from their 32 inexperienced listeners with 500–Hz and 1000–Hz tones were normally distributed, since parametric statistics were conducted on untransformed data. Further, ITD thresholds from 19 listeners using 500–Hz–centred or 4000–Hz–centred narrowband noise reported by Bernstein et al. (1998) were at least approximately normally distributed. The approach taken in the current studies was therefore a pragmatic one: if the data were positively skewed on initial examination, a logarithmic transformation was applied before parametric statistical analyses. However, sample–size calculations were based on logarithmically transformed data because of the *a priori* considerations described above. For convenience, such as when plotting the logarithmically transformed data, the transformed thresholds were expressed in decibels relative to 1 μs. Thus, an ITD threshold of 1 μs becomes 0 dB, 10 μs becomes 10 dB, 100 μs becomes 20 dB and a doubling or halving corresponds to a change of 3 dB.

The raw data from all experiments are presented in Appendix D.

Chapter 4. Exploratory study of learning at high frequencies

4.1 Experiment 1: Introduction

Interaural time difference is an important cue to horizontal–plane sound–source localisation. Although humans are insensitive to ongoing ITD with tones above approximately 1500 Hz, sensitivity is apparent with high–frequency AM stimuli. However, this can be adversely affected by an IFD, an effect that appears to arise primarily due to the envelope being represented in increasing asymmetrical populations of auditory–nerve fibres between the ears (Nuetzel and Hafter, 1976; 1981). It has been suggested that this has relevance to BICIs as asymmetrical activity may also occur, such as because of asymmetrical insertions (e.g. Wilson et al., 2003). It may therefore be necessary to consider binaural performance (e.g. ITD sensitivity) during implant tuning. Alternatively, given the body of evidence on learning and plasticity, users may learn to utilise binaural information with such novel patterns of activity with experience and/or training.

It has been noted that substantial improvements may occur with practice/training on some binaural tasks in normal–hearing listeners (e.g. Hafter, 1984; Trahiotis et al., 1990). Nuetzel and Hafter (1976) trained listeners over many days or weeks before high–frequency ITD thresholds approached asymptote, although it is unclear if this was just associated with IFD. Wright and Fitzgerald (2001) studied learning with low–frequency ITD but concluded that, overall, training *per se* did not influence performance. This appears to be at odds with Nuetzel and Hafter, although specific aspects of their stimuli may be important, such as the high frequency or the IFD. Wright and Fitzgerald also concluded that the characteristics of learning with high–frequency ILD were fundamentally different to those with low–frequency ITD but it remains unclear if this relates to the different cues or the different frequency regions. Further study of ITD discrimination learning may be useful in understanding auditory plasticity under novel conditions, as may occur with BICI.

Further information in this area may also aid the development of robust methods to evaluate ITD sensitivity in clinical groups. While many previous studies using unmodulated and AM stimuli have been conducted with hearing–impaired people (e.g. Smith-Olinde et al., 1998; Colburn and Trahiotis, 1991) and users of BICI (e.g. van Hoesel and Tyler, 2003; Wilson et al., 2003), few report providing formal training. Further, most studies using normal–hearing listeners, and even the normal–hearing 'controls' in some studies of hearing impairment, used experienced, trained listeners. The variation in the use of training and the uncertainty as to its importance complicates comparison between groups and studies.

This chapter describes an exploratory study of ITD discrimination learning in normal–hearing listeners using high–frequency AM stimuli[12]. The primary aims were to evaluate the general methodology described in Chapter 3 and to explore learning under conditions that may be relevant to BICI. It was envisaged that this would lead to the development of robust methods for future studies of learning and ITD sensitivity with BICI, and other clinical groups, and the development of hypotheses for future studies of learning. A cohort of normal–hearing, young adults were trained on ITD discrimination using a diotic or dichotic (IFD) standard, referred to as 'matched' and 'unmatched' stimuli, respectively. Measurements were completed with both stimuli at pre– and post–test to evaluate generalisation of learning and the effect of IFD in a larger number of listeners than hitherto reported.

4.1.1 Summary of objectives

The objectives of this experiment were:

- Evaluate the methodology designed to investigate discrimination learning.
- Explore the time–course of ITD discrimination learning, using stimuli more relevant to bilateral cochlear implant research than those used by Wright and Fitzgerald (2001).
- Explore the effect of IFD and generalisation of learning across IFD.

[12] Rowan D, Lutman ME. High–frequency ITD discrimination with 'transposed' stimuli in trained and untrained normal–hearing adults. Poster presented at the 27th Annual Mid Winter Research Meeting of the Association for Research in Otolaryngology; 2004 Feb 22–26; Daytona Beach, Florida, USA.

4.2 Additional methods

4.2.1 Organisation of experiment

The general organisation and salient details of this experiment are summarised in Figure 4.1. The experiment consisted of nine sessions: session one was screening and familiarisation; session two was pre–test; sessions three to eight were formal training; session nine was post–test. Familiarisation consisted of exposure to three stimuli, a 1000–Hz unmodulated tone followed by the matched and unmatched stimuli, using a 30–trial block with a fixed ITD of 700 μs for the tone and 2000 μs for the AM stimuli. This was followed by a block of the adaptive procedure with the AM stimuli without and then with the masking noise. The 1000–Hz tone was used because it was found to lead to clear changes in lateral position and support good ITD discrimination in inexperienced listeners. Pre– and post–test consisted of six measurements of ITD thresholds with each stimulus; listeners were consecutively allocated to receive testing with one or other stimulus first (the order was as described in Section 3.1). Following pre–test, listeners were randomised to one of two groups; one group received training with the matched stimulus and the other with the unmatched stimulus. Formal training consisted of 1800 trials.

4.2.2 Adaptive procedure

The 4I 1C 3AFC task with feedback was used throughout. The initial ITD in the adaptive procedure was four large step sizes above the mean ITD threshold from the previous session with an upper limit of 2000 μs. The ITD was reduced after a correct response on three trials (k=3) although these were not required to be consecutive. This rule, first described by Zwislocki et al. (1958), theoretically targets the 75% correct response level (Zwislocki and Relkin, 2001). The contribution of non–consecutive trials may make it more difficult for listeners to anticipate the decision rule. It has been suggested that such anticipation may produce biased measurements (Levitt, 1971), but presumably only occurs with experience. This rule was selected to avoid or minimise such experience–dependent bias. Preliminary tests indicated that the number of trials during the first run could be significantly reduced by initially setting k=2 whereby the ITD is reduced following a correct response on two consecutive trials (Levitt, 1971). This also had the effect of increasing, on average, the number of reversals accrued. The initial step size was 120 μs, changing to 60 μs after the first reversal and to 20 μs after the third reversal. A block was terminated after 50 trials. Threshold was defined as the arithmetic mean of the scored reversals or scored as '≥2000 μs' as described in Section 3.3.2. Computer simulations [assuming a Gaussian psychometric function (e.g. Saberi, 1995)] indicated that the maximum ITD threshold that could be measured was approximately 1950 μs. The choice of step sizes was informed by initial considerations of the ITD resolution of the equipment (see Section 3.5). These seemed appropriate for normal–hearing listeners during pilot work, each block typically lasting 4–5 minutes and accruing six or eight scored reversals.

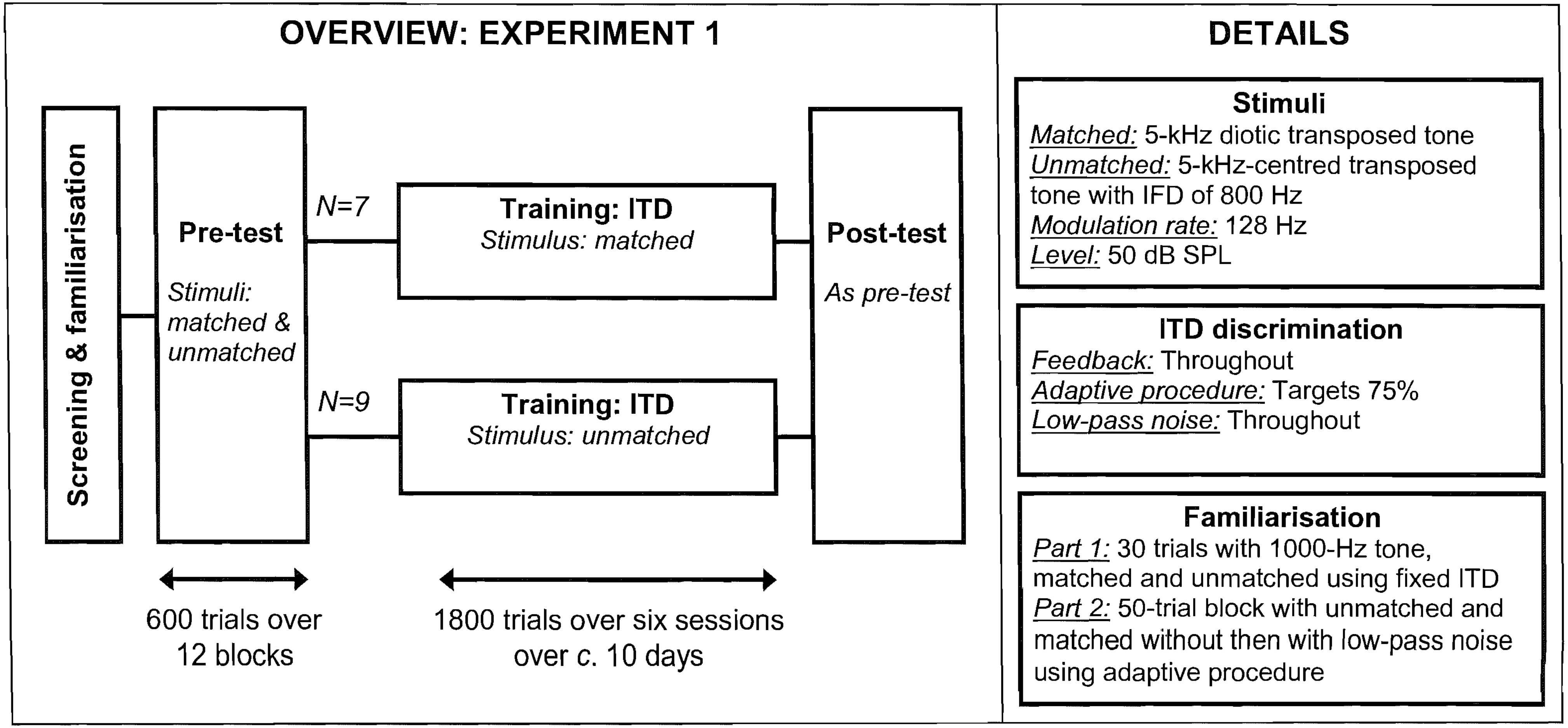


Figure 4.1. Schematic illustration of the organisation of Experiment 1. The salient details of methodology are also indicated.

4.2.3 Stimuli

This experiment used the transposed tone with n=2 in Equation 3 (i.e. restricted to the five central spectral components) and modulated at 128 Hz. The carrier frequencies were 5000 Hz in the 'matched' condition and 4600 Hz and 5400 Hz in the 'unmatched' condition, producing an IFD of 800 Hz centred on 5000 Hz. The ear receiving the lower carrier frequency was constant for each listener but varied across listeners. Stimuli were presented at 50 dB SPL. Although this was 10 dB higher than the nominal level used by Nuetzel and Hafter (1976, 1981), differences in earphones (ER–5A insert vs. TDH–39 supra–aural earphones) and couplers (Zwislocki coupler vs. unspecified coupler) make detailed comparisons difficult. Preliminary measurements indicated that the distal spectral sidebands would probably be at least 5 dB above absolute threshold in each ear.

Previous studies of IFD using a SAM tone and highly trained listeners indicated that ITD thresholds would increase by a factor of 4–10 compared to the matched condition (Nuetzel and Hafter, 1981; Henning, 1974). Preliminary measurements with 'conventional' AM stimuli indicated that few inexperienced listeners could detect ITD with an IFD of this magnitude; in fact, many were unable to reliably detect ITD with stimuli with no IFD. A transposed tone was found to enhance ITD thresholds compared to these stimuli in both the matched and unmatched conditions and inexperienced listeners appeared able to detect ITD with either stimulus. The carrier frequencies employed fall within the range where ITD thresholds are not strongly dependent on this and were centred on a frequency for which data exists on the effect of IFD with SAM (Henning, 1974; Nuetzel and Hafter, 1976; 1981). According to Greenwood's place–frequency map equation (Greenwood, 1990), the unmatched stimulus would typically produce an interaural difference in the cochlear place of maximal excitation of roughly 1 mm.

It was noticed during calibration that the waveform envelope of the transposed tone was slightly degraded. One possibility was that the frequency response of the ER–5A insert earphones, which gradually slopes off above about 3000 Hz, differentially attenuated the spectral components. Using Equation 4, the relative amplitudes of the five components were adjusted at source to produce the desired spectral envelope, resulting in an substantial enhancement of the waveform envelope. The required adjustment was as much as 6 dB. The frequency response of the earphones may also have altered the relative phases of the components, although no attempt was made to compensate for this.

4.2.4 Statistics

Although the experiment was exploratory, statistical considerations helped determine an appropriate sample size. The primary planned analysis was a mixed ANOVA on session mean ITD threshold with related factors of Session (pre– vs. post–test) and Stimulus (matched vs. unmatched) and unrelated factor of Training Group (matched–trained vs. unmatched–trained). In addition, a second mixed ANOVA was planned on session mean ITD threshold with the trained stimuli with related factor Training Session (sessions 3–8) and unrelated factor Training Group. Given the number of main effects and interactions, the criterion probability was reduced to 0.01. While this did not correspond to a full Bonferroni correction, it did reduce the possibility of erroneously rejecting a null hypothesis.

Preliminary estimates indicated pre–test ITD thresholds of around 25 dB (corresponding to approximately 315 µs) for the matched stimulus and 30 dB (1000 µs) for the unmatched stimulus. Insufficient listeners were used in preliminary tests to accurately estimate the across–listener SD, although previous studies (Wright and Fitzgerald, 2001; Bernstein et al., 1998) suggested a reasonable value was 2 dB (giving an pre–test upward standard deviation of 184 µs for the matched and 584 µs for the unmatched stimuli). It was assumed that 75% of the variance was shared by repeated measurements (i.e. the correlation of repeated measurements was 0.87). The post–test thresholds in the groups displaying least and most learning were assumed to be lower than pre–test thresholds by 1 dB (20%) and 2.5 dB (40%), respectively. At least 12 listeners per group was estimated to be required to detect this magnitude of learning within each group and differential learning between groups given a statistical power of 80% (two–tailed tests, $\alpha = 0.01$ and $\beta=0.2$ giving $\delta=3.2$; $\rho=0.87$).

4.3 Results

4.3.1 Overview

Sixteen listeners (8 females, 8 males) aged 21–32 years (mean 25.7 years) were recruited. Seven and nine listeners were trained with the matched and unmatched stimuli, respectively. The experiment was terminated after recruiting and fully testing 16 listeners, as opposed to the 24 listeners as planned (Section 4.2.4). This was because important and unanticipated trends emerged that differed substantially from the assumptions made in the *a priori* sample–size calculation. Firstly, a significant proportion of listeners had pre–test ITD thresholds scored as '≥2000 μs'; that is, thresholds could not be determined. This was despite running some blocks without the masking noise. Secondly, performance in a lesser but significant proportion of listeners was found to gradually and markedly *worsen*. These trends are described in more detail shortly. However, all listeners achieved at least 80% correct with the 1000–Hz unmodulated tone during familiarisation, with almost all errors occurring within the first 10 trials. Despite these difficulties, the experiment led to several important findings: (1) ITD thresholds improved substantially with training in some listeners over many hundreds to thousands of trials; (2) improvements with either trained stimulus generalised widely to the untrained stimulus, although it was not possible examine this statistically; (3) the magnitude of the effect of IFD and the variation of this across listeners was much smaller than anticipated from previous research.

In some listeners, thresholds both could and could not be determined within a single session. Consequently, a session mean was calculated only when three or more thresholds were determined; otherwise, the overall session threshold was also scored as '≥2000 μs'. Given these difficulties, statistical analyses were conducted on a within–individual basis and only for listeners for whom session mean thresholds could be determined. A further problem relating to the methodology was that 1.3% of blocks resulted in no scored reversals, most of which occurred during pre– and post–test. Although several reversals were accrued in all blocks, in these 1.3% of cases the data from the accrued reversals were discarded according to the method used to estimate the threshold (Section 4.2.2). When this occurred, the session mean was based on the available data.

4.3.2 Pre– and post–test ITD thresholds

Table 4.1 presents, for each listener, the allocated training group, mean pre– and post–test ITD thresholds and the mean difference between (and ratio of) these. (The numbering of listeners was unrelated to the order of testing.) Wide inter–individual variation in ITD thresholds was observed at pre– and post–test and for both stimuli, with a range of over 1000 μs. Some listeners were unable to reliably detect ITD of as large as 2000 μs and few listeners had thresholds of less than 700 μs at pre–test. However, all listeners displayed ITD sensitivity to both stimuli at post–test. The SDs of pre– and post–test ITD thresholds were generally less than 25% of the mean. Positive values for the difference scores in columns five and nine of Table 4.1 indicate that thresholds were more acute at post–test. For example, thresholds of Listener 7 improved by at least 1416 μs and 914 μs in the matched (trained) and unmatched conditions, respectively. However, some listeners had *poorer* thresholds at post–test, such as Listener 4 for whom thresholds worsened by 637 μs and 754 μs in the matched (trained) and unmatched conditions, respectively.

Twenty–one unrelated–samples t–tests were used to determine if the apparent differences in mean ITD thresholds between pre– and post–tests were statistically reliable. (Homogenous variance was not assumed following Levene's test.) A Bonferroni correction gave a criterion probability of 0.002 (0.05/21) although results were also judged against a less conservative value of 0.01. Single and double asterisks in Table 4.1 indicate statistically significant differences at the 1% and 0.2% level, respectively. Table 4.2 presents the results for the trained stimulus; a positive value of t indicates an improvement between sessions. For listeners with a statistically significant improvement, at least at the 1% level, thresholds improved by at least one third of pre–test thresholds between test sessions.

Four sub–groups of listeners were apparent in terms of how performance changed between pre– and post–tests: those with indeterminate thresholds at pre–test (but with evidence of learning), those with statistical evidence of learning, those with statistical evidence of a deterioration in performance and those with no statistically reliable difference in performance. Thresholds across the formal training sessions were explored to investigate these differences further.

Table 4.1. The trained stimulus, mean pre– and post–test ITD thresholds and the difference between (and ratio of) these for each listener and for both stimuli. Single and double asterisks indicate where the difference between mean pre– and post–test ITD thresholds was statistically significant at either the 1% or 0.2% level, respectively (see text for details).

Listener	**Trained stimulus**	**ITD Threshold (μs)**							
		Matched				**Unmatched**			
		Pre	Post	Pre–post	(Pre/Post)	Pre	Post	Pre–post	(Pre/Post)
1	Matched	1416	876	540 *	(1.62)	≥2000	1778	≥222	(≥1.14)
2	Matched	492	202	290 **	(2.44)	723	339	384 *	(2.13)
3	Matched	758	1207	–449 **	(0.63)	1666	1762	–96	(0.95)
4	Matched	822	1459	–637 *	(0.56)	1018	1772	–754 *	(0.57)
5	Matched	≥2000	1649	≥351	(≥1.23)	≥2000	1730	≥270	(≥1.17)
6	Matched	≥2000	1247	≥753	(≥1.62)	≥2000	1823	≥177	(≥1.11)
7	Matched	≥2000	584	≥1416	(≥3.46)	≥2000	1086	≥914	(≥1.86)
8	Unmatched	599	319	280 *	(1.88)	1353	455	898 **	(2.97)
9	Unmatched	834	573	261	(1.46)	1374	785	589 **	(1.75)
10	Unmatched	242	148	94	(1.64)	717	366	351	(1.96)
11	Unmatched	431	366	65	(1.18)	706	536	170	(1.32)
12	Unmatched	663	435	228	(1.52)	866	668	198	(1.31)
13	Unmatched	678	508	170	(1.33)	634	701	–67	(0.91)
14	Unmatched	697	461	236	(1.51)	780	1175	–395 *	(0.66)
15	Unmatched	≥2000	794	≥1206	(≥2.54)	≥2000	1263	≥737	(≥1.6)
16	Unmatched	≥2000	1286	≥714	(≥1.57)	≥2000	1556	≥444	(≥1.3)

4.3.3 ITD thresholds across training sessions

The Pearson correlation coefficient, r, and its statistical significance were obtained by comparing individual ITD thresholds with the trained stimulus across eight sessions (i.e. excluding familiarisation) with block number (i.e. time). The results are given in Table 4.2; r^2 is shown to provide a clearer indication of the strength of the relationship; N indicates the number of blocks for which data were available (maximum of 48). This analysis led to a stronger separation of listeners: the relationship between ITD thresholds and time was either highly or not statistically significant. Two listeners with no statistically reliable difference between pre– and post–test thresholds were found to have statistically related thresholds with time (Listeners 10 and 11). The amount of variance in ITD thresholds accounted for by time was modest, typically 30–40%.

Table 4.2. Results of statistical analyses on ITD thresholds with the trained stimulus.

Listener	Stimulus	ITD threshold with the trained stimulus					
		Pre– minus post–test			Relationship with time		
		t	df	p	r^2	N	p
1	Matched	4.2	6.8	0.004	0.51	46	<0.001
2	Matched	9.1	9.7	<0.001	0.30	47	<0.001
3	Matched	–5.6	6.1	0.001	0.34	47	<0.001
4	Matched	–3.5	7.2	0.010	0.34	48	<0.001
8	Unmatched	10.1	4.5	<0.001	0.39	45	<0.001
9	Unmatched	4.5	8.0	0.001	0.59	48	<0.001
10	Unmatched	3.0	5.3	0.028	0.39	48	<0.001
11	Unmatched	1.9	7.0	>0.05	0.24	47	<0.001
12	Unmatched	1.9	10.0	>0.05	0.01	48	>0.05
13	Unmatched	–0.9	7.3	>0.05	<0.01	47	>0.05
14	Unmatched	–4.5	7.2	0.003	0.04	47	>0.05

Mean ITD thresholds across these eight sessions (and 2400 trials) are plotted in Figure 4.2. For clarity, error bars are displayed for selected listeners only and indicate 1 SD. Listeners have been separated into four sub–groups. Overall, it was impossible to predict how a listener's performance would vary with training from his/her pre–test thresholds or knowing the trained stimulus. The upper panels plot data from 10 listeners with evidence of learning during the experiment. The upper *left* panel plots data from listeners with statistically reliably lower thresholds at post– than at pre–test. Listener 10 is also included as post–test thresholds only marginally missed being significant at the 1% level and the relationship between thresholds and time was robust. Improvements were accrued over numerous sessions for all five listeners and often over more than 1000 trials. There was a hint of a positive relationship between pre–test threshold and the magnitude of learning, but this was difficult to assess given the small number of listeners. The curves were non–linear, which at least partially accounts for finding only modest correlations. Attempts were made to fit various negative exponential functions to the data but sensible or statistically reliable fits could not be obtained using the same function for all listeners. The upper *right* panel plots data from listeners for whom pre–test ITD thresholds were indeterminate. Performance clearly improved during the experiment for all five listeners, even after many hundreds of trials of apparent insensitivity, and asymptotic performance may not have been reached even after 2400 trials.

The bottom left panel plots data from three listeners with apparently stable performance throughout the experiment. Inspection of the curve of Listener 12 suggests that performance may have gradually improved over the first half of the experiment but then reduced over the second half. The bottom right panel plots data from three listeners whose thresholds worsened between pre– and post–tests. The raw data in Appendix D indicate that the performance of Listener 4 deteriorated to such an extent that some blocks at post–test led to indeterminate thresholds! Further, not only did performance gradually worsen across sessions, but also within each session. A similar trend was also apparent for Listener 3.

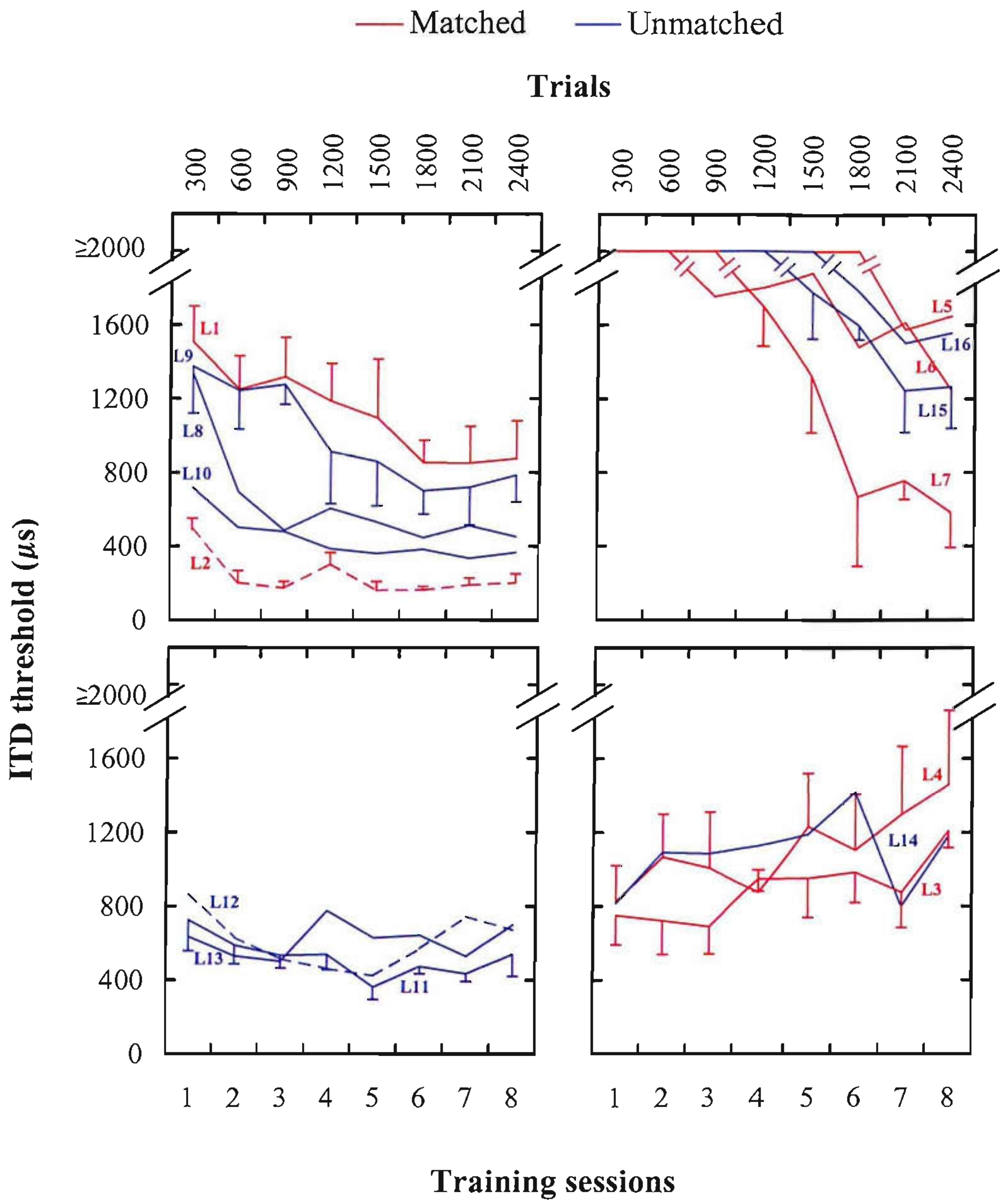


Figure 4.2. Session mean ITD threshold with the trained stimulus as a function of number of sessions and trials. Error bars on selected listeners indicate 1 SD.

4.3.4 Generalisation

Figure 4.3 plots mean pre– and post–test ITD thresholds with both stimuli in listeners with evidence of learning. Improvements in performance are apparent for the trained and untrained stimulus, providing evidence of generalisation. Table 4.1 shows that statistically significant improvements were apparent with the untrained stimulus for all but one listener (Listener 10) with measured thresholds at pre–test.

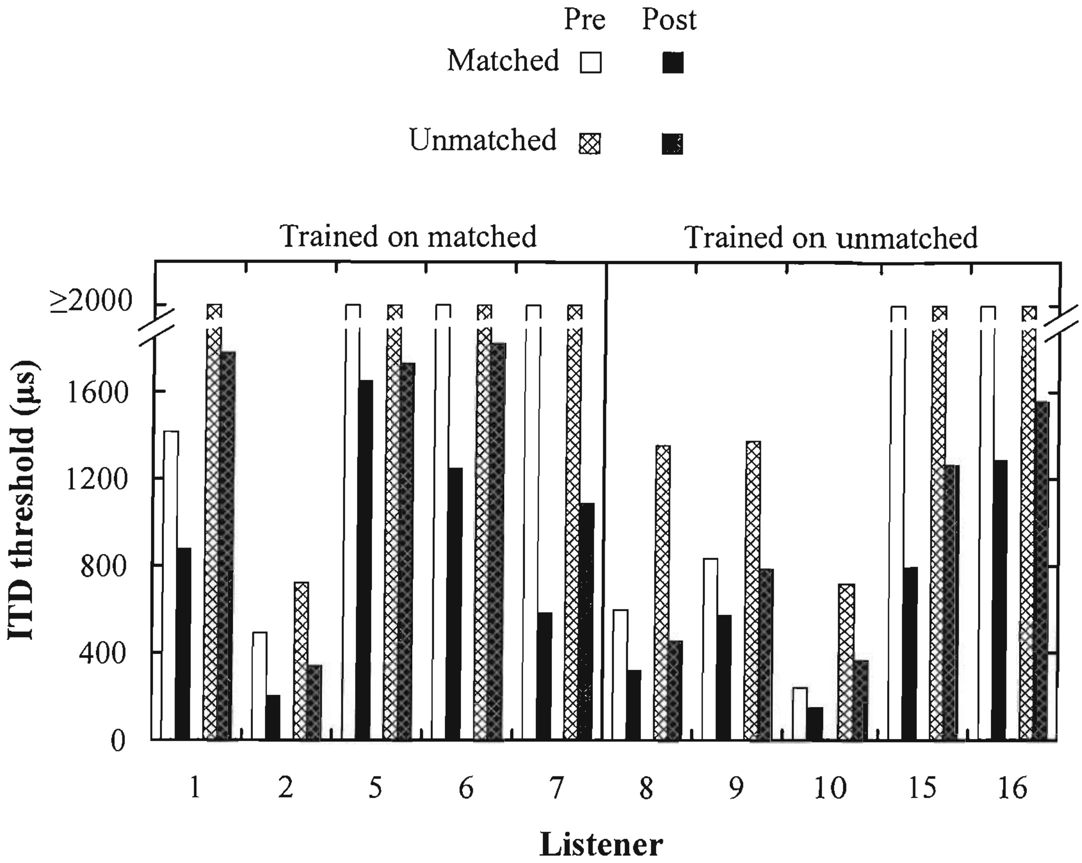


Figure 4.3. Mean pre– and post–test ITD thresholds with both stimuli in listeners with evidence of learning.

4.3.5 Interaural carrier frequency difference

Figure 4.4 displays the effect of IFD as the ratio of post–test ITD thresholds with the unmatched and matched stimuli for each listener, given the centre carrier frequency in the unmatched condition; a value of greater than unity indicates performance was poorer with the unmatched stimulus. The effect of IFD was remarkably consistent across listeners given the wide inter–individual variation in ITD thresholds and, overall, led to approximately a doubling of ITD thresholds. Figure 4.4 also compares these data with four listeners from previous studies using SAM tones[13]; stimulus bandwidth and IFD are indicated to aid comparison. It can be seen that the magnitude of and inter–individual variation in the effect of IFD found here appear less than reported previously.

[13] Thresholds for ITD were read off figures from Henning (1974) or obtained from raw data in Nuetzel (1976). Data from Long (2000) are not included as the centre carrier frequency for a comparable IFD was approximately an octave lower than used here.

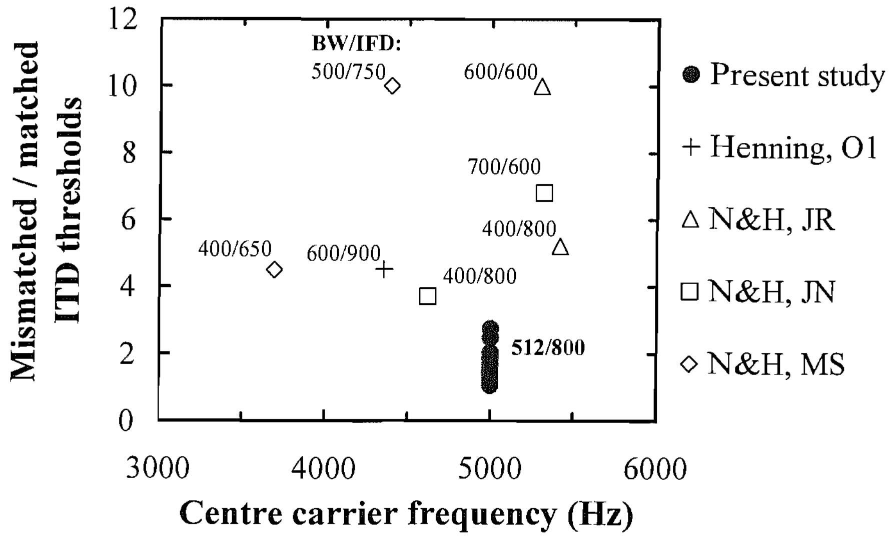


Figure 4.4. Ratio of post–test ITD thresholds with unmatched and matched stimuli for individual listeners. Data from listeners in previous experiments are also presented (Henning, 1974; Nuetzel and Hafter (N&H), 1981; the initials of their listeners are given in the legend). The bandwidth (BW, Hz) of the stimuli and IFD (Hz) of the unmatched stimulus are indicated.

4.4 Discussion

4.4.1 Overview

This preliminary study sought to evaluate the methodology for studying ITD discrimination learning and to explore learning with high–frequency AM stimuli. A number of limitations with the methodology were observed and need to be addressed before further studies are conducted. Thresholds at pre– and post–test were found to varying widely across listeners, such that some were apparently unable to reliably detect ITD of as large as 2000 μs at pre–test. Evidence of discrimination learning was observed in over 50% of listeners; this was apparent with both trained stimuli and the untrained stimulus. The magnitude and time–course of learning varied between listeners, with some requiring hundreds and others thousands of trials to reach asymptote. The effect of the IFD was consistent across listeners despite inter–individual variation in ITD thresholds and training effects. These findings and the implications for studies of learning, ITD sensitivity in clinical populations and the clinical utility of ITD discrimination, such as for BICI tuning, are discussed in detail in the following sections.

4.4.2 Methodological issues

This experiment highlighted a number of limitations with the methodology. For example, the adaptive procedure occasionally terminated with insufficient reversals to calculate a threshold. This was particularly problematic during pre– and post–tests as fewer threshold estimates were available for averaging. The problem appeared to arise at least partly from the combination of the 50–trial termination rule and the decision rule. Further refinement of the procedure was indicated. However, the proportion of blocks lost because of no scored reversals (1.3%) was still much smaller than at least one previous study, which reported loosing 6% of blocks during pre– and post–test alone for the same reason (Wright et al., 1997, p. 3975).

Another issue was the number of listeners with indeterminate thresholds at pre–test. The pre–post test design is not well suited to the study of learning in these conditions and future research with similar stimuli would require an alternative approach. Also problematic was the worsening of performance observed in some listeners. These listeners claimed afterwards to be unaware of it but reported finding the task ‘difficult’, especially given the low sensation level. Previous studies have reported irregular ‘learning curves’ that appear to show a worsening of performance over at least some sessions, if not the entire experiment (e.g. Wright and Fitzgerald, 2001). It is difficult to explain these results in terms of auditory processing and seem more likely to reflect a loss of attention and motivation, the subjective difficultly of the stimuli perhaps sensitising the experiment to this. Refinement of the methods to attempt to better control attention and motivation was indicated.

4.4.3 ITD thresholds with low–level transposed stimuli

This study appears to be the first to examine ITD discrimination with transposed stimuli in inexperienced listeners and found that some listeners are seemingly unable to detect ITDs of as large as 2000 μs with some AM stimuli. This seems consistent with the comment by Trahiotis et al. (1990) that inexpereienced listeners will ‘often be completely unable’ to perform some binaural tasks without training, although it is unclear if this comment was intended to apply to similar conditions as used in the present study. Wide inter–individual variation was apparent at pre–test but also post–test, although not all listeners appeared to have reached asymptote. Thresholds were generally much poorer than in studies with 'conventional' AM stimuli presented at a similar level but with experienced listeners (e.g. Nuetzel and Hafter, 1976, 1981; McFadden and Pasenan, 1976). However, wide inter–individual variation has also been reported by Bernstein et al. (1998) and Saberi and Antonio (2003). The former used a 4000–Hz–centered narrowband noise (presented at a higher SPL) and listeners experienced with binaural signal detection tasks, and found ITD thresholds ranging from approximately 150 to 750 μs. The latter used a single broadband click and inexperienced listeners, and found thresholds ranging from less than 100 μs to in excess of 1000 μs. Rutkowski (2005) reported ITD thresholds with transposed tones (modulation rate of 200 Hz and carrier frequency of 5000 Hz) from eight inexperienced listeners, using 75 dB SPL, and three experienced listeners, using 40 dB SPL, and found group mean thresholds of approximately 300 μs and 100 μs, respectively. Data from the former group were comparable to the pre–test thresholds of the better performers in the present study (i.e. Listeners 2, 10 and 11) despite the difference in level.

Figure 4.5. compares post–test ITD thresholds with the matched stimulus to data with transposed tones from Bernstein and Trahiotis (2002) and Oxenham and Penagos (2003; Oxenham et al., 2004)[14] using comparable modulation rates (128 Hz or 125 Hz, respectively); the carrier frequency is indicated. Data from listeners for whom thresholds worsened during the experiment are not presented. Thresholds were much poorer than those reported by Bernstein and Trahiotis although overlapped with those from Oxenham and Penagos (2003). All but one listener with thresholds of greater than 600 μs came from the sub–group with apparent insensitivity at pre–test and for whom post–test thresholds may not have reached asymptote. If these data are excluded, the distribution was comparable to that found by Oxenham and Penagos, whose listeners only had two hours of practice, despite the difference in level.

[14] The data from Bernstein and Trahiotis (2002) were kindly provided by Dr. Bernstein. The data from Oxenham and Penagos (2003) were taken from their Figure 2.

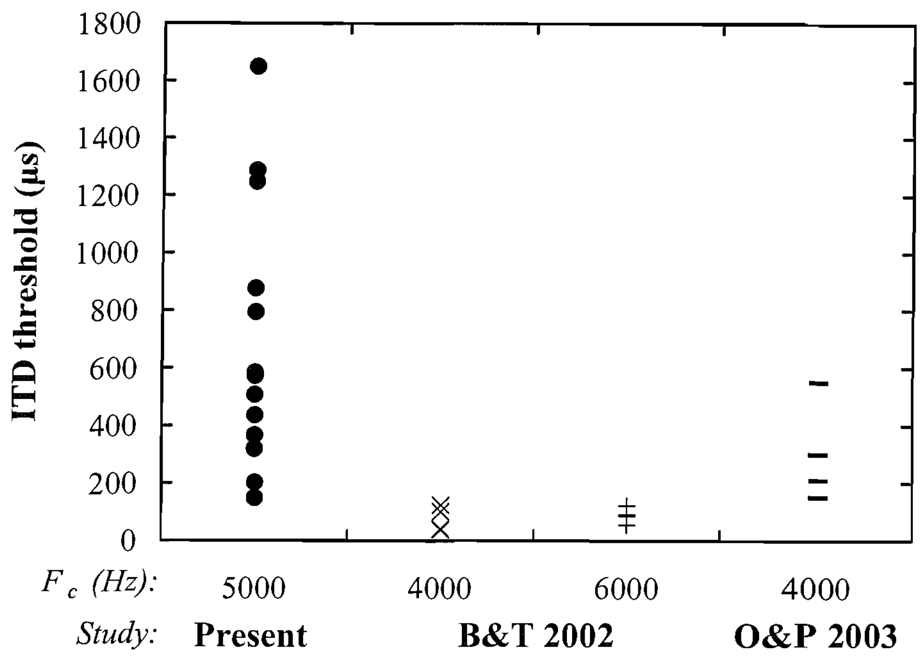


Figure 4.5. ITD thresholds with the transposed tone from the present study (post–test) and two previous studies (Bernstein and Trahiotis (B&T), 2002; Oxenham and Penagos (O&P), 2003) with comparable modulation rates; carrier frequency (F_c) is shown.

It is currently unclear why ITD thresholds could not be determined in some listeners at pre–test. That all listeners reliably detected ITD with the 1000–Hz tone during familiarisation indicates that this did not represent difficulties with the task *per se*. Further, that thresholds were determined at post–test indicates that this also did not represent inherent problems with the representation of the waveform envelope. During informal discussion with these listeners after post–test, they reported eventually being able to detect the ITD using an ephemeral cue rather than an obvious change in lateral position. It seems most parsimonious to suggest, therefore, that these listeners failed to utilise appropriate information provided by their binaural system to detect the ITD during the earlier sessions.

4.4.4 Effects of training

The observation of learning over many hundreds to thousands of trials in many listeners appears broadly consistent with Nuetzel and Hafter (1976), although it was not possible to make a more detailed comparison since they did not report their listeners' learning curves. It is also now clear that this is apparent for stimuli with and without an IFD. The magnitude of improvement, in microseconds, observed in the present study far exceeds that reported by Wright and Fitzgerald (2001). This could be interpreted as indicating differential learning effects with ITD in different frequency regions and/or with different types of ITD. However, this interpretation is confounded if asymptotic thresholds with the stimuli differ widely and if the amount of learning depends on asymptotic performance. Although little is known regarding the latter, vision research indicates that the characteristics of learning are related to task difficulty (Ahissar and Hochstein, 1997). Modelling research into the different time–courses of frequency discrimination learning with different harmonic complexes (as reported by Grimault et al., 2002) apparently suggests that these findings could arise from differential sensitivity to the stimuli rather than fundamentally different processing (Carlyon, 2004, personal communication).

Differences in asymptotic ITD thresholds with the stimuli used in each study, the IFD aside, may arise from the differences in SPL or in modulation rate and tonal frequency. A comparison across studies with tones (Zwislocki and Feldman, 1956; Hershkowitz and Durlach, 1969) and AM stimuli (Nuetzel and Hafter, 1976; McFadden and Pasanen, 1976) suggests that the effect of SPL may be stronger with the latter. This is also suggested by Hawkins and Wightman (1980) who compared the effect of SPL directly with low– and high–frequency bands of noise. However, their study also found that high–frequency stimuli are associated with poorer thresholds even at high SPL. It is unclear if stronger level effects occur with transposed stimuli, which are known to lead to comparable sensitivity with low–frequency tones at least at high levels. The importance of the difference in tonal frequency and modulation rate is more clearly understood, at least with high–level stimuli. Studies show that thresholds improve with frequency for tones between 128 and 500 Hz (e.g. Klump and Eady, 1956; Zwislocki and Feldman, 1956) whereas thresholds worsen with frequency with AM stimuli above 300 Hz (e.g. Bernstein and Trahiotis, 2002; Nuetzel and Hafter, 1981). Given these potentially important differences in stimuli used in learning studies, it is not possible to draw clear conclusions regarding differences in learning with the stimuli. Future research should attempt to resolve these issues.

It is unclear what listeners learnt during training. Given the time–course observed, and the use of familiarisation, it seems unlikely that the improvement was related to learning to make the response or adjusting to the procedures. This is supported by the study by Hawkey et al. (2004) who found that the learning over the first 500 trials on frequency discrimination was not strongly related to the procedures. Without the aid of an untrained control, it was also not possible to determine the degree to which the improvements reflected a 'consolidation' of learning, as postulated by Wright and Fitzgerald (2001). However, given the time–course of learning, it seems unlikely that this was the sole or even primary component. Future work should attempt to define more clearly training–induced ITD discrimination learning with different stimuli and then attempt to identify these components.

4.4.5 Effect of interaural carrier frequency difference

This study demonstrates again, but in a larger number of listeners, that the imposition of an IFD may impair ITD discrimination. Importantly, the magnitude of the effect at post–test was fairly consistent across listeners despite the wide variation in post–test thresholds with either stimulus or variation in 'learning curves'. As shown in Figure 4.4, the effect of IFD with transposed tones and the variation across listeners was less than some previous studies using SAM tones. It is possible that differences in the amplitude spectra of the stimuli contributed to the smaller effect of IFD measured here. Specifically, the transposed tone has twice the bandwidth of a SAM tone for a given modulation rate. However, Figure 4.4 plots data with comparable bandwidths, although consequently different modulation rates. [Nuetzel and Hafter (1981) reported that the effect of IFD was influenced by modulation rate but assumed this to be related to spectral effects.] Rutkowski (2005) compared the effect of IFD with SAM and transposed tones, and found that this effect was weaker with the transposed tones. However, the modulation rate was fixed leading to differences in the bandwidth of the stimuli, possibly exaggerated by the use of low–pass filtering at 2000 Hz in generating the transposed stimulus.

An alternative explanation for the smaller effect of IFD measured in the present study compared to previous studies relates to task differences. The 'classic' studies with SAM used the 2AFC task whereby the ITD caused a percept of movement from right to left or left to right and listeners were required to identify which had occurred. However, the use of an oddity paradigm in the present study may permit the measurement of lower ITD thresholds in the unmatched relative to the matched condition because it does not explicitly require listeners to judge lateral position. This may permit detection of the ITD based on more ephemeral cues. Measurements with the unmatched stimulus may be particularly sensitive to this given that the perceived image was only partially fused. Rutkowski (2005) used an oddity paradigm (the two–cue, 2AFC task) and compared the effect of IFD using SAM with that reported by Nuetzel and Hafter (1981). This suggests that the effect of IFD is less strong when measured using an oddity paradigm. However, other differences in the procedures used to measure and define threshold in this study may also influence this comparison. Further studies are necessary to determine the influence of the specific nature of the task on the apparent effect of IFD.

4.4.6 Applied and clinical implications

The training effects observed here might have implications for studies of clinical populations. A number of studies have investigated ITD discrimination in hearing–impaired people but did not provide formal training (e.g. Smith-Olinde et al., 1998). This is also the case for most, if not all, published studies of ITD discrimination in BICI users. Further, some users have been reported to be insensitive to ITD with AM stimuli. The present study raises the possibility that these results underestimate users' potential performance and that comparison with data from normal–hearing, and usually highly trained, listeners may overestimate the underlying 'impairment'. This study may also have implications for clinical testing based on ITD discrimination. If the aim of such testing is to obtain asymptotic performance, many hundred or thousands of trials of training may be required. However, the consistent effect of IFD across listeners suggests that reliable measurements of interaural place difference may be obtained in clinical populations given less training. However, the possible influence of the task on the effect of IFD suggests that detecting changes in lateral position may be more sensitive to this effect than detection of ITD *per se*; information on the former may be more useful in BICI tuning.

4.4.7 Conclusions

In summary, the conclusions of this experiment were:

- ITD discrimination with low–level high–frequency stimuli is associated with substantial inter–individual variation in both sensitivity and effects of training.

- The time–course of learning on ITD discrimination with some stimuli extended over hundreds to thousands of trials, which is greater than anticipated from Wright and Fitzgerald (2001). However, a detailed comparison of learning in this study and Wright and Fitzgerald (2001) was complicated by methodological differences.

- Learning generalises across IFD.

- An IFD of sufficient magnitude impairs ITD thresholds and may be smaller than indicated in previous studies (e.g. Nuetzel and Hafter, 1981).

- The provision of extensive training is important when carrying out measurements of ITD discrimination in clinical populations, at least if the aim is to determine potential capability. However, valid measurements of the effect of IFD may not require training to asymptotic levels.

Chapter 5. Differential learning at low and high frequencies?

5.1 Experiment 2: Introduction

Experiment 1 explored ITD discrimination learning using high–frequency AM stimuli with and without an IFD. Evidence was found of learning in the majority of listeners. This appeared inconsistent, in time–course and magnitude, with the general conclusions of a previous study using low–frequency unmodulated tones (Wright and Fitzgerald, 2001). However, comparisons between experiments may be complicated by differences in the stimuli, particularly the SPL. Further, while the methodology in Experiment 1 was able to elicit learning, weaknesses were apparent. In particular, the adaptive procedure occasionally accumulated insufficient reversals to calculate threshold at pre– and post–test. Some listeners were also seemingly unable to detect ITD of up to 2000 μs at pre–test, raising questions as to how appropriate the pre–post design is in studying learning in these conditions.

This chapter describes two experiments to address these issues[15], the general organisation of which is summarised in Figure 5.1. In Experiment 2a, ITD thresholds were measured in experienced listeners using modified procedures with unmodulated and transposed tones at low and high SPL. The procedures were found to be robust and the thresholds at the higher SPL were consistent with Bernstein and Trahiotis (2002). Evidence was obtained for a stronger effect of SPL with the high–frequency AM stimulus, confirming the presence of important differences in the stimuli used in previous studies of learning. In Experiment 2b, measurements were obtained from a new cohort of inexperienced listeners using modified familiarisation and pre–test procedures and the higher–level stimuli. It was possible to determine ITD thresholds in all listeners. However, thresholds were disproportionately poorer with the transposed stimulus, providing indirect evidence for differential learning with the stimuli.

[15] Rowan D, Lutman ME. High–frequency ITD discrimination with 'transposed' stimuli in trained and untrained normal–hearing adults. Poster presented at the 27th Annual Mid Winter Research Meeting of the Association for Research in Otolaryngology; 2004 Feb 22–26; Daytona Beach, Florida, USA.

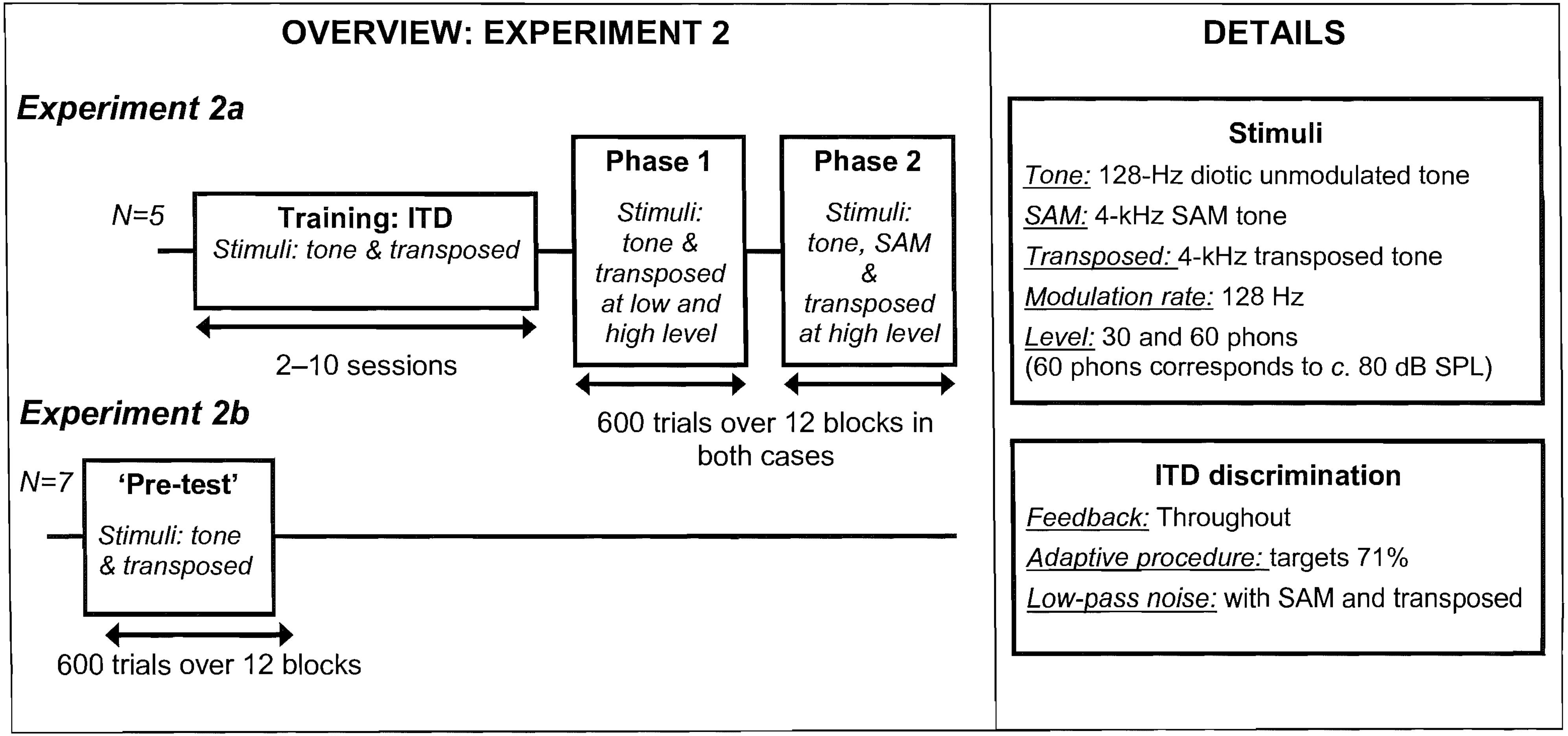


Figure 5.1. Schematic illustration of the organisation of Experiment 2 a and b. The salient details of the common methodology are also indicated.

5.2 Modified adaptive procedure

On examining adaptive tracks from Experiment 1, the problem of too few reversals appeared to be associated with a large number of trials being expended during runs and the fixed–trial termination rule. To address this problem, three changes were made: (1) the step sizes were modified to be based on fixed ratios rather than differences; (2) the two–down, one–up decision rule was used throughout; (3) blocks were terminated after 60 trials rather than 50 trials.

Previous ITD discrimination research has used step sizes based on ratios rather than differences (e.g. Saberi, 1995; Bernstein et al., 1998; Bernstein and Trahiotis, 2002; Wright and Fitzgerald, 2001). It has been argued that this is preferable because a constant ratio seems to correspond approximately to a constant change in percentage correct over most of the range of the psychometric function for ITD discrimination (Saberi, 1995). Since the corresponding change in ITD increases with increasing ITD, in contrast to Experiment 1, this may reduce the number of trials per run. This was supported by computer simulations. Following the previous studies, the step size was a factor of 1.58 (corresponding to 2 dB) for the first three reversals, reducing to a factor of 1.12 (0.5 dB) for the remainder of the block[16]. To be consistent with this, threshold was calculated as the *geometric* mean of the scored reversals. [The initial ITD was changed to three (rather than four) large step sizes above previous estimates of threshold with an upper limit of 2000 μs]. Changes to the decision and termination rules were also intended to increase the number of scored reversals obtained. These increased test time to approximately 5 minutes. Following these changes, 8–10 scored reversals were typically accumulated. (As will be described in Chapter 6 for Experiment 3, it was intended that a modification to this decision rule would be used during formal training.)

[16] One reason that this approach was not applied in Experiment 1 was because it was thought that the equipment was unable to resolve sufficiently small changes in ITD, essentially forcing the step sizes to be constant differences. This was based on the assumption that the resolution of ITD (i.e. inter–channel phase shifts) was determined by amplitude resolution, with 16 bits limiting resolution to approximately 15 μs. However, careful measurements of the output of the Extigy sound card subsequent to Experiment 1 indicated that this was not the case and that ITDs of as small as 1 μs could be produced reliably. The ITD resolution of a digital system appears to be related to the number of sample points in the signal (i.e. stimulus duration and sampling frequency) and the reconstruction filters used during digital–to–analogue conversion. It is difficult to predict the ITD resolution analytically.

5.3 Experiment 2a: Additional methods

5.3.1 Organisation of experiment

This experiment consisted of two phases. In Phase 1, a small group of normal–hearing experienced listeners were further trained and tested with a 128–Hz tone and the transposed tone, as in Experiment 1 (but no IFD), presented at the low and a higher SPL. Testing at each SPL was conducted consecutively, starting with the higher level, rather than being interleaved, to allow listeners to learn the (subtle) cues for each level. In Phase 2, listeners were retested with the unmodulated and transposed tones and also a 128–Hz–modulated SAM tone presented at the higher level. The transposed tone was identical to that used by Bernstein and Trahiotis (2002) (i.e. contained more spectral components than in Phase 1) to permit comparison. Phase 2 was conducted several weeks after Phase 1, permitting an evaluation of measurement stability.

Five listeners (two female; two male) were recruited from the University population. All had previous experience with binaural experiments including listening to transposed stimuli. Listeners were also given varying amounts of training, consisting of repeated measurements over two to ten sessions, until performance on all conditions appeared to have approached an asymptotic value. This was defined as: (1) a SD over the last six thresholds of less than 20% of the mean; (2) no evidence on visual inspection that thresholds were gradually improving over the last 10 thresholds. Formal testing was conducted during a single session and the arithmetic mean of the six thresholds was calculated for each stimulus. The first half of the session consisted of three measurements with each stimulus; the second half was identical but in reverse order (see Section 3.1). Measurements with the AM stimuli were made using the masking noise. (During training, a number of measurements were also collected without the low–pass noise, which confirmed that this had little, if any, effect on performance.)

5.3.2 Stimuli

The transposed tone used in Phase 1 (n=2 in Equation 3) will be referred to as the 'narrowband' stimulus. The transposed tone used in Phase 2 (n=8 in Equation 3) will be referred to as the 'wideband' stimulus. The carrier frequency for all AM stimuli was 4000 Hz to permit closer comparison with Bernstein and Trahiotis (2002). As noted in Section 3.5, the equipment used to generate, present and calibrate the stimuli differed slightly from Experiment 1. The ER–2 insert earphones used here have a 'flat' frequency response over the relevant range, as measured in a Zwislocki coupler. Issues regarding the accurate representation of the waveform envelope, as in Experiment 1, therefore did not arise.

Stimuli were presented at 30 and 60 phons to account for loudness differences at 128 and 4000 Hz (ISO 226, 2003). A loudness level of 30 phons corresponded to approximately the same SPL at 4000 Hz as used in Experiment 1. Target coupler SPLs corresponding to these loudness levels were derived as follows. Firstly, equal–loudness contours and the minimal audible field curve were obtained from ISO 226 (2003) (for tones presented bilaterally via a loudspeaker in the free–field and calibrated in the absence of the listener) and reference equivalent threshold SPLs (RETSPLs) were obtained from ISO 389–2 (1994) (for tones presented monotically by insert earphones and calibrated using the IEC 711 coupler). The difference, in dB, between the appropriate equal loudness contour and the minimal audible field at the frequency of interest was then determined. Finally, this value was added to the RETSPL at same frequency to provide the target coupler SPL. As the ISO standards do not provide sufficient data for 128 Hz, those at 125 Hz were used. No correction was made to account for the change from binaural to monaural loudness, as this appears to be the same at both frequencies (Marks, 1978). In addition, no correction was made for the AM used at 4000 Hz. Moore et al. (1998) found that the SPL of unmodulated and AM tones at equal loudness are within 1–2 dB, when using comparable modulation rates to that used here.

5.4 Experiment 2a: Results and discussion

Figure 5.2 plots mean ITD thresholds with all four stimuli at the higher SPL for each listener. Error bars indicate ±1 SD; this was usually close to 20% and rarely above 25% of the mean. Listener 4 dropped out after testing with the higher–level stimuli in Phase 1. It can be seen that ITD thresholds with the wideband and narrowband transposed tones (grey, middle columns) were comparable. Thresholds with the unmodulated and transposed tones were also comparable for Listeners 2 and 4; thresholds for Listeners 1, 3 and 5 were slightly lower for one or other stimulus. Thresholds with SAM were poorer than other stimuli in all listeners. These results are consistent with Bernstein and Trahiotis (2002) and Oxenham et al. (2004). They also show that the additional spectral components with the wideband transposed tone did not influence performance. This is consistent with van de Par and Kohlrausch (1997) for MLD. Figure 5.3 compares ITD thresholds between unmodulated and transposed tones across listeners in the present (using the narrowband stimulus) and previous studies; ITD thresholds and the relative thresholds with each stimulus were comparable to these studies.

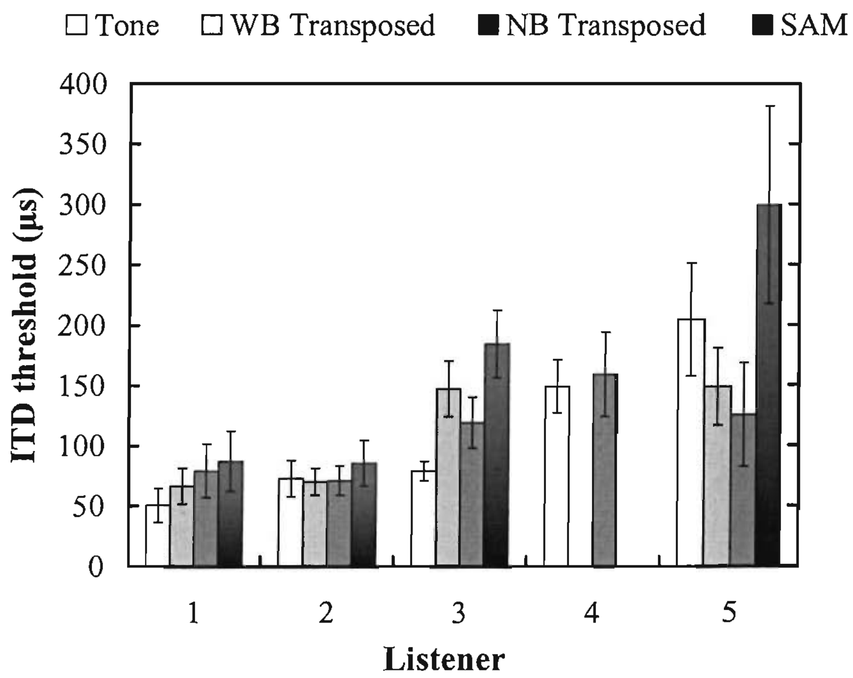


Figure 5.2. Mean ITD thresholds for each listener with the unmodulated tone, wideband (WB) and narrowband (NB) transposed tones and SAM tone presented at 60 phons. Error bars indicate ±1 SD.

It was interesting that all listeners in the present study reported finding the task subjectively more difficult with the transposed tone (and SAM tone) compared to the unmodulated tone, despite having comparable thresholds. This is counter–intuitive since the adaptive procedure ensures that listeners achieve approximately the same proportion of 'correct' responses in each block, independent of overall sensitivity. Further, Bernstein and Trahiotis (2001) have provided data showing that the transposed tone is associated with similar displacement in lateral position with ITD compared to low–frequency stimuli in trained listeners. However, Oxenham et al. (2004) have provided data indicating that the transposed tone, unlike the corresponding unmodulated tone, is associated with little if any perception of pitch. It may be that non–spatial perceptual qualities of the stimuli such as pitch influenced listeners' subjective judgements on the difficulty of the task.

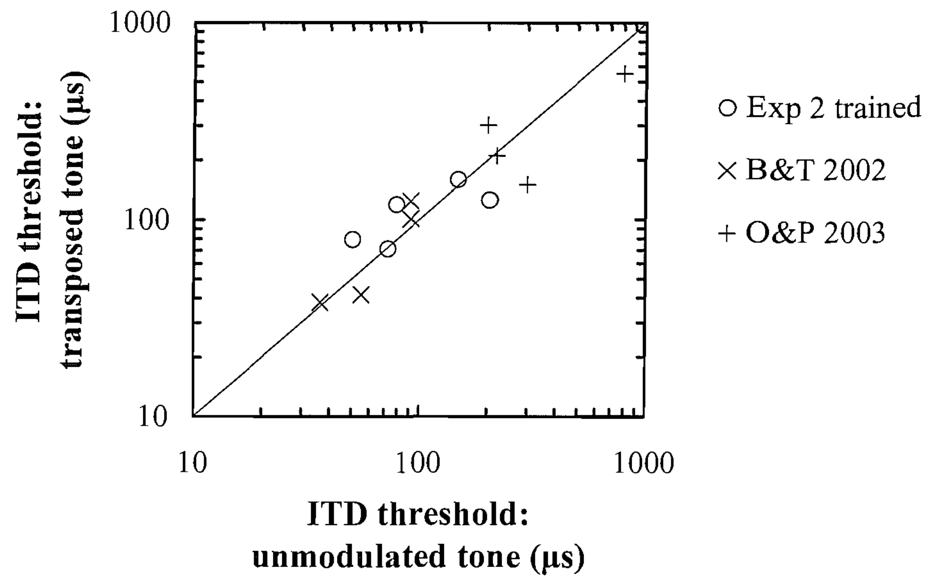


Figure 5.3. Comparison of ITD thresholds with the 128–Hz unmodulated and transposed (4000–Hz carrier frequency) tones from listeners in the present (at 60 phons) and previous studies. Abbreviations used as in Figure 4.5.

Figure 5.4 plots the ratio of thresholds with the unmodulated and transposed tones at 30 and 60 phons for each listener. A value of unity indicates that level did not have an effect; a value of greater than unity indicates that thresholds were poorer at the lower level. Any level effect was small for the unmodulated tone but stronger for the transposed tone, leading to approximately a doubling of threshold for three listeners. Interestingly, thresholds of Listener 5 with the transposed tone were over six times poorer at the lower level. However, the SD in the 30–phon condition for Listener 5 was 20% indicating that thresholds were reasonably consistent. Also, audiometric thresholds at 4000 Hz were consistent across listeners. The reasons for these apparent inter–individual differences are unclear.

In Section 4.4.4, it was suggested that the comparison of data between Experiment 1 and Wright and Fitzgerald (2001) might be complicated by differences in sensitivity to the trained stimuli, possibly resulting from differences in SPL. This has been confirmed by the present study. These data suggest that learning with low– and high–frequency ITD could be compared using unmodulated and transposed tones both presented at the higher SPL, as comparable thresholds seem apparent in most trained listeners. In fact, if these stimuli were indeed associated with different magnitudes of learning, inexperienced listeners would be expected to have disproportionately poorer ITD thresholds with the transposed stimuli. This was evaluated in Experiment 2b.

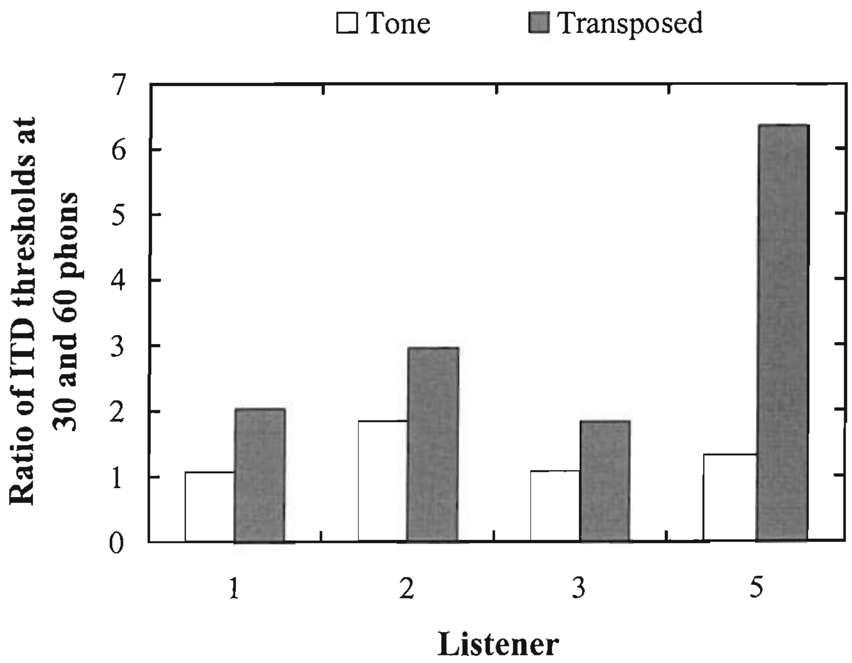


Figure 5.4. Effect of level expressed as the ratio of ITD thresholds at 30 and 60 phons for each listener with unmodulated and transposed tones.

Reducing the SPL of the stimulus towards absolute threshold is generally assumed to reduce ITD discrimination, regardless of the stimulus, due to the increase in sensory 'noise' relative to the representation of the stimulus (Hafter and Trahiotis, 1997). In physiological terms, this may be at least partially related to the reduction in synchrony of phase–locking with decreasing level (Palmer and Russell, 1986). It may also be related to the reduction in number of stimulated auditory–nerve fibres tuned to adjacent frequencies as level reduces (e.g. Palmer, 1995, Figure 4). The reasons for the differential effect of SPL with low–frequency unmodulated and high–frequency AM stimuli are less clear. One possibility is related to the level dependency of the bandwidth of the auditory filter (e.g. Moore, 1995, Figure 7), arising from that of the envelope of basilar membrane vibration (Moore, 1986). The skirts of the filter may attenuate the sidebands of AM stimuli as bandwidth reduces, essentially leading to a reduction in the modulation depth of the representation of the envelope. Physiologically, this may be associated with reduced synchrony of phase–locking. Indeed, ITD thresholds have been found to be sensitive to stimulus modulation depth (e.g. Nuetzel and Hafter, 1981). However, comparing across studies, the level effect with a two–tone complex with a bandwidth of only 89 Hz centred on 4000 Hz (McFadden and Pasanen, 1976) also seems stronger than with tones (Zwislocki and Feldman, 1956; Hershkowitz and Durlach, 1969). It has been suggested that cochlear compression may have a differential effect on unmodulated and AM stimuli with reducing level (e.g. Bernstein and Trahiotis, 2002), again expected to influence modulation depth. It may also be related to the how activity in auditory–nerve fibres tuned to adjacent frequencies changes with stimulus level. One might imagine that optimum ITD thresholds require activity in a criterion number of fibres across the cochleotopic array and that a reduction below this impairs performance. It is known that the response of the auditory–nerve fibres along the cochleotopic array is broader to low– than high–frequency stimulation (e.g. Palmer, 1995, Figure 4). It may be that at high SPL, high–frequency stimuli produce activity closer to this criterion thus making ITD thresholds more sensitive to reducing level than at low frequencies. These issues remain to be resolved.

5.5 Experiment 2b: Additional methods

In Section 5.4, it was suggested that a difference in magnitude of learning between low– and high–frequency ITD may be evaluated indirectly by comparing thresholds from the trained listeners in Experiment 2a with those from a group of untrained listeners. The expectation was that a difference would be apparent as disproportionately worse thresholds with the transposed compared to unmodulated tone in untrained listeners. This was evaluated in Experiment 2b. This experiment also served to evaluate the modified procedures with untrained listeners and provide information to assist the design of future studies of learning with these stimuli.

Seven normal–hearing listeners were recruited; these denied previous experience with psychoacoustical experiments. Listeners attended two sessions: session one was screening and familiarisation; session two was essentially pre–test. Familiarisation was similar to Experiment 1 in that listeners completed 30 trials of a fixed ITD with a 1000–Hz tone followed by the test stimuli and one adaptive block with the test stimuli. However, prior to this, monotic absolute thresholds with the test stimuli were measured using the same general trial format and adaptive procedure used in ITD discrimination, modified for signal detection. This was used to provide listeners with extra experience of the measurement procedures but not of binaural discrimination. The pre–test session was identical to the test session in Phase 2 of Experiment 1a, except that testing was only conducted with the unmodulated tone and narrowband transposed tone, presented at 60 phons. The masking noise was presented during testing with the AM stimuli. The session mean threshold with each stimulus was based on the arithmetic mean of the six estimates.

5.6 Experiment 2b: Results and discussion

Figure 5.5 reproduces Figure 5.3 and adds the ITD thresholds from the untrained listeners. It can be seen that, overall, the untrained listeners had poorer thresholds compared to the trained listeners from Experiment 1 and the two previous experiments. However, these were disproportionately poorer with the transposed tone in all but one case, thresholds being between 1.5 and 7.0 times higher than with the unmodulated tone. Figure 5.6 plots the individual estimates with both stimuli for each listener. It can be seen that a number of listeners had variation in thresholds with the transposed tone exceeding that with the unmodulated tone. Indeed, the SD for the transposed tone was greater than 40% in four listeners, and 50% in one, but no more than 33%, and usually less than 25%, with the unmodulated tone. The listener with the greatest disparity in thresholds between stimuli also showed clear evidence of learning on the transposed tone during testing (crosses in Figure 5.6). However, even if the session mean for this listener was based on the three best thresholds, performance was still 4.0 times poorer than with the unmodulated tone.

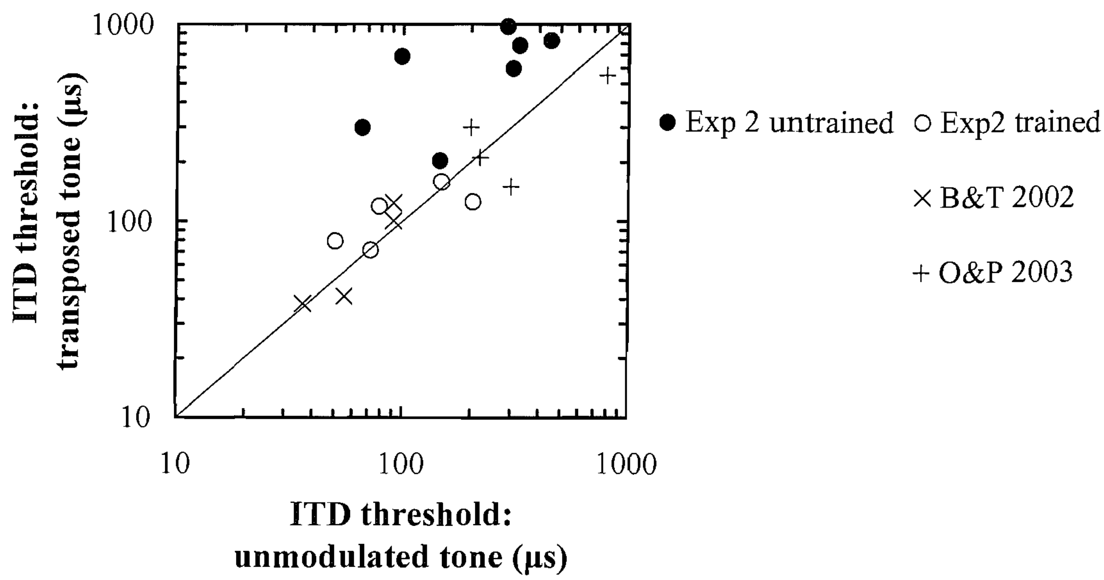


Figure 5.5. As Figure 5.3 but also including the thresholds from the untrained listeners from the present experiment.

In summary, ITD thresholds in untrained listeners with both low–frequency unmodulated and high–frequency transposed tones were worse than found in Experiment 2a with trained listeners. In fact, thresholds on the transposed tone of the untrained listeners were disproportionately worse compared to the trained listeners. The latter finding may imply that the magnitude, and possibly time-course, of learning may differ between the stimuli and, specifically, be greater for the transposed tone. This may therefore provide indirect evidence for a dependency of ITD discrimination learning on frequency region and/or type of ITD cue (fine–structure–based vs. envelope–based). This interpretation assumes that the disparity in thresholds between the two stimuli in the untrained listeners was related to their inexperience and would have 'resolved' (i.e. the thresholds would have become comparable, as with the trained listeners) following training. Other factors, however, may also have influenced performance with the transposed tone but not, or less so, with the unmodulated tone. For example, low-pass noise was used only during testing with the transposed tone and may distract some inexperienced listeners from optimally attending to ITD discrimination; indeed, some listeners said as much. Further, several listeners reported finding the transposed tones "peculiar" to listen to, which may also have influenced their attention. Consequently, a full training experiment, with modified procedures for testing inexperienced listeners, was necessary to compare directly the characteristics of learning in low– and high–frequency regions.

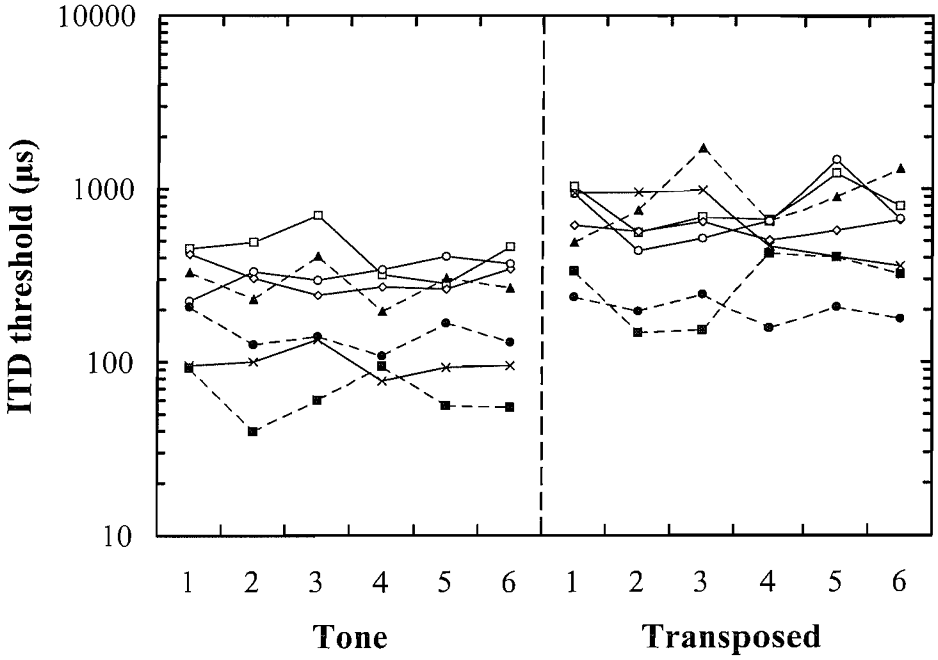


Figure 5.6. ITD threshold as a function of block number (i.e. one to six) are plotted for the unmodulated and transposed tones, and for each listener separately. Symbols represent different listeners. Thresholds for the unmodulated tone are placed arbitrarily on the left of the graph.

Chapter 6. Comparison of learning at low and high frequencies

6.1 Experiment 3: Introduction

Recent years have witnessed a growth in interest in training–induced learning on discrimination tasks in normal–hearing listeners. For example, Wright and Fitzgerald (2001) trained listeners with ongoing ITD or ILD using low– or high–frequency tones, respectively; additional listeners served in untrained, control groups. They suggested that low–frequency ITD discrimination is generally associated with rapid learning whereas high–frequency ILD discrimination is generally associated with a longer time–course of learning that only partially generalises to untrained stimuli at different frequencies. They also suggested that these differences might be associated with differences in the brainstem pathways involved in the binaural processing of the stimuli. Plasticity is known to be manifest in the brainstem subsequent to hearing loss (Illing, 2001), and recent research in humans suggests it may also be demonstrated with auditory training, under some conditions (Russo et al., 2005).

One limitation with Wright and Fitzgerald's experiment is that it is not clear if the apparent difference in learning between stimuli was related to the different binaural cues or the different frequencies, or an interaction of the two. This issue may be addressed by comparing learning with the same cue at different frequencies. For example, if Wright and Fitzgerald's results reflect the different cues, comparable time–courses of learning may be apparent with ongoing ITD discrimination at both low and high frequencies. Alternatively, if their results reflect the different frequencies of the stimuli, different time–courses may occur. If the time–course of learning is related to the degree of plasticity in binaural processing in the brainstem, as suggested by Wright and Fitzgerald, the degree to which these pathways process frequency or cue separately may be important.

There is reason to suspect that binaural pathways in the brainstem are separated by frequency rather than cue (Section 2.3.3). Most neurons in the MSO are responsive to low frequencies, have EE–type responses and are thought to process low–frequency ITDs. In contrast, most neurons in the LSO are responsive to high frequencies, have EI–type responses and are thought to process high–frequency ILDs. These LSO neurons are also sensitive to high–frequency envelope–based ITD. This suggests that low and high–frequency binaural cues are processed in distinct pathways. However, the situation is probably more complex than this. For example, the MSO also contains some neurons with EI–type responses and neurons that respond to high–frequency ITD. Further, the LSO contains some neurons with EE–type responses and neurons that respond to low–frequency ITD (e.g. Yin, 2002). Overall, the initial processing of binaural cues may not be separated by cue or frequency in a straightforward manner. However, psychoacoustical models of ITD sensitivity generally use the same binaural processing mechanism independent of frequency (e.g. cross–correlation, Colburn, 1995). If the time–course of ITD discrimination learning is frequency dependent, all other important factors being equal, this would be difficult to explain with these simplified models.

This may also have implications for clinical populations, such as BICI users. For example, BICIs may represent low–frequency information in normally high–frequency channels (Skinner et al., 2002). Although current BICI processing may not preserve low–frequency ITD, the view that it is the dominant horizontal–plane localisation cue in normal–hearing people has motivated reconsideration of BICI processing (Wilson et al., 2003). It is therefore of interest to consider the potential for plasticity associated with the processing of ITD, both in terms of the effects of auditory deprivation subsequent to hearing loss and the adaptation to novel stimulation. However, data on learning using low–frequency acoustic stimuli may not apply to BICIs.

Experiment 1 showed that ITD thresholds with high–frequency stimuli improved over many hundreds to thousands of trials during training, even for otherwise diotic stimuli. This seemed inconsistent with the time–course of learning suggested by Wright and Fitzgerald for low frequencies. However, differences in asymptotic thresholds between the stimuli used in each study complicated comparison.

Experiment 2 used low–frequency and high–frequency (transposed) stimuli producing comparable ITD thresholds in trained listeners, and found that inexperienced listeners appeared to perform disproportionately worse with the high–frequency stimulus. It was suggested that this supported the hypothesis that the magnitude and/or time–course of learning on ITD discrimination is longer at high than low frequency, although did not provide a direct test of it. This chapter describes an experiment to test directly the hypothesis that characteristics of ITD discrimination learning differ at low and high frequencies, using stimuli associated with comparable asymptotic thresholds[17].

6.1.1 Summary of objectives

The objectives of this experiment were:

- To re–evaluate ITD thresholds at low and high frequencies in inexperienced listeners, as in Experiment 2b, but using modified methodology. This is in order to determine if the disproportionately worse thresholds measured at high frequency, as found in Experiment 2b, were due to differences in auditory capabilities or potential experimental confounds.
- To determine if training–induced ITD discrimination learning is evident at low frequency. According to Wright and Fitzgerald's conclusions, no training–induced learning would be predicted at low frequency. However, there is uncertainty as to whether this is conclusion is true in general.
- To determine if training–induced ITD discrimination learning is evident at high frequency, and to compare with the results at low frequency. The hypothesis is that training–induced learning is apparent at high–frequency.

---

[17]Rowan D, Lutman ME. Effects of training on ITD discrimination using "transposed stimuli": preliminary analysis. Poster presented at the British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness; 2004 Sept 16–17; London, UK.

Rowan D, Lutman ME. Generalisation of learning with ITD discrimination across frequency and type of cue. Poster presented at the 28th Annual Mid Winter Research Meeting of the Association for Research in Otolaryngology; 2005 Feb 19–24; New Orleans, Louisiana, USA.

Rowan D, Lutman ME. Learning on binaural discrimination tasks in humans. Oral presentation at the British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness; 2005 Sept 12–13; Cardiff, Wales.

- If training–induced learning is apparent at both low and high frequencies, to compare directly the magnitude and time–courses of learning. The hypothesis is that the time–course and/or magnitude of learning are/is greater at high frequency.
- To explore across–frequency generalisation of learning.
- To explore inter–individual differences in learning.

A new cohort of normal–hearing listeners was recruited. These were trained on ITD discrimination with either low–frequency unmodulated or high–frequency transposed tones, or served in an untrained, control group. The experiment also measured ITD thresholds with both stimuli at pre– and post–test to permit the evaluation of the generalisation of learning across frequency.

6.2 Additional methods

6.2.1 Organisation of experiment

The general organisation and salient details of this experiment are summarised in Figure 6.1. Listeners attended nine sessions: session one was screening and Phase 1 of familiarisation; session two was Phase 2 of familiarisation and pre–test; sessions three to eight were formal training; session nine was identical to session two (i.e. included a replication of Phase 2 of familiarisation, see below). Pre– and post–testing consisted of four blocks of ITD discrimination with a 128–Hz unmodulated tone and 128–Hz–modulated SAM and transposed tones at 4000 Hz, all presented at the higher SPL as in Experiment 2b (the order was as described in Section 3.1). Following pre–test, listeners were randomly allocated to receive training with either a low–frequency unmodulated or high–frequency transposed tone, or served in an untrained control group. Each training session consisted of six blocks of ITD discrimination, as in Experiment 1. The adaptive procedures used in testing and training were further modified from those used in Experiment 2. To help maintain listeners' motivation, they were paid £5 an hour on completion of the experiment with a bonus payment of £10 for the five listeners with best overall mean post–test thresholds.

Phase 1 of familiarisation was testing on binaural signal detection allowing listeners to adjust to the general trial format and adaptive procedure; Phase 2 was testing on a lateral–position rating task, to encourage listeners to listen to changes in lateral position. Binaural signal detection was tested in N_0S_0 and N_0S_π conditions and at low and high frequencies, the latter with transposed stimuli. At least two blocks were completed in each condition[18]. Measurements at each frequency were completed consecutively although the first frequency and the order of each signal condition were varied across listeners. Two blocks of the rating task were completed prior to ITD discrimination. In each block, listeners judged the lateral position of the unmodulated and transposed tones, and also broadband noise, as it leads to a compact image and large changes in lateral position. These tests are described in detail in Section 6.2.2.

[18] Blocks with a SD of the reversals of greater than 4 dB were discarded and retested.

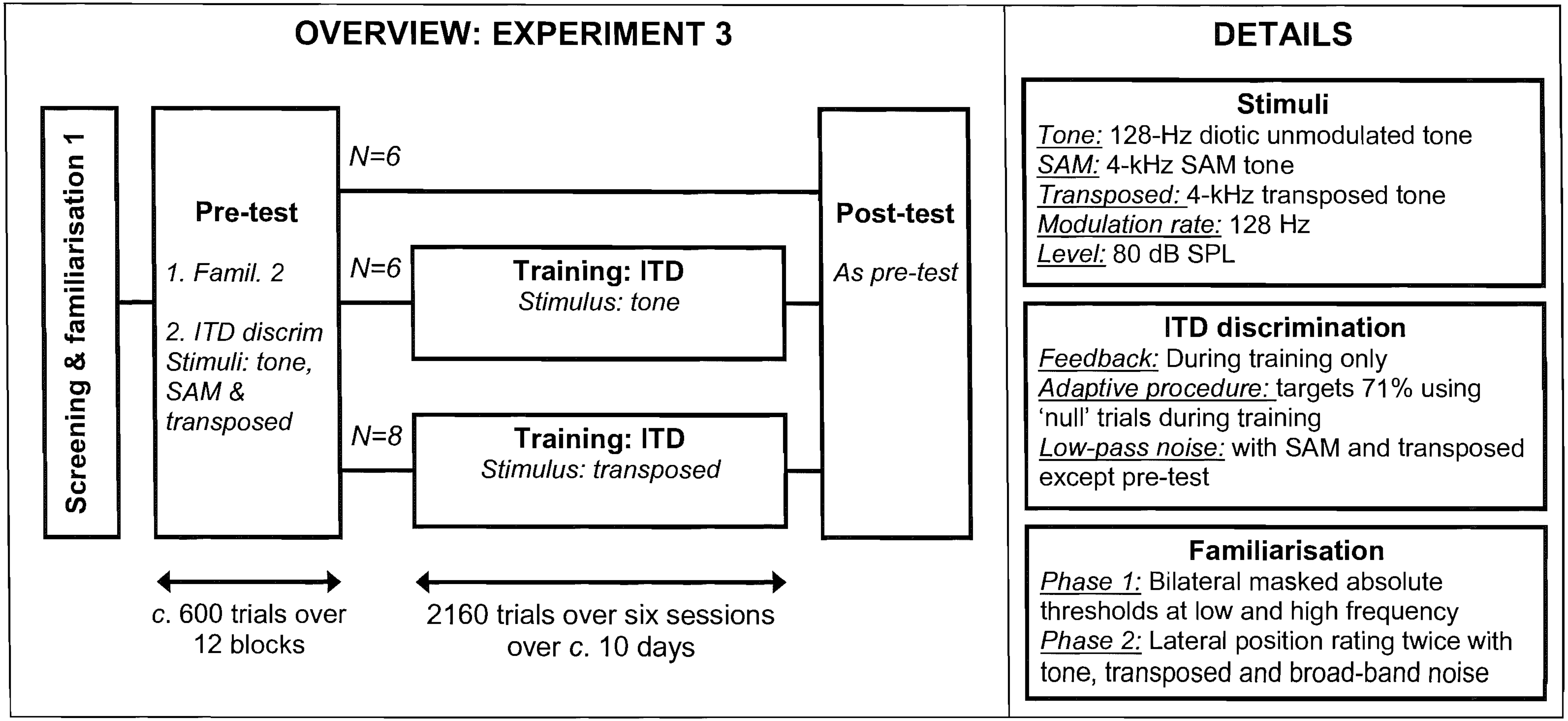


Figure 6.1. Schematic illustration of the organisation of Experiment 3. The salient details of the methodology are also indicated.

6.2.2 Procedures

6.2.2.1 Binaural signal detection

Absolute thresholds were measured using the same general trial format and adaptive procedure as with ITD discrimination during pre–test but modified for signal detection. Observation intervals were 600 ms in duration, each containing a 600–ms–long burst of the 'masker' (i.e. low–frequency narrowband noise or its transposed counterpart). One of the last three intervals, selected at random, also contained a 400–ms–long 'signal' (i.e. the 128–Hz tone or its transposed counterpart), temporally centred on the masker. Durations included 60–ms–long ramps. Listeners were required to identify the interval containing the signal and to press one of three buttons on a keypad accordingly. Feedback was not provided for consistency with the pre–test discrimination procedure (see Section 6.2.2.3). The decision rule controlled the signal–to–noise ratio by varying the level of the signal; the initial signal–to–noise ratio was 10 dB; the step size was 5 dB for the first three reversals and 2 dB thereafter. Testing was terminated after eight scored reversals and threshold was defined as the arithmetic mean of these.

6.2.2.2 Lateral–position rating

The trial format and method of scoring of the lateral–position rating procedure was based on Yost (1981), and the response method was based on Rule (1994) and Rule and Nickolaychuk (1995). Each trial comprised two sections, separated by 800 ms. The first section comprised three observation intervals, separated by 400 ms, containing diotic stimuli. The second section comprised four observation intervals, again separated by 400 ms, containing stimuli with an ITD. A single stimulus type and ITD was presented on each trial. Following a trial, listeners could either replay the trial (an unlimited number of times) or make a response and proceed to the next trial. To respond, listeners drew a vertical line, indicating perceived lateral position, on a pre–drawn horizontal line, representing the interaural axis. The pre–drawn line was 18–cm long, and seven were drawn on a single side of A4 paper (landscape). Each block consisted of 21 trials (replayed trials not included), using three stimuli and seven ITDs; the order was selected randomly prior to each block. The ITDs were 0 μs, ±350 μs, ±700 μs, ±1000 μs. The scale was divided into 21 intervals and each response scored between 1 and 21.

6.2.2.3 ITD discrimination

The ITD discrimination procedures were based on those used during Experiment 2 but modified differently for testing and training phases. During pre– and post–testing, blocks were terminated after eight scored reversals, to ensure sufficient reversals were accrued, and no trial–by–trial feedback was provided, to help reduce threshold instability. During formal training, trial–by–trial feedback was provided and blocks were terminated after 60 trials. In a change from Experiment 2, the initial ITD was 700 μs for the first block in each session and one large step size above the previous ITD threshold thereafter. The decision rule during training consisted of the modification to the two–down, one–up procedure described by Trahiotis et al. (1990). This modified rule, illustrated in Figure 6.2, includes a 'null' trial after every incorrect response, whereby the ITD is the same as the previous trial but the response is disregarded. This maintains the statistical properties of the two–down, one–up procedure, permitting comparison with pre– and post–test, and, together with the lower initial ITD, increases the proportion of trials close to listeners' thresholds. It has been suggested that this may enhance the potency of the training (Trahiotis et al., 1990). The modified rule may also make it more difficult for listeners to anticipate the decision rule.

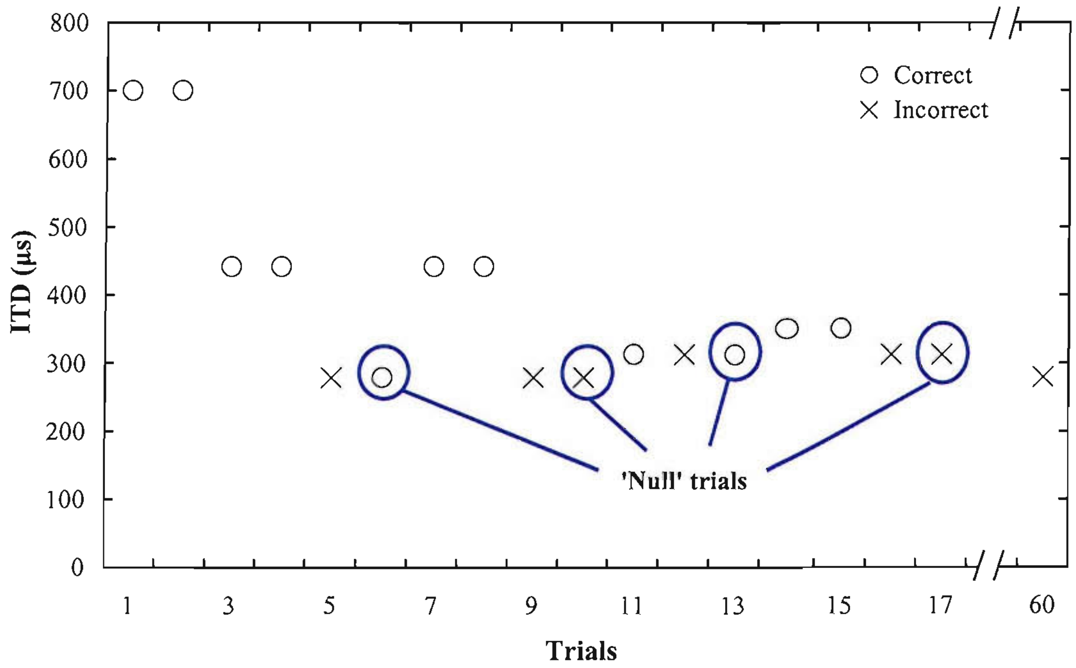


Figure 6.2. Illustration of the modified decision rule used during training.

The masking noise was presented with the high–frequency stimuli during training and post–test but not pre–test. This was to ensure that pre–test thresholds did not differ between stimuli because of a distracting effect of the noise. Although this risked performance reflecting use of low–frequency spectral information, this seemed unlikely for inexperienced listeners. This issue is revisited during the Discussion (Section 6.4.2).

6.2.3 Stimuli

Stimuli were generated, presented and calibrated as in Experiment 2.

6.2.3.1 Binaural signal detection

Masked absolute thresholds were measured in N_0S_0 and N_0S_π conditions. Maskers were presented at 70 dB SPL. In the low–frequency condition, the signal was a 128–Hz unmodulated tone and the masker was a 100–Hz–wide band of Gaussian noise centred on 128 Hz. This bandwidth permitted comparison with van de Par and Kohlrausch (1997). In the high–frequency condition, the masker alone or signal plus masker were transposed to 4000 Hz using the procedure described by Bernstein and Trahiotis (2002) (i.e. half–wave rectification, low–pass filtering at 2000 Hz and multiplication with the carrier, see Section 2.3.5). Following van de Par and Kohlrausch, the signal and masker were combined, and the signal–to–noise ratio determined, prior to transposition. A portion of the waveforms and the amplitude spectra of the low–frequency and transposed maskers are illustrated in Figures 6.3 and 6.3, respectively. Bernstein and Trahiotis (2004) have discussed the differences between the spectra of the transposed band of noise (Figure 6.4) and the transposed tone (Figure 3.5). In short, the power in the sidebands of the transposed noise is spread across a range of frequencies (and so they overlap) reducing that at the centre frequency of the sideband relative to the carrier. Twenty independent tokens of the low–frequency masker were generated offline. [According to Siegel and Colburn (1989), ten tokens are required for detection performance to be comparable to that found with 'running' noise.] At the start of each trial, four tokens were randomly selected and one was combined with the signal (including ramps). In the high–frequency condition, the four stimuli were then transposed. Final ramps were then imposed.

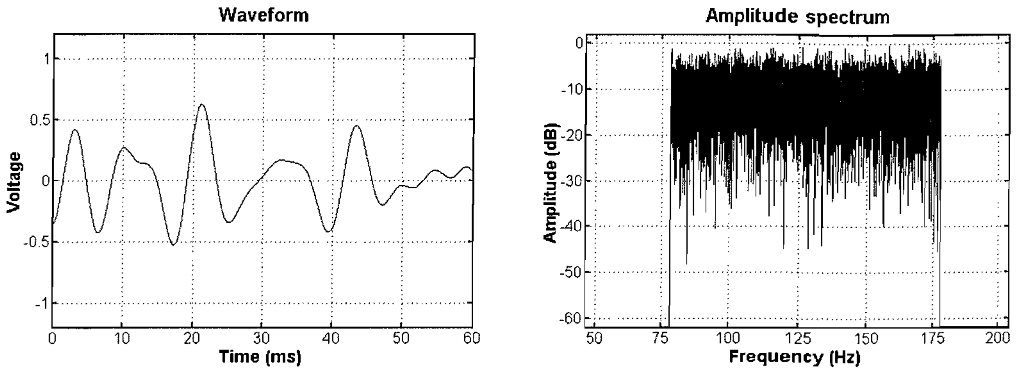


Figure 6.3. A portion of the waveform and the amplitude spectrum of the low–frequency masker.

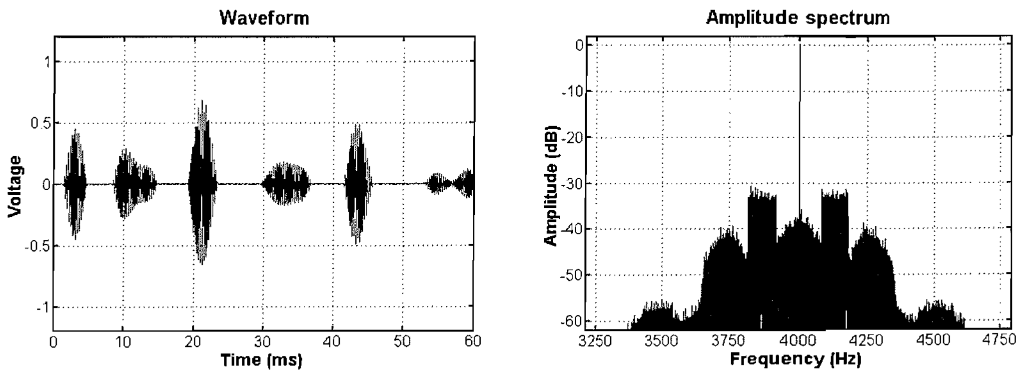


Figure 6.4. A portion of the waveform and the amplitude spectrum of the high–frequency (transposed) masker.

6.2.3.2 Lateral–position rating

Unmodulated and transposed tones were presented at 80 dB SPL during the lateral–position rating procedure. The latter was generated with n=2 in Equation 3 (Section 3.4.2). The broadband noise consisted of Gaussian noise filtered between 50 and 8000 Hz, and was presented at 60 dB SPL. An ITD was applied to the broadband noise by imposing the appropriate frequency–dependent phase shifts. Ramps were then applied to reduce the saliency of an onset or offset cue.

6.2.3.3 ITD discrimination

Unmodulated, SAM and transposed tones were presented at 80 dB SPL during ITD discrimination. The latter was generated with n=2 in Equation 3.

6.2.4 Statistics

The primary planned analysis was an ANOVA on individual ITD thresholds with:

- Related factors of:
  - *Block* (repetitions 1–4 with each stimulus, in each session)
  - *Stimulus* (unmodulated, SAM and transposed tones)
  - *Session* (pre– vs. post–test)
- Unrelated factors of:
  - *Training Group* (control, tone–trained, transposed–trained)
  - *Counterbalancing Group* (three groups receiving testing with each stimulus in a different order)

This produced five main effects and 22 interactions, but only a subset of these effects were of primary interest. Two specific hypotheses of this experiment were that thresholds with the unmodulated and transposed tone were different at pre–test, and that this difference reduces with training on the transposed stimulus. Effects involving Stimulus, particularly the main effect and interactions with Session and Training Group, permitted initial examination of these hypotheses. Another central hypothesis was that the transposed–trained but not the tone–trained group learned more than the untrained group on the trained stimulus. Effects involving Session, in particular the interaction with Training Group, permitted initial investigation of this hypothesis. Additional effects were also of interest. Effects involving Block permitted examination of the stability of thresholds within pre– and post–test sessions and effects involving Counterbalancing Group, such as the interactions with Block and Stimulus, permitted the investigation of order effects.

A further ANOVA was planned on ITD thresholds with:

- Related factors of:
  - *Block* (repetitions 1–6 with each stimulus, in each session)
  - *Session* (six formal training sessions)
- Unrelated factor of *Training Group* (tone–trained, transposed–trained)

This produced three main effects and four interactions. This analysis evaluated the other central hypothesis that the time–course of learning differed between training groups. Effects related to Block permitted investigation of whether learning occurred within as well as across sessions.

Given the number of main effects and interactions of interest with these analyses, the criterion probability was reduced to 0.01. While this does not correspond to a full Bonferroni correction, it reduced the possibility of erroneously rejecting a null hypothesis.

A sample–size calculation was conducted. The SD of thresholds across listeners was estimated to be 2.5 dB for all stimuli and all sessions, based on data from Experiment 2. The mean difference in thresholds between pre– and post–test for the unmodulated tone for all three groups was estimated to be 1.7 dB, based on Wright and Fitzgerald's (2001) untrained group. The mean difference for the transposed tone in the transposed–trained group was estimated to be 3.3 dB, based on data from Experiment 1 and assuming the modifications to the training procedures slightly increased the amount of learning. The mean difference for the transposed stimulus in the remaining groups was estimated to be 1.7 dB (i.e. assumed no training–induced learning). As in Experiment 1, it was assumed that 75% of the variance was shared by repeated measurements, and the sample size was calculated with a statistical power of 80% (two–tailed tests, $\alpha = 0.01$ and $\beta=0.2$ giving $\delta=3.2$; $\rho=0.87$). Accordingly, the sample–size calculation indicated that at least 14 listeners were required in each group to detect the differences in learning between groups on the same and different stimuli (i.e. between learning of 3.3 dB and 1.7 dB). Thus, a sample size of 42 was indicated.

6.3 Results

6.3.1 Overview of analysis

Twenty listeners (9 females, 11 males) aged 22–34 years (mean age 26 years) were recruited. An additional listener was excluded as she reported having a medical condition that affected her short–term memory of temporal order. Six listeners participated in the untrained group, six listeners were trained with the unmodulated tone and eight listeners were trained with the transposed tone.

The experiment was terminated after recruiting and fully testing 20 listeners, as opposed to the 42 listeners as planned. This was because the trends emerging in the data (reviewed mid–way through experiment) deviated in important ways from the assumptions made when conducting the sample–size calculation. The prediction of a sample–size of 12 listeners per group was based on the assumption of a difference in learning between groups of 1.7 dB (see Section 6.2.4). The differences between groups that was estimated mid–way through the experiment were either substantially greater (up to 3.9 dB) or substantially smaller (no greater than 0.3 dB) than this. The former differences were found to be highly statistically significant and the latter were found to be far from statistically significant with the smaller sample size. Regarding the latter non–significant differences, the results suggest that either no differences in learning exist or that a substantially larger sample–size than the planned size would be required to detect it statistically.

Section 6.3.2 describes the trends observed in the pre– and post–test data. It starts by considering the distributions of ITD thresholds at pre–test, possible order effects and the relative performance of training groups. These issues are important as they influence the subsequent analysis conducted into learning. Section 6.3.3 then describes statistical analyses conducted on pre– and post–test thresholds. Section 6.3.4 describes the exploration and analyses of the data from formal training and Section 6.3.5 investigates the relationship between learning and pre–test thresholds. Finally, Section 6.3.6 investigates further ITD thresholds at pre– and post–tests, and Section 6.3.7 describes the results from the additional measures of binaural hearing.

In this experiment, 1.2% of blocks resulted in no scored reversals, comparable to Experiment 1. However, in the present study, these occurred only during formal training sessions. For the purposes of the analyses described here, these missing values were replaced with the session mean value.

Table 6.1. For the trained stimulus, mean pre– and post–test ITD thresholds and the difference between (and ratio of) these for each listener, and mean data for each training group (see text for details).

Listener	Trained condition	ITD Threhold (μs)								
		Tone			SAM			Transposed		
		Pre test	Post test	Pre–Post (Pre/Post)	Pre test	Post test	Pre–Post (Pre/Post)	Pre test	Post test	Pre–Post (Pre/Post)
1	Untrained	193	153	40 (1.26)	567	434	133 (1.31)	311	188	123 (1.65)
2	Untrained	222	332	-110 (0.67)	1131	678	453 (1.67)	529	361	168 (1.47)
3	Untrained	126	115	11 (1.09)	273	396	-123 (0.69)	193	224	-31 (0.86)
4	Untrained	199	285	-86 (0.70)	726	531	195 (1.37)	561	716	-156 (0.78)
5	Untrained	338	360	-22 (0.94)	1017	941	76 (1.08)	430	596	-166 (0.72)
6	Untrained	270	561	-291 (0.48)	442	537	-95 (0.82)	220	220	0 (1.00)
7	Tone	153	90	63 (1.70)	727	439	288 (1.66)	375	139	236 (2.70)
8	Tone	138	59	79 (2.34)	155	77	78 (2.02)	62	50	13 (1.25)
9	Tone	822	362	459 (2.27)	1207	675	532 (1.79)	666	256	409 (2.60)
10	Tone	83	31	52 (2.67)	217	146	72 (1.49)	107	69	38 (1.55)
11	Tone	145	76	69 (1.91)	325	130	195 (2.50)	156	88	68 (1.77)
12	Tone	140	99	41 (1.41)	250	196	54 (1.27)	110	116	-5 (0.95)
13	Transposed	195	113	82 (1.72)	714	398	316 (1.79)	518	212	306 (2.44)
14	Transposed	236	120	116 (1.97)	561	395	166 (1.42)	494	238	256 (2.07)
15	Transposed	155	173	-18 (0.90)	180	255	-76 (0.70)	108	153	-45 (0.70)
16	Transposed	303	173	130 (1.75)	365	255	110 (1.43)	168	153	16 (1.10)
17	Transposed	178	126	52 (1.41)	293	162	131 (1.81)	180	66	115 (2.75)
18	Transposed	458	153	305 (3.00)	729	584	145 (1.25)	351	258	92 (1.36)
19	Transposed	260	78	182 (3.34)	331	125	206 (2.65)	239	62	177 (3.86)
20	Transposed	660	145	515 (4.56)	1124	323	801 (3.48)	414	197	217 (2.10)
Geometric mean	Untrained	215	264	-49 (0.86)	620	561	59 (1.16)	345	335	9 (1.08)
	Tone	176	88	88 (2.01)	366	209	156 (1.75)	175	104	71 (1.69)
	Transposed	272	131	141 (2.07)	462	281	181 (1.64)	271	149	122 (1.82)

6.3.2 Exploration of pre– and post–test ITD thresholds

Table 6.1 presents the geometric mean pre– and post–test ITD thresholds with the three stimuli for each listener and training group. The difference and ratio of pre– and post–test thresholds are also given. The difference and ratio for the training groups was based on the group mean pre– and post–test thresholds rather than an average of differences and ratios across individual listeners.

6.3.2.1 Distributions of ITD thresholds at pre–test

Twelve thresholds were measured during pre–test (four tests over three stimuli) for each listener. The distributions of thresholds across listeners were positively skewed in all 12 cases, but were at least approximately normally distributed following a logarithmic transformation. This transformation was therefore applied prior to all analyses except investigation of the relationship between learning and pre–test thresholds in Section 6.3.5.

6.3.2.2 Order effects

Figure 6.5 plots mean ITD thresholds over the four blocks with each stimulus for each training group at pre– and post–test. It can be seen that, overall, performance was generally stable in both sessions. Only at post–test with the tone–trained group (bottom row, centre column) was there a trend for performance to drift monotonically, in this case a tendency to improve, within the session. Similarly, no clear drifts were apparent when the data were grouped by counterbalancing group. However, while this was the case overall, the thresholds of some listeners did appear to gradually improve or worsen during a session. The pre– or post–test mean ITD threshold for each stimulus did not appear to depend on counterbalancing group, as might occur if performance depended on the order in which stimuli were tested. These observations were supported by statistical analyses (Section 6.3.3.2).

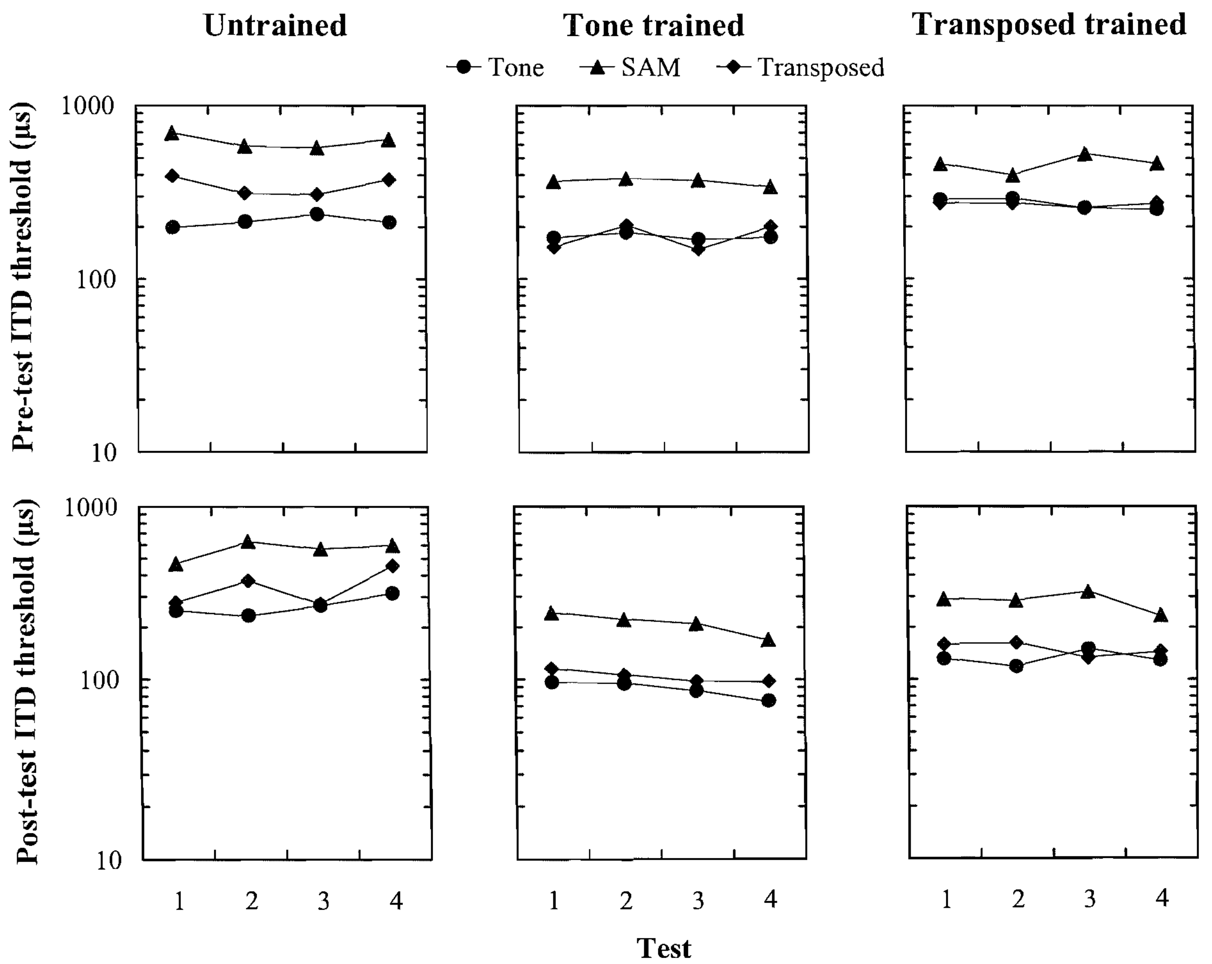


Figure 6.5. Geometric mean ITD threshold as a function of block number for each group and stimulus at pre– and post–test.

6.3.2.3 *Pre–test ITD thresholds across stimuli*

Figure 6.6 plots the distribution of mean pre–test ITD thresholds across all listeners for each stimulus, and displays the geometric means. The positive skew noted in Section 6.3.2.1 is readily apparent. As found in Experiment 1, there was a considerable spread of thresholds, particularly with SAM. The mean thresholds with the unmodulated and transposed tones were comparable and approximately half the mean threshold with the SAM tone. Table 6.1 indicates that thresholds with the SAM tone were poorer than with the other stimuli in all listeners. This is also apparent in Appendix E, which plots mean ITD thresholds with each stimulus for each listener separately. There was more variation in the relationship between thresholds with unmodulated and transposed tones across listeners, and this is described in detail in Section 6.3.6. However, listeners with relatively good or poor thresholds with one stimulus tended also to have relatively good or poor thresholds with the other stimuli. These results are inconsistent with the experimental hypothesis, and the findings of Experiment 2b, that ITD thresholds in untrained listeners are poorer with the transposed tone compared to the unmodulated tone.

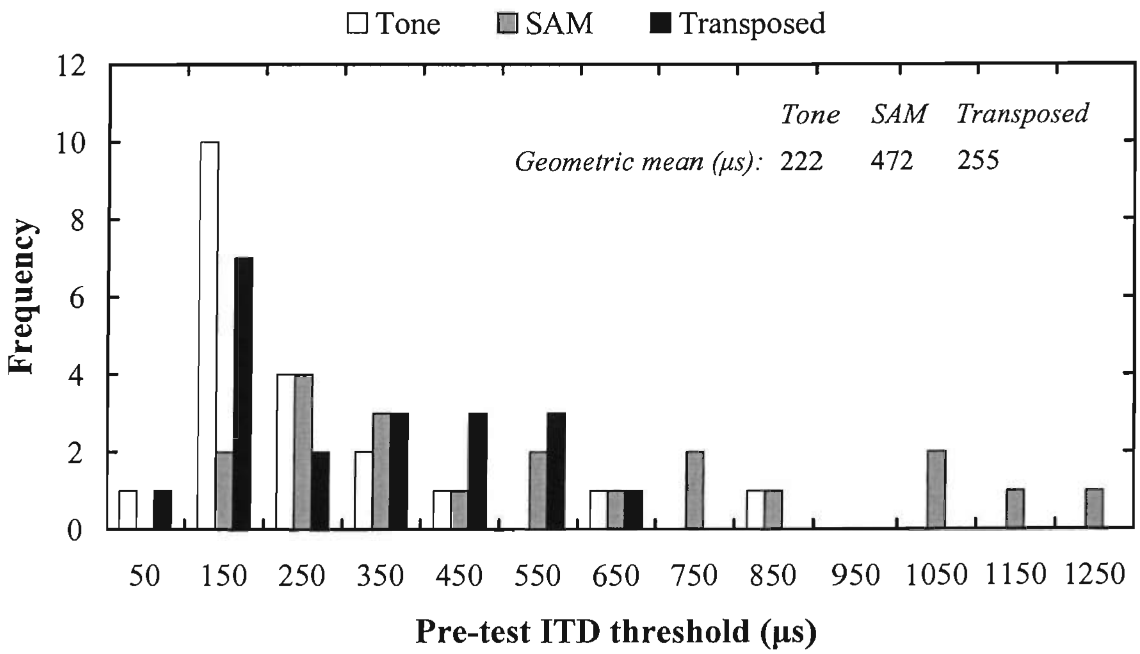


Figure 6.6. Histogram of geometric mean pre–test ITD thresholds across all 20 listeners for each stimulus. The centre values of the bins are given on the abscissa. Overall geometric mean thresholds for each stimulus are displayed.

6.3.2.4 Pre–test ITD thresholds across training groups

Figure 6.7 plots pre–test ITD thresholds for each listener grouped by training group and stimulus. While there is some similarity in the distributions across groups, the tone–trained group contained listeners both with the best (e.g. Listener 8, ×) and poorest thresholds (Listener 9, ✳). The variation also appears greater for the transposed–trained compared to the untrained group. Figure 6.8 plots the mean pre–test ITD threshold for each stimulus across each training group; error bars indicate ±1 geometric SE. It can be seen that, overall, the tone–trained group had (by chance) the lowest ITD thresholds but greatest variation across all stimuli. The untrained group also had a slightly higher mean threshold than the two trained groups with the high–frequency stimuli. These apparent differences in pre–test mean thresholds across groups were not found to be statistically significant (Section 6.3.3.4). Nevertheless, these differences are important to note as they may still influence the interpretation of apparent differences in learning.

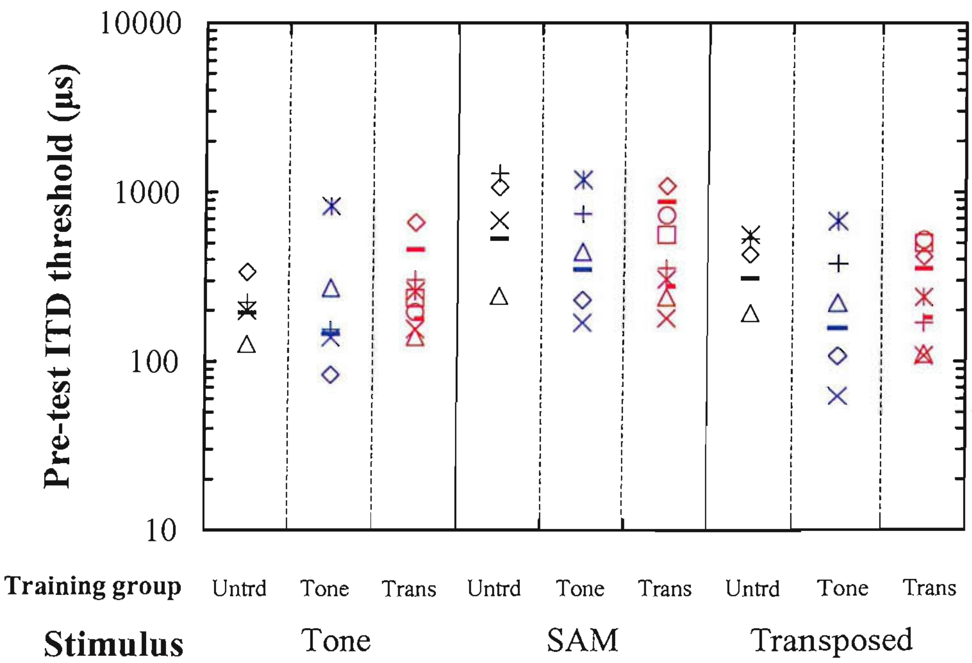


Figure 6.7. Mean ITD threshold for each listener in each training group (untrained: 'untrd'; transposed: 'trans') and with each stimulus.

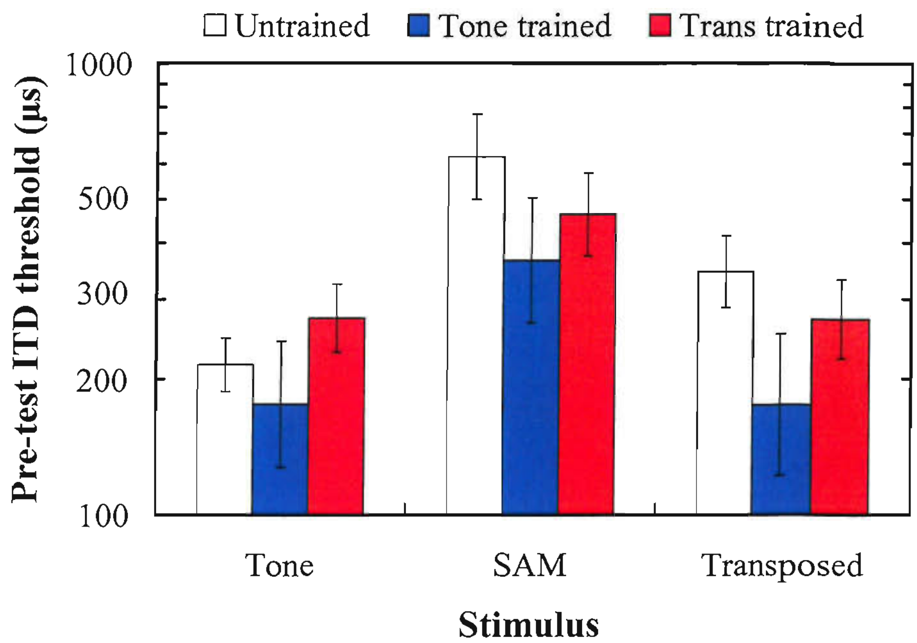


Figure 6.8. Group geometric mean pre–test ITD threshold with each stimulus; error bars indicate ±1 geometric SE.

6.3.2.5 Learning on the trained stimuli

Figure 6.9 provides scatter graphs of pre– and post–test ITD thresholds for each listener in the untrained and trained groups with the trained stimulus. Symbols lying to the right of the diagonal indicate lower thresholds at post– compared to pre–test, consistent with learning. Overall, there is a separation between the two groups, particularly with the unmodulated tone. The untrained listeners generally lie close to the diagonal whereas the trained listeners are generally displaced to the right. However, this was not universally the case. Two transposed–trained listeners (Listeners 15 and 16) had post–test thresholds that were comparable to or slightly worse than those at pre–test. Two untrained listeners (Listeners 1 and 2) appeared to improve between sessions with the transposed tone (and SAM) by an amount comparable to some trained listeners, although not with the unmodulated tone. The overall separation of the groups is clearer in Figure 6.10, which plots the group mean pre– and post– test ITD thresholds; error bars indicate ±1 standard error (SE). While there is no clear improvement between sessions for the untrained group, the trained group improved by approximately 50% (see Table 6.1). There is a slight trend for the thresholds of the untrained group with the unmodulated tone to worsen between sessions, although, as can be seen in Figure 6.9, this is largely due to one listener (Listener 6).

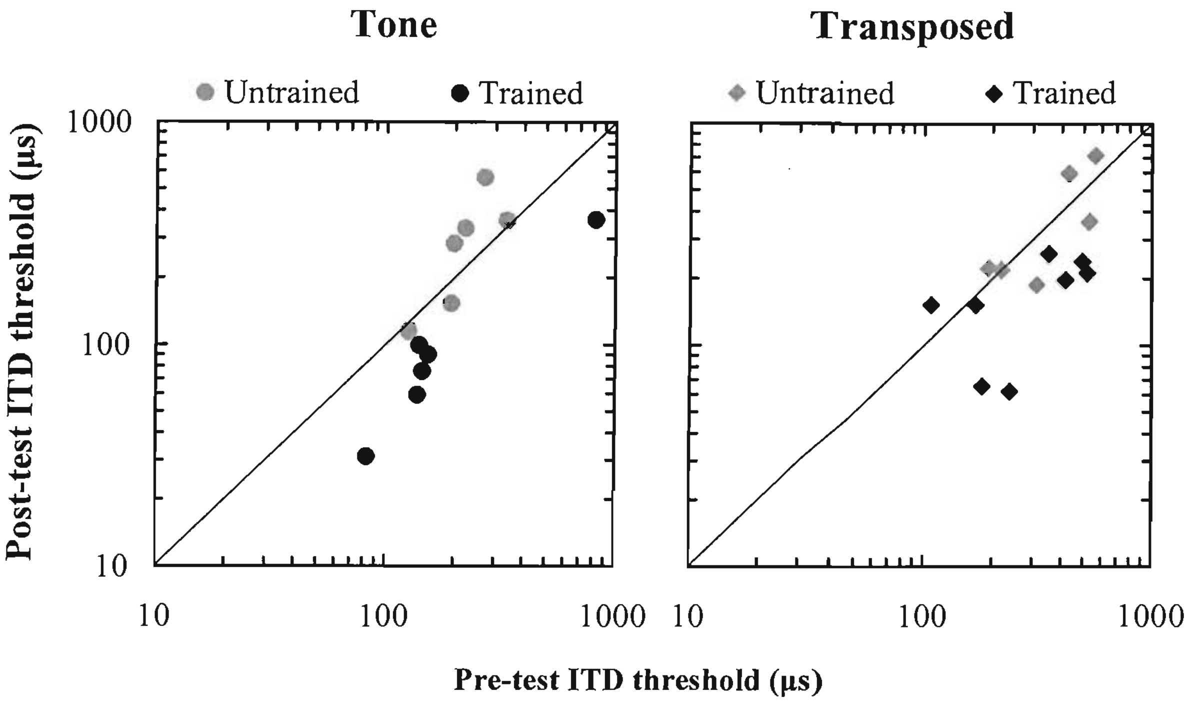


Figure 6.9. Geometric mean pre– and post–test ITD thresholds for each listener in the untrained and trained groups with each trained stimulus.

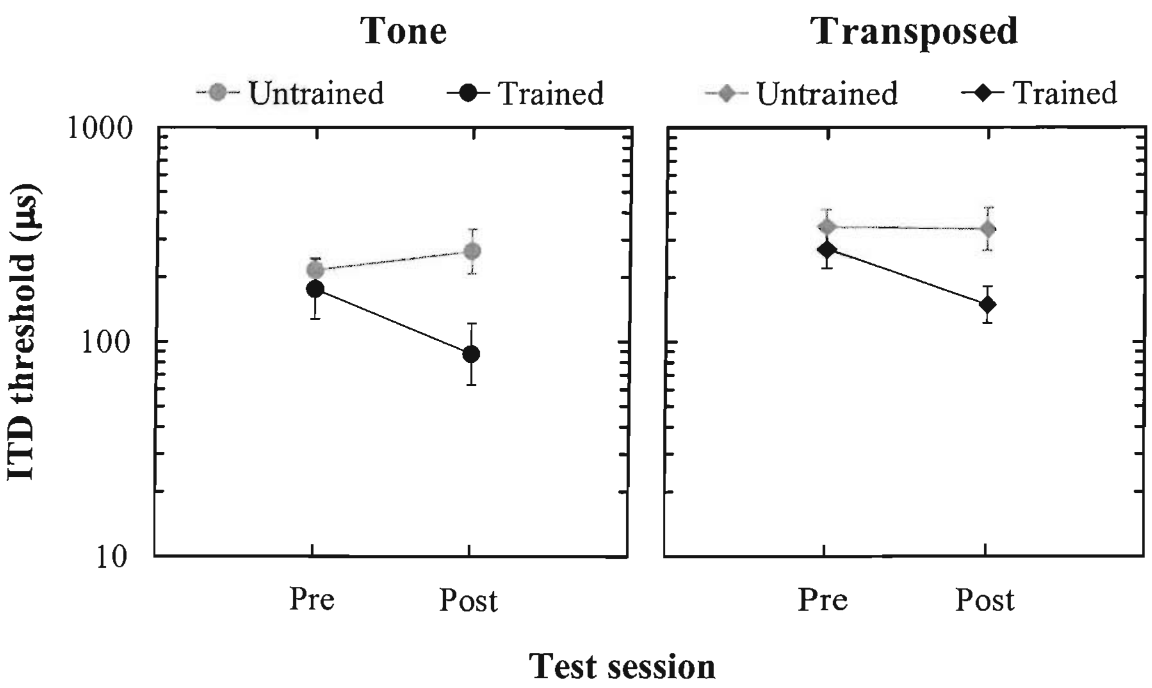


Figure 6.10. Geometric mean pre– and post–test ITD thresholds in the untrained and trained groups with each trained stimulus; error bars indicate ±1 geometric SE.

Overall, these data appear consistent with the experimental hypothesis that the transposed–trained listeners learn more than the untrained listeners. However, in contrast with the experimental hypothesis, and the conclusions by Wright and Fitzgerald (2001), this was also the case with the tone–trained listeners. Also in contrast with Wright and Fitzgerald, the untrained listeners did not appear to learn.

6.3.2.6 Learning on all stimuli

Figure 6.11 plots mean ITD thresholds at pre– (unhashed) and post–test (hashed) with all three stimuli and for each training group; the error bars indicate 1 SE. Overall, the untrained group did not appear to improve between sessions with any stimulus. However, learning was apparent with both the trained groups and the amount of learning was comparable across stimuli, suggesting generalisation of learning from the trained to untrained stimuli.

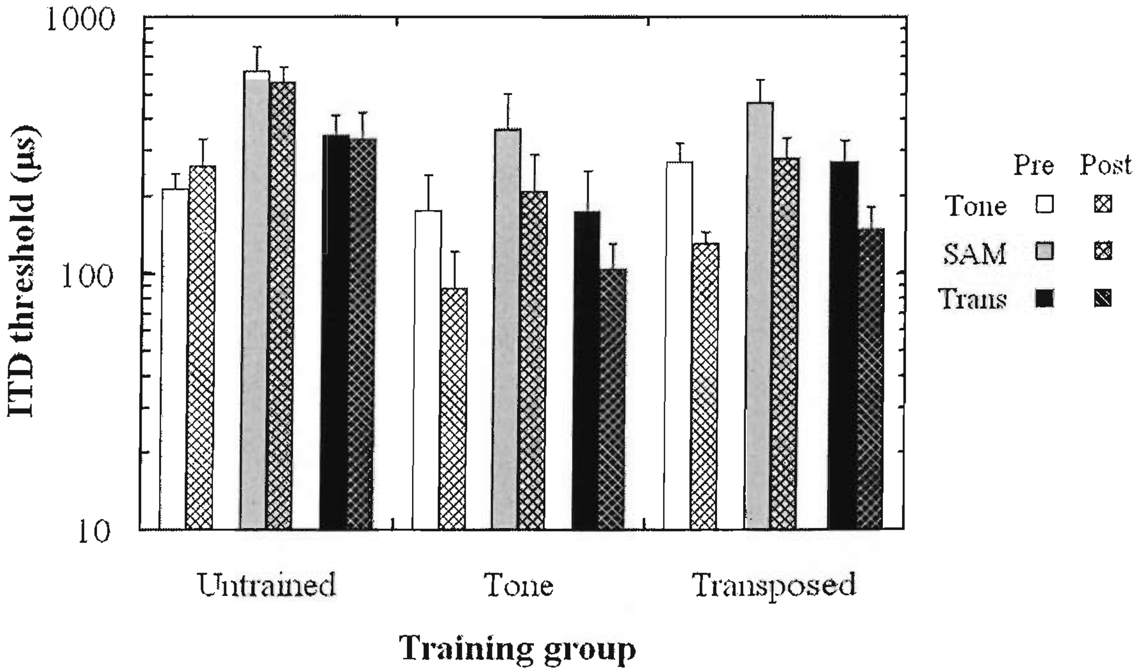


Figure 6.11. Geometric mean ITD thresholds for each training group at pre– and post–test (open and hashed columns) and with the three stimuli (white, grey and black columns); error bars indicate 1 geometric SE.

To examine the generalisation of learning in more detail, Figure 6.12 plots the *improvement* in ITD thresholds with each stimulus for the trained groups. The dependent variable is the mean difference in logarithmically transformed pre– and post–test thresholds, and error bars indicate ±1 SE. The amount of learning observed with each stimulus did not vary between training groups, indicated by the difference in heights of the columns being far smaller than the SE. This combined with the apparent absence of learning in the untrained group indicates wide generalisation of ITD discrimination learning across stimuli.

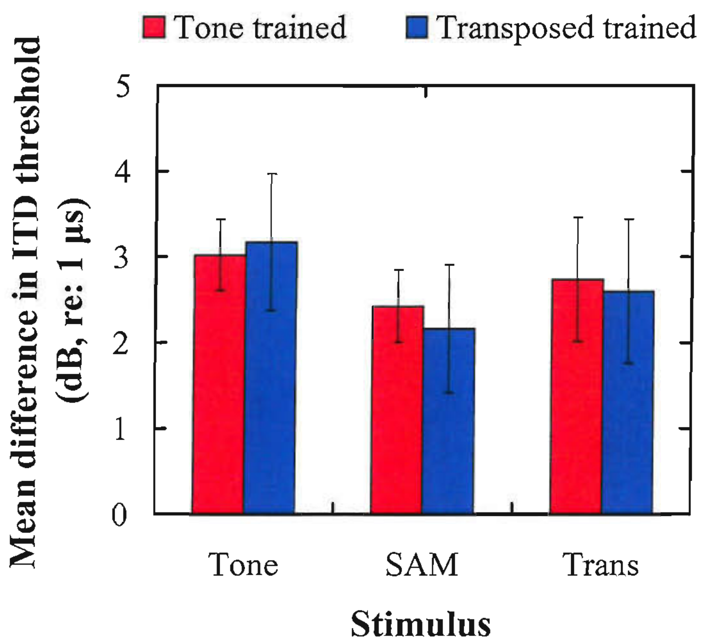


Figure 6.12. Difference in mean logarithmically transformed pre– and post–test ITD thresholds with each stimulus for each training group. Error bars indicate ±1 SE.

6.3.3 Statistical analyses of pre– and post– test ITD thresholds

6.3.3.1 Description of ANOVA

Statistical analyses examined the trends described in Section 6.3.2. The planned ANOVA, described in Section 6.2.4, was conducted on (logarithmically transformed) ITD thresholds, with related factors Block (four blocks), Session (pre– and post–test) and Stimulus (unmodulated, SAM and transposed tones), and unrelated factors Training Group (untrained, tone–trained and transposed–trained) and Counterbalancing Group (three groups). Mauchly's test of sphericity was significant for Stimulus (p=0.009). Levene's test of equality of error variances was significant (p<0.01) for only three of the 24 variables and so homogeneity was assumed. The results are presented in Table F.1 in Appendix F. In summary, the main effects of Stimulus and Session and the two–way interaction between Session and Training Group were highly significant but all other effects were not significant. Appendix F also contains ANOVA tables for *post hoc* tests. The following sections described the results of these analyses in detail.

6.3.3.2 Order effects

Section 6.2.3.2 suggested that there were no obvious order effects with ITD thresholds at pre– and post–test. This was confirmed statistically by the absence of significant effects involving Block and effects involving Counterbalancing Group.

6.3.3.3 ITD thresholds across stimuli and training groups

The significant main effect of Stimulus (p<0.001) indicates that, averaged across all other factors, ITD thresholds differed across stimuli. This effect was consistent across time, both within and between sessions, as indicated by non–significant two–way interactions with Block or Session and the three–way interaction with Block and Session. Interestingly, the trend observed in Section 6.3.2.4 for pre–test thresholds to differ across training groups did not reach statistical significance, as indicated by the non–significant main effect of Training Group, two–way interactions between this and Stimulus or Session and the three–way interaction between these factors.

To investigate the Stimulus effect further, three *post hoc* Pairwise Comparisons were conducted, using the Least Significant Difference adjustment to the degrees of freedom (equivalent to no adjustment). This confirmed the observations that ITD thresholds with the unmodulated and transposed tones were comparable (p=0.50) and more acute than with the SAM tone (p <0.001 in both cases).

6.3.3.4 Learning

The main effect of Session was statistically significant (p=0.001), as was the interaction between Session and Training Group (p=0.007). Learning was therefore evident but differed across groups. The non–significant three–way interaction between Session, Training Group and Stimulus indicates that the pattern of learning in each group was uniform across stimuli. *Post hoc* ANOVAs were conducted on mean ITD thresholds for the untrained group (Table F.2) and trained groups (Table F.3) with related factors Session and Stimulus in both cases and unrelated factor Training Group with the latter. As anticipated, the main effects of Stimulus were highly significant. With the untrained group, effects involving Session were non–significant, indicating that learning was not apparent with any stimulus. In contrast, the effect of Session with the trained groups was significant ($p<0.001$) but the interactions involving Session were not. These indicate that comparable training–induced learning was observed with all stimuli regardless of the trained stimulus.

As described in Section 6.3.2.4, the tone–trained group tended to have more acute pre–test ITD thresholds across all stimuli compared to the transposed–trained group. Although this was not statistically reliable, it could nevertheless have influenced the comparison of learning between the groups if learning was related to pre–test thresholds. Table 6.2 presents the results of correlation analyses, which compared logarithmically transformed pre–test thresholds and pre– minus post–test thresholds. There was neither a statistically significant nor robust (as indicated by r^2) relationship for any one stimulus in *both* training groups. Therefore, it is unlikely that the apparent difference in pre–test thresholds between trained groups influenced the comparison of learning. Section 6.3.2.4 also noted a trend for the untrained and trained groups to differ on pre–test thresholds with the high–frequency stimuli. However, since the untrained group had slightly *poorer* pre–test thresholds, this would have confounded interpretation only if the untrained group had learnt as much as (or more than) the trained group, which was not the case.

Table 6.2. Results of correlation analysis of pre–test and pre– minus post–test thresholds.

Training group	Stimulus	r^2	p
Tone	Tone	<0.01	0.92
	SAM	<0.01	0.85
	Transposed	0.73	0.03
Transposed	Tone	0.75	0.006
	SAM	0.34	0.13
	Transposed	0.26	0.20

6.3.4 ITD thresholds across training sessions

This section describes the performance of the trained groups during the formal training. For convenience, ITD thresholds will be plotted following the logarithmic transformation, as in Figure 6.12. Figure 6.13 plots mean ITD threshold as a function of block number across all eight sessions for each training group. The consistent separation of training groups follows the difference apparent at pre–test. Although the graphs are quite irregular, thresholds appeared to improve more often within a single training session than was apparent during pre– and post–tests.

Figure 6.14 presents the mean ITD threshold as a function of session (including pre– and post–test) for each listener separately within each training group. There is no absolute scale on the *y*–axes and the curves have been separated on this axis for clarity; however, each interval corresponds to 3 dB (i.e. a factor of 2). The percentage of the variance in ITD thresholds accounted by a linear relationship with session number (i.e. time) was calculated and shown in parentheses. In most listeners, there was a robust relationship between ITD thresholds and time, with improvements occurring over a number of training sessions. While most appeared to have approached asymptotic performance by post–test, this not universally the case (Listeners 7, 8 and 17). Some listeners also displayed irregular curves, most notably Listeners 9 despite appearing to learn over Sessions 1–5. Irregular 'learning curves' were also found in Experiment 1 and Wright and Fitzgerald (2001). Learning was not apparent for Listeners 15 and 16. Appendix G, which plots individual learning curves on an absolute scale, shows that there was no clear trend for the time–course of learning to be related to pre–test thresholds.

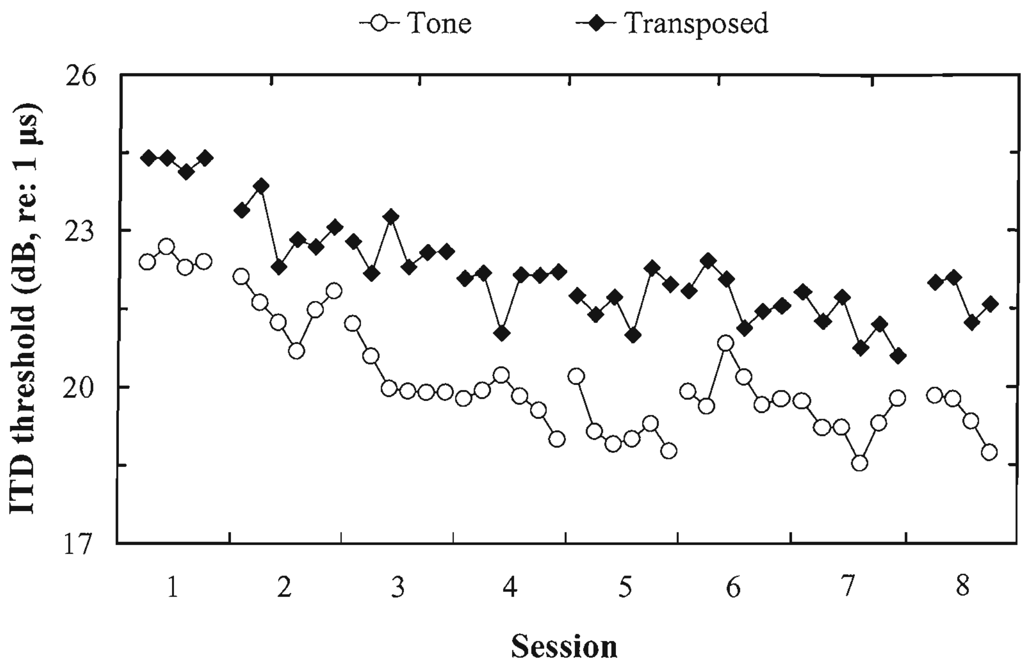


Figure 6.13. Mean ITD threshold for each block across sessions (including pre– and post–test) for the tone– and transposed–trained groups.

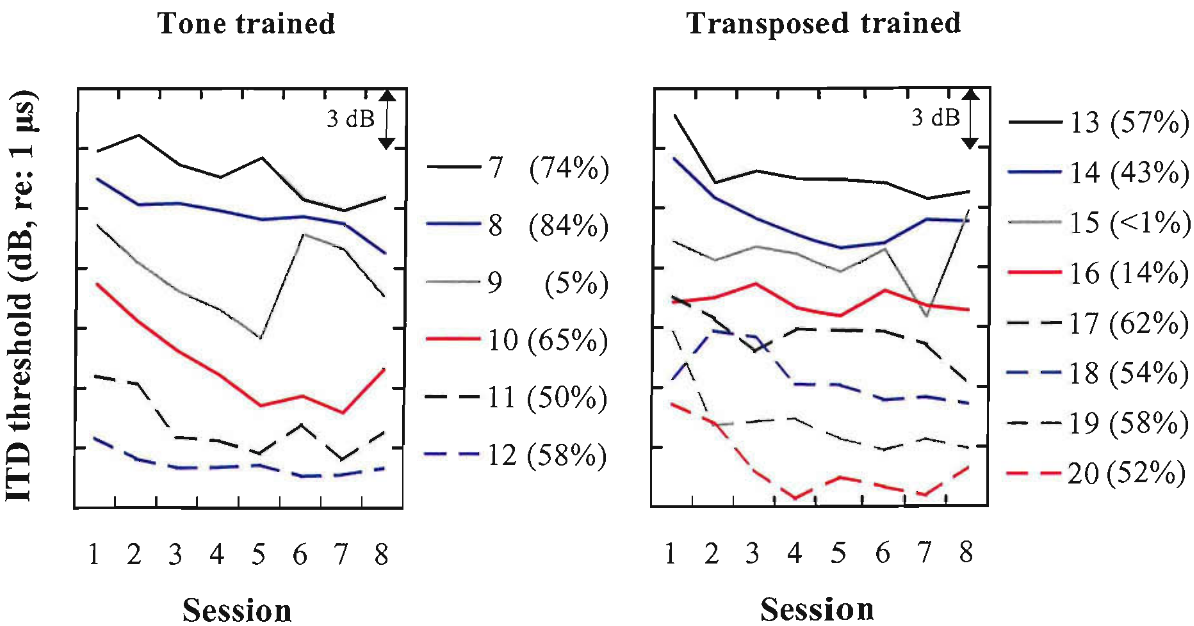


Figure 6.14. Mean ITD thresholds as a function of session (including pre– and post–test). Each line represents a separate, numbered listener. The proportion of the variance accounted by a linear relationship between the two variables is indicated parentheses.

Figure 6.15 plots the mean ITD threshold as a function of session number; error bars indicate ±1 SE. The percentage of the variance in ITD thresholds accounted by a linear relationship with session is again shown in parentheses. The curves of the two groups seem parallel, suggesting comparable time–courses of learning. Interestingly, there is a clear trend for performance in both groups to be more acute at the second session (i.e. first formal training session) than at pre–test.

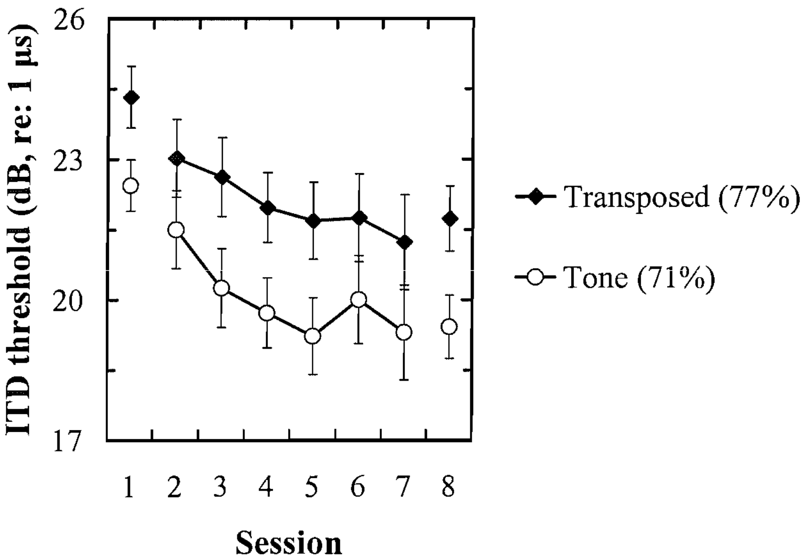


Figure 6.15. Mean ITD threshold as a function of session for both trained groups. Error bars represent ±1 SE.

The planned ANOVA was conducted on the session mean ITD threshold with related factors Block (six tests) and Training Session (six formal training sessions only) and unrelated factor Training Group (tone– and transposed–trained). The results are tabulated in Table F.4 of Appendix F. As with previous analyses, the main effect of Training Group was not statistically significant, despite the trend apparent in Figures 6.13 and 6.15. The main effects involving Block were not significant, indicating that the trends for within–session learning observed in Figure 6.13 were not statistically reliable. However, the main effect of Session was highly significant (p<0.001) confirming the observation that improvements occurred during formal training. However, the two–way interaction between Session and Training Group was not significant (p=0.78) confirming the observation that the time–courses of learning were comparable across groups.

6.3.5 Relationship between learning and pre–test thresholds

Section 6.3.3.4 showed there was no consistent relationship between learning and pre–test thresholds in the logarithmic domain. However, a constant difference in thresholds in the logarithmic domain indicates a constant *ratio* of thresholds in the untransformed data. That is, the difference in thresholds *was* related to pre–test threshold, when expressed in µs. To illustrate this, Figure 6.16 plots the untransformed pre–test threshold and pre– minus post–test thresholds for each trained listener with each stimulus. (Data were not separated by training group since previous analysis failed to identify evidence for stimulus–specificity.) The diagonal line indicates learning corresponding to a halving of pre–test thresholds; overall, the data seem well accounted by this relationship.

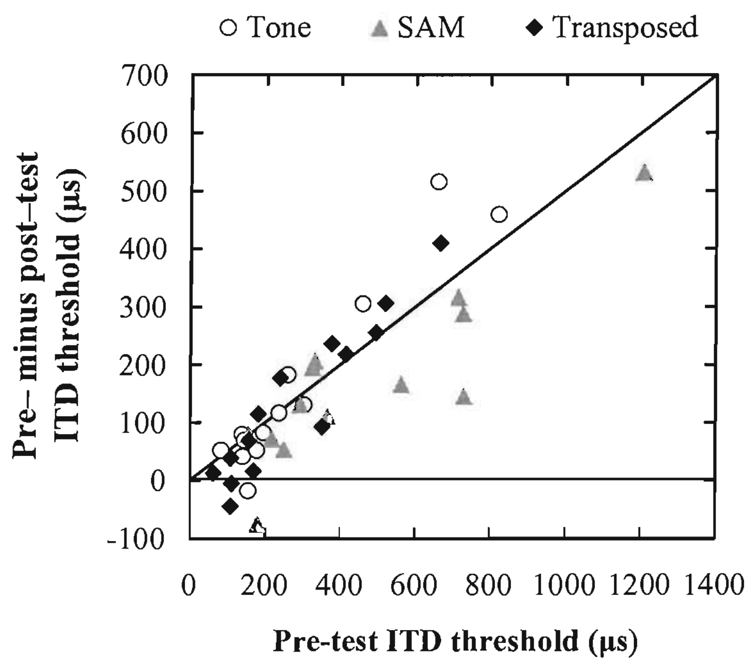


Figure 6.16. Difference in pre– and post–test ITD thresholds as a function of pre–test thresholds for each trained listener separately, and with each stimulus.

Linear regression parameters (and associated SEs) and r^2 of the 'best' fits to the data in Figure 6.16 are provided in Table 6.3. Approximately 90% of the variance in learning with the unmodulated and transposed tones could be accounted for by pre–test thresholds; this was slightly less so for the SAM tone. The slopes were comparable for unmodulated and transposed tones but larger than for the SAM tone, suggesting that learning with SAM was slightly less than with the other stimuli. However, the SE indicates that this is not a statistically reliable difference. Table 6.3 also shows the proportion of the variance accounted by the diagonal line in Figure 6.16 (i.e. the 'alternative' fit). Data from the two listeners who account for the negative values of learning (Listener 12 from the tone–trained and Listener 15 from the transposed–trained group) were excluded from this analysis. The proportion of variance accounted by this relationship is only slightly lower than for the best fit for unmodulated and transposed tones.

Table 6.3. Results of a regression analysis on the difference in pre–test and pre– minus post–test ITD thresholds in the trained listeners.

Stimulus	**'Best'**			**Alternative**		
	Slope (SE)	Intercept (SE)	r^2	Slope	Intercept	r^2
Tone	0.71 (0.06)	–46.5 (21.9)	0.91	0.50	0	0.89
SAM	0.57 (0.09)	–75.8 (53.8)	0.78	0.50	0	0.67
Transposed	0.68 (0.07)	–55.6 (22.8)	0.89	0.50	0	0.85

6.3.6 Individual differences in pre–test ITD thresholds

Figure 6.17 presents a scatter graph of ITD thresholds with the unmodulated and transposed tone for individuals in the trained groups, and at pre– and post–test. In addition, data from the untrained and trained listeners in Experiment 2 and data from trained listeners reported by two previous studies are plotted for comparison. Untrained and trained listeners are indicated by open and filled symbols, respectively. The thresholds of untrained listeners in Experiment 2 were disproportionately worse with the transposed tone, but this was not generally the case in the present experiment, consistent with previous data with trained listeners. However, Figure 6.17 shows that the relationship between stimuli varied considerably across listeners. It was of interest to ascertain if such variation in the present experiment reflected stable inter–individual differences or simply random variation (e.g. measurement error). Figure 6.18 plots the ratio of thresholds with the transposed and unmodulated tones at pre– and post–test for each listener; a value of greater than unity indicates that thresholds were higher with the transposed tone. Stable inter–individual differences would be apparent as data points lying close to the diagonal. While a relationship between values at pre– and post–test is apparent, confirmed statistically (Pearson's r=0.67, p=0.003), only 40% of the variance was shared by pre– and post–test data. This indicates that most of the variation in relative performance with unmodulated or transposed tones across listeners was not attributed to stable inter–individual differences.

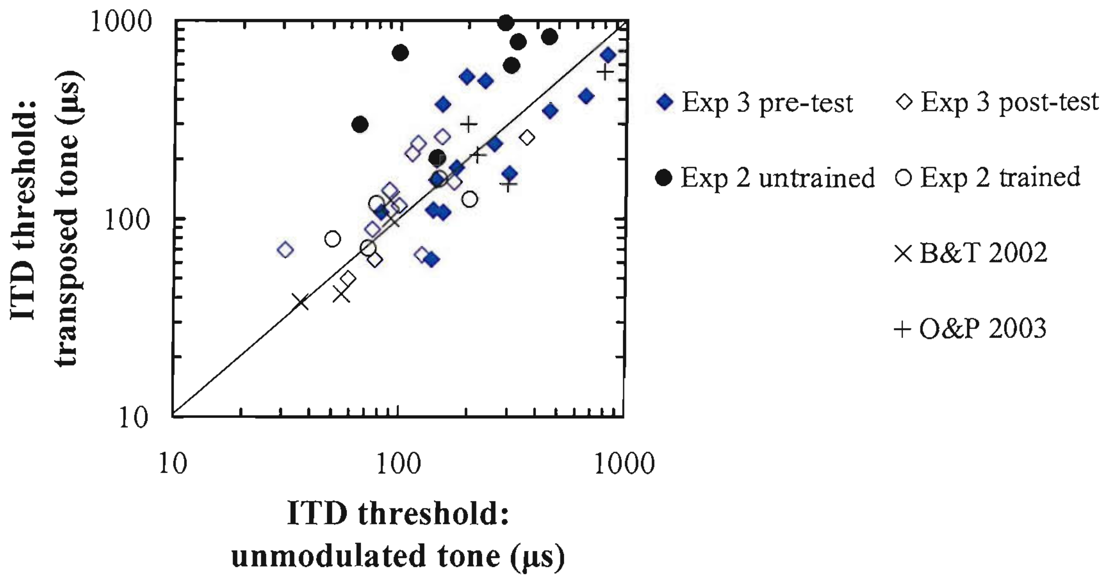


Figure 6.17. Geometric mean pre– and post–test ITD thresholds with the unmodulated and transposed tones for each listener separately. Data from Figure 5.5 are reproduced.

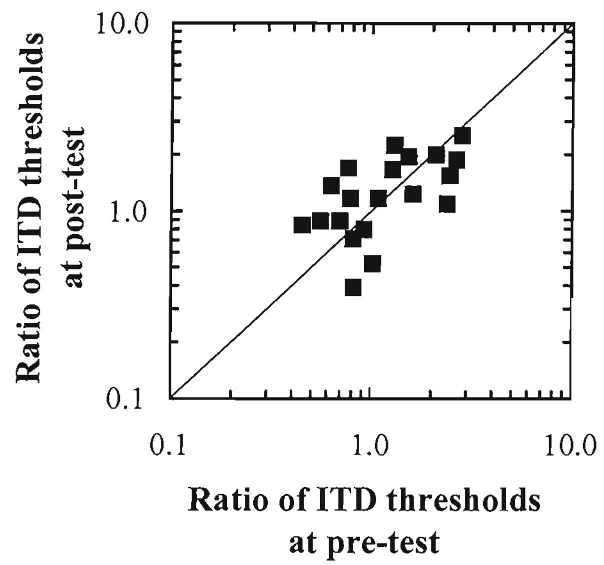


Figure 6.18. Ratio of ITD thresholds with the transposed and unmodulated tones at pre– and post–test for each listener separately.

6.3.7 Performance on different measures of binaural processing

The relative performance of inexperienced listeners with the low–frequency and high–frequency (transposed) stimuli was found to differ across measures of binaural processing. While ITD thresholds and extents of laterality were comparable between stimuli, overall, MLDs were smaller with the transposed stimuli.

Figure 6.19 displays the arithmetic mean data from the binaural signal detection tests. The arithmetic mean absolute thresholds and the difference (i.e. MLD) are plotted for each stimulus; error bars show ±1 SD. Thresholds were comparable in the N_0S_0 condition, confirmed with a two–tailed related–samples *t*–test (t_{19}=0.12, p=0.90). However, thresholds were more acute with the low–frequency stimulus in the N_0S_π condition (t_{19}=4.3, p<0.001) and hence so were MLDs (t_{19}=4.5, p<0.001). A Pearson correlation matrix was computed between absolute thresholds, MLDs and pre–testing ITD thresholds with the two stimuli. The results indicated these measures were unrelated.

The number of responses across all listeners for each rating of lateral position as a function of ITD, stimulus and session is given in Appendix H. The arithmetic mean responses are plotted in Figure 6.20. Left–leading ITDs are indicated by negative values. Ratings were generally close to 1 (i.e. towards the left ear) for the largest left–leading ITD and increased towards 21 (i.e. towards the right ear) with varying ITD, although there was considerable variation across listeners. Overall, a diotic stimulus produced a rating approximately mid–way along the scale. The effect of ITD was comparable between unmodulated and transposed tones but stronger with the broadband noise, as indicated by the slopes of the functions. The stronger effect on lateral position for broadband noise was as anticipated (Durlach and Colburn, 1978). There was no obvious trend for the slopes to differ between pre– and post–tests for either stimulus, thus providing no evidence for learning.

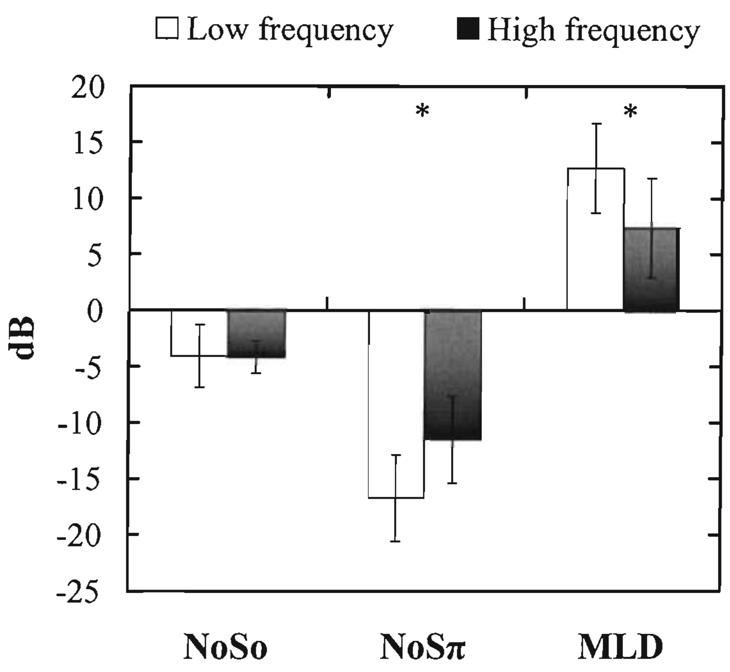


Figure 6.19. Arithmetic mean signal–to–noise ratios at absolute threshold and the masking level difference with the low–frequency and high–frequency (transposed) stimuli. Error bars indicate ±1 SD. Asterisks indicate a statistically significant difference (p<0.001) between stimuli.

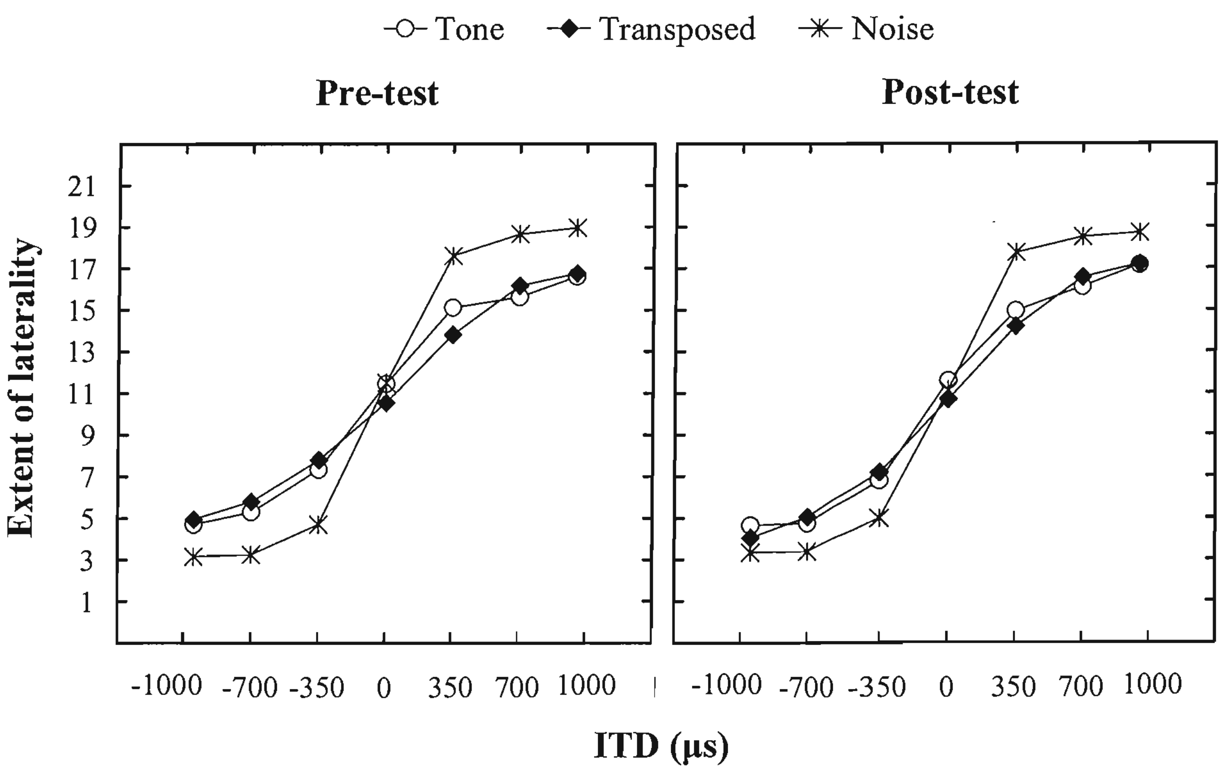


Figure 6.20. Arithmetic mean rating ('extent of laterality') as a function of ITD for each of the three stimuli during pre–and post–test. Appendix H indicates the inter–individual variation in responses.

6.4 Discussion

6.4.1 Overview

The aim of this experiment was to compare directly the time–courses of ITD discrimination learning at low and high frequency. Specifically, based on previous research, it was hypothesised that the time–course of learning was greater at the higher frequency, despite the stimuli being associated with comparable asymptotic ITD thresholds. In contrast, the time–courses were found to be comparable. Learning generally occurred over many hundreds to thousands of trials, with improvements of typically 50% of pre–test performance being obtained. These results are consistent with Experiment 1 which also used high–frequency stimuli but at a lower level. However, these results appear inconsistent with other studies. Firstly, training–induced learning was apparent with low–frequency ITD discrimination, in contrast to Wright and Fitzgerald (2001). This finding seems more closely comparable to Wright and Fitzgerald's conclusions for ILD discrimination. Secondly, thresholds of inexperienced listeners were not disproportionately worse with the high–frequency stimulus, in contrast to Experiment 2. Overall, the relative ITD thresholds between the unmodulated, SAM and transposed tones did not depend on training, confirming Bernstein and Trahiotis (2002). Measurements of extents of laterality also appeared consistent with previous research (Bernstein and Trahiotis, 2003) although MLDs did not (van de Par and Kohlrausch, 1997). The present experiment also found that ITD discrimination learning generalised between stimuli (i.e. across frequency and type of ITD), with no evidence of stimulus–specific learning. This is in apparent contrast with previous research with ILD discrimination (Wright and Fitzgerald, 2001). These findings are discussed in detail in the following sections.

Section 6.4.2 places the pre– and post–test ITD thresholds in context with previous data. Section 6.4.3 then discusses the various measures of binaural processing in inexperienced listeners. The apparent inconsistencies between the present study and Experiment 2 are discussed, followed by discussion of the apparent inconsistencies between different binaural measures. Sections 6.4.4 and 6.4.5 compare learning in so–called untrained and trained groups, respectively, across studies. It is suggested that the most striking difference in results across studies relates to the untrained, as opposed to the trained, groups. It is also suggested that possible limitations in Wright and Fitzgerald's analysis may contribute to the apparent differences in conclusions. This provoked a detailed re–examination of Wright and Fitzgerald's data, which is described in Chapter 7. Sections 6.5.6 and 6.5.7 discuss the time–course and generalisation of learning, respectively, on binaural discrimination tasks at different frequencies. Finally, Sections 6.5.8 and 6.5.9 discuss the implications of these findings for the understanding of binaural processing and for clinical applications, respectively.

6.4.2 Comparison of ITD thresholds with previous research

Figure 6.17 compared ITD thresholds measured in the present experiment to those from Experiment 2 and previous studies. Pre–test thresholds from the present study were generally poorer than those from Bernstein and Trahiotis (2002), although not uniformly so, but comparable to those from Oxenham et al. (2004). This may reflect differences in training across studies. Many listeners had post–test thresholds that were comparable to Bernstein and Trahiotis. However, perhaps the most notable feature of the graph is the wide spread of data. A number of previous studies have reported inter–individual differences on a range of binaural tests (McFadden et al., 1973; Koehnke et al., 1986; Bernstein et al, 1998; Saberi and Antonio, 2003; Saberi et al., 2004). While listeners with the best or worst performance on one test tend to be among the best or worst performers on other tests, it is generally not possible to predict performance on one test accurately from another. The origins of these inter–individual differences remain unclear.

Measurements of pre–test ITD thresholds with the high–frequency stimuli did not use low–pass masking noise. It was possible therefore that these thresholds reflected the use of low–frequency spectral information. However, this seems unlikely for two reasons. Firstly, the transposed–trained listeners generally had thresholds that were more acute during the second session (i.e. first training session) than at pre–test, despite the use of the masking noise. Yet, one may have expected performance to worsen had the noise masked information used during pre–test. Secondly, the relative performance across stimuli was comparable for all three groups and at both pre–test, where the noise was not used, and post–test, where the noise was used. Yet, one may have expected the groups to have had poorer thresholds with the high–frequency relative to the low–frequency stimuli at post–test had the noise masked information used during pre–test.

6.4.3 ITD sensitivity in inexperienced listeners

Overall, ITD thresholds at pre–test were comparable between the unmodulated and transposed tones, and roughly half the thresholds with the SAM tone. This is consistent with Experiment 2a and Bernstein and Trahiotis (2002), who used trained listeners. However, it is inconsistent with Experiment 2b also using inexperienced listeners. There are a number of possible reasons for this, for example the use of the masking noise in Experiment 2 but not the present study, and differences in the content of familiarisation. While the masking noise probably had no material effect on performance due to the masking of low–frequency spectral information, it may have interfered with performance in inexperienced listeners by simply distracting them. Differences in familiarisation, such as greater experience with the task, adaptive procedure and judging lateral position in the present experiment, may have had a stronger influence on performance with the transposed tone. This may be related to the novelty, and reported peculiarity, of this stimulus to some listeners. It is also possible that the listeners in Experiment 2b were extreme examples of listeners found in the present experiment, who had poorer thresholds with the transposed tone.

The overall results from lateral–position scaling appeared consistent with Bernstein and Trahiotis (2003), using a different task. However, interpretation of these data is complicated because the sensitivity of the measure to differences in lateral–position across stimuli is poorly understood. While a steeper function relating extent of laterality to ITD was apparent for the broadband noise, which produces more extreme changes in laterality, it is unclear if the procedure would be sensitive to differences between transposed and SAM tones. The MLD with the high–frequency transposed stimulus was smaller than with the low–frequency stimuli, inconsistent with van der Par and Kohlrausch (1997). This is somewhat surprising since their listeners also had no prior experience with binaural masking and did not appear to have received extensive training. The explanation for this disparity may be similar to that suggested above for the findings of Experiment 2b with ITD discrimination.

6.4.4 ITD discrimination learning in 'untrained' groups

Table 6.4 lists studies of auditory discrimination that have included an untrained group. The second column indicates whether learning was observed in this group and the third column indicates the parameter discriminated. It can be seen that while learning in the untrained group was apparent in some studies, this was not universally the case. Constantinides (2004) found learning in the untrained group tested on ILD discrimination, consistent with Wright and Fitzgerald (2001), but not ITD discrimination, consistent with the present study. Inconsistent findings are also apparent across studies of frequency discrimination. Interestingly, Wright and Fitzgerald (2003) found learning in a level discrimination experiment, no learning in a duration discrimination experiment and learning with only one of three stimuli in a frequency discrimination experiment. The untrained group in the present study displayed a trend for performance to slightly worsen with one stimulus, albeit mostly resulting from one listener, and Grimault et al. (2002) found a similar trend across all stimuli.

Table 6.4. A comparison of studies of auditory discrimination learning that included an untrained group (untrained: untrd; familiarisation: famil; stimulus conditions: condx). See text for details.

Study	Learning in untrd	Parameter	Famil	Feed–back	Condx	Total trials
Present	N	ITD	Y	N	3	*c.* 660
Grimault et al. 2002	N	Frequency	Y	N	6	*c.* 1800
Wright and Fitzgerald 2001	Y	ITD	Y	Y	5	1500
	Y	ILD	Y	Y	5	1500
Ari–Even Roth et al. 2003	Y	Frequency	N	Y	3	*c.* 210
Constantinides 2004 (Chap 3, Exp 1)	N	ITD	Y	Y	1	240
	Y	ILD	Y	Y	1	240
Wright and Fitzgerald 2003	N	Duration	?	Y	4	1200
	Y	Level	?	Y	5	1500
	Y and N	Frequency	?	Y	3	900

The remaining columns in Table 6.4 indicate aspects of the methodology that may be associated with learning in the untrained group. In summary, there is no clear relationship between learning in the untrained group and the aspects considered. For example, feedback was used in both studies with inconsistent findings across untrained groups. One might also imagine that the provision and nature of familiarisation may be important. For example, both the present study and Wright and Fitzgerald (2001) provided familiarisation using lateral–position scaling tasks, but the present study additionally provided familiarisation on binaural signal detection. Studies of frequency discrimination reporting learning (Ari–Evan Roth et al., 2003) or no learning (Grimault et al., 2002) also differed in terms of familiarisation. However, Wright and Fitzgerald (2003) probably employed a common general approach to familiarisation across experiments and yet found differences in learning with the untrained groups. Constantinides (2004) also used a common approach to familiarisation with both ITD– and ILD–untrained groups, but again learning was found in one group but not the other.

The order of testing during pre– and post–test may also be important. In the present study, two blocks were completed with one stimulus before proceeding to the next, each stimulus being revisited in reverse order. In Wright and Fitzgerald's (2001) study, all (five) blocks were completed with one stimulus before proceeding to the next. One could imagine that testing for longer on a single stimulus might facilitate greater learning. However, it is again difficult to account for the differences apparent in Wright and Fitzgerald (2003) with this factor alone. One difficultly in comparing across studies is that they generally do not report analyses of learning *within* test sessions. Learning within pre– or post–test is perhaps more likely to occur when testing with one stimulus is completed before proceeding to the next and when there is a relatively large number of stimulus conditions, such as in Wright and Fitzgerald's studies. However, this is difficult to ascertain from these reports.

6.4.5 ITD discrimination learning in trained groups

Attention is now turned to a comparison of the trained groups. All three studies of ITD discrimination learning concluded that the trained groups did indeed learn. However, a detailed comparison of the studies is hampered by methodological differences. This is particularly so for Constantinides (2004), since the number of trials presented during training was less than other studies presented during pre–test.

The present study also provided fewer trials during each training session and overall compared to Wright and Fitzgerald (2001). Other potentially important differences include the trial formats and the adaptive procedures. For example, it has been suggested that the oddity paradigm (Colburn and Trahiotis, 1991) and modified adaptive procedure (Trahiotis et al., 1990), as in the present study, require fewer trials to reach asymptotic performance than the standard 2AFC paradigm and the adaptive rule, as in Wright and Fitzgerald, at least with some stimuli. There is currently no evidence that this is the case in the conditions used in these studies.

Despite differences in methodology, the improvements in the trained groups on the trained stimulus between pre– and post–test were comparable across the studies; all found post–test ITD thresholds to be approximately half the pre–test thresholds. This was also the case for Constantinides's and Wright and Fitzgerald's ILD–trained groups. Comparing the mean learning curves, it is apparent that improvements occurred throughout training. However, it is interesting that Constantinides found a similar amount of learning despite providing far fewer trials than the other studies, and using the same trial format and adaptive procedure as Wright and Fitzgerald. Comparing the average learning curves between the present study and Wright and Fitzgerald's ITD and ILD experiments also suggests less disparity than implied in the conclusions. However, Wright and Fitzgerald argued that the average curve with ITD discrimination misrepresented the individual data. According to their analysis, only two of the eight ITD–trained listeners actually displayed learning during training (L9 and L10). Yet, as described in Section 2.5.2, visual inspection of the learning curves indicates that the thresholds of two additional listeners improved over multiple sessions (L11 and L12). Other listeners had highly irregular curves, which may be difficult to interpret. Nevertheless, two of these (L13 and L14) had their lowest thresholds at post–test, suggestive of training–induced learning. The individual curves were also more irregular than those of the ILD–trained group, the reasons for which are not clear.

The absence of statistical evidence for learning with some listeners may also be related to the particular analyses conducted. Wright and Fitzgerald fitted linear regression lines to the curves and learning was accepted if a statistically significant negative slope was found. However, learning curves were highly non–linear, showing a reduction in slope over time. This analysis differs from the present study in that the data were not subjected to a logarithmic transformation, which tends to make learning curves straighter. The approach used by Wright and Fitzgerald may therefore be relative insensitive to learning during training.

In summary, the data from the present study and Wright and Fitzgerald (2001) seem to differ more in terms of the untrained than trained groups, complicated by complex learning curves from Wright and Fitzgerald and differences in analysis. It is important to consider if these accounted for the present study but not Wright and Fitzgerald reporting greater learning in the ITD–trained compared to untrained group. This is further complicated by limitations in Wright and Fitzgerald's analysis of learning in the two groups. As described in Section 2.5.2, Wright and Fitzgerald did not compare learning in the two groups directly, despite a clear trend for the trained group to learn more than the untrained group. This limitation is addressed in Chapter 7.

6.4.6 Learning on binaural discrimination

The primary hypothesis of this experiment was that the magnitude and time–course of ITD discrimination learning was greater at high compared to low frequencies, given stimuli that produce comparable asymptotic thresholds. This was investigated by comparing the time–courses of learning in two groups of listeners, each group trained with a different stimulus. The degree to which the time–courses measured in each group reflect the different stimuli as opposed to the different listeners depends on how comparable the groups are in terms of other important factors that might influence learning. One such factor is pre–test performance, and the two groups did appear to differ in this regard, albeit not statistically reliably so. However, this factor was only found to be consistently related to learning before, but not after, the logarithmic transformation, and so did not appear to influence the analysis across groups. Nonetheless, future studies of learning may need to adopt a different approach to assigning listeners to each group, to further guard against this.

The time–courses of learning in the present experiment were broadly consistent with the results of Experiment 1 using lower level stimuli, although subtle differences may exist. The magnitude of learning was generally greater in Experiment 1 when expressed in microseconds, but broadly comparable when expressed in terms of a ratio. The time–courses in the present study, as in Experiment 1, seemed to vary across listeners, although did not obviously appear to depend on pre–test performance. However, it is unclear if the time–course depends on the stimulus used. For example, the amount of learning with the SAM tone was slightly less than with the other stimuli. While this was not statistically reliably so, it may reflect a longer time–course of learning. It is not clear why the time–course would vary across stimuli given the same frequency and modulation rate, although the associated asymptotic thresholds, which differed, may be important.

The comparison of time–courses of ITD discrimination learning at different frequencies was motivated by Wright and Fitzgerald's (2001) conclusion that the time–course of learning differed for ITD and ILD discrimination at different frequencies. While the time–courses of ITD discrimination learning in the present study were comparable across frequency, they were also broadly consistent with Wright and Fitzgerald's data on ILD discrimination. As described in the previous section, there are some uncertainties with their interpretation regarding low–frequency ITD discrimination. Further, it was noted in Section 2.5.2 that pre–test thresholds from Wright and Fitzgerald's untrained and trained groups in the ITD and ILD experiments might not be representative of most inexperienced listeners. Specifically, listeners in the ITD experiment had relatively good and listeners in the ILD experiment had relatively poor thresholds on the 'trained' stimuli, compared to previous studies (e.g. Bernstein et al., 1998). It is unclear if these findings reflect the particular test conditions or listeners, since Wright and Fitzgerald do not indicate the pre–test ITD thresholds on the stimuli common to both experiments.

Although this thesis has concentrated on ongoing ITD, it is interesting to consider the time–course of learning with onset ITD. For example, it has been suggested that onset ITD may not be processed in the same way as either ongoing ITD or ILD (Batra and Fitzpatrick, 2002). However, there are at least two difficulties with comparisons between onset and ongoing ITD. Firstly, ITD thresholds with abrupt stimuli, such as clicks, are generally much poorer than with sustained stimuli (Durlach and Colburn, 1978), and this may influence learning. Secondly, differentiating between onset and ongoing ITD is less straightforward physiologically than acoustically. This is because a single click may produce ongoing information in the auditory nerve due to 'ringing' on the basilar membrane (Palmer, 1995). Some experiments using pairs of clicks, under conditions of the so–called precedence effect, have demonstrated learning over much longer time–courses than found in the present study. It is unclear if this reflects differences in processing under these conditions (cf. Litovsky et al., 1999, and Hartung and Trahiotis, 2001) or inherently poorer asymptotic thresholds with the stimuli, coupled perhaps with inter–individual differences in learning.

6.4.7 Generalisation

Learning in the present study was found to generalise widely across frequency (5 octaves) and type of cue (fine–structure–based and envelope–based cue). This is in apparent contrast to ILD discrimination learning, which Wright and Fitzgerald (2001) reported to generalise incompletely across a small frequency range (0.6 octaves). One possible reason for not observing stimulus specificity is that the present study provided insufficient training. Stimulus–specific learning has generally been found to be a small component of learning and is generally assumed to occur during the latter stages. However, as described in Section 2.5.2, there are also limitations with Wright and Fitzgerald's approach to studying stimulus–specificity, which complicates comparison. Specifically, their conclusions rely on comparisons with an untrained group rather than comparison across groups trained on the various stimuli (such as in Irvine et al., 2000). This approach also requires assumptions regarding the inherent potential learning with the untrained stimuli. As discussed in Section 6.4.4, the interpretation of learning in the untrained group may also not be straightforward. It is therefore not possible to draw firm conclusions regarding these potential differences in the pattern of generalisation across frequency between ITD and ILD discrimination.

The present study indicates that the mechanisms involved in the observed learning were not dependent on frequency region. One possibility is that listeners simply learned general skills that could have been obtained with training with an alternative discrimination task, such as frequency discrimination. However, the results from Hawkey et al. (2004) suggest this is not entirely the case. They found that a large component of the learning that occurred over the first 500 trials on frequency discrimination could not be acquired from training on alternative auditory or visual tasks. It seems likely that the broad findings of this research apply to ITD, and ILD, discrimination. However, the details of what is learned remain unclear. For example, it is of interest to determine if listeners learn to attend to specifically ITD–related cues or cues that are available for both ITD and ILD. For example, some models of ITD processing assume that ITD is detected by a decorrelation of the representation of the signals (Bernstein and Trahiotis, 2002). If so, the full extent of the learning on ITD may not occur with training on ILD discrimination. However, other models assume that both ITD and ILD are detected by a change in lateral position (Stern and Trahiotis, 1995), in which case, learning on ILD discrimination would be expected to generalise to ITD discrimination. While Wright and Fitzgerald (2001) found that the learning on ILD discrimination did not lead to greater improvements on ITD discrimination than in the untrained group, this is difficult to interpret since they also argued that listeners trained on ITD discrimination did not learn more than untrained group.

6.4.8 Implications for binaural processing

This experiment has added to the data on the relative ITD sensitivity with low–frequency and high–frequency stimuli in trained listeners, confirming the findings from Bernstein and Trahiotis (2002, 2003) and Oxenham et al. (2004) in a larger sample. These findings are extended to inexperienced listeners, and show that the enhancement of ITD sensitivity with transposed stimuli is not dependent on experience. The comparable time–courses of learning at the different frequency regions is also consistent with the notion that ITD processing is functionally uniform at low and high frequencies, as suggested by Colburn and Esquissaud (1976). However, these data do not imply a particular processing mechanism, neither do they rule out the possibility that there are important neurophysiological differences in ITD processing at different frequencies. Further research into the particular features of the stimuli that were learned is required in order to determine the implications of the learning *per se* and the generalisation of learning for binaural models. For example, while the model of Breebaart et al. (2001) apparently captures an initial, extremely rapid phase of learning, this seems to relate to attention and decision–making rather than binaural processing. If the learning on binaural discrimination is found to be specific to the cue and apparent difference in pattern of generalisation between ITD and ILD discrimination is confirmed, this may have implications for binaural processing. These findings would seem difficult to account for in terms of simple attention or decision–making processes.

6.4.9 Clinical and applied relevance

This experiment adds to evidence indicating that studies of binaural discrimination in clinical populations should consider training effects. While the time–course of learning in this experiment was in the order of hundreds to thousands of trials, numerous factors might influence this. This may include elements of the methodology, such as the oddity paradigm and the adaptive procedure (Section 6.4.6). Interestingly, Koehnke et al. (1995) also selected an oddity task for an experiment involving binaural sensitivity in hearing–impaired listeners, following experience of larger training effects with other tasks. However, few experiments using BICI users have used an oddity paradigm. Constantinides (2004) found that presenting stimuli at and close to threshold, using an adaptive procedure, is more effective in eliciting learning than presenting the same number of trials with a stimulus far above threshold. Constantinides also found that completing training in one day is as effective as distributing training over many days. Other factors may influence the time–course of learning in clinical studies. For example, the asymptotic performance of the stimulus of interest may be important, as discussed in Section 6.4.6. Further, studies using young adults from a University population may not reflect most hearing–impaired people (Moore et al., 2003; van der Elst et al., 2005). Hearing impaired people may also be more motivated to learn due to the functional limitations associated with the loss. This may be at least partially responsible for apparent differences in the results of studies of learning with intensity discrimination (Wright and Fitzgerald, 2003, cf. Robinson and Gatehouse, 1995, 1996). Neurophysiological evidence, such as relating to the function of the nucleus basalis, supports the importance of the salience of stimulus being important in learning (e.g. Gilbert, 1998; Bakin and Weinberger, 1996).

One motivation for studies of discrimination learning is the possibility that they may lead to the development of therapeutic tools. Three findings from the present study are encouraging in this regard. Firstly, multi–hour training leads to improvements in performance; this suggests that the learning reflects more than experience with the task and procedures. Secondly, the time–course of ITD discrimination learning is not inherently frequency dependent, which suggests that the mechanisms of learning are also not frequency dependent. Thirdly, learning generalises across frequency. As pointed out by Moore et al. (2003), the usefulness of training as a therapeutic tool depends on the degree to which the learning generalises to untrained stimuli. However, questions remain to be answered before training on binaural discrimination can be used therapeutically. For example, as discussed in previous sections, more information is required on what listeners learn and the conditions that are required for learning. Another important question is the impact of binaural discrimination learning on global spatial abilities. Emerging evidence suggests that learning on auditory training using relative simple stimuli may generalise to more global tasks, and can be used to treat language–based disorders (e.g. Merzenich et al., 1996; Moore et al., 2005). However, little is known regarding binaural and spatial perception, and an initial experiment by Constantinides (2004) found that training on ITD or ILD discrimination did not influence one measure of sound–source localisation ability in children and adults.

6.4.10 Conclusions

In summary, the conclusions of this experiment were:

- High–frequency transposed tones enhance ITD thresholds over SAM tones and produce comparable thresholds with low–frequency unmodulated tones in both experienced *and* inexperienced listeners.
- Training–induced learning is apparent with ITD discrimination at low–frequency, in contrast to the conclusions of Wright and Fitzgerald (2001).
- Training–induced learning is also evident with ITD discrimination at high–frequency.
- The time–course of learning with ITD discrimination is not inherently dependent on frequency.
- ITD discrimination learning generalises across frequency.

- The magnitude of ITD discrimination learning is related to initial threshold.
- The findings of comparable overall sensitivity and time–courses of learning and the generalisation of learning is consistent with the notion that a common mechanism is responsible for ITD processing across frequency.
- The findings have implications for the design of future studies of ITD discrimination in clinical populations.

Chapter 7. Examination of data from Wright and Fitzgerald

7.1 Introduction

Two studies have arrived at different conclusions regarding the effect of training over thousands of trials on ongoing ITD discrimination using low–frequency tones. Wright and Fitzgerald (2001) suggested that, overall, learning occurred rapidly and was not influenced by training. In contrast, Experiment 3 suggested that learning generally occurred over a longer time–course and was dependent on training. These conclusions were partly dependent on comparisons of ITD thresholds before and after training from trained and untrained (control) groups. In Experiment 3, learning was directly compared between trained and untrained groups, and a statistically reliable difference was found. In contrast, Wright and Fitzgerald only compared the groups before and then after training, and no statistically reliable effects were found. However, the trained group had slightly poorer thresholds than the untrained group before training, but better thresholds after training; that is, the trained group appeared learn more than the untrained group. Further, the difference between groups before training was only marginally non–significant at the 5% level (p=0.078). In the discussion of Experiment 3, it was noted that the ratio of thresholds before and after training appeared broadly comparable between the groups trained on low– and high–frequency ITD in Experiment 3 and the groups trained on low–frequency ITD and high–frequency ILD in Wright and Fitzgerald (2001). The differences between untrained groups were more striking. It is possible that the differences in analysis and data from the untrained groups between studies may at least partly account for the different conclusions.

The conclusions of the two studies were also dependent on comparisons of thresholds in the trained groups across training sessions. While listeners in Experiment 3 generally appeared to improve over multiple sessions, statistical evidence of this was apparent in only two of the eight trained listeners in Wright and

Fitzgerald's study. However, as noted in the discussion of Experiment 3, the limited statistical evidence for learning during training in this study may have been related to the particular analysis conducted. In addition, some listeners had highly irregular 'learning curves', which may be difficult to interpret.

Wright and Fitzgerald (2001) also presented data from an experiment on ILD discrimination learning, and concluded that learning was dependent on training and generalised only partially across frequency, in apparent contrast with ITD discrimination. However, Wright and Fitzgerald noted that the ILD thresholds of listeners in the ILD experiment were generally poor compared to previous studies (e.g. Bernstein et al., 1998), whereas the ITD thresholds of listeners in the ITD experiment were generally more acute in comparison (also see Hinton et al., 2004). If listeners in the ITD experiment had generally more acute binaural hearing than those in the ILD experiment, they may have had less to gain from training. This is difficult to judge, since Wright and Fitzgerald did not report pre–test thresholds from conditions common to both experiments. On the other hand, as described above, the amount of learning in the trained groups appeared to be comparable across ITD and ILD experiments.

This chapter describes a re–examination of Wright and Fitzgerald's data. The principal aims were to compare the data to Experiment 3, such as to examine differences in the untrained groups, and to address the possible limitations in Wright and Fitzgerald's analysis of thresholds before and after training[19]. The implications of this re–analysis for the interpretation of generalisation were also explored.

7.2 Methods

Wright and Fitzgerald (2001) conducted two experiments. In the 'ITD Experiment', eight listeners were trained on ITD discrimination with a 500–Hz tone (the 'ITD–trained' group) and five served in an untrained group. In the 'ILD Experiment', eight listeners were trained on ILD discrimination with a 4000–Hz tone (the 'ILD–trained' group) and eleven served in an untrained group. Both groups were tested

[19]Rowan D, Lutman ME. Learning on binaural discrimination tasks in humans. Oral presentation at the British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness; 2005 Sept 12–13; Cardiff, Wales.

with the trained condition at pre– and post–test, as well as in four other conditions. Of the five conditions used within each experiment, three were common to both: ITD discrimination at 500 Hz, and ILD discrimination at both 500 Hz and 4000 Hz. In addition, the ITD Experiment involved testing on ITD discrimination at 1000 Hz and at 500 Hz with a baseline, or 'standard', ITD of 100 μs or 150 μs. The ILD Experiment involved testing on ILD discrimination at 6000 Hz and at 4000 Hz with a standard ILD of 6 dB. Five 60–trial blocks were completed with each stimulus at pre– and post–test. Training was conducted over a nominal interval of 14 days, within which time listeners completed 6480–7200 trials. Thresholds were measured using a 2AFC task and a 2–down, 1–up adaptive procedure (Levitt, 1971).

Dr. Wright kindly provided the arithmetic means of the five blocks from pre– and post–test for all listeners. The principal planned statistical analysis was an ANOVA on ITD thresholds on the trained stimulus with related factor Session (pre– and post–tests) and unrelated factor Training Group, and the same but with ILD thresholds on the trained stimulus. Following Wright and Fitzgerald (2001), it was envisaged that the data from the untrained groups would be pooled across common conditions. The criterion probability was set at 0.05.

7.3 Results

Section 7.3.1 describes important trends in the pre–test thresholds across trained and untrained stimuli. Sections 7.3.2, 7.3.3 and 7.3.4 consider learning on the trained stimuli. Section 7.3.2 compares the learning observed in untrained and trained groups and with the findings of Experiment 3. Section 7.3.3 explores the trends in Wright and Fitzgerald's data and Section 7.3.4 describes statistical analyses of these. Finally, Section 7.3.5 investigates generalisation of learning across trained and untrained stimuli.

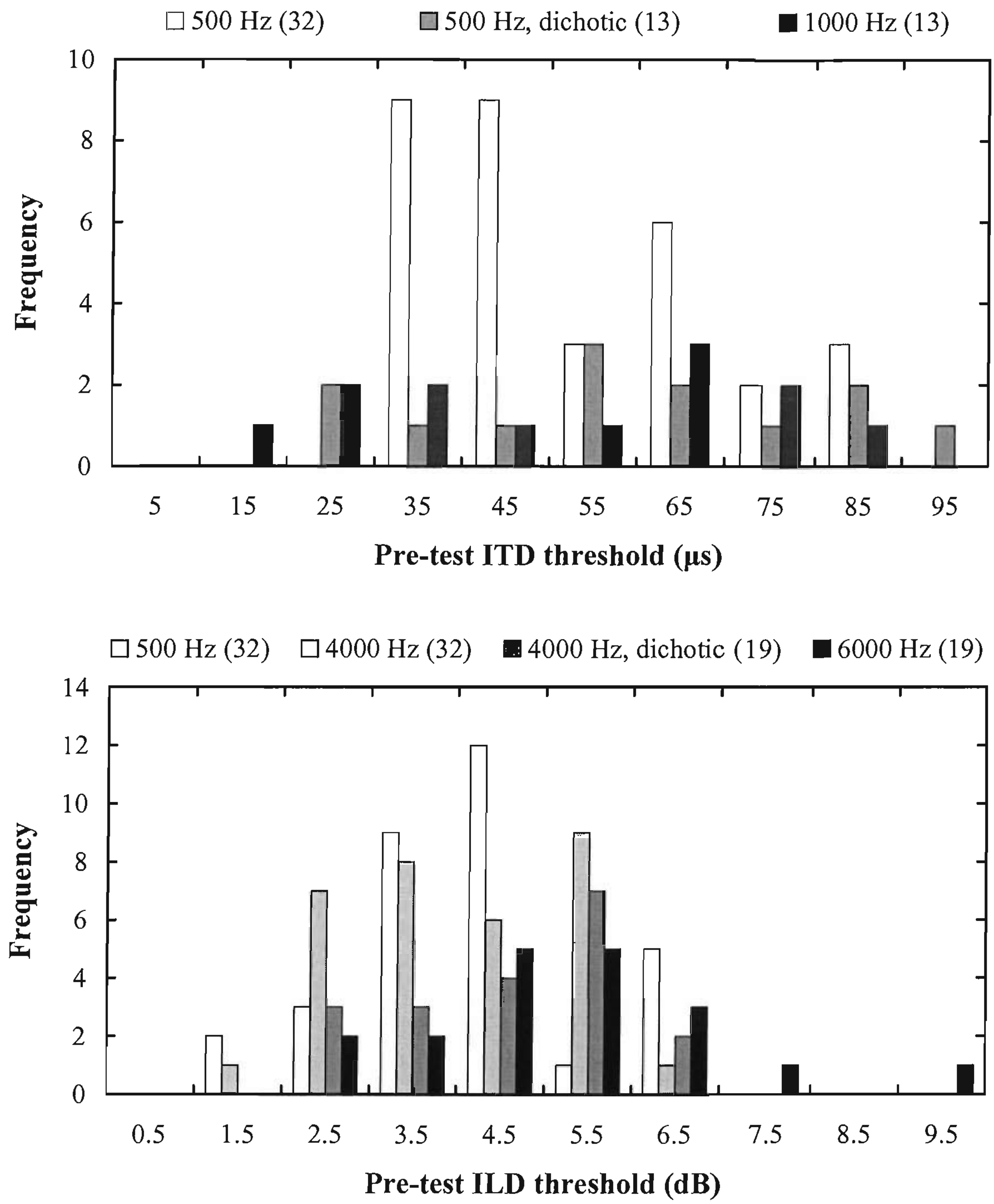


Figure 7.1. Histograms of pre–test ITD (upper panel) and ILD (lower panel) thresholds for all listeners with each stimulus. Stimuli were diotic unless indicated otherwise. The centre values of the bins are given on the abscissa. The sample size with each stimulus is indicated in parentheses in the legend.

7.3.1 Pre–test thresholds across stimuli

The distributions of pre–test ITD and ILD thresholds for all stimuli are shown in Figure 7.1. The ITD thresholds with the diotic 500–Hz tone appeared slightly positively skewed, broadly consistent with Experiment 3 and Hinton et al. (2004), although this was of marginal statistical significance (e.g. Kolmogorov–Smirnov normality test: p=0.09; Shapiro–Wilk normality test: p=0.03). The positive skew was reduced following a logarithmic transform (e.g. Kolmogorov–Smirnov normality test: p>0.20; Shapiro–Wilk normality test: p=0.05). However, the thresholds in the other ITD conditions (and in all ILD conditions) were at least approximately normally distributed, but became negatively skewed following the logarithmic transform. For this reason, together with the results of a comparison of variances across conditions described in Section 7.3.4, the logarithmic transform was not applied to ITD thresholds.

Figure 7.1 also shows that the mean ITD and ILD thresholds were broadly comparable across stimuli. However, relative performance on different stimuli varied considerably across listeners, particularly for ITD thresholds. This is illustrated in Figure 7.2, which provides scatter graphs of pre–test ITD and ILD thresholds with the trained stimuli and the various untrained stimuli. The correlations between the thresholds with the various stimuli were calculated and examples of the results are given in Table 7.1. It was found that ILD thresholds were modestly but significantly related across stimuli. In contrast, ITD thresholds were unrelated across stimuli. This is also inconsistent with Experiment 3, which found that ITD thresholds at 128 Hz and 4000 Hz were related.

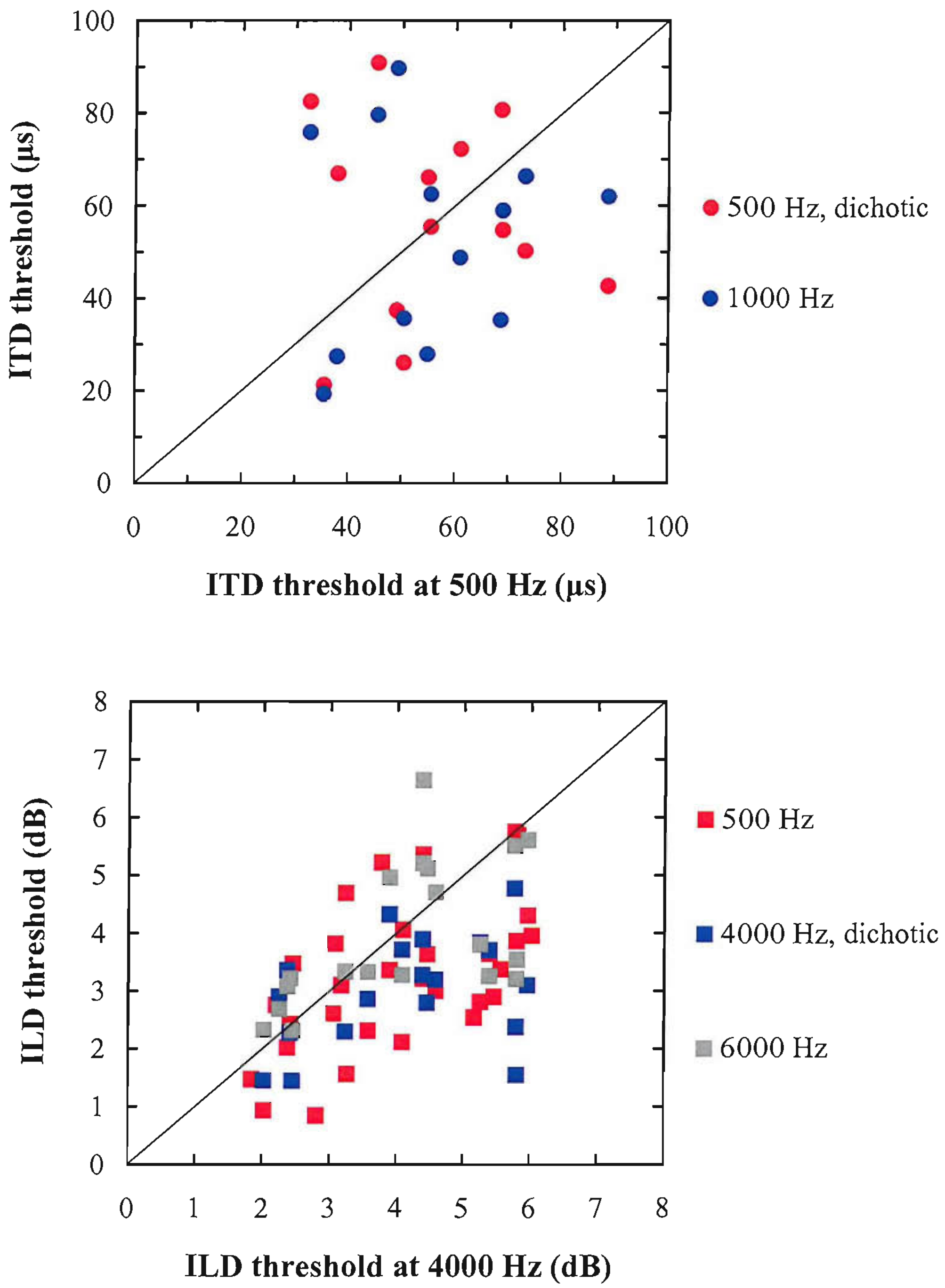


Figure 7.2. Pre–test ITD (upper panel) and ILD (lower panel) thresholds with the trained stimulus and the various untrained stimuli for each listener separately. Stimuli were diotic unless indicated otherwise.

Table 7.1. Examples of the results of analysis of correlation in thresholds between trained and untrained stimuli. The comparisons shown are commensurate with Figure 7.2.

Cue	Stimulus 1	Stimulus 2	n	r^2	p
ITD	500 Hz,	500 Hz, dichotic	13	0.01	0.79
		1000 Hz	13	0.02	0.68
ILD	4000 Hz	500 Hz	32	0.42	<0.001
		4000 Hz, dichotic	19	0.44	0.002
		6000 Hz	19	0.32	0.01

Figure 7.3 plots the mean pre–test ITD and ILD thresholds with each stimulus; error bars indicate 1 SD. One aim of this Chapter was to determine if listeners in the ITD Experiment had generally more acute binaural hearing than listeners in the ILD Experiment. This can be evaluated on the three conditions that are common to both experiments (i.e. ITD thresholds with the diotic 500–Hz tone, and ILD thresholds with the diotic 500–Hz and 4000–Hz tones). It can be seen from Figure 7.3 that thresholds in these conditions were broadly comparable, although the trained group in each experiment tended to have slightly poorer thresholds. Comparing to previous research (e.g. Bernstein et al., 1998; Hinton et al., 2004), it would appear that ITD thresholds measured by Wright and Fitzgerald were relatively good in general, whereas ILD thresholds were relatively poor in general, rather than reflecting inter–individual differences. It is unclear why this was the case.

Figure 7.3 also reveals that the mean and SD of the thresholds with the trained stimuli (indicated by asterisks) were comparable between the two untrained groups. This supports the validity of pooling the data from these groups in the upcoming analyses.

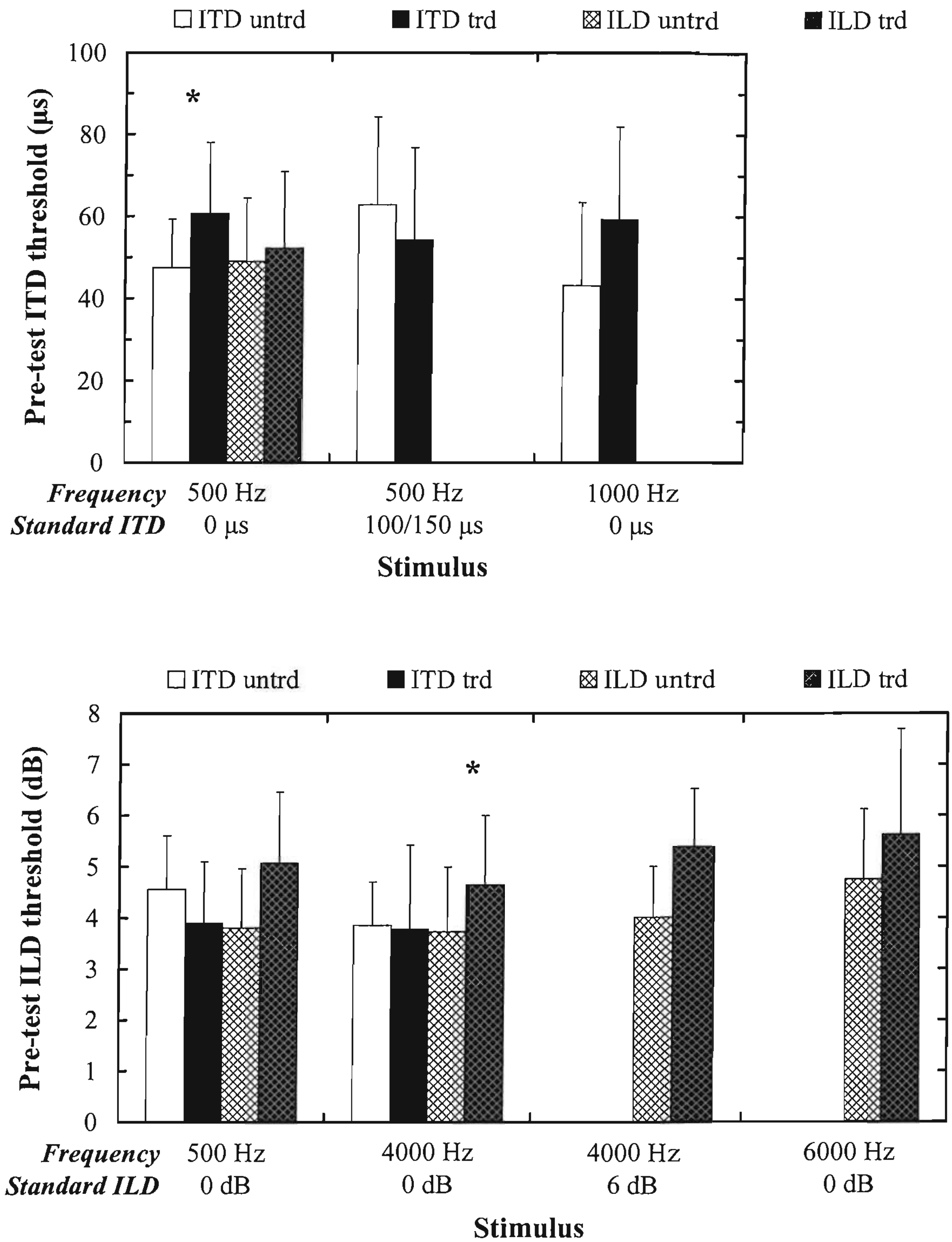


Figure 7.3. Mean pre–test ITD (upper panel) and ILD (lower panel) threshold with each stimulus, across the four groups (untrained: untrd; trained: trd). The frequency and 'standard' ITD or ILD are indicated (see Section 7.2). Error bars indicate 1 SD. The trained stimulus in each experiment is indicated by an asterisk.

7.3.2 Comparison of learning across studies

Figure 7.4 plots a measure of learning (the ratio of pre– and post–test thresholds) with the trained stimuli between untrained and trained groups from Wright and Fitzgerald (2001) and Experiment 3. Data from Wright and Fitzgerald's untrained groups were pooled. A value of greater than unity indicates that thresholds were lower at post– compared to pre–test. Learning expressed in this way was broadly comparable across all four trained groups, and thus across stimuli, binaural cues and experiments. The amount of learning in the *untrained* groups with the high–frequency conditions was also comparable. In contrast, the performance of the untrained group from Experiment 3 with low–frequency ITD discrimination slightly reduced between sessions. The learning in Wright and Fitzgerald's untrained group on low–frequency ITD discrimination was also inconsistent with the learning on the high–frequency conditions. Specifically, it was greater than in all other conditions. Consequently, the difference in learning between untrained and trained groups on this condition was smaller than in the other conditions.

The disparity in learning in Wright and Fitzgerald's untrained group on the ITD and ILD conditions, in contrast with the trained groups, can also be seen in Figure 7.5. This re–plots data in Wright and Fitzgerald's Figure 2. The measure of learning here, referred to as '*z* score', is the difference between pre– and post–test thresholds for each group divided by the SD of pre–test thresholds across both groups. A value of greater than zero indicates lower thresholds at post– compared to pre–test. Again, greater differences between conditions are apparent with the untrained rather than the trained groups.

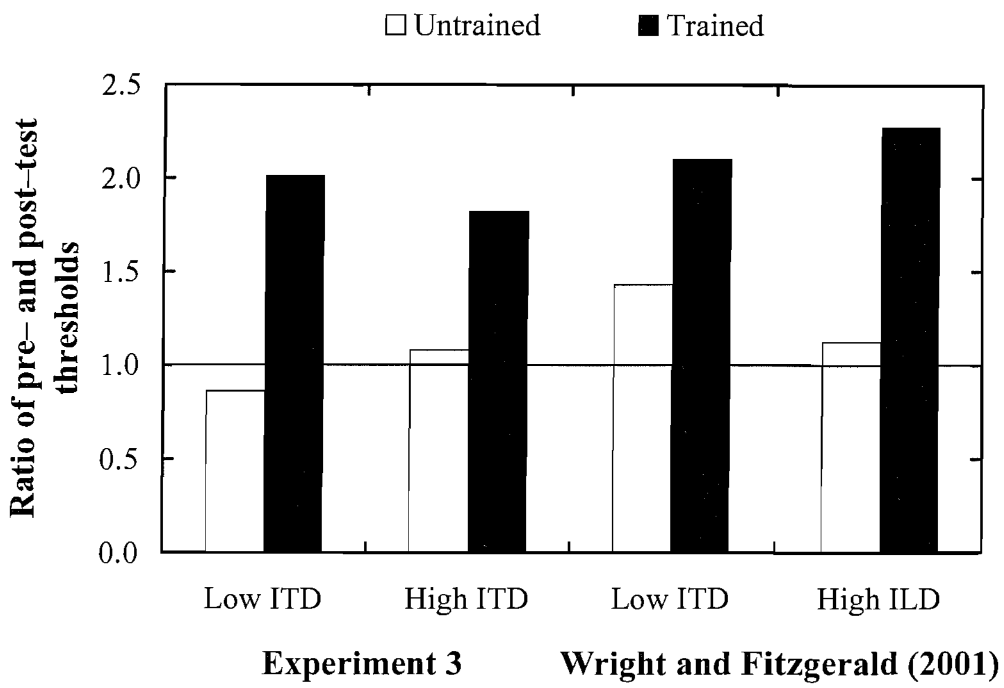


Figure 7.4. Ratio of pre– and post–test thresholds with the trained stimulus in the untrained and trained groups from two studies.

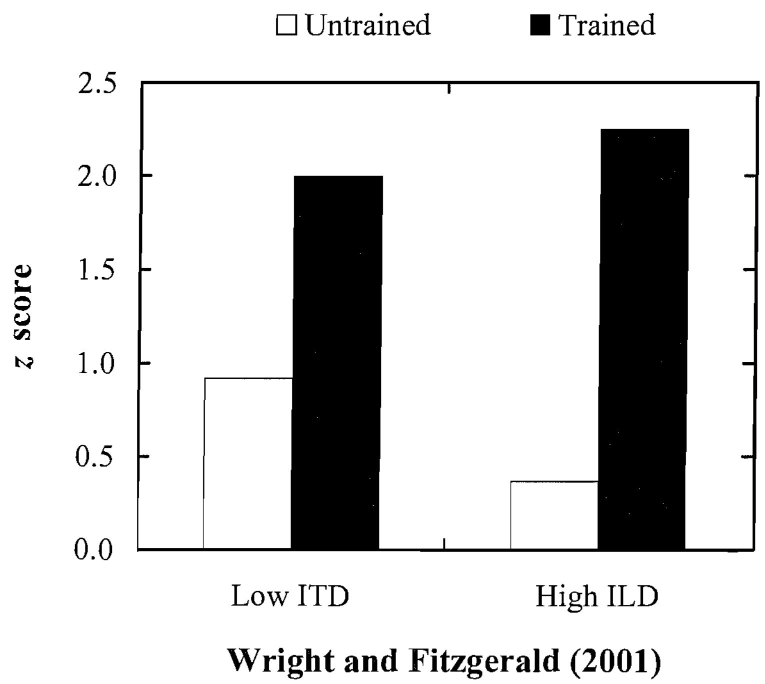


Figure 7.5. A measure of learning on the trained stimulus in untrained and trained listeners from Wright and Fitzgerald (2001). See text for details.

7.3.3 Pre– and post–test thresholds

Figure 7.6 plots the pre– and post–test ITD and ILD thresholds from listeners in the untrained and trained groups with the trained stimuli. These replicate panels from Wright and Fitzgerald's Figure 4. With the ITD condition, most trained listeners were further to the bottom–right corner of the figure than most untrained listeners, indicating greater learning. However, two trained listeners appeared to learn as little as most untrained listeners, and three untrained listeners (indicated by arrows) appeared to learn as much as most trained listeners; put differently, these three listeners appeared to learn disproportionately more than other untrained listeners. One of the latter three untrained listeners was from the ITD Experiment, the others from the ILD Experiment. There was no clear statistical reason to label these listeners as 'outliers'. For example, the difference in pre– and post–test scores were not greater than three SDs from the mean, a common criterion for an outlier, although one listener did differ by 2.6 SDs from the mean. This lack of statistical evidence, despite the visual observations of the graph, may at least partly reflect the relatively small sample (n=16): the results from these listeners may have strongly influenced the overall mean and SD. Indeed, the mean difference and SD in pre– and post–test thresholds including and excluding reduced from 15 μs to 7 μs and from 18 μs to 8 μs (i.e. approximately halved), respectively.

There was also no clear relationship between pre– and post–test thresholds for either groups, as noted by Wright and Fitzgerald. With the ILD condition, less overlap was apparent between groups, and a clearer relationship was apparent between pre– and post–test thresholds, again as noted by Wright and Fitzgerald. In general (that is, across all stimuli), pre– and post–test thresholds were related only in the ILD conditions with the untrained group. It is particularly curious that a relationship was apparent for the untrained group in the ILD condition (r^2=0.76, p<0.001) but not with the ITD condition. However, when the seemingly outlying data from the three untrained listeners indicated by arrows in Figure 7.6 are excluded, a relationship is apparent (r^2=0.78, p<0.001).

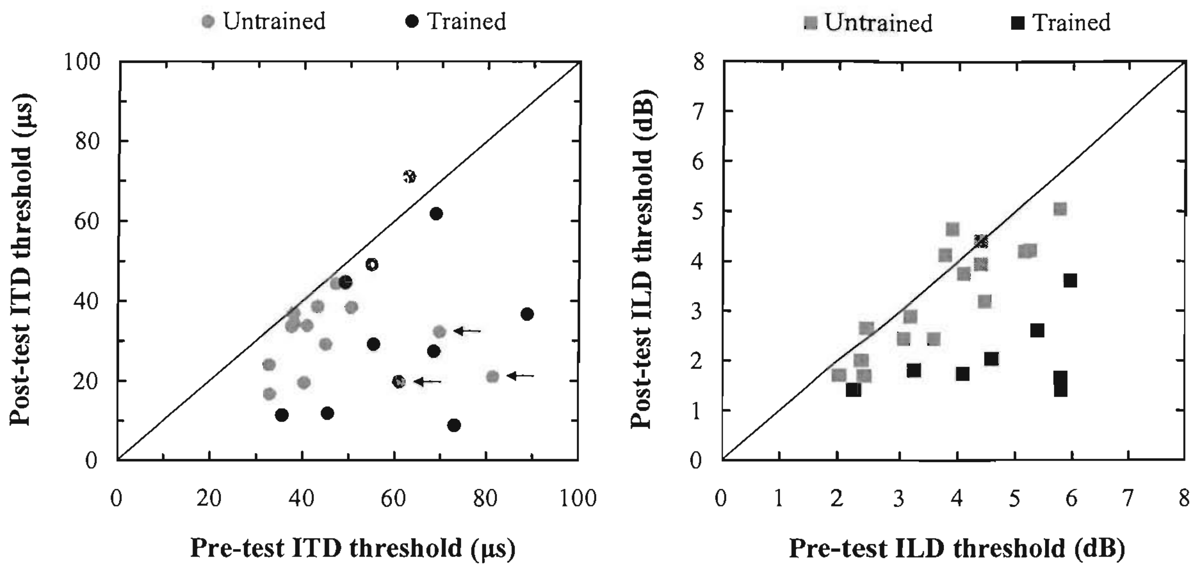


Figure 7.6. Pre– and post–test ITD (left) and ILD (right) thresholds for each listener separately from the untrained and trained groups from Wright and Fitzgerald (2001).

The effect of the exclusion of the data from these three untrained listeners on the overall amount of learning in the untrained group with the ITD–trained stimulus can be seen by comparing Figures 7.7 and 7.8, which plot the arithmetic mean data from Figure 7.6 including and excluding, respectively, these three listeners. (Figure 7.7 re–plots data from Wright and Fitzgerald's Figure 1.) Excluding these listeners had the main effect of reducing the mean pre–test threshold of the untrained group in the ITD condition, which also reduced the mean difference between sessions. The mean difference in the untrained group on ITD then became more comparable to that on ILD. Initial investigations also indicated that these three untrained listeners had a material influence of the statistical analyses of learning in the ITD condition, and so the author (DR) excluded these from subsequent analysis. Interestingly, though, the one ITD–trained listener could not be viewed as an outlier in the other ITD conditions.

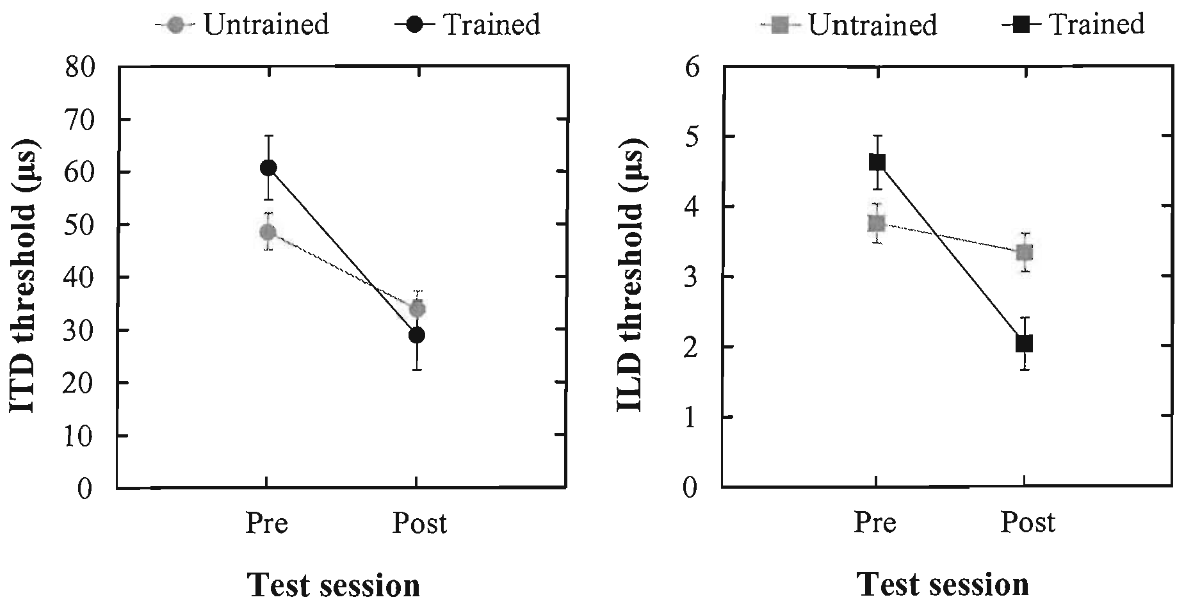


Figure 7.7. Overall arithmetic mean pre– and post–test ITD (left) and ILD (right) thresholds from the untrained and trained groups from Wright and Fitzgerald (2001). Error bars indicate ±1 SE.

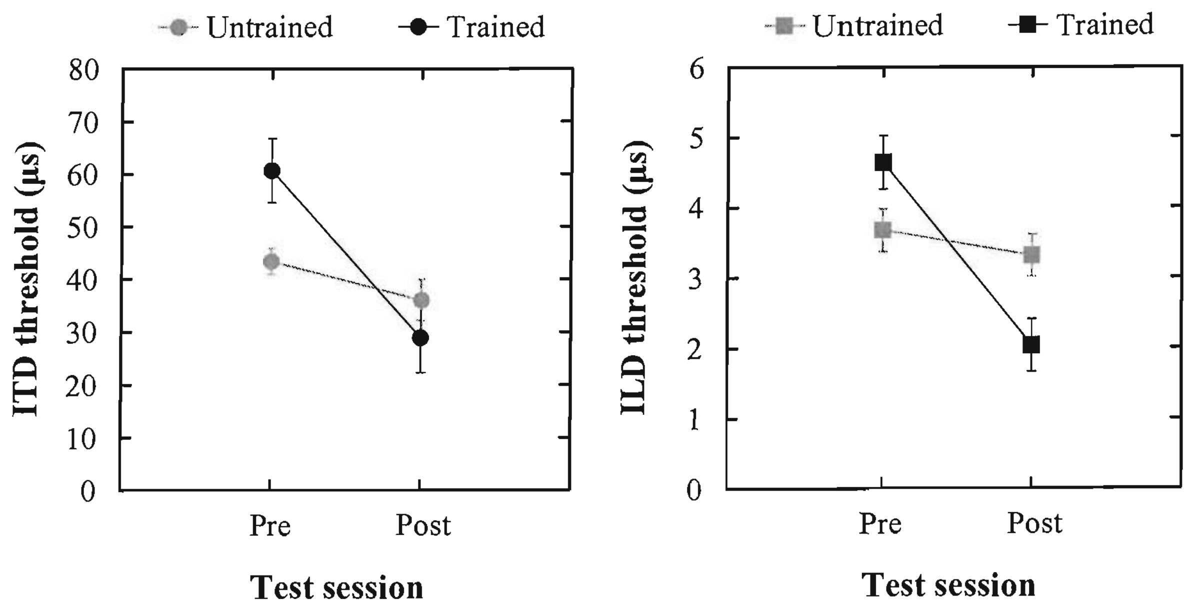


Figure 7.8. As with Figure 7.6 but with data from three untrained listeners removed.

7.3.4 Analysis of learning on the trained stimulus

In analysing learning, the data from the two untrained groups were pooled. On one hand, this was desirable because it increased the number of listeners in the untrained group, which seemed to increase the statistical power. Otherwise, the ITD–untrained group would have had only four listeners. On the other hand, the number of untrained listeners was then twice the number of trained listeners, which could be viewed as statistically undesirable. The analyses were also complicated by other features of the data. As described in the previous section, thresholds at pre– and post–test were unrelated for both trained groups, despite being paired. This questioned the validity of treating Session as a related factor. As described in Section 7.3.1, the distribution of pre–test ITD thresholds was slightly positively skewed. Although applying a logarithmic transform produced a normal distribution, it also resulted in the variances in thresholds at pre– and post–test being unequal for the ITD–trained group (Levene's test: $p=0.01$). Unequal variances were also apparent between ITD–trained and untrained groups at pre–test (Levene's test: $p=0.02$) or at post–test (Levene's test: $p=0.02$) before or after the logarithmic transform, respectively. These considerations questioned the validity of parametric analysis (i.e. ANOVA). However, it is difficult to compare learning in the two groups directly using non–parametric analyses given Session as an unrelated factor.

As a compromise, both parametric and non–parametric approaches were therefore conducted and the results compared. The three 'outliers' in the untrained group were discarded for the ITD condition but not the ILD condition. Then, ANOVAs were conducted on ITD and then ILD thresholds with unrelated factors Session and Training Group. The results of these are provided in Tables F.5 and F.6 of Appendix F. (Analysis on log transformed ITD thresholds yielded essentially identical results.) The effect of Session was significant ($p<0.001$) in both cases indicating that, overall, thresholds were lower at post–test. The effect of Training Group was not significant in either case indicating that, overall, thresholds did not differ between groups. Importantly, the interaction between Session and Training Group was significant for both ITD ($p=0.01$) and ILD ($p=0.002$) conditions, indicating that both trained groups learned more than the untrained group.

Non–parametric analysis consisted of Mann–Whitney tests on ITD and then ILD thresholds between pre– and post–test for both the untrained and trained groups. This assumed that pre– and post–test thresholds were unrelated. The results are provided in Table 7.2. This indicates that pre– and post–test thresholds did not differ in either ITD or ILD conditions for the untrained group, although the result in the ITD condition was marginal. In contrast, the differences between pre– and post–test thresholds in both trained groups were highly significant. It is worth noting that while the untrained and trained groups differed in size, statistical significance was not reached with the group with the larger number of listeners. These results are consistent with the parametric analysis, suggesting that the trained group learned more than the untrained group in both ITD and ILD conditions. These findings are in contrast to the conclusions reached by Wright and Fitzgerald (2001), but the results with ITD are consistent with Experiment 3.

Table 7.2. Results of Mann–Whitney tests on learning.

	ITD thresholds			**ILD thresholds**		
	U	*n*	*p*	*U*	*n*	*p*
Untrained	49	26	0.07	105	32	0.39
Trained	6	16	0.006	3	16	0.002

However, a further complication with these data is that untrained and trained groups appeared to have different pre–test thresholds, as is apparent in Figure 7.8. Two–tailed, unrelated–samples *t*–tests (homogeneous variances not assumed) indicated this was not statistically significant for either ITD ($t_{11.9}$=0.95, p=0.36) or ILD (t_{22}=1.67, p=0.12) conditions. Nevertheless, a relationship between pre–test thresholds and learning could still account, at least partially, for the greater learning in the trained compared to untrained groups. To evaluate this, Figure 7.9 plots the difference in thresholds between sessions as a function of pre–test threshold for listeners in both groups, and for ITD and ILD conditions. Regression lines are plotted and the proportion of variance account by each line is provided in parentheses in the legends. Overall, there was no clear relationship between variables for any conditions. Further, the slopes of the regression lines differed between the trained and untrained groups in both ITD and ILD conditions. An analysis including pre–test thresholds as a covariate was therefore not indicated. However, the ITD–trained listeners with poorest pre–test thresholds did appear to learn more than the others. It was therefore difficult to draw firm conclusions on whether the apparent difference in learning between ITD–trained and untrained groups reflected the effects of training, or the differences in pre–test thresholds.

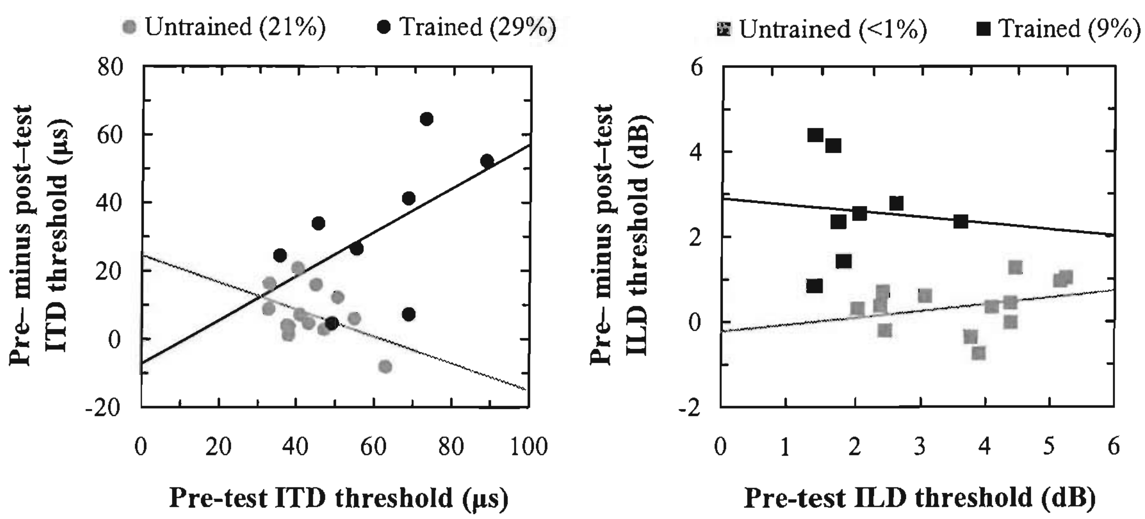


Figure 7.9. Pre– minus post–test ITD (left panel) and ILD (right panel) thresholds against pre–test thresholds for each individual in the untrained and trained groups. Regression lines are plotted and the proportion of variance accounted by each line is shown in parentheses in the legends.

7.3.5 Generalisation

Figure 7.10 presents pre– minus post–test ITD and ILD thresholds for all stimuli from the untrained (no exclusions) and trained groups in the two experiments; error bars indicate ±1 SE. Data from the untrained groups on common stimuli are pooled. The learning on ILD discrimination in the untrained groups was consistent across stimuli, typically 0.50–0.75 dB. The learning on ITD discrimination was more variable, due to the data with the untrained group with the dichotic 500–Hz tone. The ITD–trained group appeared to learn more than the untrained group on ITD discrimination at 1000 Hz, although Wright and Fitzgerald (2001) indicated that this was not statistically significant. However, Figure 7.10 indicates that the learning in the ITD–trained group at 500 Hz and 1000 Hz was also not statistically significant, as the height of the columns differed by barely more than 1 SE. It is therefore difficult to interpret the data with the untrained stimuli on ITD discrimination.

Considering ITD discrimination at 500 Hz (i.e. leftmost column of Figure 7.10), the ILD–trained group appeared to learn less than the ITD–trained group. Assuming that training–induced learning occurred with both ITD and ILD discrimination, this suggests that learning on high–frequency ILD did not generalise to low–frequency ITD. Wright and Fitzgerald indicate that the ILD–trained group also did not learn more than the untrained group on this ITD condition. Considering now ILD discrimination at 4000 Hz (i.e. rightmost column of Figure 7.10), the ITD–trained group appeared to learn less than the ILD–trained group. Again, assuming that training–induced learning occurred with ITD and ILD discrimination, this suggests that learning on low–frequency ITD did not generalise to high–frequency ILD. To determine if these differences between trained groups were statistically reliable, two ANOVAs were conducted on the thresholds with the two trained stimuli with unrelated factors Training Group (ITD–trained vs. ILD–trained) and Session. (Analysis on logarithmically transformed ITD thresholds yielded essentially identical results.) The results are summarised in Tables F.7 and F.8. The effect of Session was significant in both cases ($p \leq 0.008$) but the effect of Training Group was not. The interaction, relating to differential learning between groups, was also not significant. Although this was only marginally so for ILD thresholds ($p=0.081$), Levene's test was also marginally significant ($p=0.046$) indicating that an assumption of the ANOVA may not have been adequately met. Thus, while there were indications that learning with low–frequency ITD did not generalise to high–frequency ILD, and vice versa, these were not statistically reliable. This may reflect insufficient numbers in each group. Although not statistically significant, apparent differences in the pre–test thresholds across groups may also have been of influence.

These trends are also difficult to interpret because learning with ILD may not have completely generalised across frequency. As noted by Wright and Fitzgerald (2001), the ILD–trained group did not learn significantly more than the untrained group with ILD discrimination at 500 Hz or 6000 Hz. They interpreted this as indicating frequency–specific learning at 4000 Hz. An alternative approach to evaluating frequency–specific learning is to compare learning across stimuli in the ILD–trained group. Overall, pre–test ILD thresholds from the ILD–trained group seemed comparable across these three stimuli from Figure 7.3, although slightly lower at 4000 Hz (see Figure 7.3). Previous research suggests that this may also have been the case with asymptotic thresholds (Grantham, 1984; Yost and Dye, 1988). Frequency–specific learning may therefore be apparent as less learning on the untrained compared to the trained stimuli. Two ANOVAs were conducted on ILD thresholds from the ILD–trained group with related factor Stimulus (trained vs. untrained) and unrelated factor Session, the results of which are presented in Tables F.9 and F.10. The effect of Stimulus was statistically significant in both cases, indicating that thresholds were lowest at 4000 Hz. The effect of Session was also significant, indicating learning overall. The interaction between these factors, related to differential learning, was marginally significant at 500 Hz (p=0.03) but not at 6000 Hz (p=0.24). However, the apparent differences in pre–test thresholds may have reduced apparent differences in learning between stimuli. In summary, evidence for frequency–specific learning between 4000 Hz and 6000 Hz seems less compelling than between 4000 Hz and 500 Hz.

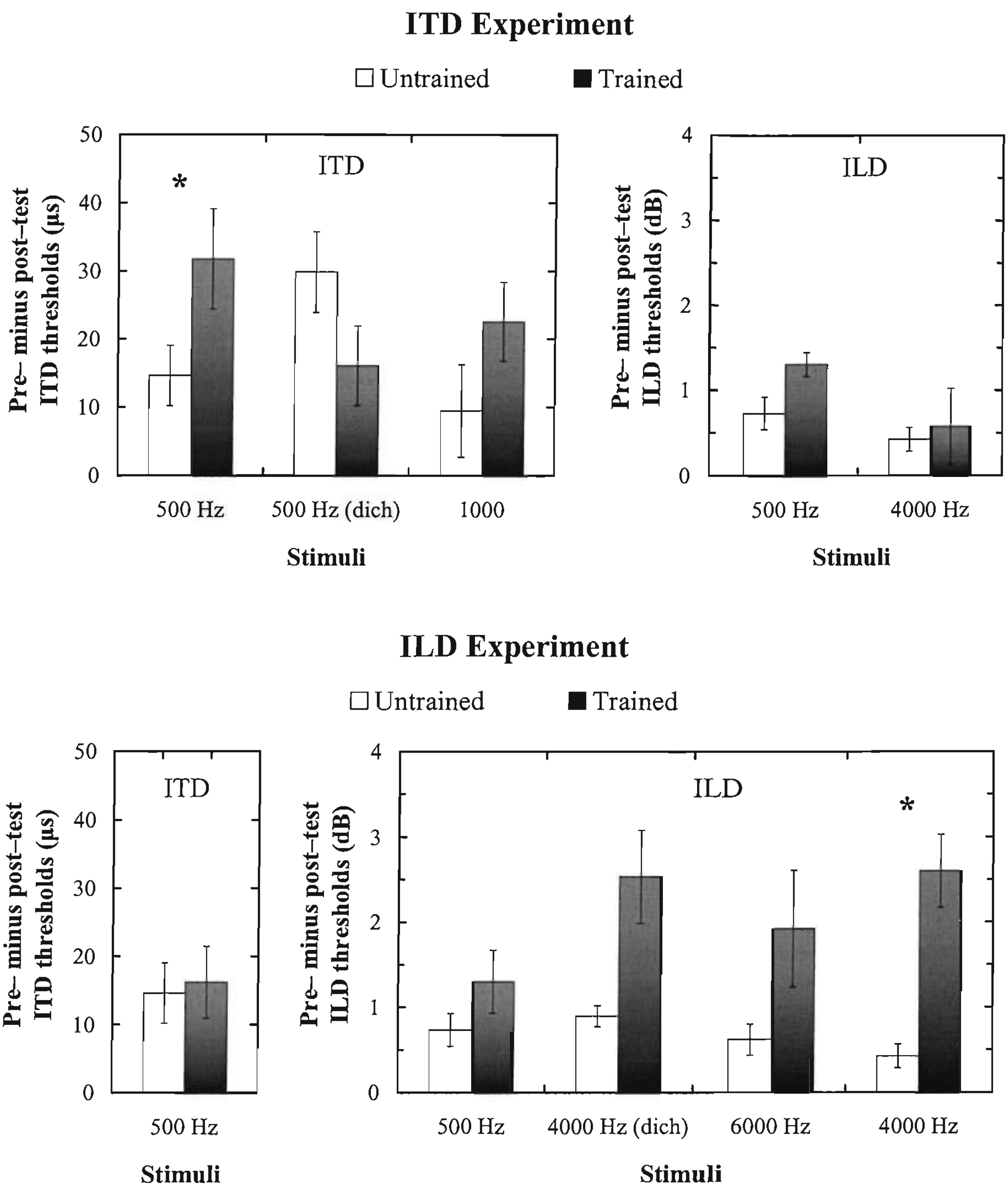


Figure 7.10. Learning with all stimuli for the untrained and trained groups in ITD and ILD Experiments. The data from the untrained groups are pooled on stimuli common to both stimuli. An asterisk indicates the trained stimulus in each experiment. Error bars represent ±1 SE.

7.4 General discussion

Wright and Fitzgerald (2001) suggested that low–frequency ITD and high–frequency ILD discrimination are characterised by different time–courses of learning. It is suggested here that this conclusion is not justified by their data. Specifically, it is suggested that the data on learning with ITD discrimination are difficult to interpret. Comparing between this study and Experiment 3, the trends relating to learning in the untrained groups differed more than in the trained groups. This appeared to arise mainly from ITD thresholds from three listeners in Wright and Fitzgerald's untrained group. When these data were removed, the trends in the untrained groups became more consistent across conditions and experiments. Subsequent analysis indicated that the trained group learned more than the untrained group with ITD discrimination, as with ILD discrimination. However, it was unclear if this result was due differences in pre–test thresholds between groups or the effect of training. However, as noted in the Introduction, there was evidence that the thresholds of some trained listeners reduced over the course of multiple training sessions. Although only two of these listeners had statistical evidence of learning according to Wright and Fitzgerald's analysis, this may partly reflect limitations in their analysis, such as the use of linear regression with highly non–linear functions. Another difficulty was that the 'learning curves' of some listeners were irregular, more so than the ILD curves, and so perhaps difficult to interpret.

There were other curious differences between the ITD and ILD data, such as the lack of a relationship between ITD thresholds across similar stimuli. These, and the irregular learning curves, may suggest that measurements of ITD thresholds were associated with greater variability than those of ILD thresholds. However, the reasons for this are not clear. One possible explanation relates to difference in the initial ITD, or ILD, on each block. Although Wright and Fitzgerald presented a small number of 'dummy' trials at the start of the block using an ITD, or ILD, above threshold, the initial ITD of the adaptive procedure was always below threshold whereas the initial ILD was always above threshold. Consequently, with the former, listeners were required to guess many trials during the early stages of a block before the ITD could be reliably detected. This may have adversely influenced measurement reliability, perhaps due to an effect of listeners' attention.

The re–examination of the generalisation of learning revealed some interesting trends, although no firm conclusions could be drawn. In particular, the learning on one cue may generalise incompletely to another. However, data from listeners trained with ITD discrimination and listeners trained with ILD discrimination at the same frequency are required before firm conclusions can be drawn. This is because learning on ILD discrimination may generalise incompletely across at least wide frequency intervals. This pattern of across–frequency generalisation appears different to that found with ITD discrimination in Experiment 3, although differences in the methodology between studies may also be important. There are perhaps clearer reasons to suspect limited generalisation across cues than differences in across–frequency generalisation between cues. For example, the subjective cue for ITD discrimination may involve detection of decorrelation of representations of the stimuli, which presumably could not be used for ILD. Plastic changes in the binaural pathways may also influence the processing of ITD and ILD differently. However, differences in across–frequency generalisation may suggest differences in across–frequency pooling of information between different spatial cues, which seems counterintuitive and counter to physiological studies of spatial processing in the SOC (Section 2.3.4).

In summary, Wright and Fitzgerald's data on ITD discrimination are difficult to interpret. Importantly though, these do not rule out multi–hour training–induced learning on ITD discrimination. These data were therefore compatible with the results of Experiment 3, and Wright and Fitzgerald's data on ILD discrimination. However, subtle but important differences in the pattern of generalisation may exist, and learning with ITD may not generalise to ILD discrimination. Future research is required to confirm these hypotheses. However, a general problem with the approach to studying learning used by Wright and Fitzgerald (2001) and in Experiment 3 is that important differences in pre–test thresholds may arise across training groups. Such differences can considerably complicate the interpretation of learning. Future studies should consider alternative approaches in order to avoid this problem, whilst controlling the interval between pre– and post–test across listeners and avoiding introducing further bias when allocating listeners to groups.

Chapter 8. Conclusions and future research

8.1 Conclusions

The present study investigated learning on ITD discrimination in normal–hearing adults. This was motivated by the possibility of providing novel insight into binaural hearing and the potential for clinical populations to learn to utilise distorted representations of binaural cues. A previous study appeared to show differences in the characteristics of learning with ITD at low frequencies and ILD at high frequencies (Wright and Fitzgerald, 2001). According to the authors, ILD was associated with a longer time–course of learning and incomplete generalisation across frequency, and so only learning with ILD was potentially related to auditory plasticity. However, there were uncertainties regarding whether these conclusions were justified. Further, it was not clear if the apparent difference in time–courses was related to the different binaural cues or frequencies and if the findings with ITD using low–frequency tones would generalise to conditions relevant to clinical populations, such as users of cochlear implants.

The present study showed that learning over many hundreds to thousands of trials is apparent with ITD discrimination with a variety of stimuli (e.g. low–frequency tones and high–frequency AM stimuli leading to fused and unfused images) and with conditions thought to minimise learning (e.g. oddity tasks and adaptive procedures). It was also found that the magnitude and time–course of learning is not inherently strongly dependent on frequency and is broadly comparable with Wright and Fitzgerald's data on ILD. While these results initially seem inconsistent with Wright and Fitzgerald's conclusions, a re–analysis of their data indicates there may be less disagreement on this issue. Overall, it is concluded that the time–course of learning is broadly comparable for ongoing ITD at low and high frequencies and ILD at high frequency (and probably low frequency), at least with the simple stimuli used in these experiments. The present study also shows that learning on ITD generalises across frequency. While this also initially appears to differ from Wright and Fitzgerald's conclusion that learning on ILD displays frequency–specificity, this may reflect differences between studies in the amount of training provided and complexities in interpreting Wright and Fitzgerald's results.

The present study provides further support for the notion that ITD processing at low and high frequencies is functionally homogenous. This comes from findings that ITD thresholds are comparable with unmodulated and transposed stimuli (under specific stimulus conditions) in both inexperienced and experienced listeners and that the time–courses of learning with these stimuli are comparable. However, new evidence for a differential effect of SPL with these stimuli has been found, although it is currently unclear if this is related to the different frequencies or the amplitude modulation and if it reflects peripheral or central processes. This study also shows that the deleterious effect of IFD is smaller than previously reported, although there may be some dependency on the trial format.

This study also indicates that training should be considered carefully when measuring binaural discrimination, at least if the aim is to measure listeners' asymptotic performance. Extensive training is usually provided in studies with normal–hearing listeners but not with hearing–impaired listeners. Consequently, the magnitude of impairment in ITD discrimination from such studies may be over–estimated. However, the findings of the present study indicate that valid measurements of the effect of IFD, and potentially other stimulus parameters, may not require training to asymptotic levels. This is because the deleterious effect of IFD was found to be consistent across listeners despite wide inter–individual differences in absolute performance and training effects. The findings that training–induced learning occurs with different binaural cues in different frequency regions with broadly comparable time–courses and, at least for ITD, generalises across frequency are also encouraging for the development of therapeutic training tools. However, there are many issues that need to be addressed before such clinical procedures can be developed.

8.2 Future directions

The pre–post paradigm potentially provides a powerful approach to studying discrimination learning since it permits a direct comparison of learning across stimuli and groups. However, some advantages of comparing learning, rather than simply post–test thresholds, across groups are lost when learning depends strongly on pre–test thresholds and when pre–test thresholds differ between groups. This severely complicates the interpretation of differences in learning across groups, as was the case with the re–analysis of Wright and Fitzgerald's (2001) data. While applying a logarithmic transform to ITD thresholds may reduce the dependency of learning on pre–test thresholds, this may not always be appropriate and difficulties with interpretation may remain. Future studies must, therefore, attempt to ensure pre–test thresholds are comparable between groups. The pre–post paradigm does at least provide the experimenter with an opportunity to observe listeners' initial ability prior to allocation and training. However, it may not be possible to obtain pre–test thresholds for all listeners prior to training *and* ensure the temporal structure of the experiment is controlled between listeners. It may also be undesirable to only use listeners with a narrow arbitrary range of pre–test thresholds, since the results may not generalise to listeners with substantially different pre–test thresholds. One alternative may be to allocate listeners to sub–groups defined by various narrow ranges of pre–test thresholds. This may have additional benefits such as permitting a clearer evaluation of inter–individual differences in learning. Amitay et al. (2005) have recently employed this approach. However, one difficulty with this approach is that a larger cohort of listeners is required than is typical of training experiments, which may have implications for the feasibility of such experiments. However, significant future advances in this area may require such designs.

Another difficulty with the pre–post paradigm is the experience provided during test sessions. The aim of these sessions is to take a 'snap–shot' of discrimination capability at two points in time without actually influencing it. Results from so–called untrained listeners in some studies indicate that this is not always the case. Indeed the use of untrained groups indicates the uncertainty as to the influence of the experience obtained during test sessions. However, learning in untrained listeners appears to vary considerably between studies, the reasons for which are unclear. The contribution of learning within and between test sessions is also unclear from most studies. Variation in learning in untrained listeners may not be strongly related to that in trained listeners, for example, due to the use of multiple stimuli during test but not training sessions. Future research is required in order to provide a clearer understanding of these issues.

One factor that may be important in learning in untrained listeners is the number of trials completed during test sessions. Studies using the pre–post design attempt to strike a balance between providing enough trials to obtain sufficiently reliable measurements whilst minimising the amount of learning induced. However, it is possible that most studies to date, including the present study, have not minimised the number of trials required to achieve the desired reliability. For example, commonly used adaptive procedures (based on Levitt–type decision rules and averaging reversals) may be less reliable over a given number of trials than alternatives (e.g. Marvit et al., 2003; Leek, 2001). However, it is unclear if these advantages occur with untrained humans on tasks not involving intensity (e.g. ITD). Further research is therefore required to identify methods that may minimise the number of trials required during pre– and post–test.

The present study found that learning on ITD discrimination generalised across frequency and found no evidence for frequency specificity. However, there is uncertainty if learning with ILD discrimination generalises across wide frequency intervals. Further research with across–frequency generalisation with ILD and ITD is important to determine if there are important differences in this regard between binaural cues. Other important and related issues regard what listeners actually learn during training on binaural discrimination and the conditions required for learning. It is likely there are various components to this and an ultimate aim of research should be to develop a model of learning that accounts quantitatively for the various components. Hawkey et al. (2004) provide an initial step towards this goal by providing an empirical definition of 'perceptual learning' and developing an approach to modelling. It is important that this research is extended both to other stimulus dimensions and by breaking down the components of 'perceptual learning' further. For example, learning with binaural discrimination is particularly interesting because it may reflect enhancement in the use of general binaural information but possibly cue–specific information also. Thus, future research is required to determine if learning on binaural discrimination displays specificity to the trained cue, such as ITD vs. ILD, or perhaps even ongoing ITD vs. onset ITD. Other important outstanding questions are whether listeners are required to attend to the changes in the stimulus or whether passive exposure is sufficient, and to degree to which learning occurs during training *per se* or is 'consolidated' between training sessions. These issues may have important implications both for our understanding of the processes underlying binaural discrimination and the development of therapeutic training tools. However, research into the influence of improvements on discrimination tasks on more global spatial tasks is also required to determine the potential usefulness of clinical discrimination–based training.

The development of a stronger body of evidence on binaural discrimination learning should also motivate research into the underlying neural mechanisms. This may benefit from recent advances in combining behavioural and physiological measurements in animals apparent in studies of ear–canal plugging (e.g. King et al., 2001) and training on frequency discrimination (e.g. Brown et al., 2004). However, recent advances with electrophysiological measures, such as using 'mismatch negativity' (e.g. Tremblay et al., 1997), and imaging techniques, such as using fMRI (e.g. Golestani and Zatorre, 2004), may also permit investigations of learning and plasticity in humans.

Further research into inter–individual differences in learning is also required in order to enhance understanding of the potential for learning and plasticity in hearing–impaired populations. For example, it has been suggested that IQ and age may influence learning on discrimination tasks (e.g. Moore et al., 2003). If this is confirmed, results from studies of learning in University students, which includes almost all studies of discrimination learning, may not apply to general hearing impaired populations. This present study should, however, prompt re–evaluation of ITD discrimination in hearing–impaired groups, such as users of cochlear implants. Previous research that has attempted to determine impairments in binaural discrimination may have been confounded by inadequate training, compared to normal–hearing studies.

Appendix A. Questionnaire

What is your sex (e.g. female)?	
What is your age?	

Ears and Hearing

Please answer the following questions by ticking the appropriate box to the right.

	Yes	No
Do you think that your hearing to quiet sounds is as good as other people of your age?		
Do you wear or have you ever been advised to wear a hearing aid?		
Does your hearing fluctuate other than when you have a cold?		
Have you ever had surgery to either ear?		
Do you suffer from tinnitus (noises, such as ringing, whistling or shushing, in the ears)?		
Do you have trouble with your balance or do you get vertigo?		
Are you experiencing or have you recently had any of the following: • Pain in either ear • Discharge (running) from either ear • Inflammation in either ear • A blockage in either ear • A injury to either ear • A cold or flu		
Are you currently on medication or have been recently taking medication for a problem related to your ears or hearing?		
Have you ever had a head injury requiring a stay in hospital?		
Have you been exposed to loud noise in the past 2 days?		

History of Listening Tests

Do you have any previous experience with listening or hearing tests?

Do you have any previous experience with localisation tests in particular?

Are you involved in or will you be involved in, in the near future, any activities requiring listening to subtle changes in sound direction?

Are you planning to have any loud noise exposure in the following month?

Appendix B. Instructions

The following text formed the basis of written instructions given to listeners prior to ITD discrimination testing.

You will hear a sequence of four sounds in both ears simultaneously. The first sound is called the 'reminder' and the last three sounds are called A, B and C.

The reminder will always sound roughly straight ahead. Two of sounds A, B or C will also appear straight ahead. However, one will sound different, it will be the 'odd-one-out'. Sounds A, B or C will be chosen as the odd-one-out completely at random. The odd-one-out will not differ in loudness or pitch but in a spatial sense, for example if may sound closer to one ear. Your job is to identify the odd-one-out. The reminder is always the same and may help you to identify the odd-one-out. It will become more difficult for you to detect the odd-one-out as the test proceeds but keep listening hard. Guess if you are not sure which to choose.

When you are ready, press any button to start the test after the prompt appears on the screen. A green light will appear on the screen just before and during the sounds. When all three sounds have been played, the light will turn blue indicating that you can respond. To respond, press one of the three buttons on your keypad depending on whether you think the odd-one-out was A, B or C. Immediately after pressing a button, the computer will flash the correct answer in red. This will then be repeated a number of times over about 4 minutes. Speak clearly into the microphone to communicate with me during or between tests, such as if you want to terminate testing.

IMPORTANT!

The sounds should never be more than comfortably loud. If at any time the sounds become too loud or uncomfortably loud, say "STOP!" clearly into the microphone and I shall stop the test.

Appendix C. Response bias

Preliminary experiments of ITD discrimination using a standard 3AFC task indicated that inexperienced listeners tend to be biased away from selecting the first interval, despite knowing that the target appears in that interval 33% of the time. In addition, some listeners explicitly reported that they found it more difficult to hear the ITD when the target occurred first. In order to reduce or eliminate this interval selection bias, the 3AFC task was supplemented with an initial 'cue' interval, which always contained the standard. This is similar to the two–cue, two–alternative forced choice task used by Bernstein and Trahiotis (e.g. 1982, 2002), except that a final cue was not employed since there was no indication of selection bias away from the third primary interval. Initial experiments indicated that the bias was far less apparent when using the cue. However, data from Experiment 1 were analysed in order to determine if there was any residual and consistent bias across inexperienced listeners.

Sixteen listeners completed 12 blocks of 50 trials at pre– and post–test and the frequencies of responses towards each primary interval were obtained for each listener. Data were not available for one block at pre– and/or post–test for a small number of listeners. The expected frequencies for each interval were then computed by dividing the total responses for each listener by three. The two sets of data were compared using a chi–squared test. Following a Bonferroni correction, the criterion probability was 0.0016.

Table C.1 displays the results for four listeners having statistically significant results at this level. Of these listeners, one also had a marginally statistically non–significant result at post–test. Data from two other listeners with marginally non–significant results at pre–test are also shown. These six listeners came from all four categories of listeners described in Chapter 4 in terms of their learning curves. The second to fourth columns indicate the proportion of total responses towards each interval, as a percentage, referred to as 'incidence'. If responses were equally distributed between the three intervals, as was the presence of the target, one would expect an incidence of 33% for each interval. In fact, the disparity between actual and expected incidence was as much as 10%. Three listeners, Listeners 4, 5 and 8, displayed a bias away from the first interval whilst the other listeners displayed a bias *towards* the first interval. Only Listeners 6 and 7 displayed a bias away from the final interval. The remaining two columns provide the values of chi–squared and the probability value.

Table C.1. Results of the analysis of response bias. The incidence of responses to each interval are presented for each listener with results of chi–squared tests (*df* = 4).

Listener	**Incidence (%)**			χ^2	*p*
	Interval 1	**Interval 2**	**Interval 3**		
L4	27	41	33	18.3	<0.001
L5	23	31	45	42.1	<0.001
L6	39	34	27	11.7	0.003
L7	40	32	28	11.6	0.003
L8					
pre–test	27	34	39	13.3	0.001
post–test	29	31	40	11.7	0.003
L12	41	27	32	16.8	<0.001

Approximately one third of inexperienced listeners displayed an interval selection response bias but its direction was not consistent across listeners and it generally resolved with training. It would be difficult to resolve this by further altering the task since the most common bias was away from the first primary interval despite the use of the cue. There was also little justification for a final cue interval in terms of reducing bias when also considering the implication for test duration. An unresolved question concerns the impact of this response bias on ITD thresholds, particularly as bias was dependent on training. However, given that the learning curves of these listeners were like those of listeners without clear response bias, it is unlikely that the resolution of bias with training had an important effect on ITD thresholds during training.

Appendix D. Raw data

Experiment 1

Listener	Trained stimulus	ITD Threshold (μs)											
		Pre–test											
		Matched						Unmatched					
1	Matched	≥2000	1686	≥2000	1298	1426	1252	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000
2	Matched	508	445	495	586	503	414	464	634	623	1100	836	679
3	Matched	603	816	959	825	586	-	1708	1782	1660	1569	1612	-
4	Matched	575	690	744	842	1109	974	922	681	1003	1191	1061	1248
5	Matched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000
6	Matched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000
7	Matched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000
8	Unmatched	638	786	471	603	694	401	1552	1168	1526	1376	-	1141
9	Unmatched	790	742	1063	1115	720	571	1547	1545	1599	1289	916	1348
10	Unmatched	375	273	282	132	167	224	1105	920	836	469	369	603
11	Unmatched	590	519	397	256	367	454	-	951	675	540	573	790
12	Unmatched	456	766	625	781	608	744	946	1000	918	966	844	521
13	Unmatched	725	638	493	762	755	697	868	477	775	569	651	462
14	Unmatched	530	920	875	911	443	503	-	881	1070	564	673	712
15	Unmatched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000
16	Unmatched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000

		Session 2						Session 3					
1	Matched	1298	1404	1118	1508	1107	1061	1586	1469	1107	1461	1196	1081
2	Matched	323	208	208	139	169	169	211	193	193	161	106	178
3	Matched	820	968	818	616	658	451	859	812	799	527	577	566
4	Matched	655	944	1159	1309	1176	1148	586	694	1144	1035	1252	1330
5	Matched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000
6	Matched	≥2000	≥2000	≥2000	≥2000	≥2000	1565	1658	1684	1916	1693	1669	1915
7	Matched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	1720
8	Unmatched	1204	673	-	477	399	469	349	523	382	829	473	358
9	Unmatched	1580	1330	1102	1013	1111	1315	1176	1400	1248	1413	1248	1161
10	Unmatched	608	473	562	482	503	382	647	660	391	436	482	256
11	Unmatched	592	751	462	642	549	521	540	573	610	540	525	406
12	Unmatched	618	744	484	653	605	668	475	401	521	599	638	436
13	Unmatched	488	781	842	371	286	406	495	549	406	536	378	634
14	Unmatched	1285	1031	1024	816	1259	1120	1239	927	1063	1215	911	1150
15	Unmatched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000
16	Unmatched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000

		Session 4						Session 5					
1	Matched	1552	1141	1081	1287	996	1066	1094	1417	1378	1176	964	549
2	Matched	224	336	319	234	382	332	137	176	111	137	165	247
3	Matched	1005	1026	931	961	833	938	1063	736	658	1059	1209	972
4	Matched	747	987	918	710	987	905	684	1313	1133	1411	1476	1356
5	Matched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000
6	Matched	1758	≥2000	1478	≥2000	1916	1660	≥2000	1899	1890	1678	1912	1922
7	Matched	≥2000	1810	1853	1319	1762	1638	1823	1291	1300	1467	883	1200
8	Unmatched	632	538	625	647	558	629	408	536	510	651	649	434
9	Unmatched	1100	1209	1115	579	562	914	1280	916	688	896	584	790
10	Unmatched	456	362	397	321	477	306	447	295	360	417	256	391
11	Unmatched	636	477	471	486	645	506	436	382	256	410	299	384
12	Unmatched	438	599	464	230	475	551	237	328	339	612	538	480
13	Unmatched	1241	1196	716	595	538	352	540	690	497	623	775	632
14	Unmatched	929	1148	1298	844	1018	1510	1497	922	957	1059	1378	1300
15	Unmatched	≥2000	≥2000	≥2000	≥2000	≥2000	1751	≥2000	1645	1345	1866	≥2000	1812
16	Unmatched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	1825

		Session 6						Session 7					
1	Matched	898	862	697	892	1033	740	883	629	1016	614	1105	857
2	Matched	139	167	191	169	148	169	228	-	224	167	158	148
3	Matched	1007	1278	942	872	981	807	868	794	749	734	1248	855
4	Matched	534	1003	1363	1211	1228	1274	636	1085	1500	1467	1474	1617
5	Matched	≥2000	≥2000	≥2000	≥2000	≥2000	≥2000	1879	1866	1875	1612	940	1285
6	Matched	1656	1345	-	1387	1534	1476	1808	1391	1753	1467	1795	1474
7	Matched	1326	903	519	336	430	499	857	781	855	777	660	608
8	Unmatched	486	471	447	356	525	393	634	432	412	612	512	469
9	Unmatched	703	675	777	888	655	510	1085	731	514	558	770	658
10	Unmatched	543	284	271	566	332	299	388	375	362	234	334	317
11	Unmatched	454	525	469	501	412	467	423	399	471	382	488	428
12	Unmatched	551	582	521	538	662	551	770	788	803	857	454	777
13	Unmatched	545	484	688	807	651	664	458	525	642	699	369	467
14	Unmatched	1441	1484	1411	1558	1189	1413	629	961	705	571	974	944
15	Unmatched	1706	1593	1578	1649	1582	1478	1146	1437	1024	981	1343	1530
16	Unmatched	≥2000	≥2000	1879	1526	1576	1704	1322	1445	1617	1701	1372	1558

Post–test

		Matched						Unmatched					
1	Matched	781	760	1042	1209	777	684	1747	1790	1862	1662	1784	1823
2	Matched	299	193	187	206	167	161	271	291	330	451	384	306
3	Matched	1350	1126	1163	1107	1241	1252	1634	1829	1753	1927	1521	1905
4	Matched	764	1417	1280	1775	1825	1693	1886	≥2000	≥2000	1317	1658	≥2000
5	Matched	1669	1736	1606	1758	1519	1606	≥2000	≥2000	1580	1912	1795	1634
6	Matched	1684	1523	1411	1007	935	924	1760	1790	1897	1758	1860	1875
7	Matched	825	428	836	454	490	471	694	992	1194	1120	1133	1385
8	Unmatched	380	273	330	382	306	245	473	530	406	436	428	-
9	Unmatched	393	438	462	859	738	549	940	905	699	842	538	788
10	Unmatched	208	158	104	174	130	115	365	371	356	447	356	302
11	Unmatched	386	306	382	419	408	297	664	608	614	503	339	490
12	Unmatched	460	178	299	645	454	571	684	414	668	924	512	803
13	Unmatched	610	378	623	516	412	-	638	829	697	686	653	-
14	Unmatched	443	349	302	532	573	564	1415	1194	-	1339	1309	1057
15	Unmatched	738	890	738	747	762	890	1137	1421	1552	1083	1400	983
16	Unmatched	1667	1048	1319	1011	1456	1215	1721	1549	1419	1771	1430	1447

A score of ≥2000 indicates that ITD thresholds were not measurable due to the staircase reaching the maximum permissible ITD.

Missing data (-) arose either because no scored reversals were obtained.

Experiment 2a

Listener	ITD Threshold (μs)											
	Pure tone at 60 phons						SAM at 60 phons					
1	64	56	41	81	64	52	63	89	115	65	70	120
2	64	74	81	107	70	78	68	103	66	110	92	70
3	60	43	63	61	52	54	206	170	190	196	207	135
4	125	156	167									
5	195	177	171	241	110	232	365	355	262	392	219	200
	'Wideband' transposed at 60 phons						'Narrowband' transposed at 60 phons					
1	46	85	59	83	57	66	128	65	102	111	87	109
2	67	65	53	83	76	76	74	62	60	84	88	61
3	128	187	141	155	125	146	84	123	108	137	124	141
4							196	156	126			
5	186	159	105	114	169	161	167	119	244	192	186	216
	Pure tone at 30 phons						'Narrowband' transposed at 30 phons					
1	54	42	37	92	56	41	181	145	124	176	159	
2	134	124	125	142	124	156	218	223	203	215	195	
3	91	99	101	47	90	81	218	290	160	178	234	
4												
5	207	329	334	197	231	315	780	787	1049	1146	798	

Experiment 2b

Listener	ITD Threshold (µs)											
	Pure tone at 60 phons						Transposed at 60 phons					
1	451	490	701	318	282	462	1027	557	684	662	1240	796
2	92	40	60	94	56	55	334	147	153	422	407	322
3	418	303	242	270	263	344	613	564	643	503	577	662
4	328	230	408	195	304	267	491	749	1722	648	906	1312
5	223	331	296	340	406	368	935	437	516	647	1482	672
6	207	126	140	108	167	130	236	195	245	157	209	178
7	95	100	134	78	93	95	945	951	982	464	407	358

Experiment 3

Listener	Lead ear	Order	Training condition	ITD Threshold (µs)											
				Pre–test											
				Pure tone				SAM				Transposed			
1	Right	1	Untrained	186	135	271	206	449	411	810	691	321	332	263	335
2	Left	1	Untrained	282	259	241	139	1174	1420	1284	764	634	361	598	573
3	Right	2	Untrained	140	120	128	117	278	245	215	377	249	184	179	169
4	Left	2	Untrained	189	274	166	181	1062	846	351	880	681	581	496	503
5	Right	1	Untrained	215	383	324	489	1153	949	1128	868	526	372	313	557
6	Right	3	Untrained	203	215	388	312	616	350	400	442	200	196	194	306
7	Right	1	Tone	132	138	169	179	518	1003	776	691	291	475	193	741
8	Left	3	Tone	164	138	146	111	166	194	147	122	64	77	78	39
9	Right	2	Tone	482	846	754	1483	1380	1020	1144	1318	270	616	701	1683
10	Left	1	Tone	84	111	75	67	253	197	241	186	114	107	78	140
11	Right	3	Tone	194	166	128	106	417	329	303	267	154	223	131	132
12	Left	1	Tone	161	138	132	132	197	245	282	287	150	138	98	73
13	Right	1	Transposed	327	178	164	152	711	581	922	681	539	565	449	526
14	Left	3	Transposed	218	231	215	287	477	526	691	573	383	539	598	482
15	Right	2	Transposed	133	171	186	137	203	221	127	184	121	97	99	115
16	Left	1	Transposed	255	429	303	255	332	270	496	400	127	248	161	158
17	Right	3	Transposed	186	108	290	171	327	165	394	345	308	181	128	148
18	Left	2	Transposed	406	517	467	449	743	754	1195	422	286	303	322	541
19	Right	1	Transposed	234	245	274	291	303	312	300	423	278	212	231	238
20	Left	3	Transposed	1107	1322	266	488	1280	919	1077	1262	423	383	489	372

			Session 2						Session 3					
1	Right	Untrained												
2	Left	Untrained												
3	Right	Untrained												
4	Left	Untrained												
5	Right	Untrained												
6	Right	Untrained												
7	Right	Tone	140	155	232	167	299	155	152	137	83	133	185	117
8	Left	Tone	197	112	92	100	92	64	129	105	97	112	108	83
9	Right	Tone	296	396	424	472	735	1299	475	431	283	244	570	381
10	Left	Tone	74	80	53	44	48	38	49	55	47	27	30	30
11	Right	Tone	200	141	109	89	69	273	91	83	81	96	27	82
12	Left	Tone	154	121	106	84	115	95	129	80	111	95	94	96
13	Right	Transposed	225	470	30	450	322	405	264	257	267	200	251	454
14	Left	Transposed	520	325	332	165	277	375	341	262	269	236	228	180
15	Right	Transposed	-	61	85	81	121	94	97	69	158	97	101	101
16	Left	Transposed	221	185	165	152	195	157	197	197	258	209	222	176
17	Right	Transposed	100	200	154	154	138	115	94	80	121	99	104	81
18	Left	Transposed	688	839	808	516	367	549	712	534	685	501	492	491
19	Right	Transposed	97	108	114	67	52	69	72	70	83	69	120	107
20	Left	Transposed	345	390	352	373	253	290	206	204	204	212	148	194

			Session 4						**Session 5**					
1	Right	Untrained												
2	Left	Untrained												
3	Right	Untrained												
4	Left	Untrained												
5	Right	Untrained												
6	Right	Untrained												
7	Right	Tone	130	74	109	127	157	101	-	107	202	156	134	127
8	Left	Tone	89	105	124	107	93	69	119	69	61	58	169	89
9	Right	Tone	290	324	383	356	226	305	263	242	240	246	181	179
10	Left	Tone	19	36	36	26	37	26	26	27	14	18	30	14
11	Right	Tone	88	79	74	67	55	56	79	62	45	61	39	79
12	Left	Tone	131	128	99	93	80	81	145	103	118	103	79	82
13	Right	Transposed	197	288	268	217	177	421	203	222	267	211	316	273
14	Left	Transposed	332	267	137	196	233	138	263	215	193	112	172	139
15	Right	Transposed	82	91	73	99	90	133	87	87	109	28	96	83
16	Left	Transposed	171	178	191	186	120	120	144	121	186	131	173	117
17	Right	Transposed	112	65	112	160	142	200	119	89	113	127	151	139
18	Left	Transposed	335	467	284	251	355	307	328	315	189	357	428	377
19	Right	Transposed	70	73	94	102	123	77	54	61	64	88	58	103
20	Left	Transposed	192	204	44	164	186	133	179	150	167	182	198	188

			Session 6						Session 7					
1	Right	Untrained												
2	Left	Untrained												
3	Right	Untrained												
4	Left	Untrained												
5	Right	Untrained												
6	Right	Untrained												
7	Right	Tone	77	81	121	94	79	81	94	67	83	68	83	72
8	Left	Tone	113	116	81	75	74	91	90	87	77	51	106	101
9	Right	Tone	398	580	858	909	844	1051	348	516	641	671	753	1004
10	Left	Tone	26	23	28	27	20	16	-	17	21	22	19	16
11	Right	Tone	-	94	130	77	61	69	86	66	58	27	43	71
12	Left	Tone	119	51	104	98	104	85	143	101	69	96	71	89
13	Right	Transposed	164	182	179	251	329	391	164	253	243	156	276	132
14	Left	Transposed	200	216	176	195	174	158	376	219	232	241	193	230
15	Right	Transposed	110	166	176	47	73	76	57	36	65	23	54	48
16	Left	Transposed	194	329	254	204	128	114	150	154	167	160	163	168
17	Right	Transposed	94	127	140	105	109	155	108	131	103	89	98	92
18	Left	Transposed	338	-	268	201	251	308	295	-	321	308	306	190
19	Right	Transposed	111	65	52	44	58	51	105	66	81	72	53	50
20	Left	Transposed	121	182	165	185	171	134	161	137	147	147	124	142

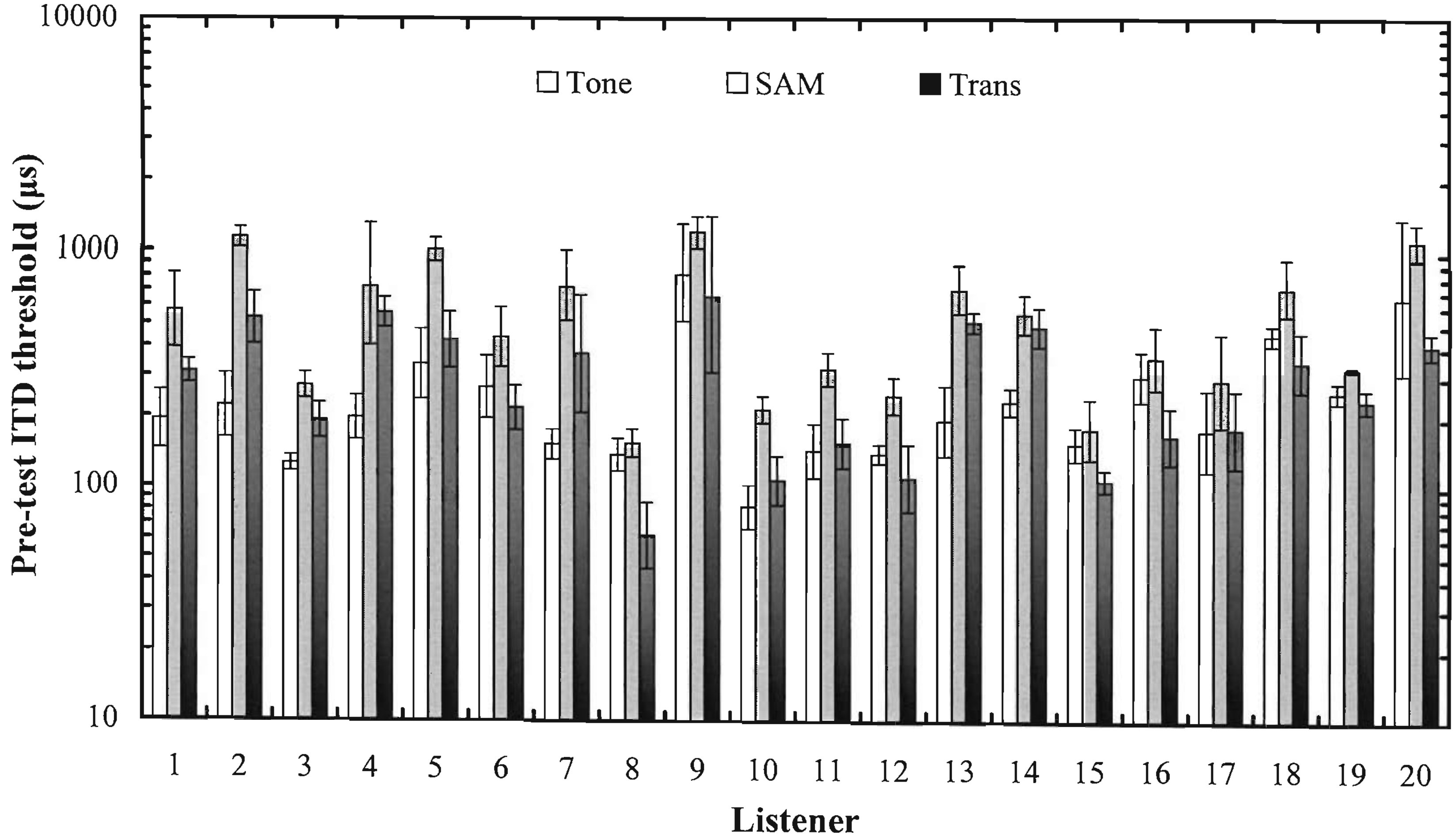
				Post–test											
				Pure tone				SAM				Transposed			
1	Right	1	Untrained	119	138	186	181	366	394	423	581	221	218	116	224
2	Left	1	Untrained	341	236	318	475	573	810	864	526	212	1211	152	436
3	Right	2	Untrained	112	106	93	159	271	536	405	417	121	238	309	282
4	Left	2	Untrained	238	283	255	382	541	625	361	652	841	454	784	880
5	Right	1	Untrained	351	361	467	283	710	1248	1005	881	372	607	404	1379
6	Right	3	Untrained	625	449	549	643	467	455	634	616	255	148	234	263
7	Right	1	Tone	94	79	105	84	417	518	691	248	235	87	113	160
8	Left	3	Tone	82	52	41	70	71	113	44	98	56	54	49	41
9	Right	2	Tone	318	573	331	286	1141	517	743	475	228	468	199	203
10	Left	1	Tone	39	41	24	24	183	135	146	125	70	58	87	65
11	Right	3	Tone	80	77	99	54	157	158	178	64	81	114	99	66
12	Left	1	Tone	104	98	119	80	214	189	148	248	139	97	91	146
13	Right	1	Transposed	114	61	164	144	340	317	565	411	224	372	215	113
14	Left	3	Transposed	151	107	111	115	423	346	518	322	255	218	274	212
15	Right	2	Transposed	133	189	245	146	271	322	224	218	148	136	125	215
16	Left	1	Transposed	133	189	245	146	271	322	224	218	148	136	125	215
17	Right	3	Transposed	119	150	128	110	187	173	184	116	70	66	93	43
18	Left	2	Transposed	126	153	159	178	511	662	722	475	286	309	197	255
19	Right	1	Transposed	114	63	68	75	146	107	155	100	61	78	44	70
20	Left	3	Transposed	166	114	161	144	356	322	377	253	263	200	131	218

Order 1: tone, SAM, transposed then reverse; 2: SAM, transposed, tone then reverse; 3: transposed, tone, SAM then reverse.

Missing data (-) arose because no scored reversals were obtained; the session geometric mean was inserted in analysis.

Appendix E. Pre–test ITD thresholds from Experiment 3

Geometric mean pre–test ITD threshold for each listener in Experiment 3 and with each stimulus; error bars indicate ±1 SD.



Pre-test ITD threshold (µs)
10000
1000
100
10
□ Tone
□ SAM
■ Trans
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Listener

Appendix F. ANOVA tables

Table F.1. Summary of results of ANOVA on ITD thresholds from Experiment 3, described in Section 6.3.3.1, continued overleaf. Factors *Training Group* and *Counterbalancing Group* are summarised as *Training* and *Counter*, respectively. Significant effects at the *a priori* significance level ($p \leq 0.01$) are highlighted.

Source of variation	Sum of squares	Degrees of freedom	Mean square	F	p
Related					
Main effects:					
Stimulus[a]	748.9	1.2	601.5	41.8	<0.001*
Block	0.9	3	0.3	0.2	0.87
Session	254.2	1	254.2	21.6	0.001*
Interactions:					
Stimulus × Counter[a]	87.1	2.59	35.0	2.4	0.12
Stimulus × Training[a]	12.3	2.5	5.0	0.3	0.76
Stimulus × Counter × Training[a]	119.5	5.0	24.0	1.7	0.21
Block × Counter	15.1	6	2.5	1.9	0.10
Block × Training	19.1	6	3.2	2.4	0.05
Block × Counter × Training	22.4	12	1.9	1.4	0.21
Session × Counter	11.4	2	5.7	0.5	0.63
Session × Training	192.1	2	96.1	8.2	0.007*
Session × Counter × Training	68.9	4	17.2	1.5	0.28
Stimulus × Block	17.6	6	2.9	2.1	0.07
Stimulus × Block × Counter	28.5	12	2.4	1.7	0.10
Stimulus × Block × Training	10.5	12	0.9	0.6	0.83
Stimulus × Block × Counter × Training	55.1	24	2.3	1.6	0.07
Session × Block	3.5	3	1.2	0.7	0.56
Session × Block × Counter	5.2	6	0.9	0.5	0.79
Session × Block × Training	21.2	6	3.5	2.1	0.08
Session × Clock × Counter × Training	38.8	12	3.2	2.0	0.06
Stimulus × Block × Session	4.7	6	0.8	0.5	0.84
Stimulus × Block × Session × Counter	9.8	12	0.8	0.5	0.92
Stimulus × Block × Session × Training	13.5	12	1.1	0.7	0.79
Stimulus × Block × Session × Counter × Training	28.6	24	1.2	0.7	0.84

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F*	*p*
Error:					
Stimulus[a]	197.0	13.7	14.4		
Block	44.0	33	1.3		
Session	129.2	11	11.7		
Stimulus × Block	94.1	66	1.4		
Session × Block	54.7	33	1.7		
Stimulus × Block × Session	114.0	66	1.7		
Unrelated					
Main effects:					
Training	328.1	2	164.0	1.8	0.21
Counter	316.9	2	158.4	1.7	0.22
Interactions:					
Training × Counter	873.7	4	218.4	2.4	0.11
Error:					
Error	1006.3	11	91.5		

[a] Mauchly's test of sphericity gave $p = 0.009$

Table F.2. Summary of results of ANOVA on ITD thresholds in untrained group from Experiment 3, described in Section 6.3.3.4. Significant effects at the *a priori* significance level ($p \leq 0.01$) are highlighted.

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F*	*p*
Related					
Main effects:					
Stimulus	94.7	2	47.4	16.1	0.001*
Session	0.1	1	0.1	0.07	0.80
Interaction:					
Stimulus × Session	2.9	2	1.4	1.6	0.25
Error:					
Stimulus	29.4	10	2.9		
Session	7.7	5	1.5		
Stimulus × Session	9.1	10	0.9		

Table F.3. Summary of results of ANOVA on ITD thresholds in trained groups from Experiment 3, described in Section 6.3.3.4. Factor *Training Group* is summarised as *Training*. Significant effects at the *a priori* significance level ($p \leq 0.01$) are highlighted.

Source of variation	**Sum of squares**	**Degrees of freedom**	**Mean square**	***F***	***p***
Related					
Main effects:					
Stimulus[a]	164.9	1.151	143.3	31.1	<0.001*
Session	139.8	1	139.8	34.4	<0.001*
Interactions:					
Stimulus × Training[a]	1.9	1.151	1.6	0.4	0.71
Session × Training	0.02	1	0.02	0.0	0.94
Stimulus × Session	2.5	2	1.3	1.7	0.20
Stimulus × Session × Training	0.3	2	0.2	0.2	0.81
Error:					
Stimulus[a]	63.7	13.817	4.6		
Session	48.7	12	4.1		
Stimulus × Session	17.5	24	0.7		
Unrelated					
Training	50.9	1	50.9	1.4	0.25
Error	421.2	12	35.1		

[a] Mauchly's test of sphericity gave $p = 0.001$

Table F.4. Summary of results of ANOVA on ITD thresholds in trained groups from Experiment 3, described in Section 6.3.4. Significant effects at the *a priori* significance level ($p \leq 0.01$) are highlighted.

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F*	*p*
Related					
Main effects:					
Training Session	218.3	5	43.7	7.5	<0.001*
Block	21.7	5	4.3	1.4	0.25
Interactions:					
Training Session × Training Group	14.4	5	2.9	0.5	0.78
Block × Training Group	2.7	5	0.54	0.2	0.97
Training Session × Block	28.3	25	1.1	0.8	0.79
Training Session × Block × Training Group	40.2	25	1.6	1.1	0.37
Error:					
Training Session	349.8	60	5.8		
Block	189.5	60	3.2		
Training Session × Block	448.1	300	1.5		
Unrelated					
Training Group	517.3	1	517.3	1.5	0.24
Error	4072.1	12	339.3		

Table F.5. Summary of results of ANOVA on ITD thresholds in ITD–trained and untrained groups (pooled, less three listeners) with trained stimulus, described in Section 7.3.4. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted.

Source of variation	**Sum of squares**	**Degrees of freedom**	**Mean square**	***F***	***p***
Unrelated					
Main effects:					
Session	3781.0	1	3781.0	18.6	<0.001*
Training Group	249.1	1	249.1	1.2	0.28
Interaction:					
Session × Training Group	1477.9	1	1477.9	7.3	0.01*
Error:					
Error	7733.0	38	203.5		

Table F.6. Summary of results of ANOVA on ILD thresholds in ILD–trained and untrained groups (pooled) with trained stimulus, described in Section 7.3.4. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted.

Source of variation	**Sum of squares**	**Degrees of freedom**	**Mean square**	***F***	***p***
Unrelated					
Main effects:					
Session	24.4	1	24.4	20.0	<0.001*
Training Group	0.5	1	0.5	0.4	0.53
Interaction:					
Session × Training Group	12.6	1	12.6	10.4	0.002*
Error:					
Error	53.6	44	1.2		

Table F.7. Summary of results of ANOVA on ITD thresholds in ITD–and ILD–trained groups with ITD–trained stimulus, described in Section 7.3.5. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted.

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F*	*p*
Unrelated					
Main effects:					
Session	4600.2	1	4600.2	14.9	0.001*
Training Group	3.3	1	3.3	0.01	0.92
Interaction:					
Session × Training Group	483.5	1	483.5	1.6	0.22
Error:					
Error	8629.1	28	308.2		

Table F.8. Summary of results of ANOVA on ILD thresholds in ITD–and ILD–trained groups with ILD–trained stimulus, described in Section 7.3.5. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted.

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F*	*p*
Unrelated					
Main effects:					
Session	20.2	1	20.2	8.1	0.008*
Training Group	0.2	1	0.2	0.07	0.97
Interaction:					
Session × Training Group	8.2	1	8.2	3.3	0.08
Error:					
Error	69.7	28	2.5		

Table F.9. Summary of results of ANOVA on ILD thresholds in ILD–trained group between diotic 500– and 4000–Hz tones, described in Section 7.3.5. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted.

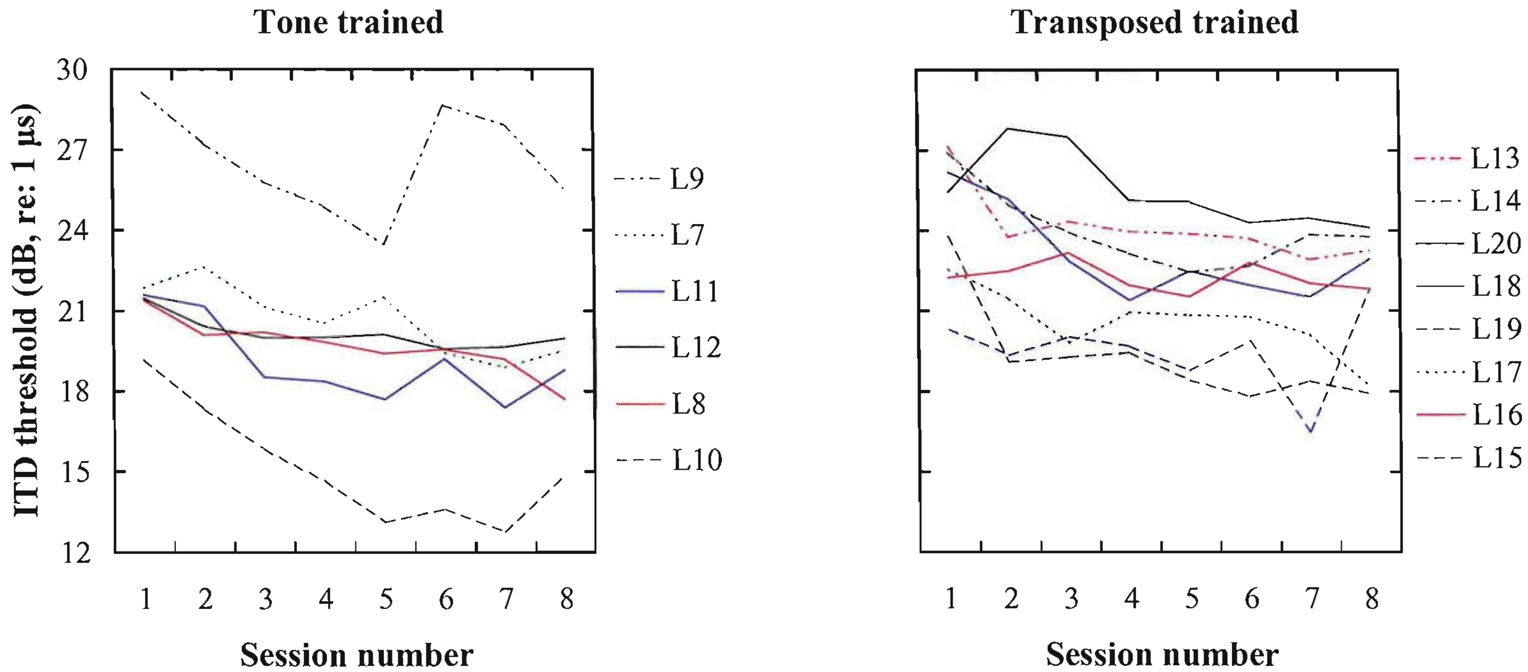
Source of variation	Sum of squares	Degrees of freedom	Mean square	*F*	*p*
Related					
Main effect:					
Stimulus	9.3	1	9.3	15.4	0.002*
Interaction:					
Stimulus × Session	3.4	1	3.4	5.6	0.03*
Error:					
Stimulus	8.5	14	0.6		
Unrelated					
Session	30.4	1	30.4	13.8	0.002*
Error	30.9	14	2.2		

Table F.10. Summary of results of ANOVA on ILD thresholds in ILD–trained group between diotic 4000– and 6000–Hz tones, described in Section 7.3.5. Significant effects at the *a priori* significance level ($p \leq 0.05$) are highlighted.

Source of variation	Sum of squares	Degrees of freedom	Mean square	*F*	*p*
Related					
Main effect:					
Stimulus	14.0	1	14.0	22.9	<0.001*
Interaction:					
Stimulus × Session	0.9	1	0.9	1.5	0.24
Error:					
Stimulus	8.6	14	0.6		
Unrelated					
Session	30.4	1	30.4	13.8	0.003*
Error	30.9	14	2.2		

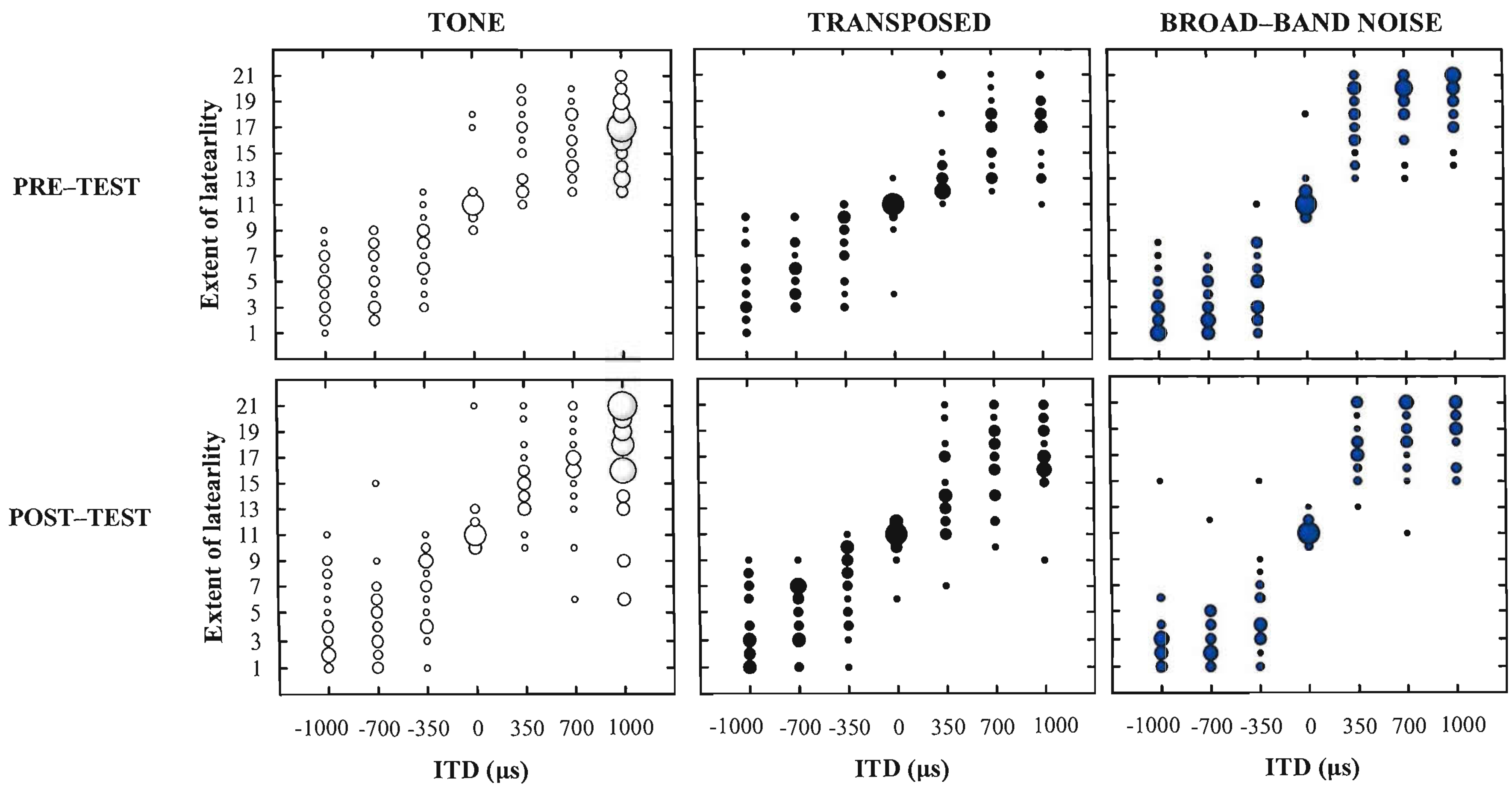
Appendix G. Individual 'learning curves' from Experiment 3

Individual learning curves from Experiment 3 (see Section 6.3.4). Session–mean logarithmically transformed ITD threshold is plotted as a function of session number (excluding familiarisation) with the trained stimulus for the tone–trained and transposed–trained groups. Lines represent different listeners.


Tone trained
Transposed trained
ITD threshold (dB, re: 1 µs)
30
27
24
21
18
15
12
1 2 3 4 5 6 7 8
Session number
L9
L7
L11
L12
L8
L10
1 2 3 4 5 6 7 8
Session number
L13
L14
L20
L18
L19
L17
L16
L15

Appendix H. Lateral–position scaling data

Ratings of lateral position, as measured in Experiment 3 (see Section 6.3.7), are plotted (overleaf) as a function of ITD for three stimuli and for measurements at pre– and post–test. The area of the circle is proportional to the number of responses across all listeners.


TONE
TRANSPOSED
BROAD–BAND NOISE
PRE–TEST
POST–TEST
Extent of latearlity
21
19
17
15
13
11
9
7
5
3
1
-1000 -700 -350 0 350 700 1000
ITD (μs)

References

Ahissar M and Hochstein S. Task difficulty and the specificity of perceptual learning. Nature 1997;387:401–6.

Ahissar M and Hochstein S. The reverse hierarchy theory of visual perceptual learning. Trends Cogn Sci 2004;8:457–64.

Alain C, Arnott SR, Hevenor S, Graham S, Grady CL. “What” and “where” in the human auditory system. Proc Natl Acad Sci USA 2001;98:12301–6.

Amitay S, Hawkey DJC, Moore DR. Auditory frequency discrimination learning is affected by stimulus variability. Percept Psychophys 2005;67:691–8.

Ari–Even Roth D, Amir O, Alauf L, Buchsenspanner S, Kishon–Rabin L. The effect of training on frequency discrimination: generalization to untrained frequencies and to the untrained ear. J Basic Clin Physiol Pharmacol 2003;14:137–50.

Ari–Even Roth D, Avrahami T, Sabo Y, Kishon–Rabin L. Frequency discrimination training: is there ear symmetry? J Basic Clin Physiol Pharmacol 2004;15:15–27.

Bakin JS, Weinberger NM. Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. Proc Natl Acad Sci USA 1996;93:11219–24.

Batra R, Fitzpatrick DC. Processing of interaural temporal disparities in the medial division of the ventral nucleus of the lateral lemniscus. J Neurophysiol 2002;88:666–75.

Batra R, Kuwada S, Fitzpatrick DC. Sensitivity to interaural temporal disparities of low– and high–frequency neurons in the superior olivary complex. I. Heterogenity of responses. J Neurophysiol 1997;78:1222–36.

Batra R, Kuwada S, Stanford TR. High–frequency neurons in the inferior colliculus that are sensitive to interaural delays of amplitude–modulated tones — evidence for dual binaural influences. J Neurophysiol 1993;70:64–80.

Bernstein LR. Auditory processing of interaural timing information: new insights. J Neurosci Res 2001;66:1035–46.

Bernstein LR. Sensitivity to interaural intensive disparities: listeners' use of potential cues. J Acoust Soc Am 2004;115:3156–60.

Bernstein LR, Trahiotis C. Detection of interaural delay in high–frequency noise. J Acoust Soc Am 1982;71:147–52.

Bernstein LR, Trahiotis C. Lateralization of sinusoidally amplitude–modulated tones: effects of spectral locus and temporal variation. J Acoust Soc Am 1985;78:514–23.

Bernstein LR, Trahiotis C. Detection of interaural delay in high–frequency sinusoidally amplitude–modulated tones, two–tone complexes, and bands of noise. J Acoust Soc Am 1994;95:3561–7.

Bernstein LR, Trahiotis C. Binaural interference effects measured with masking–level difference and with ITD– and ILD–discrimination paradigms. J Acoust Soc Am 1995;98:155–63.

Bernstein LR, Trahiotis C. Using transposed stimuli to reveal similar underlying sensitivity to interaural timing information at high and low frequencies: support for the Colburn–Esquissaud hypothesis. In: Breebaart DJ, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R, editors. Physiological and psychophysical bases of auditory function. Proceedings of the 12th International Symposium on Hearing; 2000 Aug 4–9; Mierlo, The Netherlands. Maastricht: Shaker; 2001. p. 137–44.

Bernstein LR, Trahiotis C. Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli". J Acoust Soc Am 2002;112:1026–36.

Bernstein LR, Trahiotis C. Enhancing interaural–delay–based extents of laterality at high frequencies by using "transposed stimuli". J Acoust Soc Am 2003;113:3335–47.

Bernstein LR, Trahiotis C. The apparent immunity of high–frequency "transposed" stimuli to low–frequency binaural interference. J Acoust Soc Am 2004;116:3062–9.

Bernstein LR, Trahiotis C, Hyde EL. Inter–individual differences in binaural detection of low–frequency or high–frequency tonal signals masked by narrow–band or broadband noise. J Acoust Soc Am 1998;103:2069–78.

Blauert J. Spatial hearing: the psychophysics of human sound localisation. Cambridge (MA): MIT Press; 1997.

Brand A, Behrend O, Marquardt T, McAlpine D, Grothe B. Precise inhibition is essential for microsecond interaural time difference coding. Nature 2002;417:543–7.

Breebaart J, van de Par S, Kohlrausch A. Binaural processing model based on contralateral inhibition. I. Model structure. J Acoust Soc Am 2001;110:1074–88.

British Society of Audiology. Recommended procedures for pure–tone audiometry using a manually operated instrument. Br J Audiol 1981;15:213–6.

Brown M, Irvine DRF, Park VN. Perception learning on an auditory frequency discrimination task by cats: association with changes in primary auditory cortex. Cereb Cortex 2004;14:952–65.

Buonomano DV, Merzenich MM. Cortical plasticity: from synapses to maps. Ann Rev Neurosci 1998;21:149–86.

Butler RA. An analysis of the monaural displacement of sound in space. Percept Psychophys 1987;41:1–7.

Byrne D, Dirks D. Effects of acclimatization and deprivation on non–speech auditory abilities. Ear Hear 1996;17 Suppl 1:29S–37S.

Calford MB. Dynamic representational plasticity in sensory cortex. Neuroscience 2002;111:709–38.

Casseday JH, Fremouw T, Covey E. The inferior colliculus: a hub for the central auditory system. In: Oertel D, Fay RR, Popper AN, editors. Integrative functions in the mammalian auditory pathway. New York: Springer–Verlag; 2002. p. 238–318.

Clark GM. Learning to understand speech with the cochlear implant. In: Fahle M, Poggio T, editors. Perceptual learning. Cambs, Mass; MIT Press; 2002. p. 147–60.

Colburn HS. Computational models of binaural processing. In: Hawkins HL, McMullen TA, Popper AN, Fay RR, editors. Auditory computation. New York: Springer–Verlag; 1995. p. 332–400.

Colburn HS, Durlach NI. Models of binaural interaction. In: Carterette EC, Friedman MP. Handbook of perception. Vol 4. Hearing. New York: Academic Press; 1978. p. 365–466.

Colburn HS, Esquissaud P. An auditory–nerve model for interaural time discrimination of high–frequency complex stimuli. J Acoust Soc Am 1976;59 Suppl 1:S23.

Colburn HS, Trahiotis C. Effects of noise on binaural hearing. In: Dancer A, Henderson D, Salvi R, Hamernik R. Noise induced hearing loss. New York: Mosby Yearbook; 1991. p. 293–302.

Constantinides H. Binaural hearing and auditory training [dissertation]. Bristol: University of Bristol; 2004.

Delgutte B. Physiological models for basic auditory percepts. In: Hawkins HL, McMullen TA, Popper AN editors. Auditory computation. New York: Springer–Verlag; 1995. p. 157–220.

Delhommeau K, Micheyl C, Jouvent R. Generalization of frequency discrimination learning across frequencies and ears: implications for underlying neural mechanisms in humans. J Assoc Res Otolaryngol 2005;6:171–9.

Delhommeau K, Micheyl C, Jouvent R, Collet L. Transfer of learning across durations and ears in auditory frequency discrimination. Percept Psychophys 2002;64:426–36.

Demany L. Perceptual learning in frequency discrimination. J Acoust Soc Am 1985;78:1118–20.

Demany L, Semal C. Learning to perceive pitch differences. J Acoust Soc Am 2002;111:1377–88.

Dillon H. Hearing aids. Sydney: Boomerang Press; 2000.

Drennan WR, Watson CS. Sources of variation in profile analysis. I. Individual differences and extended training. J Acoust Soc Am 2001;110:2491–7.

Duda RO. Elevation dependence of the interaural transfer function. In: Gilkey R, Anderson T, editors. Binaural and spatial hearing. Mahwah (NJ): Lawrence Erlbaum Assoc.; 1996. p. 49–76.

Durlach NI, Colburn HS. Binaural phenomena. In: Carterette EC, Friedman MP, editors. Handbook of perception. Vol 4. Hearing. New York: Academic Press; 1978. p. 365–466.

Fahle M. Perceptual learning: a case for early selection. J Vis 2004;4:879–90.

Fahle M. Perceptual learning: specificity versus generalization. Curr Opin Neurobiol 2005;15:154–60.

Fahle M, Poggio T, editors. Perceptual learning. Cambs, Mass; MIT Press; 2002.

Fine I, Jacobs RA. Comparing perceptual learning across tasks: a review. J Vis 2002;2:190–203.

Fitzpatrick DC, Kuwada S, Batra R. Transformations in processing interaural time differences between the superior olivary complex and inferior colliculus: beyond the Jeffress model. Hear Res 2002;168:79–89.

Flor H, Hoffmann D, Struve M, Diesch E. Auditory discrimination training for the treatment of tinnitus. Appl Psychophysiol Biofeedback 2004;29:113–120.

Florentine F. Relation between lateralization and loudness in asymmetrical hearing loss. J Am Audiol Soc 1976;1:243–51.

Friauf E. Developmental changes and cellular plasticity in the superior olivary complex. In: Parks TN, Rubel EW, Fay RR, Popper AN, editors. Plasticity of the auditory system. New York: Springer–Verlag; 2004. p. 49–95.

Gabriel KJ, Koehnke J, Colburn HS. Frequency dependence of binaural performance in listeners with impaired binaural hearing. J Acoust Soc Am 1992;91:336–47.

Ghose GM. Learning in mammalian sensory cortex. Curr Opin Neurobiol 2004;14:513–8.

Gilbert CD. Adult cortical dynamics. Physiol Rev 1998;78:467–85.

Goldstone RL. Perceptual learning. Annu Rev Psychol 1998;49:585–612.

Golestani N, Zatorre RJ. Learning new sounds of speech: reallocation of neural substrates. Neuroimage 2004;21:494–506.

Gottselig JM, Brandeis D, Hofer-Tinguely G, Borbely AA, Achermann P. Human central auditory plasticity associated with tone sequence learning. Learn Mem 2004;11:162–71.

Grantham DW. Interaural intensity discrimination: insensitivity at 1000 Hz. J Acoust Soc Am 1984;75:1191–4.

Grantham DW. Spatial hearing and related phenomena. In: Moore BCJ, editor. Handbook of perception and cognition, Hearing. 2nd ed. San Diego: Academic Press; 1995. p. 297–345.

Greenwood DD. A cochlear frequency–position function for several species — 29 years on. J Acoust Soc Am 1990;87:2592–605.

Griffin S, Bernstein LR, Ingham NJ, McAlpine D. Neural sensitivity to interaural envelope delays in the inferior colliculus of the guinea pig. J Neurophysiol 2005;93:3463–78.

Grimault N, Micheyl C, Carlyon RP, Bacon SP, Collet L. Learning in discrimination of frequency or modulation rate: generalization to fundamental frequency discrimination. Hear Res 2003;184:41–50.

Grimault N, Micheyl C, Carlyon RP, Collet L. Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. Percept Psychophys 2002;64:189–197.

Hafter ER. Spatial hearing and the duplex theory: how viable is the model? In: Edelman GM, Gall EW, Cowan WM, editors. Dynamic aspects of neocortical function. New York: John Wiley and Sons; 1984. p. 425–48.

Hafter ER, Carrier SC. Masking–level differences obtained with a pulsed tonal masker. J Acoust Soc Am 1970;47:1041–47.

Hafter ER, Carrier SC. Binaural interaction in low–frequency stimuli: the inability to trade time and intensity completely. J Acoust Soc Am 1972;51:1852–62.

Hafter ER, Trahiotis C. Functions of the binaural system. In: Crocker MJ. Encyclopedia of Acoustics. Vol 3. New York: John Wiley and Sons; 1997. p. 1461–79.

Hall G. Perceptual and associative learning. New York: Oxford University Press; 1991.

Hancock KE, Delgutte B. A physiologically based model of interaural time difference discrimination. J Neurosci 2004;24:7110–17.

Hartmann WM. Signals, sound, and sensation. New York: Springer–Verlag; 1998.

Hartmann WM, Wittenberg A. On the externalization of sound images. J Acoust Soc Am 1996;99:3678–88.

Hartung K and Trahiotis C. Peripheral auditory processing and investigations of the "precedence effect" which utilize successive transient stimuli. J Acoust Soc Am 2001;110:1505–13.

Häusler R, Colburn S, Marr E. Sound localization in subjects with impaired hearing. Acta Otolaryngol Suppl 1983;400:1–62.

Hawkey DJC, Amitay S, Moore DR. Early and rapid perceptual learning. Nature Neurosci 2004;7:1055–6.

Hawkins DB, Wightman FL. Interaural time discrimination ability of listeners with sensorineural hearing loss. Audiol 1980;19:495–507.

Henning GB. Detectability of interaural delay in high–frequency complex waveforms. J Acoust Soc Am 1974;55:84–90.

Henning GB. Some observations on the lateralization of complex waveforms. J Acoust Soc Am 1980;68:446–54.

Hershkowitz RM, Durlach NI. Interaural time and amplitude jnds for a 500–Hz tone. J Acoust Soc Am 1969;46:1464–7.

Hinton K, Rowan D, Mackenzie E. Comparison of adaptive procedures for binaural discrimination: test-retest reliability and bias. Poster presented at the British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness; 2004 Sept 16–17; London, UK.

Hofman PM, Vlaming MSMG, Termeer PJJ, van Opstal AJ. A method to induce swapped binaural hearing. J Neurosci Methods 2002;113:167–79.

Howell DC. Statistical methods for psychology. 4th ed. Belmont, CA, USA: Duxbury Press; 1997.

IEC 711. Occluded–ear simulator for the measurement of earphones coupled to the ear by ear inserts. Geneva: International Electrotechnical Commission; 1981.

Illing RB. Activity–dependent plasticity in the adult auditory brainstem. Audiol Neurootol 2001;6:319–45.

Irvine DRF. Physiology of the auditory brainstem. In: Popper AN, Fay RR, editors. The mammalian auditory pathway: neurophysiology. New York: Springer–Verlag; 1991. p. 153–231.

Irvine DRF, Martin RL, Klimkeit E, Smith R. Specificity of perceptual learning in a frequency discrimination task. J Acoust Soc Am 2000;108:2964–8.

ISO 226. Acoustics. Normal equal–loudness–level contours. Geneva: International Organization for Standardization; 2003.

ISO 389–2. Acoustics. Reference zero for the calibration of audiometric equipment. Part 2: reference equivalent threshold sound pressure levels for pure tones and insert earphones. Geneva: International Organization for Standardization; 1994.

Javer AR, Schwarz WF. Plasticity in human directional hearing. J Otolaryngol 1995;24:111–7.

Jeffress LA. A place theory of sound localization. J Comp Physiol Psychol 1948;41:35–9.

Joris PX, Carney LH, Smith PH, Yin TCT. Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristics frequency. J Neurophysiol 1994;71:1022–35.

Joris PX, Schreiner CE, Rees A. Neural processing of amplitude–modulated sounds. Physiol Rev 2004;84:541–77.

Joris PX, Smith PH, Yin TCT. Coincidence detection in the auditory system: 50 years after Jeffress. Neuron 1998;21:1235–38.

Joris PX, Yin TCT. Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences. J Neurophysiol 1995;73:1043–62.

Karmarkar UR, Buonomano DV. Temporal specificity of perceptual learning in an auditory discrimination task. Learn Mem 2003;10:141–7.

King AJ, Schnupp JWH, Doubell TP. The shape of ears to come: dynamic coding of auditory space. Trends Cog Sci 2001;5:261–70.

Klump RG, Eady HR. Some measurements of interaural time difference thresholds. J Acoust Soc Am 1956;28:859–60.

Koehnke J, Besing J. Binaural performance in listeners with impaired hearing: aided and unaided results. In: Gilkey R, Anderson T, editors. Binaural and spatial hearing. Mahwah (NJ): Lawrence Erlbaum Assoc.; 1996. p. 725–51.

Koehnke J, Colburn HS, Durlach NI. Performance in several binaural–interaction experiments. J Acoust Soc Am 1986;79:1558–62.

Koehnke J, Culotta CP, Hawley ML, Colburn HS. Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing. Ear Hear 1995;16:331–53.

Kuhn GF. Physical acoustics and measurements pertaining to directional hearing. In: Yost WA, Gourevitch G, editors. Directional hearing. New York: Springer–Verlag; 1987. p. 3–25.

Kujala T, Karma K, Ceponiene R, Belitz S, Turkkila P, Trevaniemi M, Näätäven R. Plastic neural changes and reading improvement caused by audiovisual training in reading–impaired children. Proc Natl Acad Sci USA 2001;98:10509–14.

Leakey DM, Sayers B McA, Cherry C. Binaural fusion of low– and high–frequency sounds. J Acoust Soc Am 1957;30:222.

Leek MR. Adaptive procedures in psychophysical research. Percept Psychophys 2001;63:1279–92.

Levitt H. Transformed up–down methods in psychoacoustics. J Acoust Soc Am 1971;49:467–77.

Litovsky RY, Colburn HS, Yost WA, Guzman SJ. The precedence effect. J Acoust Soc Am 1999;106:1633–54.

Litovsky, RY, Hawley ML, Fligor BJ, Zurek PM. Failure to unlearn the precedence effect. J Acoust Soc Am 2000;108:2345–52.

Long CJ. Bilateral cochlear implants: basic psychophysics [dissertation]. Cambridge (MA): MIT; 2000.

Macmillan NA, Creelman CD. Detection theory: a user's guide. New York: Cambridge University Press; 1991.

Macpherson EA, Middlebrooks JC. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. J Acoust Soc Am 2002;111:2219–36.

Marks LE. Binaural summation of the loudness of pure tones. J Acoust Soc Am 1978;64:107–13.

Marvit P, Florentine M, Buus S. A comparison of psychophysical procedures for level–discrimination thresholds. J Acoust Soc Am 2003;113:3348–61.

McAlpine D, Grothe B. Sound localization and delay lines — do mammals fit the model? Trends Neurosci 2003;26:347–50.

McAlpine D, Jiang D, Palmer AR. A neural code for low–frequency sound localization in mammals. Nature Neurosci 2001;4:396–401.

McFadden D, Jeffress LA, Russell WE. Individual differences in sensitivity to interaural differences in time and level. Percept Motor Skills 1973;37:755–61.

McFadden D, Pasanen EG. Lateralization at high frequencies based on interaural time differences. J Acoust Soc Am 1976;59:634–9.

Merzenich MM, Jenkins WM, Johnston P, Schreiner C, Miller SL and Tallal P. Temporal processing deficits of language–learning impaired children ameliorated by training. Science 1996;271:77–81.

Middlebrooks JC. Auditory mechanisms underlying a neural code for space in the cat's superior colliculus. In: Edelman GM, Gall EW, Cowan WM, editors. Auditory function. Neurological bases of hearing. New York: Wiley; 1988. p. 431–55.

Middlebrooks JC, Green DM. Directional dependence of interaural envelope delays. J Acoust Soc Am 1990;87:2149–62.

Middlebrooks JC, Xu L, Furukawa S, Mickey BJ. Location signalling by cortical neurons. In: Oertel D, Fay RR, Popper AN, editors. Integrative functions in the mammalian auditory pathway. New York: Springer–Verlag; 2002. p. 319–57.

Mollon JD, Danilova MV. Three remarks on perceptual learning. Spat Vis 1996;10:51–8.

Moore BCJ. Parallels between frequency selectivity measured psychophysically and in cochlear mechanics. Scand Audiol Suppl 1986;25:139–52.

Moore BCJ. Frequency analysis and masking. In: Moore BCJ, editor. Handbook of perception and cognition. Hearing. 2nd ed. San Diego: Academic Press; 1995. p. 161–205.

Moore BCJ, Launer S, Vickers D, Baer T. Loudness of modulated sounds as a function of modulation rate, modulation depth, modulation waveform and overall level. In: Palmer AR, Rees A, Summerfield AQ, Meddis R, editors. Psychophysical and physiological advances in hearing. Proceedings of the 11th International Symposium on Hearing; 1997 Aug 1–6; Grantham, UK. London: Whurr Publishers; 1998. p. 465–71.

Moore DR, Amitay S, Hawkey DCJ. Auditory perceptual learning. Learn Mem 2003;10:83–5.

Moore DR, King AJ. Plasticity of binaural systems. In: Parks TN, Rubel EW, Fay RR, Popper AN, editors. Plasticity of the auditory system. New York: Springer–Verlag; 2004. p. 96–172.

Moore DR, Rosenberg JF, Coleman JS. Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children. Brain Lang 2005;94:72–85.

Nuetzel JM. Spectral and temporal factors in the lateralization of complex auditory signals [dissertation]. Berkeley: University of California; 1976.

Nuetzel JM, Hafter ER. Lateralization of complex waveforms: effects of fine structure, amplitude and duration. J Acoust Soc Am 1976;60:1339–46.

Nuetzel JM, Hafter ER. Discrimination of interaural delays in complex waveforms: spectral effects. J Acoust Soc Am 1981;69:1112–8.

Oertel D. The role of timing in the brain stem auditory nuclei of vertebrates. Annu Rev Physiol 1999;61:497–519.

Oxenham AJ, Bernstein JGW, Penagos H. Correct tonotopic representation is necessary for complex pitch perception. Proc Natl Acad Sci USA 2004;101:1421–5.

Oxenham AJ, Penagos H. Pitch perception of transposed stimuli. Poster presented at the 26th Annual Mid Winter Research Meeting of the Association for Research in Otolaryngology; 2003 Feb 22–27; Daytona Beach, Florida, USA.

Palmer AR. Neural signal processing. In: Moore BCJ, editor. Handbook of perception and cognition. Hearing. 2nd ed. San Diego: Academic Press; 1995. p. 75–121.

Palmer AR, Russell IJ. Phase–locking in the cochlear nerve of the guinea–pig and its relation to the receptor potential of the inner hair–cells. Hear Res 1986;24:1–15.

Park TJ, Klug A, Holinstat M, Grothe B. Interaural level difference processing in the lateral superior olive and the inferior colliculus. J Neurophysiol 2004;92:289–301.

Plenge G. On the differences between localization and lateralization. J Acoust Soc Am 1974;56:944–51.

Rayleigh Lord (Strutt JW, 3rd Baron of Rayleigh). On our perception of sound direction. Phil Mag 1907;13:214–32.

Recanzone GH, Makhamra SDDR, Guard DC. Comparison of relative and absolute localization ability in humans. J Acoust Soc Am 1998;103:1085–97.

Recanzone GH, Schreiner CE, Merzenich MM. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J Neurosci 1993;13:87–103.

Robinson K, Gatehouse S. Changes in intensity discrimination following monaural long–term hearing aid use. J Acoust Soc Am 1995;97:1183–90.

Robinson K, Gatehouse S. The time course of effects on intensity discrimination following monaural fitting of hearing aids. J Acoust Soc Am 1996;99:1255–8.

Robinson K, Summerfield AQ. Adult auditory learning and training. Ear Hear 1996;17 Suppl 1:51S–65S.

Rule SJ. Lateralization of high frequency, amplitude modulated sounds as a function of carrier frequency. Can J Exp Psychol 1994;48:435–43.

Rule SJ, Nickolaychuk BR. Lateralization of high frequency sounds as a function of interaural amplitude disparity. Can J Exp Psychol 1995;49:368–75.

Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. Auditory training improves neural timing in the human brainstem. Behav Brain Res 2005;156:95–103.

Rutkowski RG. An investigation into the effects of interaural place mismatches on the detection of interaural timing cues with sinusoidally amplitude–modulated and transposed stimuli [dissertation]. Southampton: University of Southampton; 2005.

Saberi K. Some considerations on the use of adaptive methods for estimating interaural–delay thresholds. J Acoust Soc Am 1995;98:1803–6.

Saberi K, Antonio JV. Precedence–effect thresholds for a population of untrained listeners as a function of stimulus intensity and interclick interval. J Acoust Soc Am 2003;114:420–9.

Saberi K, Antonio JV, Petrosyan A. A population study of the precedence effect. Hear Res 2004;191:1–13.

Saberi K, Perrott DR. Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate. J Acoust Soc Am 1990;87:1732–7.

Schiano JL, Trahiotis C, Bernstein LR. Lateralization of low–frequency tones and narrow bands of noise. J Acoust Soc Am 1986;79:1563–70.

Schulze H, Neubauer H, Ohl FW, Hess A. Representation of stimulus periodicity and its learning induced plasticity in the auditory cortex: recent findings and new perspectives. Acta Acustica 2002;88:399–407.

Seidl AH, Grothe B. Development of sound localization mechanisms in the Mongolian gerbil is shaped by early acoustic experience. J Neurophysiol 2005;94:1028–36.

Shackleton TM, Skottun BC, Arnott RH, Palmer AR. Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of guinea pigs. J Neurosci 2003;23:716–24.

Siegel RA and Colburn HS. Binaural processing of noisy stimuli: internal/external noise ratios for diotic and dichotic stimuli. J Acoust Soc Am 1989;86:2122–8.

Skinner MW, Ketten DR, Holden LK, Harding GW, Smith PG, Gates GA, et al. CT–derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus–22 recipients. J Assoc Res Otolaryngol 2002;3:332–50.

Skottun BC, Shackleton TM, Arnott RH, Palmer AR. The ability of inferior colliculus neurons to signal differences in interaural delay. Proc Natl Acad Sci USA 2001;98:14050–4.

Smith PH, Joris PX, Yin TCT. Projections of physiologically characterised spherical bushy cells axons from the cochlear nucleus of the cat: evidence for delay lines to the medial superior olive. J Comp Neurol 1993;331:245–60.

Smith–Olinde L, Koehnke J, Besing J. Effects of sensorineural hearing loss on interaural discrimination and virtual localization. J Acoust Soc Am 1998;103:2084–99.

Smoski WJ, Trahiotis C. Discrimination of interaural temporal disparities by normal–hearing listeners and listeners with high–frequency sensorineural hearing loss. J Acoust Soc Am 1986;79:1541–7.

Stern RM, Trahiotis C. Models of binaural interaction. In: Moore BCJ, editor. Handbook of perception and cognition. Hearing. 2nd ed. San Diego: Academic Press; 1995. p. 347–86.

Syka J. Plastic changes in the central auditory system after hearing loss, restoration of function, and during learning. Physiol Rev 2002;10:601–36.

Terhune JM. Localization of pure tones and click trains by untrained humans. Scand Audiol 1985;14:125–31.

Tollin DJ. The lateral superior olive: a functional role in sound source localization. Neuroscientist 2003;9:127–43.

Trahiotis C, Bernstein LR. Lateralization of bands of noise and sinusoidally amplitude–modulated tones: effects of spectral locus and bandwidth. J Acoust Soc Am 1986;79:1950–57.

Trahiotis C, Bernstein LR, Buell TN, Spektor Z. On the use of adaptive procedures in binaural experiments. J Acoust Soc Am 1990;87:1359–61.

Tremblay K, Kraus N, Carrell TD, McGee T. Central auditory system plasticity: generalization to novel stimuli following listening training. J Acoust Soc Am 1997;102:3762–73.

Trussell LO. Synaptic mechanisms for coding timing in auditory neurons. Annu Rev Physiol 1999;61:477–96.

Tyler RS, Summerfield AQ. Cochlear implantation: relationships with research on auditory deprivation and acclimatization. Ear Hear 1996;17 Suppl 1:38–50.

van de Par S, Kohlrausch A. A new approach to comparing binaural masking level differences at low and high frequencies. J Acoust Soc Am 1997;101:1671–80.

van der Elst W, van Boxtel MP, van Breukelen GJ, Jolles J. Rey's verbal learning test: normative data for 1855 healthy participants aged 24–81 years and the influence of age, sex, education, and mode of presentation. J Int Neuropsychol Soc 2005;11:290–302.

van Hoesel RJM. Exploring the benefits of bilateral cochlear implants. Audiol Neurootol 2004;9:234–46.

van Hoesel RJM, Clark GM. Psychophysical studies with two binaural cochlear implant subjects. J Acoust Soc Am 1997;102:495–507.

van Hoesel RJM, Tyler RS. Speech perception, localization, and lateralization with bilateral cochlear implants. J Acoust Soc Am 2003;113:1617–30.

Watanabe T, Náñez JE, Sasaki Y. Perceptual learning without perception. Nature 2001;413:844–8.

Watson CS. Time course of auditory perceptual learning. Ann Otol Rhinol Laryngol Suppl 1980;74:96–102.

Watson CS. Auditory perceptual learning and the cochlear implant. Am J Otol 1991;12:S73–9.

Webster FA. The influence of interaural phase on masked threshold. I. The role of interaural time–deviation. J Acoust Soc Am 1951;23:452–62.

Wightman FL, Kistler DJ. The dominant role of low–frequency interaural time differences in sound localization. J Acoust Soc Am 1992;91:1648–61.

Wilson BS, Lawson DT, Müller JM, Tyler RS, Kiefer J. Cochlear implants: some likely next steps. Annu Rev Biomed Eng 2003;2:207–49.

Wright BA. Why and how we study human learning on basic auditory tasks. Audiol Neurootol 2001;6:207–10.

Wright BA, Bumonomano DV, Mahncke HW, Merzenich MM. Learning and generalization of auditory temporal–interval discrimination in humans. J Neurosci 1997;17:3956–63.

Wright BA, Fitzgerald MB. Different patterns of human discrimination learning for two interaural cues to sound–source location. Proc Natl Acad Sci USA 2001;98:12307–12.

Wright BA, Fitzgerald MB. Sound–discrimination learning and auditory displays. Proceedings of the International Conference on Auditory Displays; 2003 July 6–9;

Boston, MA, USA. Available from: http://www.icad.org/websiteV2.0/Conferences/ICAD2003/index.html

Yin TCT. Neural mechanisms of encoding binaural localization cues in the auditory brainstem. In: Oertel D, Fay RR, Popper AN, editors. Integrative functions in the mammalian auditory pathway. New York: Springer–Verlag; 2002. p. 99–159.

Yin TCT, Chan JCK. Interaural time sensitivity in medial superior olive in cat. J Neurophysiol 1990;64:465–88.

Yost WA. Lateral position of sinusoids presented with interaural intensive and temporal disparities. J Acoust Soc Am 1981;70:397–409.

Yost WA, Dye RH. Discrimination of interaural differences of level as a function of frequency. J Acoust Soc Am 1988;83:1846–51.

Yost WA, Wightman FL, Green DM. Lateralization of filtered clicks. J Acoust Soc Am 1971;50:1526–31.

Zwislocki J, Feldman RS. Just noticeable differences in dichotic phase. J Acoust Soc Am 1956;28:860–4.

Zwislocki J, Maire F, Feldman, AS, Rubin H. On the effect of practice and motivation on the threshold of audibility. J Acoust Soc Am 1958;30:254–62.

Zwislocki JJ, Relkin EM. On a psychophysical transformed–rule up and down method converging on a 75% level of correct responses. Proc Natl Acad Sci USA 2001;98:4811–4.