Perceptual Adaptation to Binaurally Mismatched Frequency-to-Place Maps: Implications for Bilateral Stimulation with Cochlear Implants

Catherine McKenna Siciliano

UCL

Submitted for the degree of Doctor of Philosophy

2010

Declaration

I, Catherine McKenna Siciliano, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

The following publication has originated from this work:

Siciliano, C., Faulkner, A., Rosen, S. and Mair, K. (*Accepted*). Resistance to learning binaurally mismatched frequency-to-place maps: Implications for bilateral stimulation with cochlear implants. *Journal of the Acoustical Society of America*.

Statement of Conjoint Work

One experiment of this thesis was conducted by another UCL student, K. Mair, while she was working on a Wellcome Trust-funded Summer Vacation Scholarship [VS/06/UCL/A12]. The idea for the experiment was directly based on findings from my own work, and K. Mair was brought in to run this experiment in the interests of speeding up data collection, since the experimenter effort was very high.

I created all testing stimuli for the experiment and performed all statistical analyses based on the data that are included in the thesis. I contributed to the design of the experiment, though some of the details were finalized in my absence by my supervisors Dr. Andrew Faulkner and Professor Stuart Rosen. K. Mair was responsible for recruiting subjects and conducting the experiment. The data from the experiment is crucial to this thesis, and K. Mair is clearly credited in the thesis.

Signed ____

Abstract

Simulations of monaural cochlear implants in normal-hearing listeners have shown that the deleterious effects of upward spectral shifting on speech perception can be overcome with training. This thesis examines whether the same is true when simulating bilateral stimulation. Can listeners adapt to upward-shifted speech information presented together with contralateral unshifted information? In two series of experiments, perceptual adaptation was investigated for both speech in quiet with a large interaural spectral mismatch, and speech in noise with a moderate interaural spectral mismatch.

For speech in quiet, a six-channel dichotic sine-carrier vocoder simulated the binaurally mismatched frequency-to-place map. Odd channels were presented to one ear with an upward shift equivalent to a 6 mm basilar membrane distance, while even channels were presented to the contralateral ear unshifted. For speech in noise, the number of vocoded channels was increased to ten, and the upward spectral shift applied to the odd channels was decreased to 3.8 mm. Prior to vocoding, speech was combined with speech-shaped noise at a signal-to-noise ratio of 10 dB (or 0 dB for vowels). Listeners were trained with Connected Discourse Tracking for 5.3 hours or 10 hours, with the binaurally mismatched processor and/or just the shifted monaural bands. Speech perception was tested with sentence and vowel tests before, during and after training.

Listeners showed adaptation to the upwardly shifted speech, but for nearly every speech test, intelligibility with the binaurally mismatched processor matched intelligibility with just the unshifted bands. Consistent with earlier findings with monaural spectral shifts, then, this research suggests that listeners are capable of adapting to a spectral shift, even in the presence of background noise. However, they appear to be resistant to integrating mismatches in frequency-place maps between the ears. A theory of “better ear” listening is proposed to account for this resistance. The findings are consistent with psychophysical studies of binaural hearing, which show maximal ITD and ILD sensitivity for similar interaural cochlear places. In optimizing bilateral cochlear implants for speech perception, it may thus be important to keep frequency-to-place maps similar in the two ears.

Table of Contents

Figures

Tables

Acknowledgements

First and foremost I would like to thank my supervisor, Dr. Andrew Faulkner, for all of his support and guidance before and throughout this work. I am also grateful to Dr. Stuart Rosen, my second supervisor, for his advice and especially his technical support. I would also like to thank the following: Katharine Mair for her contribution to some of the work in this thesis and her enduring patience in recording the stimuli for the latter part of this work; Steve Nevard for his technical support, especially in setting up the recordings; and Mary Wykes for admistering subject payments.

For funding all or part of this work, including dissemination, I am grateful to the European Commission [FP6 – 004171 HEARCOM], the UCL Staff Development and Training Unit, the Wellcome Trust [VS/06/UCL/A12], Deafness Research UK and the US National Institute on Deafness and Other Communication Disorders (NIDCD).

Finally, I thank Kim for his steadfast support, and my parents for their encouragement throughout my academic endeavours.

Chapter 1. Introduction

The advantages of binaural hearing have been well established in normal hearing listeners, yet the possibility of binaural rehabilitation for those with a profound hearing loss has only been realised in recent years. Normal hearing listeners show a binaural advantage for speech that arises from binaural squelch, the head-shadow effect and binaural redundancy, and all three abilities enhance the perception of speech in noise. For those with a cochlear implant (CI), the restoration of binaural hearing is especially important because many CI users perform poorly when listening to speech in noise. Limitations in cochlear implant technology and/or the listener's auditory processing capacities may preclude the re-establishing of all of these benefits in CI users, but a binaural advantage for speech perception has been reported in many patients with bilateral cochlear implants (Litovsky, Parkinson, Arcaroli, and Sammeth, 2006; Dorman and Dahlstrom, 2004; Dunn, Tyler, Oakley, Gantz, and Noble, 2008; Tyler, Gantz, Rubinstein, Wilson, Parkinson, Wolaver, Preece, Witt, and Lowder, 2005; Tyler, Dunn, Witt, and Noble, 2007; Wackym, Runge-Samuelson, Firszt, Alkaf, and Burg, 2007).

While the potential benefits are manifold, the use of bilateral devices raises new questions regarding the appropriate cochlear frequency-to-place mapping for CI speech processors. Cochlear implant electrode arrays are designed for an insertion of 25 mm into the typically 35 mm cochlea, but the insertion achieved at surgery varies widely and is often much shallower than 25 mm. Estimates based on in vivo computed tomography measurements from 26 Nucleus-22 implant recipients showed cochlear lengths ranging between 29.1 – 37.4 mm, and electrode array insertion depths ranging from 11.9 – 25.9 mm (Ketten, Skinner, Wang, Vannier, Gates, and Neely, 1998; Skinner, Ketten, Holden, Harding, Smith, Gates, Neely, Kletzker, Brunsden, and Blocker, 2002). For shallow electrode insertions, the resulting tonotopic misalignment between the analysis filter of the implant speech processor and the characteristic frequency (CF) of the basilar membrane at a given electrode contact means that the most apical electrode of a cochlear implant will stimulate nerves normally "tuned" to higher frequencies (i.e. a basalward shift). Previous research exploring the effect of a monaural shallowly inserted electrode array on the perception of speech has shown that listeners are able to adapt to the tonotopic misalignment, at least to some extent (Faulkner, Rosen, and Norman, 2006; Fu, Shannon, and Galvin III, 2002; Fu and Galvin III, 2003; Rosen, Faulkner, and Wilkinson, 1999). However, bilateral cochlear implants are implanted and operated independently, and are therefore likely to introduce mismatches not only relative to the natural frequency-place mapping in the normal cochlea but also in relation to each other. If CI electrode arrays are inserted to different depths, then what effect does such a mismatch in frequency-place mapping between the ears have on speech perception? Can listeners learn to accommodate binaurally mismatched frequency-to-place maps in the same way they have been shown to for monaural basalward spectral shifts?

This thesis presents evidence from four experiments exploring perceptual adaptation to binaurally mismatched frequency-to-place maps in simulations of bilateral CI processing in normally hearing listeners. It draws on the same methods used commonly in researching adaptation for the monaural case (Faulkner et al., 2006; Rosen et al., 1999). In the first two experiments, perceptual adaptation is explored for a large spectral mismatch between the ears, such as might arise from one fully inserted and one shallowly inserted CI electrode, for speech in quiet. In the final two experiments, adaptation is explored for a more moderate spectral mismatch and for speech in noise. Secondary aims of the research were to examine the effects of training condition on adaptation and also to explore whether adaptation is based on peripheral or more central auditory cues.

The thesis begins in Chapter 2 with an introduction to speech perception, normal hearing and cochlear implants and a review of perceptual adaptation to monaural frequency shifts. Chapter 3 introduces the significance of binaural hearing and its importance for cochlear implant patients, and also the research problems that will form the basis of the experimental research. Chapter 4 presents evidence from the first two of the four experiments, where perceptual adaptation is explored for speech in quiet with a large degree of binaural mismatch. This chapter also delves into the issues of training condition and duration. In Chapter 5, the research is extended to examine adaptation to a more moderate degree of mismatch in the presence of background noise. Further issues explored include the role of specific ear cues in adaptation, the retention of adaptation in the absence of training and, again, the role of the speech training condition in facilitating adaptation. The thesis concludes in Chapter 6 with a summation of the evidence from the four experiments, a working hypothesis that accounts for the findings, limitations of the results and their implications for bilateral cochlear implant fittings, and finally areas for future work.

Chapter 2. Speech Perception and Perceptual Adaptation with Cochlear Implants

2.1 Perceptually important acoustic properties of speech

Speech sounds can be classified into vowels and consonants according to the way they are produced and thus their acoustic properties. The source filter theory of speech production is a model to describe how speech is produced. In the model, a sound source, originating from the ex-flow of air from the lungs, is modified by the resonance of the vocal tract to produce the sounds of speech, with the cavities in the vocal tract taking on the properties of simple Helmholtz resonators. In this model, individuals can vary the rate of vocal fold vibration (associated with relative pitch within an individual) and the particular articulation (eg. timbre) independently to create speech utterances with varying intonation contours and speech sounds, respectively. As with any resonator, the larger the tube, the lower the resonant frequencies. Thus, males will have lower frequency resonances than females, and both males and females will have lower frequency resonances than children. Each speaker can individually alter the rate of their vocal fold vibration to create relative changes in pitch contours that are important in the perception of speech.

For vowels, the sound source is the vibration of the vocal folds. When a person articulates a vowel, they move their tongue and mouth which changes the shape of the vocal tract and thus modifies its resonant properties. The vowel that is produced is then characterised by its *formant frequencies*, which are the resonant frequencies of the vocal tract for a given vowel articulation. The frequencies most important for speech perception fall within the range of 100 – 10,000 Hz, and the most important

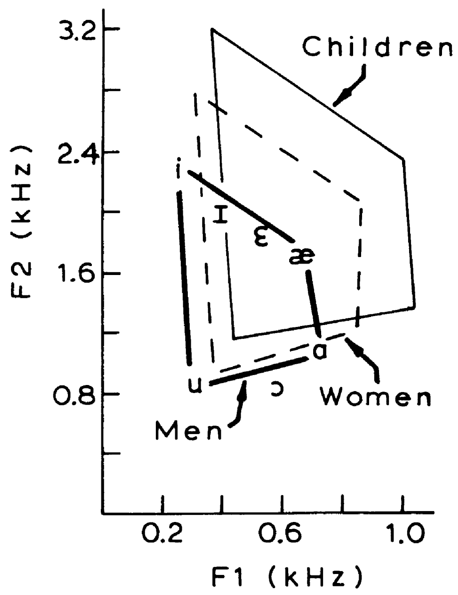

Figure 2-1. A vowel "quadrilateral" depicting the F1 and F2 resonant frequencies for English vowels. The dark solid line indicates men, the dashed line indicates women, and the solid line indicates children. This figure is reprinted with permission from R. Kent, C. Read and Singular Publishing Group (now Cengage). (Kent and Read, 1992). © 1992, Singular.

acoustic characteristics for vowels are the first two formant frequencies. The first formant frequency (F1) increases with an increasing opening of the mouth, while the second formant frequency (F2) increases with increasing advancement of the tongue root. The first two vowel formant frequencies occur in the lower frequencies for most talkers, up to a maximum of 700 – 1000 Hz for F1, and a maximum of 2400 – 3200 Hz for F2. Figure 2-1 shows the F1 and F2 coordinates of English vowel productions for men, women and children.

Despite the large variability of vowel formant frequencies both across and within speakers, speech recognition remains highly robust, suggesting that absolute vowel formant frequencies are not likely to underpin the accurate recognition of vowels. Theories of speech recognition that account for this variability posit some form of speaker normalisation, such as computing the relative differences between formants before categorization (Fant, 1976).

Consonants are classified according to three dimensions which include the *manner* in which they are produced, their *place* of articulation, and their *voicing,* or whether vocal fold vibration is sustained throughout their articulation. Manner is a way of describing the type of vocal tract constriction of a consonant, and manner consonant classes include stops, fricatives, affricates, nasals, glides and liquids. As an example, stop consonants [b, d, g, p, t, k] are produced by making then releasing a complete closure somewhere in the vocal tract. Fricatives [f, v, s, sh, ɵ, ð] are produced by forming a partial closure in the vocal tract and forcing air through the closure, which creates a turbulence of noise. Fricatives are relatively high in frequency, and they increase in frequency as the location of the closure is moved forward in the vocal tract, since the cavity of resonance for fricatives occurs in the front of the vocal tract after the point of constriction. Affricates [ʧ, ʤ] are comprised of a complete closure followed by frication, and thus have similar acoustic properties to both the stop and fricative components of which they are comprised. Nasals [m, n, ŋ] are produced in a manner similar to stop consonants, except that the nasal cavity remains open so that resonance in the nasal cavity occurs. Glides [w, j] have properties of both vowels and consonants, and are sometimes referred to as semi-vowels. They are produced by a marked narrowing then gradual release of the articulators. They are characterized primarily by a large change in F2 into the following vowel. Liquids [l, r] have properties of both consonants and glides, and are produced rapidly. [r] is characterized by a sharp drop in F3.

Within each manner class, consonants can be further classified according to their place of articulation and whether they are produced with voicing. For example, within the stop category, [p, t] are bilabials, [t, d] are alveolars and [k, g] are velars. The voiced stops are [b, d, g] and the voiceless stops [p, t, k]. Consonant perception relies on both slow spectral changes such as formant transitions and rapid spectral transitions such as those that occur at the release bursts of stops. For example, the perception of the stop manner class is cued by a relatively silent period corresponding to the period of closure in the vocal tract, followed by the rapid release burst. The perception of the place of a stop consonant is further cued by formant transitions from the previous vowel leading into the stop, and from the stop leading into the succeeding vowel. The perception of voicing of a stop is cued by the voice onset time (VOT), or the time elapsed between the release of the stop and the start of vocal fold vibration for the succeeding vowel. Though VOT is used to distinguish voiced and voiceless consonants, it is the relative length of the VOT within the phonetic repertoire of a given language that cues voicing, and not an absolute time period.

2.2 Hearing in the normal cochlea

Sound is processed by the ear in a series of signal transformations that convert acoustic vibrations into an electrical signal that is then interpreted by the brain. Figure 2-2 shows a schematic representation of the human hearing mechanism. Sound is modified by the pinna of the outer ear and then passes through the ear canal to the middle ear. The middle ear contains three bones (ossicles) that, when vibrated by the acoustic signal, convert the acoustic vibration of sound into mechanical vibrations of the fluid in the inner ear through vibration of the eardrum. The displacement of the fluid in the inner ear in turn causes the basilar membrane to vibrate. The inner hair cells, which lie on the basilar membrane, bend according to the motion of the basilar membrane, and as they bend they release neurotransmitters which cause the afferent neurons to fire.

The cochlea is the snail-shaped cavity of the inner ear which contains the basilar membrane. Pioneering work by von Bekesy (1960) showed that the basilar membrane is responsible for analyzing the incoming signal into its frequency

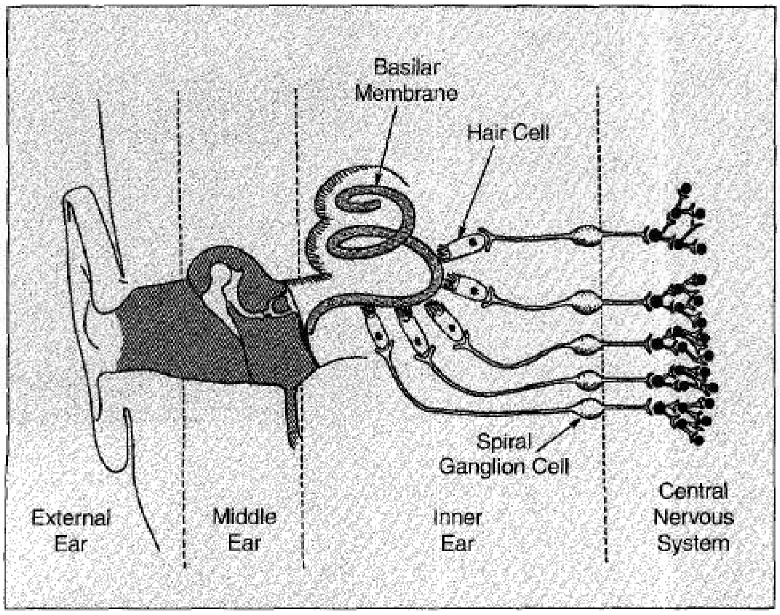

Figure 2-2. Diagram of the human hearing mechanism. Sounds pass from the external ear to the middle ear, where the movement of the ossicular chain transduces mechanical energy from air to the fluids of the inner ear. The activation of the hair cells on the basilar membrane is transmitted to the spiral ganglion cells, which connect to higher centres in the brain. This figure is reprinted with permission from P. Loizou and IEEE. (Loizou, 1998). © 1998, IEEE.

components. This is largely due to the resonant properties of the membrane itself and has been called the *passive mechanism*. The basilar membrane is tautly connected to the basal end (oval window) of the cochlea, but at the apex, the membrane is loose. When the fluid inside the cochlea vibrates according to the acoustic signal transmitted at the middle ear, it creates a ‘travelling wave’ along the basilar membrane. According to the resonant properties of the basilar membrane, the travelling wave created in response to high frequencies will have maximal displacement at the basilar end, while a wave created in response to lower frequencies will have maximal displacement at the apical end. For complex signals which contain multiple frequencies, the basilar membrane will be displaced in multiple locations. In essence, then, the basilar membrane performs a Fourier analysis on the incoming signal. A schematic representation is depicted in Figure 2-.

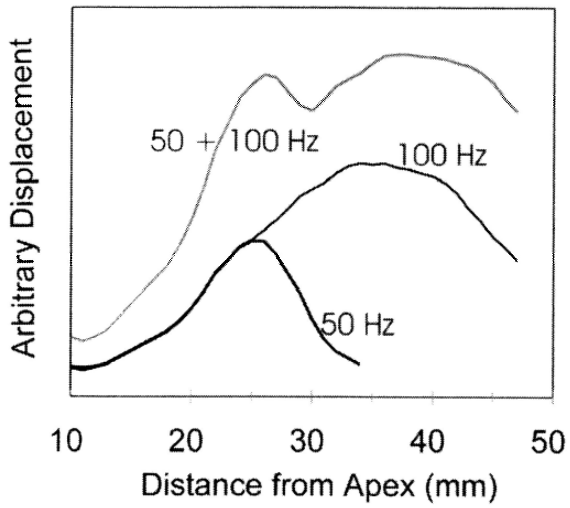

Figure 2-2. Basilar membrane displacement as a function of distance from the apex for simple (50, 100 Hz) and complex (50 + 100 Hz) tones. This adapted figure appears in (Yost, 2000) and is based on (Tonndorf, 1962). The adapted figure is reprinted with permission from the W. Yost, Elsevier, and the Acoustical Society of America. ©2000 Elsevier; © 1962, Acoustical Society of America.

As the membrane responds selectively to certain frequencies, hair cells are bent and stimulate nerve structures at that location along the basilar membrane which are also selectively responsive to these certain frequencies. This frequency-tuning, or *tonotopicity*, is carried on through afferent innervation to higher centres in the brain, which have also been shown to be similarly tonopically organized (Merzenich and Reid, 1974). Because each place along the basilar membrane can be described as best responding to a particular frequency, this tonotopicity of the basilar membrane has been called *place theory* or *place coding* (Loizou, 1998). The characteristic frequency (CF) is defined as the frequency for which a given cochlear place is most responsive. It is this tonotopicity of the basilar membrane that retains the information about the acoustic sound signal, and that is ultimately critical to our perception of speech and other sounds (Loizou, 1998).

Information about the frequency of a sound is further encoded by the rate of firing of the action potentials along the auditory nerve, which occurs at an interval proportional to the period of the incoming signal for frequencies up to 5000 Hz. *Phase locking*, as this type of frequency-encoding has been called, is thought to be used in the perception of pitch (Moore, 2003). The information conveyed through the rate of nerve firing along the basilar membrane is referred to as the *temporal code* (Moore, 2003).

2.3 Sensorineural hearing loss and cochlear implants

There are two types of hair cells in the inner ear which appear to have different roles in hearing. The inner hair cells (IHCs), depicted in Figure 2-2, are thought to be directly responsible for transmitting information about sound frequency to higher centres in the brain, as shown in the figure by their connection to the spiral ganglia cells and then the central nervous system. The outer hair cells (OHCs) appear to play a more supportive role, acting to enhance sensitivity and frequency resolution within the cochlea which, in turn, modulates the transduction of small vibrations into neural impulses via the inner hair cells (Yost, 2000). Outer hair cells are motile, and in changing shape they actively influence the mechanics of the basilar membrane by sharpening its frequency-tuning and increasing overall sensitivity. This action contributes to what has been referred to as the *active mechanism* of hearing, and it is highly dependent on the physiological condition of the inner ear (Moore, 1998).

Sensorineural hearing loss occurs as a result of hair cell death in the inner ear. When IHCs die, the mechanism for transduction of the mechanical energy into electrical nerve stimulation is damaged. Perceptually, IHC death is characterised by a rise in the thresholds of hearing for sounds over a range of frequencies. This is because a greater amount of activation is needed to achieve a “threshold of neural activity” in basilar membrane areas with IHC death (Moore, 1998). IHC death in the absence of OHC death is rare, and indeed it is the OHCs that are more vulnerable. As the OHCs contribute to the active mechansim of hearing, this mechanism is disrupted when they die. Perceptually, the disruption of the active mechanism resulting from OHC death is characterised by a loss of sensitivity over a range of frequencies, a ‘broadening of auditory filters’ so that spectral detail is smeared when passed on to higher centres in the brain, and a reduction in the dynamic range of hearing so that audible sounds become uncomfortably loud over a smaller range of levels.

Sensorineural hearing loss is caused by genetic factors, aging, ototoxic medication and illness such as meningitis, and can occur at any age. Because hair cell death is more common in the basal end of the cochlea (this end is more closely linked to other structures and is thus more susceptible to infections, ototoxic drug takeup, etc.), sensorineural hearing loss is typically worse in the higher frequencies. The most common intervention for sensorineural hearing loss is an acoustic amplification aid, which amplifies the signal reaching the ear to make it more audible to the listener. While acoustic hearing aids can partially overcome the issue of audibility for more moderate losses, they cannot improve the spectral resolution lost due to the broadening of auditory filters caused by OHC death. For listeners with more profound impairments, where IHC death may cause a total loss of sensitivity for a given cochlear place (e.g. above 1-2 kHz), the benefit from an acoustic hearing aid may be minimal.

A cochlear implant (CI) is an implantable electronic device that can restore the sensation of hearing to individuals with a profound sensorineural loss. An electrode array is surgically inserted into the cochlea and electrically stimulates structures in the inner ear, thus bypassing the dead cochlear hair cells which are no longer

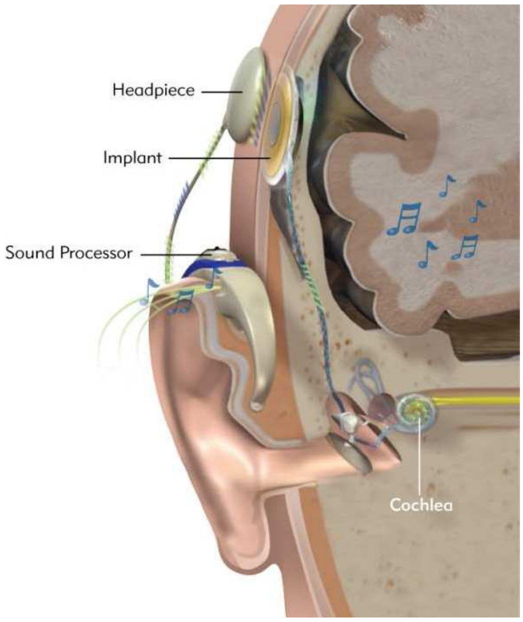

Figure 2-3. A modern cochlear implant system. Figure reproduced with permission from Advanced Bionics. © Advanced Bionics.

responsive to acoustic stimulation. Figure 2-3 shows a diagram of a modern cochlear implant system.

A microphone is worn externally which picks up the incoming sound signal and delivers this to a speech processor. The speech processor encodes the time-varying spectrum of the input from the microphone, thus mimicking the auditory filtering of the basilar membrane. This information is then transmitted transdermally to the cochlear implant via a radio wave. The implant electrode array then directly stimulates the spiral ganglion cells in the auditory nerve, thus bypassing the dead IHCs, and this information is then transmitted to higher centres in the brain via the normal pathway (Advanced Bionics, 2008). While normal hearing is often referred to as *acoustic* hearing, hearing through a cochlear implant is called *electric* hearing. Most individuals receiving cochlear implants receive a single monaural implant, and monaural implants will be the focus of this introductory chapter. More recently, individuals are beginning to be implanted with two cochlear implants, and the implications of bilateral implants will be the focus of the remainder of the thesis.

Modern cochlear implants exploit the tonotopicity of the cochlea. Like the basilar membrane, the cochlear implant speech processor analyses the incoming signal into its frequency components and encodes this for delivery to the individual electrodes: low frequency information is delivered to electrodes that have been placed towards the apical end of the cochlea, while high-frequency information is delivered to more basally situated electrodes.

An example of one speech processing system used in modern cochlear implants, the *continuous interleaved sampling* (CIS) processor, is outlined in Figure 2-4. The signal is first pre-emphasised to attenuate the low frequencies and amplify the high frequencies so that in later stages of processing, low-energy, high-frequency speech information (e.g. consonants) is not discarded as noise. Next, the signal is band-pass filtered into a number of frequency bands (in this case six), rectified and low-pass filtered to extract the envelope, and then compressed. Finally, the signal is used to amplitude-modulate a series of electrical pulses which are sent through the electrode contacts. Stimulation at each electrode is interleaved in time in order to

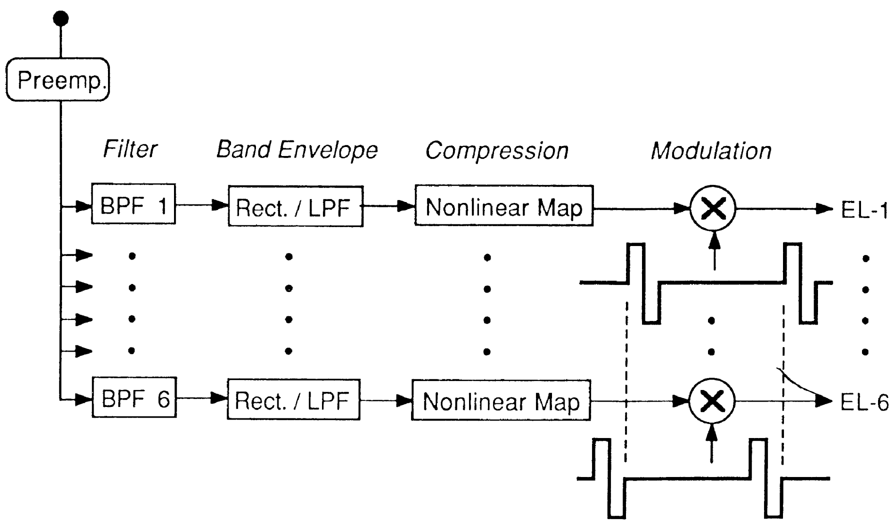

Figure 2-4. Outline of the continuous interleaved sampling (CIS) speech processing strategy for cochlear implants. This figure is reproduced with permission from B. Wilson and Macmillan Publishers Ltd. (Wilson, Finley, Lawson, Wolford, Eddington, and Rabinowitz, 1991). © 1991, Macmillan.

minimize interactions among the electrodes.

2.3.1 Signal distortion with cochlear implants

The success of cochlear implants is often heralded, but outcomes do vary widely, and many patients do not achieve optimal benefit (Tyler and Summerfield, 1996). While CIs are designed to exploit the natural hearing mechanism as much as possible, there are several aspects of the processing that distort the signal relative to normal hearing, and this can limit their utility for some individuals. These distortions, described below, include a basalward spectral shift of the speech information, band-limiting of the signal, reduced frequency resolution and loss of temporal fine structure.

2.3.1.1 Basalward spectral shift

During cochlear implant surgery, the narrowness and tight curvature of the cochlea prevents the electrode array from reaching the apex. While the average length of a human cochlea is 35 mm, cochlear implant electrode arrays are typically designed for an insertion depth of only 25 mm, and in many cases, the insertion achieved is shallower than this. Estimates based on *in vivo* computed tomography from twenty-six Nucleus-22 implant recipients showed insertion depths ranging from 11.9 – 25.9 mm, as well as considerable variation of cochlear length from 29.1 – 37.4 mm (Ketten et al., 1998; Skinner et al., 2002).

Even though the cochlear implant electrode array is not inserted fully into the cochlea, the speech processor delivers time-varying spectral information to the electrode array that represents the almost the entire range important for speech – 150-8000 Hz. The CF of the site of stimulation of a CI electrode is thus likely to be situated more basally than the band of frequencies delivered to that site. This has

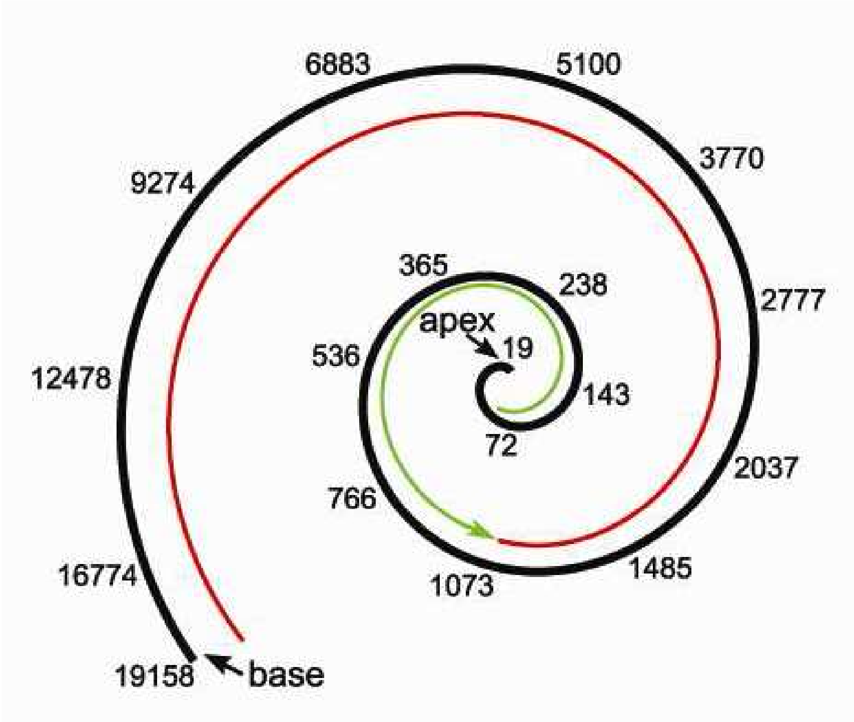

Figure 2-5. Characteristic frequencies along the basilar membrane. The superimposed red line represents a cochlear implant electrode array, and the green arrow indicates the shallow insertion. Note that the most apical electrode reaches a point along the basilar membrane with a characteristic frequency of around 1000 Hz. Despite this, the speech information delivered to that electrode will typically be 150 – 300 Hz, resulting in a basalward shift. This figure has been adapted from (Loizou, 1998) by the thesis author with permission from P. Loizou and IEEE. © 1998, IEEE.

been called a *basalward shift*. An electrode array inserted to a typical depth of 25 mm is depicted schematically in Figure 2-5.

The topography of the neural elements stimulated by the electrode contacts is not at present completely understood. The standard approach for estimating the effective CF at each electrode contact has followed the frequency-to-place mapping of the organ of Corti established by Greenwood (1990). Using a median insertion depth of 20 mm, as found by Ketten et al., this mapping estimates a CF of 1000 Hz at the most apical electrode contact. However, the Greenwood map may not be a realistic model for more modern modiolus-hugging CIs (Sridhar, Stakhovskaya, and Leake, 2006; Stakhovskaya, Sridhar, Bonham, and Leake, 2007), which place electrode contacts close to the modiolus and thus sprial ganglion cells. Stakhovskaya and colleagues (2007) argue that a spiral ganglion map may be more appropriate for these CIs, which assigns substantially lower CFs than an Organ of Corti map, especially for more apical electrode locations. It may therefore be important to consider proximity to the modiolar wall when estimating the effective CF of an electrode contact.

Notwithstanding our incomplete knowledge of the effective CFs along an electrode array, the altered frequency-to-place mapping resulting from incomplete insertion of the implant electrode array has demonstrable and sizable effects on speech perception (Skinner et al., 2002). At the time of implant switch-on, many implant listeners report that the speech sounds ‘munchkin-like’ (Eddington, Tierney, Noel, Herrmann, Whearty, and Finley, 2002), indicating the altered high frequency sensation from the upwardly shifted speech. This sensation is likely to be more pronounced in CI users whose implants have been implanted more “shallowly” than 25 mm.

2.3.1.2 Band-limited signal and reduction in frequency selectivity

Normal hearing spans the range of around 2 – 40,000 Hz, but most of the information important for speech occurs between 100 and 10,000 Hz. Since the primary goal of restoring the sensation of hearing through cochlear implants is to restore speech recognition capability, most current speech processing strategies band-limit the incoming signal to the range of frequencies important for speech. Furthermore, modern cochlear implants contain between twelve and twenty-two distinct channels (depending on the implant and stimulation mode), but the estimated number of auditory filters in the normal human cochlea is around thirty. Most CI users do not benefit from an increase in the number of frequency bands beyond 6 to 8 channels, even for speech in noise (Friesen, Shannon, Baskent, and

Wang, 2001). Frequency selectivity through cochlear implants is thus greatly reduced compared to normal hearing.

The discrepancy between the number of available electrodes and the number of distinguishable frequency bands may result from a spread of excitation at a given electrode contact. The interaction between the electrode contact and the neural structures which it stimulates is not well understood. An additional complicating factor may be the pattern of nerve survival at the point of electrode contact, which there is presently no way to accurately determine. The electrical stimulation provided at a given electrode may spread to other structures along the basilar membrane, and the surviving structures at a given site of stimulation may be tuned to frequencies not typical of the CF at that site. The result from either or both of these cases is a ‘spread of excitation’, and the perceptual consequence is a lack of ability to distinguish between stimulation at different electrode contacts. Thus, adjacent electrodes may elicit the same pitch percept in a given CI recipient, resulting in a reduced number of effectively distinct auditory channels.

2.3.1.3 Loss of temporal fine structure

In normal hearing, complex sounds such as speech are decomposed by the auditory filters into a series of signals, each of which can be described as a slowly fluctuating “envelope” superimposed on rapidly fluctuating “temporal fine structure” cues (Moore, 2008). In approximating the auditory filtering performed by the human cochlea, CI speech processing divides the speech spectrum into a number of bands that can be thought of as the auditory filters. However, because of limitations in the technology of electrical stimulation, most speech processing strategies discard temporal fine structure information altogether, and instead extract only envelope information in a given band to be presented to the electrode contact (Zeng, 2004). Remarkably, for speech in quiet, envelope information from just a few spectral bands is often sufficient for speech recognition (Shannon, Zeng, Kamath, Wygonski, and Ekelid, 1995). However, temporal fine structure signals rapid fluctuations in the speech signal which appear to be important for speech discrimination in noise as well as the processing of pitch (Moore, 2008), both of which appear problematic for CI users.

Current systems attempt to present some of the more rapidly fluctuating temporal information, such as fundamental frequency, by providing high rates of stimulation (e.g. through temporal coding). However, CI listeners are not able to use this information in the way that normal hearing listeners do. The reasons for this are not at present understood, but may reflect an incomplete understanding of the biophysical properties of direct electrical stimulation. Indeed, experimental attempts at restoring pitch perception through the introduction of more rapid stimulation rates have not had much success, which may reflect some sort of neural saturation with electrical stimulation not seen in normal hearing (Wilson, Lawson, Muller, Tyler, and Kiefer, 2003). For cochlear implant users, the signal distortion resulting from a loss of temporal fine structure ultimately means a lack of adequate pitch perception and the diminished ability to recognise speech in a noisy environment.

2.4 Simulations of cochlear implants in normal hearing listeners

Outcomes with cochlear implants vary vastly, with some listeners achieving near normal speech recognition after only limited experience with the implant, and others achieving sub-optimal performance after years of experience (Tyler and Summerfield, 1996). Factors that may affect outcomes with cochlear implants include the pattern of neural survival at the electrode contacts, implant insertion depth, neural plastic changes during the period of deafness and the cognitive capacities of the person receiving the implant. It is difficult to conduct controlled research about the impact of various aspects of signal distortion, since sub-optimal performance may be indicative of any of these factors, most of which are completely independent of the speech processing itself. For this reason, simulations of cochlear implants in normal hearing listeners have been used commonly in research on speech perception with cochlear implants.

Shannon and colleagues (1995) were the first to develop a simulation of the cochlear implant in normal hearing listeners. They broke the speech signal down into a varying number of input bands, extracted the information in each band via filtering, half-wave rectification and smoothing, multiplied a white noise signal by the time-varying amplitude envelope in each band, re-filtered the output to be band-limited to the same frequency range as the input band, and finally summed the bands together and presented these to the listener. A schematic representation of this processing is shown in Figure 2-6. This processing is very similar to that carried out by CI speech processors (eg. the CIS strategy outlined in 2.3), with the exception that a band of noise was used as a carrier in each band, whereas implant processors use a series of electrical pulses to excite electrodes. As with many cochlear implant patients, after a short period of acclimatization, the listeners in this experiment were remarkably adept at perceiving this speech which contained

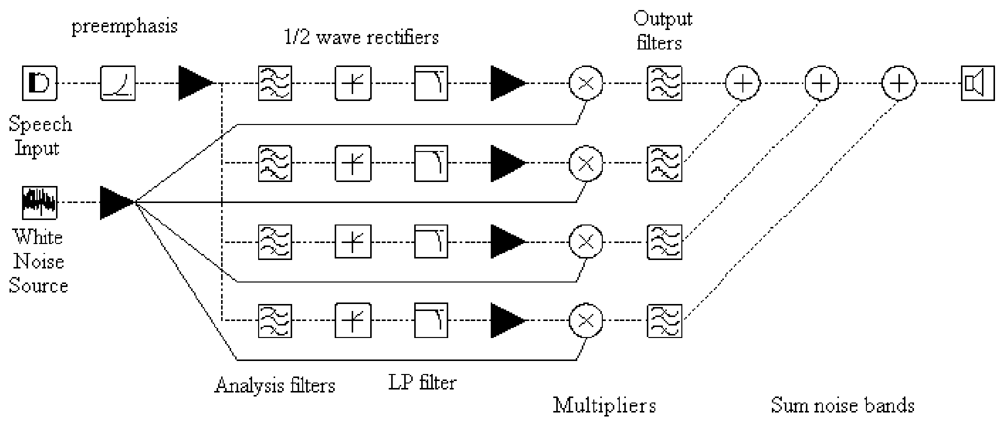

Figure 2-6. Speech processing simulating a cochlear implant. Figure is taken from Rosen, Faulkner and Wilkinson (1999) and is reprinted with permission from S. Rosen and the Acoustical Society of America. © 1999, Acoustical Society of America.

greatly reduced spectral detail, and speech perception with sentences, vowels and consonants improved when the number of frequency bands was increased.

This type of speech processing has been called noise-vocoded speech, and is now commonly used to simulate CI speech processing. By changing the parameters at various points in this speech processing scheme, one can approximate the signal distortion seen in a cochlear implant. For example, increasing the number of speech bands approximates an increase in effective electrodes; changing the carrier from a noise-band to a tone approximates a more finely-tuned electrode stimulation site; increasing the low-pass filter frequency approximates the provision of an increasing amount of fine-structure variation in the signal; and varying the frequency of the output filter can approximate upward or downward spectral shifts relative to the natural tonotopic alignment along the basilar membrane.

2.5 Frequency-place mapping and speech perceptual adaptation

Certain aspects of signal distortion with cochlear implants are largely the result of the physiological effects of electrical stimulation and nerve survival in the individual patient, and there is therefore little that can be done to make improvements once an individual has been implanted. One exception to this is the frequency-place mapping of the speech information delivered to each electrode. While the insertion depth of the implant electrode array is determined at surgery, the clinician is capable of manipulating the frequency-place map so that the most informative information can be delivered. For this reason, considerable effort at improving outcomes in those who have already received cochlear implants has focused on frequency-place mapping, since this is something that can be straightforwardly implemented. By contrast, current attempts at improving the delivery of temporal fine structure through, for example, rapid rate stimulation, have had only limited success.

As depicted in Figure 2-5, cochlear implant electrode arrays are not inserted fully into the cochlea. Because the frequency allocation of CI speech processing typically assigns frequency information in the range of 150 – 8,000 Hz[1], this often results in a basalward shift of spectral information. Basalward spectral shift has been associated with a decrement in speech intelligibility in cochlear implant patients, especially in the case of shallow electrode insertions (Skinner et al., 2002). This has led to an ongoing debate about whether it is best to preserve tonotopic matching at the expense of a frequency-shifted map that is maximally informative for speech when fitting CI processors. At the crux of this problem is whether listeners are able to adapt their internally stored speech representations to reflect the novel, basally shifted information being provided by the implant, or whether they would benefit more from a signal that is reduced in speech cues but is easily mapped on to internal speech representations because it follows the natural tonotopic mapping.

It seems obvious that some level of adaptation to the cochlear implant is possible, given the long history of their success despite the highly distorted signal they deliver. Similarly, normal hearing listeners demonstrate a remarkable ability to adapt to many types of novel speech. Some examples include within-language accented speech (Evans and Iverson, 2004); foreign-accented speech (Clarke and Garrett, 2004); sine-wave speech, where a series of tones are synthesized to track the movement of formants (Remez, Rubin, Pisoni, and Carrell, 1981); noise-vocoded speech (Rosen et al., 1999; Shannon et al., 1995); and even spectrally rotated speech, where the spectrum of a signal is rotated about a set frequency (Azadpour, 2008; Blesser, 1972). Potter and Steinberg (1950) maintained that “a certain spatial pattern of stimulation on the basilar membrane may be identified as a given sound

[1] This will vary according to the speech processing strategy used and the clinically determined frequency allocation.

regardless of position along the membrane.” To what extent is this true for users of cochlear implants? Are they able to adapt to basally shifted speech?

2.5.1.1 Acute studies of altered frequency-place mapping

Several earlier studies of basalward spectral shift suggested that listeners’ performance with such speech was greatly reduced compared with tonotopically aligned processing, suggesting that basally shifted speech may be difficult for people to recognise. However, for more shallow electrode insertions and simulations thereof, preserving tonotopic alignment came at the expense of providing maximally useful speech information. Speech delivered tonotopically through a shallowly inserted electrode would also be difficult for subjects to understand because speech limited to high-frequencies lacks much of the information important for robust speech recognition. This is what might be predicted from the articulation index, which assigns a significant importance value to frequency bands of speech below 1000 Hz and thus below the most apical point of stimulation in a cochlear implant (French and Steinberg, 1947).

Dorman and colleagues (1997a) used sine-vocoded simulations of CI processing to simulate the effects of electrode insertion depth on speech understanding. They mapped the same frequency information onto simulations of increasingly more shallow electrodes, in 1 mm increments, thus introducing incrementally larger degrees of basalward shift. While performance was near-normal with vowel, consonant and sentence recognition with a simulated insertion depth of 25 mm, performance with all speech materials declined rapidly with each decrement in insertion depth, with recognition of words in sentences dropping from near 100% at 25 mm to ~ 45% at a 22 mm simulated insertion. This study had the advantage of exploring the role of CI insertion depth without the confounding factors seen in CI patients (eg. nerve survival), and at the time seemed to suggest that for individuals with more shallow electrode insertions, a tonotopic alignment may be preferable, even if some low-frequency information is lost.

A similar result with vowel recognition was shown for both cochlear implant patients and normal hearing listeners (Fu and Shannon, 1999). For the normal-hearing subjects who listened to noise-vocoded simulations in that study, both upward and downward shifts of up to 3 mm (using a Greenwood map) had only a small effect on vowel recognition, but shifts of 4 mm and larger produced a significant drop in vowel recognition. An upward shift of 3 mm applied to adult speech may approximate the formant frequencies of a small child. Since these are in the normal range for human speech, this shift may have been easily tolerated. On the other hand, the larger 4 mm shift would have resulted in formant frequencies outside the natural speech range typical in humans, so would be difficult to map on to existing speech representations. Similarly, Assmann and Nearey (2003) found that normal hearing listeners' recognition of noise-vocoded, frequency-shifted vowels was better when the shifted vowel formant frequencies fell within the range of the listeners' prior exposure to naturally varying speech. Thus, upwardly shifted male vowels were more easily recognised than upwardly shifted female vowels, and downwardly shifted children's vowels more easily recognised than those of male or female adults.

However, in a later study with CI patients, Baskent and Shannon (2005) showed a strong interaction between frequency-place mapping and electrode insertion depth. They simulated increasingly shallow insertions in CI users by deactivating apical electrodes, and varied the frequency allocation to either provide a matched map or a compressed but basally shifted map. For deeper electrode insertions, listeners performed better with a tonotopic frequency allocation, but for more shallow electrode insertions, the listeners benefitted from a frequency allocation that compressed a larger frequency range on to the smaller number of electrodes. For example, vowel recognition showed a peak improvement of 35% compared to the matched map when frequency information as low as 700 Hz was included; however, by including increasingly lower frequencies below 700 Hz by further compression and basalward shift, performance began to decline again. The results suggest that some listeners may benefit from a basally shifted map that is more informative for speech.

2.5.1.2 Adaptation to shifts in frequency-place mapping

A major caveat of these earlier studies reporting on speech perception with a basalward shift is that they were carried out acutely, giving the listener little time to adapt to the novel speech signal. By contrast, cochlear implant users show improvements in speech perception months and even years after implantation (Tyler and Summerfield, 1996). More recent studies with both simulations of CI processing in normal hearing listeners (Rosen et al., 1999; Fu, Nogaki, and Galvin III, 2005b; Faulkner et al., 2006; Faulkner, Rosen, and Stanton, 2003; Fu and Galvin III, 2003) and cochlear implant patients (Fu et al., 2002; Fu, Galvin III, Wang, and Nogaki, 2005a; Fu and Galvin, 2007; Harnsberger, Svirsky, Kaiser, Pisoni, Wright, and Meyer, 2001; Svirsky, Silveira, Neuburger, Teoh, Helms, and Suarez, 2004; Svirsky, Silveira, Suarez, Neuburger, Lai, and Simmons, 2001) have allowed listeners time to adapt to altered frequency-place maps, and have shown improved performance after a period of training and/or passive listening experience.

Using noise-vocoded simulations of CI processing, Rosen et al. (1999) trained normal hearing listeners with noise-vocoded simulations of speech that had been shifted upwards an equivalent of 6.5 mm using a Greenwood map. When the listeners were tested prior to any experience with the spectrally-shifted speech, their performance was at floor level with easy sentence material. Yet, with only 3 hours' training, recognition of words improved to 30%, clearly demonstrating the importance of adaptation in examining speech perception with altered frequency-place maps.

In another simulation study, Faulkner et al. (2003) examined whether listeners could accommodate tonotopically matched processors with varying degrees of simulated electrode insertion depths.[2] Speech perception of consonants, vowels and words in sentences all declined when the simulated insertion depth decreased, suggesting that for more shallow electrode insertions, tonotopically aligned frequency-place mapping may not be ideal.

In a further study that directly compared basally shifted and frequency-aligned processors, Faulkner and colleagues (2006) trained listeners with two 8-band noise-excited vocoders to simulate a shallowly inserted electrode. In one simulation, the CFs of the analysis filters matched those of the output filters; in the other simulation, CFs of analysis filters were shifted the equivalent of 6 mm downwards along a 35mm cochlea, thus mimicking a basalward shift. After 3 hours of training with each processor, for a total of 6 hours of training, subjects showed improvements with both processors, but greater improvement with the shifted processor, indicating learning specific to the frequency shifting (and thus not simply accommodation to spectrally limited speech). Post-training performance on sentences and vowels was, in some conditions, better with the shifted than the matched processor, suggesting that the initial decrement in speech perception resulting from basally shifting speech can be overcome.

[2] Because no spectral shift was introduced in this processing, only brief familiarization was given, so this was not an adaptation study *per se*. However, the findings serve as a relevant comparison to data from similar experiments with upward spectral shifts for shallow insertions.

Fu et al. (2003) also demonstrated shift-specific learning in a simulation study. They first trained listeners with speech that had been shifted upwards, and then tested their perception on speech that had been shifted both upwards and downwards. Subjects showed improved performance relative to baseline with the upwardly shifted speech only. Subjects were then further trained with speech that had been shifted downwards. At the final test period, subjects showed improved performance, relative to baseline, with both upwardly shifted and downwardly shifted speech. Because improvements in downwardly shifted speech were apparent after but not before training with downwardly shifted speech, the study confirmed that adaptation took place specific to the shift being trained. The study also suggests that listeners can develop alternate representations of sounds while still retaining previously learned internal representations.

Fu et al. (2002) examined whether cochlear implant listeners could also adapt to distortions of frequency-place mapping imposed by implant speech processors. They fitted 3 patients with experimental processors with analysis filters 0.68–1 octave downwards in frequency from their clinically fitted processors, and examined their ability to perceive speech over a time period of three months. At the time the experimental processors were switched on, speech perception dropped dramatically, but all three subjects showed continuing improvements in speech recognition throughout the three month period of wearing the experimental processors. Despite the improvements through experience with the experimental processors, subjects still performed better with their clinical processors at the end of the three months. It is not clear whether this demonstrated a limit on adaptation to the shifted processing, or whether the three months' experience with the experimental processor was insufficient to allow complete adaptation.

More recently, researchers have shown that targeted auditory training can improve speech perceptual performance with altered frequency-place maps in both simulation studies with normal hearing listeners (Fu et al., 2005b; Stacey and Summerfield.Q., 2007) and cochlear implant patients (Fu et al., 2005a; Fu and Galvin III, 2007a; Fu and Galvin III, 2007b; Stacey, Raine, O'Donoghue, Tapper, Twomey, and Summerfield, 2008). In Fu et al.'s studies, normal hearing listeners were trained with spectrally shifted speech with an 8-band sine vocoder simulation of CI speech processing. After computer-based training of vowel contrasts, identification of medial vowels increased from a baseline of 13.0% correct to 28.5% correct after 5 consecutive days of training. Implant patients were also trained for an hour a day, 5 days per week, for an entire month or longer and demonstrated significantly increased speech perception performance following the training.

2.5.1.3 Evidence of adaptation in users of CIs: Vowel recognition

Evidence from adaptation studies clearly suggests that, within limits, listeners are able to accommodate shifts in frequency-place mapping after a period of adaptation. However, the process by which this adaptation occurs is not presently clear. One method researchers have used to explore this process is to examine vowel identification and categorization in users of cochlear implants. As outlined in 2.1, the recognition of vowels is largely dependent on the first and second formant frequencies (F1 and F2); duration and formant movement also appear important in vowel recognition (Iverson, Smith, and Evans, 2006). Most speakers' first and second vowel formants occur in the low frequencies and are thus likely to be affected by any basalward shift applied by cochlear implant speech processing. If basalward shift were present, then it might be expected that the preferred best exemplar locations of vowel formant frequencies of cochlear implant users would be shifted downwards so as to re-create the internal representation of the vowel from before the onset of deafness. On the other hand, if listeners have adapted to the CI processing, then their preferred vowel formant frequencies may have changed over time to match those of normal hearing listeners, thus reflecting the speech being heard through the implant. A further possibility is that the cochlear implant listeners' vowel categories may spread over the course of adaptation, demonstrating the need to map a wider range of vowel formant frequencies onto a given category. They retain their internal representations from before the period of deafness but adapt by mapping the novel CI speech onto these same categories.

Harnsberger et al. (2001) studied the perceptual vowel spaces of cochlear implant listeners who had been using their implant for at least one year. Synthesized vowels were created in a two dimensional F1 – F2 space and assigned to a grid on a computer screen. Subjects were then asked to find the synthetic stimuli that matched the orthographical labels of ten monophthong American English vowels. Of the eight cochlear implant users tested, only one showed a systematic shift of F1-F2 vowel space relative to normal hearing listeners that may have indicated a lack of adaptation to basalward shift. Interestingly, there was also no correlation between the predicted basalward shift based on CI insertion depth and the difference between CI and normal hearing listeners' vowel categories. Relative to normal hearing listeners, however, the CI users did show enlarged vowel categories, which may lend support to the 'spread of vowel spaces' hypothesis. For most of these CI listeners, adaptation to any basalward shift appeared to be complete after as little as a year of experience with their implant. The researchers hypothesized that difficulties in vowel recognition were to do with difficulties discriminating between formant frequencies, which may be indicative of spread of excitation rather than basalward shift.

In a later longitudinal study, they found that Spanish speakers' improvements in vowel recognition with experience of a cochlear implant were associated with better labeling rather than better electrode discrimination (Svirsky et al., 2001). In that study, they used a model of phoneme identification to determine the best possible recognition given the subjects' electrode configurations and their ability to discriminate differences in formant frequencies. They then monitored subjects' formant discrimination and vowel identification abilities from immediately after implant switch on for several months. They found that subjects' vowel recognition improved to the best possible ceiling performance over the course of the study, suggesting that improvements were being made based on a better ability to label the vowel stimuli.

In a separate longitudinal study, the same group of researchers monitored vowel best exemplar locations in four English and Spanish CI users over a two year period from the day of implant switch-on (Svirsky et al., 2004). Surprisingly, the listeners in this study did not demonstrate the expected signs of basalward shift at implant switch-on, since their vowel categories were not shifted downwards. However, their vowel categories were abnormal relative to normal-hearing listeners' in ways unique to each subject. Over the two year period, though, all of the CI users showed significant adaptation since their vowel categories were near to normal after two years. The time needed to achieve this level of adaptation was markedly different for each subject, with the best achieving adaptation after three months and the worst after two years. The authors suggested that the reason the subjects did not demonstrate the effects of basalward shift was a decay in the memory of how a vowel should sound, perhaps making vowel categorization on the day of implant switch-on especially difficult. This may also explain the wide variability of vowel categories among these subjects at the time of implant switch-on.

Iverson et al. (2006) demonstrated that normal hearing listeners listening to noise-vocoded simulations and cochlear implant users are similarly sensitive to vowel formant movement and vowel duration in addition to F1 and F2 target trajectories, providing further evidence that CI listeners remap their vowel spaces after implantation to resemble those of normal hearing listeners. Both groups of subjects relied on all four cues to recognise vowels, and both groups of listeners used all four cues when mapping out best exemplar locations for vowels.

Given the similarities of vowel categories between normal hearing and CI listeners, it seems plausible to question whether the effect of basalward shift exists at all. Without imaging data from each CI patient, precise information about the CI depth of insertion and the length of the cochlea is unknown, so predictions of basalward shift based on the Greenwood formula may overestimate the degree of basalward shift. Smaller basalward shifts have been shown to be more easily accommodated (Fu and Shannon, 1999). Blamey et al. (Blamey, Dooley, Parisi, and Clark, 1996) compared the pitch sensations elicited by electric and acoustic stimulation in patients with cochlear implants who wore an acoustic hearing aid in the contralateral ear. They found that the acoustic pitch match corresponding to an electric stimulus was much lower than would be predicted from a normal frequency-place map, at least for frequencies up to 2 kHz. One possible caveat with these results, however, was that the listeners had impaired hearing in the acoustic hearing ear, so a loss of frequency selectivity may have led to less reliable pitch matching results.

However, more recent data (Boex, Baud, Cosendai, Sigrist, Kos, and Pelizzone, 2006; Dorman, Spahr, Gifford, Loiselle, McKarns, Holden, Skinner, and Finley, 2007) has also suggested lower pitch matches for more apical electrodes than would be predicted by a Greenwood map. Dorman and colleagues' study was particularly interesting because the patient had mild-moderate residual hearing across all frequencies in the non-implanted ear. While these data seem to suggest that the effect of basalward shift may have been overestimated (especially in the

Boex study, the CIs were implanted near to the modiolus and may have been better represented by a spiral ganglion map), it is equally possible that the participants in these studies adapted the percept in the CI ear to match that of the acoustic hearing ear, since the participants had been using their implants and hearing aids for some time.[3] Indeed, a theory that does not include adaptation of some sort could not account for the changes in speech perception and vowel best exemplar locations seen in the first months and years following cochlear implantation.

2.6 A note on perceptual adaptation in children versus adults

The preceding discussion on speech perceptual adaptation has focused entirely on post-lingually deafened adult users of cochlear implants who have had the sensation of hearing restored after a period of deafness. However, a large percentage of the people who receive cochlear implants are pre- and peri-lingually deafened children. While it is clear that children who receive cochlear implants undergo many cortical changes indicative of adaptation to electric hearing (Doucet, Bergeron, Lassonde, Ferron, and Lepore, 2006; Dinse, Godde, Reuter, Cords, and Hilger, 2003), both during the period of deafness and after implantation, they have not undergone the processes of speech and language acquisition at the time they receive their implant. The only exposure to speech these children will ever receive is through electric hearing, and thus their internal speech representations will be formed based on the electric signal. By contrast, post-lingually deafened adults will have developed speech perception in the normal way, so their internal speech representations should be similar to those of normal hearing listeners, barring any possible memory decay in these representations during the period of deafness (cf. (Svirsky et al., 2004)).

[3] Blamey's paper only specifies that the participants were "regular cochlear implant users who wore their hearing aids most of the time."

Hence, while the electric stimulation will introduce the same signal distortions in children and adults, the adaptation processes to accommodate these distortions are likely to be very different for these two groups. Specifically, the issue of *adaptation* to basalward shift is unlikely to be relevant to children, since their only exposure to speech will be of basally shifted speech. For this reason, this thesis will be limited to the consideration of perceptual adaptation in adults.

2.7 Summary

This chapter outlined evidence in support of the notion that many users of cochlear implants are able to adapt to upward spectral shifts of speech. Almost all of the studies considered thus far have examined the case of monaural cochlear implants and/or monotic stimulation. However, it is becoming increasingly common to implant patients bilaterally, in order to improve sound localization and speech reception in noise, among many other potential benefits. The next chapter will review bilateral cochlear implants and raise questions about the effect of basalward shift when stimulating bilaterally.

Chapter 3. Bilateral Stimulation with Cochlear Implants

As is the case with vision, it has long been acknowledged that hearing with two ears is better than one, yet despite the routine fitting of corrective lenses for both of the eyes, the history of hearing loss correction was long confined to a single ear. This changed with E. R. Libby's publication *Binaural Hearing and Amplification* (1981), which argued that hearing aids should be fitted bilaterally when individuals present with a bilateral loss. It has only been in the past few years that bilateral stimulation with cochlear implants has received significant consideration, but this is likely to change as better implants become available and the processes that maximise outcomes with bilateral stimulation are better understood. While the potential benefits are manifold, the use of bilateral devices raises new questions about the perceptual processes underlying adaptation to cochlear implants. This chapter will outline binaural hearing and bilateral stimulation with cochlear implants, and introduce the potential problems with bilateral stimulation which will be the focus of this thesis.

3.1 Properties of binaural hearing

In normal hearing, listeners use similarities and differences in the sound signals received at each ear to perceive the location of sounds in space and to improve the perception of speech in both quiet and background noise. These advantages result from both physical differences in the sound signals reaching each ear, such as amplitude and phase (important for localization), and more central neurological processes such as binaural summation and binaural squelch (important for our improved speech recognition in noise) (Dunn and Ou, 2008).

The physical properties of sound used to aid localization in the horizontal plane are the interaural time difference (ITD) and interaural level difference (ILD). The cue that will dominate varies according to the frequency of a given sound, with ITDs dominating in the lower frequencies and ILDs dominating in the higher frequencies (Moore, 2003), and this is due to the different physical properties of high and low frequency sound waves. Consider a sound originating to one side of the head. For both high and low frequency sounds, there will be a time lag between the sound reaching the nearer ear and the sound reaching the farther ear, since the sound has further to travel to reach the farther ear. Since the speed of sound in air is relatively constant across frequencies, this temporal difference will be the same for both high and low frequency sounds originating from a given location, and for low frequency sounds, this ITD is key to localization. However, at higher frequencies, the ITD becomes ambiguous, since the period of the sound approaches or becomes more rapid than the ITD itself, making it impossible to determine how many periods have occurred, and hence the delay, between the sounds at each ear. For a 1000 Hz sinusoid, with a period of 1 ms and an ITD of .5 ms, the tone will reach the farther ear with a lag of half a period, which gives an ambiguous interaural phase difference (IPD) of 180^o. The tone at the farther ear could be perceived as *either lagging or leading* the tone at the nearer ear by a phase shift of 180^o. This is reflected in perceptual experiments, where normal hearing listeners make errors in localizing sinusoids with ambiguous IPD cues (Moore, 2003).

Considering again a sound originating to one side of the head, an ILD will arise from the *head shadow effect*, where the head attenuates the sound as it travels to the farther ear. Since the wavelength of a high frequency sound is shorter than the size of the head, the sound will reach the nearer ear relatively unmodified, but as it travels to the farther ear, the waves will be ‘shadowed’ by the head, causing the amplitude of the sound to diminish slightly as it travels around the head. This difference in level of the sound at each ear can be used to determine where the sound occurred. However, for low frequency sounds, the wavelength will be longer than the size of the head, the sound will travel to the opposite ear without being shadowed, and there will be no amplitude difference between the sound signals at each ear. The head shadow effect is also important for improving the perception of speech in noise. A noise source originating to one side of the head will also be shadowed by the head as it travels to the opposite ear, thus allowing for a better signal-to-noise ratio (SNR) at that ear. The improved performance results from the ability to tune in to the ear with the better SNR.

For sound localization in the vertical and front/back directions, ITD and ILD cues may be ambiguous: many vertical locations can produce the same ITDs and ILDs. Localization in these planes is dependent on spectral cue enhancement. Modifications to the spectrum of a sound introduced by the external ear can be used to localize sounds, since certain spectral features vary systematically according to sound location. While these cues are also available in monaural hearing, they are not available through older cochlear implant systems which place the microphone over the ear (Advanced Bionics, 2004). Incidentally, normal hearing listeners show a remarkable ability to adapt to changes in these cues (Hofman, van Riswick, and van Opstral, 1998). In one study, listeners who wore ear moulds that changed the shape of their outer ear for a period of weeks adapted to their ‘new ears’, improving performance in sound localization from near floor when first worn to near normal upon completion. Interestingly, when the ear moulds were removed, their sound localization ability was intact, suggesting that learning how to localize sound may be like learning a language rather than a hard-wired ability. This is promising for bilateral cochlear implant users, who must also re-learn to localize sounds with the novel auditory input delivered by the CIs.

Central auditory processes are also involved in producing a binaural advantage, and these include binaural squelch and binaural summation. Binaural squelch results from the ability to improve speech recognition in noise by combining the noise at the ear with the better SNR and the noise at the ear with the worse SNR to form an auditory object, which is then ‘tuned out’ (squelched) in favour of the speech signal (Dunn and Ou, 2008). Binaural summation, also referred to as binaural redundancy, results from the summation and integration of redundant information reaching each ear. For a signal originating directly in front of the listener, the signal reaching the two ears will be identical, the summation will result in an overall level increase, and this will improve the detection of small differences in intensity and frequency, thereby improving recognition. For a sound originating at any other location, redundant cues can be attuned to and differences discounted to allow for the improved recognition of speech. Note that binaural summation is important for optimal speech recognition in both noise and in quiet (Dunn and Ou, 2008).

3.2 Bilateral stimulation with cochlear implants

While cochlear implantation has been hailed a success, many listeners face difficulties with their implant, especially when in a noisy environment. The restoration of binaural hearing to users of CIs could provide a range of binaural advantages, such as those outlined in 3.1. Interest in bilateral stimulation with cochlear implants has taken two forms. As criteria of candidacy for cochlear implantation have been relaxed due to more widespread success, more individuals are being implanted who have some residual hearing in the non-implanted ear. For these listeners, binaural hearing can be restored through the use of an acoustic hearing aid in the contralateral ear (CI+HA). Others have been implanted with two cochlear implants in order to restore sound localization and improve speech recognition in noise.

While the potential benefits for bilateral stimulation with cochlear implants are many, clinical evidence in favour of bilateral stimulation has been mixed. A binaural advantage for speech perception has been reported in patients with bilateral cochlear implants (Dorman and Dahlstrom, 2004; Litovsky et al., 2006; Tyler et al., 2005; Tyler et al., 2007; Wackym et al., 2007), and for implants used in conjunction with acoustic hearing aids (Ching, Incerti, and Hill, 2004; Ching, Psarros, Hill, Dillon, and Incerti, 2005; Ching, 2005; Ching, Incerti, Hill, and van Wanrooy, 2006; Hamzavi, Pok, Gstoettner, and Baumgartner, 2004; Iwaki, Matsushiro, Mah, Sato, Yasuoka, Yamamoto, and Kubo, 2004; Litovsky et al., 2006; Offeciers, Morera, Muller, Huarte, A., Shallop, and Cavalle, 2005).[4] Users of bilateral CIs appear able to take advantage of the head shadow effect (Litovsky et al., 2006; van Hoesel, Ramsden, and O'Driscoll, 2002) and many show an improved ability to recognise speech in both quiet and noise, suggesting the restoration of binaural summation. Of the 37 patients reported on in Litovsky et al., for example, 58% demonstrated a binaural advantage for sentences in quiet, with the mean improvement in key words correct for the group of around 10%. On the other hand, evidence for an improved ability to localize sounds with bilateral implants is mixed (Tyler et al., 2007; Verschuur, Lutman, Ramsden, Greenham, and O'Driscoll, 2009), and there is only limited evidence for the restoration of binaural squelch (van Hoesel et al., 2002).

Data on speech perception from these studies can be difficult to interpret. Most used a within-subjects design, where data for monaural conditions is collected by asking the patient to switch off an implant. While this has the advantage of controlling for inter-subject variability, it is also likely to introduce confounding effects. While with

[4] This thesis is primarily concerned with bilateral cochlear implants, though the findings may have some relevance to CI + HA fittings. Another form of combining CI + HA combines a short-electrode CI with an acoustic HA in the same ear. Implanting with a short electrode allows for the preservation of residual low-frequency hearing in the implanted ear. An HA thus amplifies low-frequency information, and the CI provides electrical stimulation for high frequency information. While the resulting advantages of binaural hearing are not realised with this configuration, frequency-place mapping issues may have some relevance to the work described in this thesis.

normal hearing listeners we can generally assume a cochlea with full nerve survival and similar stimulation in each ear (barring localization cues such as ITDs and ILDs), users of bilateral cochlear implants are likely to receive very different signals at each ear. This will arise from the (possibly) different cochlear implant systems being worn at each ear, the different insertion depths of the implants and the varying pattern of nerve survival. It therefore seems likely that asking a bilateral CI user to turn off an ear may put them in the very unnatural situation of listening unilaterally despite having adapted to asymmetrical bilateral stimulation. None of the studies reported allowing the listener time to adapt to the unilateral condition, so making acute unilateral measures may have impacted on the results. The disadvantages of acute measurements with cochlear implant users are discussed in detail in the previous chapter.

Another complication may be whether the implants were received sequentially or simultaneously. Until only very recently, all those receiving bilateral CIs would have been implanted sequentially, meaning an asymmetrical period of adaptation to each implant. While there is evidence to suggest that both types of implantation can be successful (Tyler et al., 2007; Litovsky et al., 2006), the issue of sequential implantation makes it difficult to determine the potential for binaural advantage in people who may not show one. Any lack of improvement with the addition of a second implant may result from insufficient experience with the second implant or a failure to provide the right training procedures (Tyler et al., 2007).

One study demonstrating success with bilateral CIs, reported by Dunn et al. (2008), aimed to control for some of these factors. They compared speech perception and sound localization ability between bilateral cochlear implant users and a group of unilateral CI users who had been matched with the bilateral CI group for duration of deafness and age at implantation. The data show clear and compelling evidence for better localization ability in the bilateral CI group compared to the unilateral group, with a large reduction in localization error (25 degrees in RMS error) for the bilateral group compared to the unilateral group. They also argue that the bilateral CI subjects showed a binaural advantage for sentence and word recognition in quiet, though the data here seem less compelling.

First, for both sentence and word recognition tests, the bilateral CI subjects were significantly better than the unilateral CI subjects on each of the unilateral conditions. This suggests that this bilateral CI subject group may be better performers overall, which may or may not have resulted from the bilateral implantation. In determining binaural advantage, then, making a between-groups comparison seems unfair since the groups do not appear equal for unknown reasons. The comparison to be made, which is what they argue *against* because of problems with acute measurement of unilateral conditions, is the within-subject comparison of unilateral and bilateral conditions. However, sentence recognition scores for this group were near ceiling for both unilateral and bilateral conditions, and only showed a small, significant binaural advantage after conversion to Rationalised Arcsine Unit scores (Studebaker, 1985). For the word recognition scores, which they did not transform, the difference in intelligibility between the best CI ear and the bilateral condition was also very small. Therefore, this study does not provide clear evidence for a substantial binaural advantage over “better ear” effects alone.

In summary, while bilateral stimulation with cochlear implants appears to restore binaural advantages to many who receive them, this is not true for everyone, especially in terms of speech perception. For many subjects, bilateral speech perception scores match scores for the “better ear”, suggesting that some listeners may be attending to the “better ear” rather than integrating the signals as in normal binaural hearing. This is especially true for speech in quiet (Dorman and Dahlstrom, 2004; Tyler et al., 2007; Wackym et al., 2007), which suggests that bilateral stimulation may not always restore binaural summation. Considering the recent attention and the consumer push for bilateral cochlear implantation for all, it therefore seems a crucial consideration to determine why a binaural advantage may not be shown in some patients.

3.3 Binaural integration and interaural frequency-place mapping

Chapter 2 presented evidence about the issue of frequency-place mapping with cochlear implants, and the problem of incomplete electrode insertions and a basalward spectral shift of speech information. The studies presented examined listeners who were implanted with a single cochlear implant, or normal-hearing listeners presented with the same signal to both ears. However, in the case of bilateral implants, the frequency-place maps in the two ears may be very different. The two electrodes may be inserted to different depths, resulting in different degrees of basalward shift. This may also interact with varying patterns of nerve survival in the two ears. A schematic representation of such a situation is highlighted in Figure 3-1.

Adaptation to bilateral cochlear implants presumably involves an adaptation to the novel signals presented to each ear, and also adaptation to the interaction and/or integration of these signals. To what extent is this possible when the signals are very different, such as when frequency-place maps differ between the ears? Does an inability to adapt to mismatched frequency-to-place maps underpin the lack of binaural advantage for speech seen in many bilaterally implanted patients?

In users of bilateral cochlear implants, a binaural advantage for speech in noise with directional cues might arise from the head shadow effect and an improved ability to

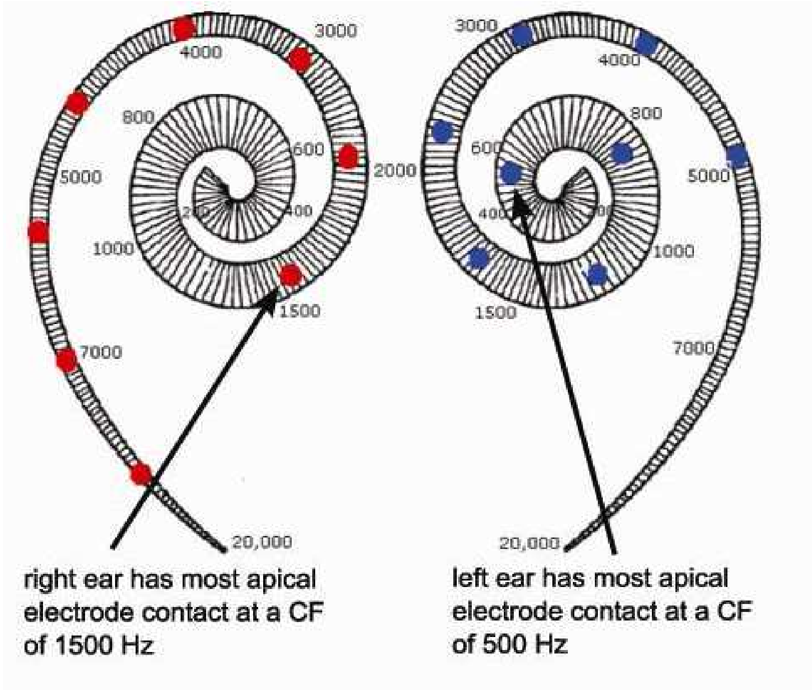

Figure 3-1. Schematic representation of right (red) and left (blue) bilateral cochlear implant electrode array contacts. Note the different insertion depths in each ear. This figure is from (Stuhlman, 1943) and is adapted by the thesis author with permission from Wiley Publishing. © 1943, Wiley.

attend to the ear with the better SNR. However, for speech in quiet or speech in noise presented at 0^o azimuth, if a binaural advantage beyond that of the "better ear" is apparent, then this implies that the user is able to *binaurally integrate* the two signals at each ear, since the advantage would result from binaural summation. Models of binaural integration are heavily tilted toward the discussion of the integration of psychophysical cues, such as ITDs and ILDs (Colburn and Durlach, 1978; Durlach and Colburn, 1978; Moore, 2003; Yost, 2000). However, research on dichotic listening also sheds light on the process of binaural integration for speech. Both schools of thought would seem to suggest that frequency-place map alignment between the ears may be important for optimal integration.

3.3.1 Binaural integration for ITDs and ILDs

As described in detail in Chapter 2, the cochlea is tonotopically organized, and this tonotopicity is maintained in ascending pathways through to the primary auditory cortex. Early connections between auditory neurons on the two sides of the head also appear to be tonotopically arranged, with binaural neurons having very similar CFs as for monaural stimulation of each ear (Smith and Delgutte, 2007). For ITDs, models of binaural hearing suggest that the firing patterns of neurons in each ear are compared between channels with the same CF to determine a common delay between the ears at each CF. Coincidence detectors then count the number of spikes arriving at the same time from the two ears. The peak in binaural interaction can be used to calculate the delay between one ear relative to the other, and thus indicates localization (Moore, 2003b). For ILDs, the models suggest that differences in sound levels within each ear are compared at the monaural processing stage, and these feed in to binaural interaction computations about ITDs, possibly as a weighting function. As for ITDs, this ILD comparison is made for neurons at similar tonotopic locations in each ear (i.e. within a CF) (Colburn and Stern, Jr., 1978; Durlach and Colburn, 1978; Stern, Jr. and Trahiotis, 1995).

Binaural coincidence detectors have been shown to be present at birth (Furst, Bresloff, Levine, Merlob, and Attias, 2004), suggesting that the mechanisms for binaural interaction that appear reliant on comparisons within a single frequency band are hard-wired. Evidence from psychophysical studies of both normal hearing listeners (Francart and Wouters, 2007; Nuetzel and Hafter, 1981; Blanks, Buss, Grose, Fitzpatrick, and Hall, 2008) and users of bilateral cochlear implants (Long, Eddington, Colburn, and Rabinowitz, 2003; Grose and Buss, 2007) also suggests that ITD and ILD sensitivity is maximal when stimuli are delivered to matched interaural cochlear places. Since a key objective of bilateral stimulation with CIs is to improve sound localization through the provision of ITDs and ILDs, this would suggest that interaural frequency-place map alignment is important for bilateral CI fittings.

3.3.2 Dichotic listening and the binaural integration of speech

In normal hearing, the binaural perception of speech can be assumed to be relatively *diotic,* since the signals reaching the two ears are the same but for fine-grained localization cues and information about the signal-to-noise ratios at each ear. By contrast, the perception of speech through bilateral cochlear implants is likely to be *dichotic.* Even at 0^0 azimuth, where localization cues and SNRs at each ear should be equivalent, the tonotopic representation at each ear will vary because of different insertion depths of the implants and different surviving nerve structures at each ear. Cutting (1976) identified six binaural fusions, and three levels of processing at which these might occur, that are important in the dichotic perception of speech. The levels of processing and the fusions which comprise them are outlined in Table 3-1. Briefly, the levels of processing include fusion by integration of interaural phase and amplitude information, fusion by integration of acoustic features, and fusion by disruption and recombination of linguistic information. Fusion by integration of ITD and ILD information is important in sound localization and was outlined in 3.3.1. Fusion by disruption and recombination of linguistic information is the result of fine-grained acoustic manipulations that seem unlikely to either arise from or be preserved in cochlear implant processing. Particularly relevant in the consideration of speech perception through bilateral cochlear implants, however, is integration by acoustic features, and especially psychoacoustic fusion and spectral fusion.

Table 3-1. Levels of processing and binaural fusions outlined in Cutting (1976).

Level of processing	**Binaural Fusion**
Integration of phase and amplitude cues	localization
Integration of acoustic features	psychoacoustic fusion spectral fusion spectral temporal fusion
Integration by disruption and recombination of linguistic features	phonetic feature fusion phonological fusion

Evidence from dichotic listening experiments suggests that information presented in complement across the two ears is easily integrated. In a landmark study, Broadbent and Ladefoged (1957) showed that listeners presented with the F1 and F2 of a /da/ syllable separately to opposite ears perceived the syllable as /da/. This process, later termed *spectral fusion* by Cutting (1976), is robust to interaural differences in level and fundamental frequency, but not relative onset time; the majority of listeners appear able to integrate (tonotopically matched) acoustic cues efficiently (Rubin, Uchanski, and Braida, 1992), though reduced spectral resolution may decrease this ability (Loizou, Mani, and Dorman, 2003).

To the extent that the information delivered to each ear with bilateral CIs is in complement, then, we might expect binaural integration to be relatively straightforward, and to yield a binaural advantage. However, this might be achieved through various possible mechanisms. For example, if the frequency-place maps are similar in each ear, then listeners may be able to integrate these relatively easily, even if there is a basalward shift. Alternatively, if listeners have completely adapted to different degrees of basalward shift in each ear, then they should be able to efficiently integrate the signals to show a binaural advantage. But, given that many bilateral CI users do not achieve this advantage, it seems plausible that misaligned frequency-place maps between each ear may underpin this problem. Evidence in favour of such a hypothesis comes from the several subjects who show performance in the bilateral condition that mirrors that in the “better ear” for speech in quiet (Tyler et al., 2007) or for speech in noise lacking in directional cues (van Hoesel et al., 2002; Wackym et al., 2007).

The issue is complicated further because of the likely differences in nerve survival at each ear, which results in signal differences at the ears that extend beyond mismatches in frequency-to-place maps. In the *psychoacoustic fusion* described by

Cutting (1976), a /ba/ presented to one ear and a /ga/ presented to the opposite ear fuse to form the percept /da/, similar to the auditory-visual fusion in the McGurk effect (McGurk and MacDonald, 1976). To the extent that bilateral implants may deliver completely different signals to each ear, then, it is feasible that psychoacoustic fusion may occur, which may ultimately lead to a decrement in performance over the “better ear” alone. Higher level speech processing resulting from the combination and fusion of very different signals at each ear could supersede the reasonably accurate perception of speech delivered monaurally through a single implant.

3.3.3 Binaural integration of speech in bilateral cochlear implants

The ability to integrate dichotic speech information has been exploited in an experimental approach to bilateral CI processing. So called “*zipper processors”* excite alternate electrodes at each ear in an interleaved fashion (Lawson, Brill, Wolford, Wilson, and Schatzer, 2000; Wilson, 2005). The information in the two implants is combined across the two ears, thereby increasing the number of effective electrodes and reducing electrode interaction. These rely on the listeners’ ability to spectrally fuse the incoming signal into a single percept.

The researchers used a behavioural pitch-ranking scheme to map out interleaved (between the ears) electrodes that successively increased in pitch. Behavioural pitch-ranking tasks are inexact, since they rely on the listeners’ ability to distinguish between stimulation at adjacent electrodes which often interact with electrical stimulation. However, the interaural interleaving of channels would have reduced channel interaction within each ear, and may have made the procedure somewhat more reliable. While the absolute characteristic frequencies of the stimulating electrodes were unknown, this type of frequency-place mapping may also have ensured that spectral information was presented in complement rather than conflict across the two ears, and hence opportunity for spectral fusion may have been optimal. By contrast, standard bilateral CI fittings are not interleaved, which increases the likelihood of within-ear channel interactions from spread of excitation, for errors in pitch-ranking both within and between ears, and thus the possibility of frequency-place maps that are mismatched between the ears.

Yet, Dorman and Dahlstrom (2004) reported a binaural advantage for speech perception in two bilateral implant patients who had different cochlear implants in each ear, supporting the hypothesis that information from mismatched frequency-to-place maps can be fused. Patients showed improvements of 32 – 34% on HINT sentences with the addition of the second implant over performance with the “better ear” alone. However, the study only included two subjects, and the method for determining mismatch between the ears, pitch-ranking of electrodes, may not have been accurate. The authors conceded that the degree of mismatch between the ears was unknown, and thus the question of adaptation to such a mapping remains open. Furthermore, only one subject showed a binaural advantage for quiet speech, which suggests that binaural summation may not have been restored in the other subject.

Tyler et al. (2007) also showed binaural benefit for speech perception in sequentially implanted bilateral CI users. Remarkably, the six subjects reported in this study had implants that were inserted to different depths, that used different processing algorithms, that had different numbers of channels and pulse rates and that had been implanted as many as seventeen years apart, which suggests that these listeners were able to integrate mismatched signals at each ear. However, they do not report on any pitch-ranking of electrodes, so the extent of mismatch between the ears was unknown. Also, the binaural benefit for speech recognition was mixed: for words in quiet, only one subject showed a significant binaural benefit, which suggests better-ear listening and a lack of binaural summation. On the other hand, for sentences in noise presented at 0^{o} azimuth, most subjects showed a significant binaural benefit. However, the small number of subjects and varied results make it difficult to draw clear conclusions.

The extent to which listeners can learn to accommodate mismatches in frequency-place maps between the ears may ultimately depend on the extent to which the information presented in each ear can be perceived as complementary. In most cases, stimulation with bilateral cochlear implants is as for the unilateral case, with the frequency range important for speech being mapped onto both ears. If listeners are able to integrate this information, then, the implication is that either the information is relatively similar, or that they have adapted their perception of the signals at each ear so that integration of the signals can ultimately occur.

3.3.4 A note on pitch-matching in acoustic and electric hearing

Section 2.5.1.3 presented evidence from cochlear implant users who wore an acoustic hearing aid in the contralateral ear (Blamey et al., 1996). In that study, the pitch elicited by the most apical electrode was lower than would be predicted from the insertion depth and a Greenwood map of frequency to cochlear place. More recent studies are also consistent with these findings (Boex et al., 2006; Dorman et al., 2007; James, Blamey, Shallop, and Incerti, 2001). For example, James et al. used contralateral masking to estimate the pitch matches between acoustic and electric hearing, and the results were consistent with Blamey et al.'s earlier study. One possible explanation for this was that the listeners had adapted their perception of the electrical stimulation over time to match the acoustic frequency stimulated by the same region. Evidence supporting this hypothesis comes from a longitudinal study of electric-acoustic pitch matches in users of the hybrid short-electrode cochlear implant/acoustic hearing aid (Reiss, Turner, Erenberg, and Gantz, 2007).

While these are typically worn ipsilaterally, so issues of binaural integration do not apply, contralateral-ear pitch matches were measured in this study. The evidence suggests that CI users change their pitch percepts gradually over long periods, although this does not appear to be correlated with speech perception. Interestingly, in a much earlier report on one of the only patients to be implanted with an implant who had normal acoustic hearing in the contralateral ear, an estimation of pitch match between acoustic and electric hearing made during implant surgery suggested that the matching acoustic pitch was as would be predicted from the electrode insertion depth and a Greenwood map (Eddington, Dobelle, Brackmann, Mladejovsky, and Parkin, 1978).

Taken together, these studies indicate that the pitch percepts of electrically evoked stimuli may change over time, but this seems only clear for unilateral electrical stimulation, where the contralateral (acoustic) ear retains an intact (although possibly band-limited) tonotopic map. It is plausible that in these circumstances the natural tonotopic map serves as a sort of target for adaptation. For example, two subjects in Reiss et al.'s study showed electric pitch matches changing over time towards the corresponding CF of the frequency range assigned to that electrode, which may have been made possible through listening to the corresponding frequency in the contralateral acoustic ear. On the other hand, there was no clear pattern across all of the subjects, and the large within-subject variability between sessions (pitch matches varying by up to two octaves), highlights the poor reliability of such measures. Even still, in the case of bilateral implants, there is unlikely to be a *current* tonotopically matched map to serve as a target, though one may exist in memory. How do listeners adapt to these signals in the absence of a "tonotopically matched target"? Are changes in pitch ranking with experience reflected in speech perception?

3.3.5 Speech training with bilateral cochlear implants

Chapter 2 presented evidence about the importance of supervised learning in adaptation to cochlear implants, though this was always for monaural/monotic stimulation. It can be argued that targeted speech training may be even more important for bilateral stimulation than for monaural stimulation, because listeners must adapt to not one but two altered frequency-place maps, and they must also learn to integrate the two mismatched signals. In cases of large interaural asymmetries in speech intelligibility, listeners may adapt by tuning in to the “better ear” only. This seems an especially important consideration for bilateral implants that are received sequentially, since the novel stimulation may prove difficult to adapt to in the presence of electrical stimulation in the contralateral ear that already provides an altered frequency-place map. Anecdotal evidence from one subject in Tyler et al.’s (2007) study suggests that targeted training with the later implant alone was essential for adaptation, and similar evidence exists from a study of sequentially implanted children (Kuhn-Inacker, Shehata-Dieter, Muller, and Helms, 2004).

For simultaneous bilateral implantation, if no targeted training is provided, “better ear” effects may also dominate in the case of very different frequency-place maps, and the patient may learn to simply ignore the “worse ear”. Indeed, some simultaneously implanted bilateral cochlear implant listeners with asymmetrical speech scores in each ear do not show a binaural benefit beyond that with the “better ear” (Mosnier, Sterkers, Bebear, Godey, Robier, Deguine, Fraysse, Bordure, Mondain, Bouccara, Bozorg-Grayeli, Borel, mbert-Dahan, and Ferrary, 2009; Litovsky et al., 2006), and it is possible that a lack of appropriate training underpins this asymmetry.

However, if training is to be provided for bilateral implants, it is unclear whether explicit training time with each implant individually is necessary, or whether some combination of binaural and monaural training is best. There are no systematic studies in the clinical literature that explore the issue of training condition.

3.4 Can mismatched frequency-place maps be learned?

This thesis presents research from simulations of bilateral cochlear implant processing in normal-hearing listeners. The aim was to extend previous findings on adaptation to monaural basalward shift to the binaural case, to see whether listeners can learn to tolerate mismatches in frequency-place mappings at each ear. While evidence seems clear that listeners can adapt to new frequency-place maps while retaining natural and other frequency-place maps (Fu et al., 2002; Fu and Galvin III, 2003; Rosen et al., 1999; Hofman et al., 1998), this is the first attempt to look at adaptation to mismatched maps that are presented simultaneously.

This is also one of the first attempts to clarify how the perception of mismatches in frequency-place maps changes over time in the case of (simulated) bilateral stimulation with CIs. Though 3.3.4 presented some longitudinal evidence about the pitch of electrical stimulation, this was obtained through acoustic pitch matches where some residual hearing was intact. The situation may be different when both ears are stimulated electrically, and it remains unclear how changes in pitch percepts affect speech perception. Most of the studies reporting on speech perception in bilateral cochlear implants have examined experienced users[5], making it difficult to understand the impact of mismatches in frequency-place maps, and the reasons why some bilateral implant users do not gain a binaural benefit. Does the lack of binaural advantage seen in bilateral CI users result from an inability to adapt to mismatches in frequency-place maps?

[5] See Chapter 6 for a discussion of newly-implanted bimodal users (Simpson, McDermott, Dowell, Sucher, and Briggs, 2009).

There has also been little attempt in the clinical literature to clarify pitch matches for users of bilateral cochlear implants, probably because of the difficulty in doing so accurately. The clinical literature suggests that bilateral implants are fitted separately in the standard way without attention to interaural pitch ranking (Muller, Schon, and Helms, 2005; Gantz, Tyler, Rubinstein, Wolaver, Lowder, Abbas, Brown, Hughes, and Preece, 2002; Litovsky et al., 2006). Without this knowledge, it is difficult to know whether the bilateral CI users who show binaural advantage have adapted to any interaural differences in frequency-to-place maps, or whether such differences did not materialize in these listeners. Using simulations in normal hearing listeners thus allows for the examination of the adaptation process under tight experimental control, in order to determine the theoretical constructs under which such adaptation may or may not be possible.

If listeners are able to adapt to mismatches in frequency-to-place maps between the ears, then we would expect to see a binaural advantage, and such adaptation would prove promising for the increased interest in bilateral implantation. On the other hand, if no binaural advantage is apparent, then listeners may not be able to adapt, and this has striking implications for bilateral stimulation with cochlear implants. Evidence from psychophysics already suggests that frequency-place alignment between the ears is important for optimal binaural perception with bilateral CIs (Long et al., 2003). If this can also be shown for speech perception, then there is a clear and present need for more accurate techniques to align frequency-place maps between the ears when fitting bilateral cochlear implants.

The work in this thesis assumes that supervised learning plays a key role in adaptation to the processed speech, as outlined in Chapter 2. However, as has been shown in studies of sequential bilateral cochlear implantation, the situation for bilateral stimulation may be quite different to the monaural case. (Tyler et al., 2007)

This thesis also examines the effects of monaural and binaural training on adaptation to interaural frequency-place mismatches.

Perceptual adaptation is explored for a large frequency-place mismatch for speech in quiet, and a moderate frequency-place mismatch for speech in noise. Much of the literature on bilateral cochlear implantation has focused on the restoration of binaural hearing benefits such as sound localization and binaural squelch. This thesis takes a different approach, examining rather the *consequences* of bilateral stimulation with mismatched frequency-place maps on speech perception and adaptation.

Chapter 4. Perceptual Adaptation to Binaurally Mismatched Frequency-to-Place Maps I: Speech in Quiet

The studies reported in this chapter were subsequently published as Siciliano, C.M., Faulkner, A., Rosen, S., & Mair, K.(2010), Journal of the Acoustical Society of America, vol 127(3), pp 1645-1660.

This chapter reports on two experiments examining speech perceptual adaptation to sine-excited vocoder simulations of bilateral cochlear implantation. The primary aim was to examine whether normally hearing listeners can adapt to a binaurally mismatched (and unilaterally shifted) cochlear frequency-to-place map after a period of training. The extent of mismatch between the ears was large—6 mm—in order to simulate the effect of a shallowly implanted electrode in one of two stimulated ears. A secondary aim of these experiments was to explore the role of the training condition in facilitating adaptation to binaurally mismatched frequency-to-place maps. Finally, the idea that unfavourable conditions such as a large mismatch between the ears may lead listeners to learn to ignore an ear was explored.

In Experiment 1, one group of listeners was trained with a binaurally mismatched processor over eight 40-minute training sessions, and speech perceptual adaptation was measured before, during and after training. It was hypothesized that the presence of moderately intelligible unshifted frequency components may hinder adaptation to the shifted frequency-place map, and thus the binaurally mismatched processor. Consequently, another group was trained and tested with just the shifted components to examine whether adaptation to the binaurally mismatched map could be facilitated by training with the shifted processor alone, in the absence of the conflict between the two maps.

In Experiment 2, the time course for training and training condition were further explored. Perceptual adaptation to the binaurally mismatched map was monitored over an extended course of training, and the question of whether adaptation could be facilitated by training with both the shifted alone and binaurally mismatched frequency-to-place maps was also tested. The length of training was doubled and the training condition alternated between that of the two groups from Experiment 1.

4.1 Speech processor design

Compared to monaural stimulation with cochlear implants, bilateral stimulation introduces complicating factors that may lead to decrements in speech perception. Consider the simplest case where a certain number of frequency bands are vocoded and presented to each ear. Introducing a spectral shift to one of two ears, or introducing different degrees of spectral shift into each ear, may result in at least three factors that may adversely affect speech perception:

1) the frequency-place maps in each ear are no longer aligned

2) the *same spectral information* is delivered to *different tonotopic places* in each ear

3) *different spectral information* is delivered to the *same cochlear place*, resulting in energetic masking

Figure 4-1 depicts this type of processing for a six channel sine vocoder. For each analysis band, the same spectral content is being delivered to two different tonotopic locations in each ear. Furthermore, sine carriers at the CF locations of frequency bands 3, 4, 5 and 6 are carrying *different* spectral information in each ear. If the simulations were processed in this way, and subjects either failed to show a binaural advantage, or else demonstrated intelligibility with the combination of shifted and unshifted channels that was *worse* than that with the monaural unshifted channels,

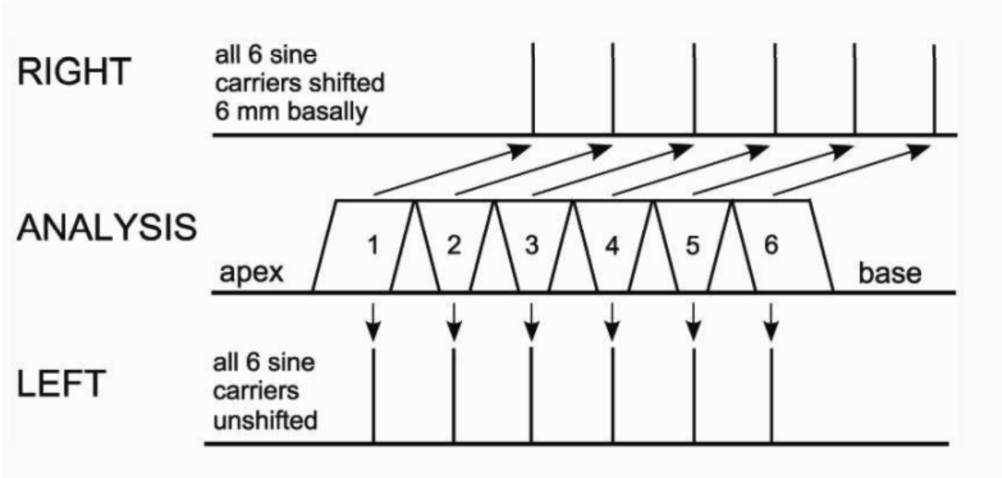

Figure 4-1. Processing depicting sine-vocoded simulations of bilateral cochlear implants. Here, all analysis channels are presented to both ears, with the sine carriers at the right ear being upwardly shifted by an equivalent of 6 mm basilar membrane difference.

then there would be no way to know which of these possible factors was responsible.

Emerging evidence suggests that the second factor may have an adverse effect on speech perception. Smith and Faulkner (2007) used noise-vocoding to simulate bilateral electro-acoustic stimulation. One ear was presented with low-frequency information that approximated residual hearing in the non-implanted ear, while the contralateral ear was presented with noise-vocoded simulations of a shallowly implanted CI. After several hours of speech training, they found that subjects had adapted to the speech processing, but that intelligibility of sentences was significantly worse when there was a region of conflict between the two ears. When the same spectral information was delivered to both the acoustic hearing ear unshifted, and to the electric hearing ear with a basalward shift, i.e. to *different cochlear locations*, speech perception was worse (by 10% of keywords) than if the conflicting information was excluded from presentation. There is also clinical evidence from monaural electro-acoustic stimulation (EAS) patients that a region of conflict between acoustic and electric stimulation may adversely affect speech perception, though this is only from a small number of subjects (Vermeire, Anderson, Flynn, and Van de Heyning, 2008).

In examining speech perception in bilateral implant patients, then, the failure of some patients to receive binaural benefit may result from regions of conflict of stimulation rather than mismatched frequency-to-place maps *per se*, but there is no clear way of knowing this without prior knowledge of the exact frequency-place alignment in each ear, which is not, at present, feasible.

Another complicating issue with this type of processing scheme is the likelihood of a dominance of “better ear” effects. Normal-hearing listeners adapt readily to unshifted vocoder simulations with six channels (Dorman, Loizou, and Rainey, 1997), making the prospect of a binaural advantage for speech in quiet rather unlikely with this processing scheme, since the additional shifted ear adds information that is both inferior and redundant. The intelligibility decrements seen when the output has been upwardly shifted by 6 mm and more may be mitigated by speech training, but the decrements with shifts this large are not eliminated (Rosen et al., 1999; Fu et al., 2002). Decreasing the number of channels may be one way to limit the ceiling of performance for the unshifted speech, but it is still unlikely to lead to binaural advantage, since the smaller number of shifted channels will always have the same spectral information as the channels in the unshifted ear, but are likely to remain less intelligible, even after training.

In real users of bilateral cochlear implants, the binaural advantage for speech in quiet may be contingent on the varying patterns in nerve survival in the two ears: the information being received at each ear may not be as extensive as the information being delivered, so the information at the two ears is not always redundant. While it is difficult to simulate this effect in normal hearing listeners, presenting frequency information as in the example above would almost certainly result in better-ear effects, especially for speech in quiet. For these reasons, the processing used in these experiments departed from what may be typical of bilateral cochlear implants in a clinical situation, but it was nevertheless carefully designed to yield results that would be clearly interpretable and free from these confounding effects. Using simulations in this way allowed for a highly-controlled study of the effects of binaurally mismatched frequency-to-place maps that is not possible in implant patients.

All speech processing was based on interleaved six-channel sine-excited vocoding (Loizou, Mani, and Dorman, 2003). Figure 4-2 shows a schematic representation of the binaurally mismatched processor. From low frequency (apical end) to high frequency (basal end), odd-numbered bands 1, 3 and 5 were presented to the right ear with the equivalent of a 6 mm basalward basilar membrane shift (assuming a 35 mm cochlea). Even-numbered bands 2, 4 and 6 were presented to the left ear without a shift. This interleaved binaural configuration is similar to the zipper processor described by Lawson et al. (Lawson, Brill, Wolford, Wilson, and Schatzer, 2000; Wilson, 2005), which was shown to be beneficial for consonant recognition for

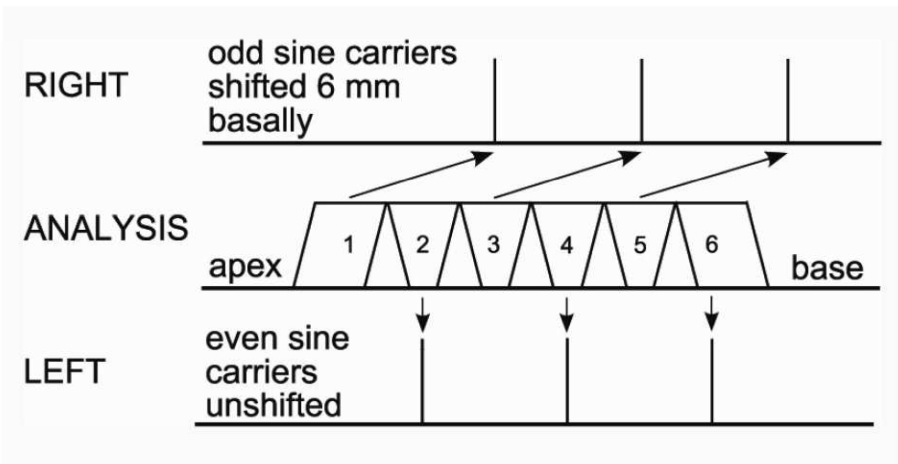

Figure 4-2. Speech processing design of the binaurally mismatched frequency-to-place maps (Experiments 1 and 2).

some CI listeners. In tests of dichotic speech perception in noise, Loizou et al. (2003) showed that for dichotically presented (unshifted) noise-vocoded speech, intelligibility was higher (and thus presumably binaural integration stronger) when channels were interleaved as here, rather than split according to low and high frequency.

The processing was carefully designed so that different spectral information was presented to each ear. In an abstract sense, this aspect of the processing may approximate the sparse representation of spectral information at each ear sometimes seen in implant patients. If adaptation to the shifted channels involves a remapping of the percept to sound more like tonotopically aligned speech, as is suggested by Reiss et al. (2007), then to the extent that the information at each ear is perceived as complementary, binaural integration should be facilitated.

The interleaved configuration also prevented the listener from achieving ceiling performance with the unshifted ear from the outset, which would have been likely had all six channels been presented to both ears. The number of channels – six – was carefully selected to minimize the chance of ceiling performance in unshifted conditions, but also to provide a shifted signal that could still be learned. It has been widely found that an unshifted 3-channel vocoder leads to substantially poorer scores for most speech materials than an unshifted 6-channel vocoder (Dorman et al., 1997; Fishman, Shannon, and Slattery, 1997; Loizou, Dorman, and Tu, 1999). Thus, any improvements gained through integration of the two ears should be easily detected. A pilot study showed that unshifted conditions were prone to ceiling effects with some easier speech materials, so the use of more than six channels in quiet was not explored (Faulkner, 2006).

4.2 Experiment 1

4.2.1 Method

4.2.1.1 Subjects

Twelve normally hearing speakers of British English took part, and each was paid for his or her participation. All had pure tone audiometric thresholds better than 20 dBHL at octave frequencies from 250 to 8000 Hz. Ethical approval for this study was granted by the UCL Research Ethics Committee.

4.2.1.2 Test conditions

Table 4-1 summarizes the six conditions tested in Experiment 1. These conditions were designed to assess what aspects of the signal were involved in any learning that took place. Test conditions were composed of monotic, dichotic and diotic combinations of the spectral components outlined in Figure 4-2 for the binaurally mismatched processor. Condition names were coded by number of channels, dichotic presentation, even- or odd-numbered channels, and shifted or unshifted presentation. When basalward shift was applied, it was always to the odd-numbered channels. Even-numbered channels were always unshifted.

Table 4-1. Conditions for Experiment 1

<table>
<tr><th rowspan="2">condition</th><th rowspan="2">abbreviation</th><th colspan="2">component bands and shift</th></tr>
<tr><th>right</th><th>left</th></tr>
<tr><td>six dichotic unshifted</td><td>6DU</td><td>odd</td><td>even</td></tr>
<tr><td>six dichotic odd shifted</td><td>6DOS</td><td>odd → 6 mm</td><td>even</td></tr>
<tr><td>six odd shifted</td><td>6OS</td><td>odd → 6 mm
even</td><td>odd → 6 mm
even</td></tr>
<tr><td>three even unshifted</td><td>3EU</td><td></td><td>even</td></tr>
<tr><td>three odd shifted</td><td>3OS</td><td>odd → 6 mm</td><td></td></tr>
<tr><td>three odd unshifted</td><td>3OU</td><td>odd</td><td></td></tr>
</table>

The **six dichotic unshifted (6DU)** condition served as a control to assess maximal intelligibility. Here, unshifted odd- and even-numbered channels were interleaved between the ears. A pilot study indicated that performance with this dichotic unshifted processor yielded intelligibility equivalent to a 6-channel diotic processor. The main experimental condition (Figure 4-2) was the **six dichotic odd shifted (6DOS)** condition. Here, odd-numbered bands were shifted an equivalent of 6 mm basally and presented to the right ear, while even-numbered bands were presented to the left ear without a shift. This binaurally mismatched processor was also used for training one group of subjects. The **six odd shifted (6OS)** condition contained the same information as the 6DOS condition, but was instead diotic: all bands were summed together and presented to both ears. In contrast to 6DOS, 6OS lacked any cue from ear of presentation to the carrier bands that were shifted. If performance here differed markedly from the 6DOS condition, this would suggest that listeners were learning to attend to information in an ear-specific manner. Conversely, if there was no difference, then it could be inferred that learning was occurring on the basis of the carrier frequencies. In the **three even unshifted (3EU)** condition, even-numbered channels alone were presented to the left ear unshifted. This was equivalent to the unshifted components of the 6DOS processor. In the **three odd shifted (3OS)** condition, odd-numbered channels were presented alone to the right ear with a 6 mm basalward shift. This comprised the shifted components of the 6DOS processor. This processor was also the trained processor for a second group of subjects. In the **three odd unshifted (3OU)** condition, odd-numbered channels were presented to the right ear unshifted. This condition was only tested for the group trained with the 3OS processor, and allowed a comparison of the information provided by the odd and even numbered bands in the absence of shifting.

4.2.1.3 Signal processing

Centre and crossover frequencies for the analysis and output filters were calculated using Greenwood's equation and its inverse, relating position along the basilar membrane to characteristic frequency. Table 4-2 shows input and output CFs as well as filter cutoffs. The assumed cochlear length was 35 mm (Greenwood, 1990). While the Greenwood equation may not be the most accurate representation of frequency-place mapping for more modern cochlear implants (Sridhar et al., 2006), it is nevertheless the most widely used mapping for simulations of basalward spectral shift. Hence, this mapping was chosen so that the results would be more comparable to previous studies.

$$frequency = 165.4(10^{0.06x} - 1)$$
$$x = \frac{1}{0.06}\log\left(\frac{frequency}{165.4} + 1\right)$$

The amplitude envelope of each band was extracted with an analysis filter, full-wave rectification, and a smoothing filter. The envelope was then multiplied by a sinusoid with frequency matching the centre frequency (CF) of the band (or shifted equivalent). Finally, the requisite bands were summed and presented to the left and/or right ears as determined by processor condition. All processor conditions used the same six analysis filters and sine carriers at either the shifted or unshifted CFs of these analysis filters.

A real-time implementation of the vocoder processor was used for live training, while offline processing of the test material was implemented in MATLAB. This ensured identical repetition of the test materials for each subject. Offline processing was executed at a 44.1 kHz sampling rate. Analysis bands were determined by a serial implementation of high-pass and low-pass third-order Butterworth IIR filters.

Table 4-2. Analysis band cutoff and centre frequencies, and carrier frequencies for each band in the unshifted (6DU, 3EU, 3OU) and odd-band shifted (6DOS, 6OS, 3OS) conditions

Band	Analysis band cutoff (Hz)		Analysis band CF (Hz)	Carrier frequency (Hz)	
	lower	Upper		unshifted	odd-shifted
1	200	403	290	290	878
2	403	718	543	543	543
3	718	1208	936	936	2359
4	1208	1971	1547	1547	1547
5	1971	3157	2498	2498	5937
6	3157	5000	3977	3977	3977

Adjacent filter responses crossed at 3 dB down from the peak of the pass-band. Envelope smoothing used second-order low-pass Butterworth filters with a 32 Hz cutoff. Note that the use of a low-frequency smoothing filter with sine-vocoding meant that any cues to temporal fine structure were removed. This was essential for keeping the spectral information delivered to each ear relatively independent, but it also better replicates the situation seen in real CIs. Real-time processing was implemented using the Aladdin Interactive DSP Workbench (Hitech Development AB) and ran on a DSP card (Loughborough Sound Images TMSC31). The computational power of the DSP was limited so the sampling rate was restricted to 16 kHz and elliptical rather than Butterworth filter designs were used with the same 3 dB crossover frequencies as for the offline processing. Analysis filters consisted of fourth order band-pass designs, while third order low-pass filters were used for envelope smoothing.

In both testing and training, an equal loudness correction was applied to each of the shifted bands to preserve relative loudness across the spectra of unshifted and shifted speech. The correction was set to half the difference (in dB) between the minimal audible field threshold of the analysis filter and that at the centre frequency of the shifted output filter. Minimal audible field values were taken from Robinson and Dadson (1956) and interpolated using a cubic spline fit to log frequency.

4.2.1.4 Training

Subjects were trained with Connected Discourse Tracking (CDT) (De Filippo and Scott, 1978). In this method, the experimenter reads successive phrases from a text to the subject, who then repeats back what he or she heard. This allows the listener to acclimatize to the spectrally distorted speech while engaging in a communication task that is similar to a conversation. The number of words repeated back correctly per minute provides a measure of progress throughout training. CDT has been shown to be an effective training method for spectrally shifted speech (Faulkner et al., 2006; Rosen et al., 1999).

The talker for the CDT portion of this experiment was the author of this thesis. Although she is a native speaker of a northeastern dialect of American English, she had been living in the UK for 5 years at the time of testing, and has been judged to have an accent similar to Standard Southern British English. CS's speech was not used for any of the testing. The talker read from the text in short phrases, and the listener repeated back what he or she heard. If the listener's response matched what the talker had said, the talker would move on to the next phrase. Otherwise the phrase was repeated. If after the third presentation, the listener could still not reproduce the phrase, the listener was presented the phrase as unprocessed speech (to the left ear only). Texts for CDT were chosen from the Heinemann Guided Readers series (Elementary Level). These texts are designed for learners of English as a second language and make use of controlled vocabulary and syntactic complexity. During training, the talker and subject were situated in adjacent soundproof rooms. The room had a double-glazed window that enabled auditory-visual training. During auditory-only training, the window was blinded. A constant pink masking noise at 45 dBA was played in the listener's room to mask any speech that might be transmitted through the wall and window.

Of the twelve subjects, six were trained with the 6DOS processor (6DOS-trained group), and the remaining six were trained with the 3OS processor (3OS-trained group).

4.2.1.5 Test materials

a. Sentence perception

The IEEE/Harvard sentence lists (Rothauser, Chapman, Guttman, Nordby, Silbiger, Urbanek, and Weinstock, 1969) were used, which have very little contextual information. Digital recordings of the sentences were from one male and one female talker of British English (16-bit, 48 kHz downsampled to 44.1 kHz). The 72 lists in the set each contained ten sentences with five keywords in each sentence. The first 36 lists were designated for the female talker, and the remaining 36 lists were designated for the male. A subset of 32 lists from each talker was used for the 6DOS-trained group, who were tested with fewer conditions. For each test session, two lists per condition were chosen from each talker set in a pseudo-random manner. No list appeared more than twice in the same condition across all of the subjects, and subjects never heard a list more than once. No feedback was given for the sentence material.

It was anticipated that the 3OS training processor would be much more difficult to learn than the 6DOS processor, and that IEEE sentence scores would be subject to a floor effect, so the group trained with this processor was also tested with the easier BKB (Bench, Kowal, and Bamford, 1979) and IHR (MacLeod and Summerfield, 1990) sentences after the final test session. BKB lists were used for the male talker, and the IHR sentences for the female talker. These two sentence sets have similar syntactic constructions and are essentially equivalent in intelligibility, with any difference in materials being less important than the expected talker differences. The female talker for the IHR lists was the same as that for the IEEE sentences, but the male talker was different. The BKB sentence lists consist of 16 sentences per set with either 3 or 4 keywords per sentence, making 50 keywords per list. The IHR sentences consist of 15 sentences of 3 keywords each, hence 45 keywords per list.

For sentence tests, the subject was asked to repeat back to the experimenter as many words as he or she could. Words were counted correct when the word root was repeated correctly.

b. Vowel identification

Nine b-vowel-d words in the carrier sentence ‘Say bVd again’ were recorded by a male and female speaker of British English in anechoic conditions at a 48 kHz sampling rate, and subsequently down-sampled to 44.1 kHz. The male talker was the same as for the sentence test, although the female talker was different. Five tokens of each bVd sentence were recorded from each talker, so that in an individual test of either a male or female talker in a given condition, there were 45 items. Vowels were restricted to monophthongs of similar duration, so that listeners would need to rely on spectral cues for identification: /æ/ (bad); /ɑː/ (bard); /iː/ (bead); /e/ (bed); /ɪ/ (bid); /ɜː/ (bird); /ɔː/ (board); /ɑ/ (bod); /uː/ (booed); /ʌ/ (bud). A grid with all nine words appeared, and the subject clicked with the computer mouse on the button displaying the word they perceived. The vowels were represented on the buttons in the orthographic form given above. Before testing, the subject was given a practice session in which the vowel material was presented unprocessed, with a single token for each vowel and each talker. This enabled the subjects to familiarise themselves with the software and the task.

4.2.1.6 Procedure

Subjects were tested before training commenced, halfway through training, and at the end of training. All subjects completed the entire cycle of training and testing within a maximum of two weeks, with no more than a two day gap between successive sessions. Sentence and vowel test presentation was counterbalanced across the group. Within each test, stimuli were pseudo-randomized by block of condition and talker. Table 4-3 describes the sequence of training and testing for Experiment 1.

In the first session, subjects were acclimatized to the speech processing with a ten-minute block of dichotic unshifted speech with CDT prior to the pre-training test session. Previous experiments with (unshifted) vocoded speech have demonstrated that listeners require a short period of acclimatization before they can reliably perceive speech through unshifted vocoders with a limited number of channels (Davis, Johnsrude, and Hervais-Adelman, 2005). As in subsequent CDT training sessions, the first five minutes of the familiarization block were auditory-visual (AV), while the remaining five minutes were auditory alone (AA). Following the pretest, subjects were trained with CDT in four 40-minute training sessions with either the dichotic odd-shifted speech (6DOS-trained group) or the monaural 3 shifted channels speech (3OS-trained group). Subjects were tested after the first four training sessions. If the testing session did not take place immediately following a training session, then the experimenter administered a ten-minute (5 AV, 5 AA) CDT block with the 6DOS or 3OS processor, which was not counted towards the total hours of training. Following this mid-training test session, subjects underwent four

Table 4-3. Sequence of training and testing conditions for Experiment 1

Session	Training	Processor	Testing
1	5 minutes audio visual (AV) 5 minutes auditory alone (AA)	6DU	*Familiarization:* unprocessed vowels *Pre-test*: IEEE sentences (2 lists x 5(6) conditions x 2 talkers) bVd identification (5 tokens x 5(6) conditions x 2 talkers)
2	5 minutes AV 35 minutes AA	6DOS or 3OS	None
3	5 minutes AV 35 minutes AA	6DOS or 3OS	None
4	5 minutes AV 35 minutes AA	6DOS or 3OS	None
5	5 minutes AV 35 minutes AA	6DOS or 3OS	None
6	Familiarization if not immediately after session 5 (see text)		*Mid-test:* IEEE sentences (2 lists x 5(6) conditions x 2 talkers) bVd identification (5 tokens x 5(6) conditions x 2 talkers)
7	5 minutes AV 35 minutes AA	6DOS or 3OS	None
8	5 minutes AV 35 minutes AA	6DOS or 3OS	None
9	5 minutes AV 35 minutes AA	6DOS or 3OS	None
10	5 minutes AV 35 minutes AA	6DOS or 3OS	None
11	Familiarization if not immediately after session 10 (see text)		*Post-test:* IEEE sentences (2 lists x 5(6) conditions x 2 talkers) bVd identification (5 tokens x 5(6) conditions x 2 talkers) IHR/BKB sentences (2 lists x 6 conditions x 2 talkers) (3OS-trained group only

more 40-minute training sessions with the same processor as in the first four sessions, and then completed the post-training testing.

Testing took place in a sound-proofed room with presentation of the processed speech over Sennheiser HD280 headphones. The level was set by the experimenter to a comfortable listening level of approximately 65 dB. Subjects could request for the level to be adjusted, but none did, so this level was used by all participants.

4.2.2 Results

Test data was analyzed using repeated measures analysis of variance (ANOVA), with within-subject factors of test session, condition and talker, and where relevant, a between-subjects factor of trained processor. Hyunh-Feldt epsilon corrections were applied to all *F* tests for factors with more than one degree of freedom. Hyunh-Feldt adjusted degrees of freedom have been rounded to the nearest integral value, and the significance criterion was $p = 0.05$. *A priori* hypotheses were tested using planned contrasts, and *post-hoc* testing was carried out using Bonferroni-adjusted paired comparisons.

4.2.2.1 IEEE sentence perception

a. 6DOS-trained group

Keywords correct for the IEEE sentence test across training sessions are shown in the left-hand panels of Figure 4-3. For the 6DOS-trained group, performance with the 3OS processor remained close to floor throughout training, while intelligibility with all other processors tended to increase. This was indicated by the significant interactions of talker with processor [$F(3, 17) = 6.57$, $p = 0.003$] and training session with processor [$F(8, 40) = 445$, $p < 0.001$]. When the data were reanalyzed excluding the 3OS condition, these interactions were no longer significant. Scores were slightly better with the male talker—not shown—at all test points (except for the 3OS condition).

Post-hoc testing on the post-training sentence scores revealed three key findings. First, performance with the 3 unshifted channels (3EU) was significantly worse than with the dichotic unshifted (6DU) condition (p = 0.001), which is a clear indication that there was room for improvement by integrating the odd and even channels, at least when they were unshifted. Second, there was no significant difference between the 6DOS condition and the 3EU condition, which indicates that subjects

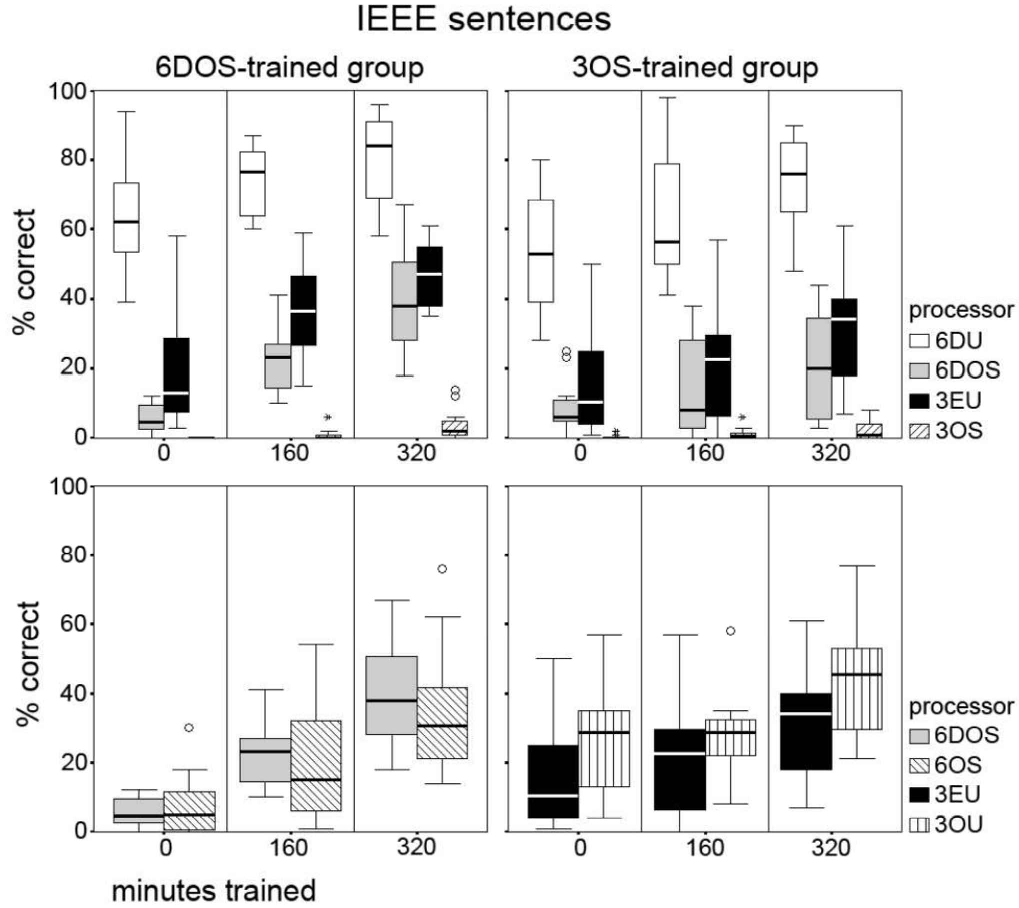

Figure 4-3. Experiment 1 box-and-whisker plots of IEEE sentence scores as a function of training time. Plots are shown for the two talkers combined. A vertical line separates each training time period. Scores for the 6DOS-trained group are in the left hand panels, while scores for the 3OS-trained group are in the right panels. The top panels show the four main conditions – 6DU, 6DOS, 3EU and 3OS. The bottom left panel shows a comparison of 6DOS and 6OS, which did not differ significantly after training. The bottom right-hand panel shows a comparison of odd and even unshifted channels (3EU and 3OU), where 3OU was significantly better than 3EU after training.

did not learn to integrate spectral cues across the two ears with the 6DOS processor. If they had, we would expect to see performance with this processor exceeding that in the 3EU condition, as was found for the 6DU condition. Finally, no significant difference was found between the 6DOS and 6OS conditions. That intelligibility was not affected by dichotic versus diotic presentation indicates that listeners were not simply ignoring information at one ear, but rather tuning into the unshifted frequencies. This finding is highlighted in the bottom left-hand panel of Figure 4-3.

b. 3OS-trained group

Scores for IEEE sentence recognition for the 3OS-trained group are shown in the right-hand panels of Figure 4-3. It is clear that learning for this group was not as great as for the 6DOS-trained group. Despite showing equivalent intelligibility to the 6DOS-trained group for all conditions before training, post-training performance was lower in all conditions. However, a repeated measures ANOVA showed significant main effects of talker [$F(1, 5) = 35.1$, $p = 0.002$)], number of training sessions [$F(2, 10) = 36.0$, $p < 0.001$] and processor [$F(3, 16) = 85.9$, $p < 0.001$]. In contrast to the 6DOS-trained group, there were no significant talker by processor or training session by processor interactions. This is an indication that there was improvement across training and also that this improvement was relatively similar in all conditions, including 3OS.

Post-hoc comparisons of post-training sentence scores for the 3OS-trained group were similar to the 6DOS-trained group across conditions. Performance with 6DU was significantly better than 3EU ($p < 0.001$), there was no significant difference in intelligibility with 6DOS and 6EU, and there was also no significant difference between 6DOS and 6OS. Thus while the intelligibility scores for this group were generally lower and improvements with training smaller than those seen with the 6DOS group, the pattern between conditions was similar. Subjects adapted to the vocoder processing, but there was no evidence of integration of the mismatched frequency-place maps.

4.2.2.2 Vowel identification

a. 6DOS-trained group

Vowel identification scores are summarized in the top panel of Figure 4-4. The overall pattern of results across the sessions and conditions for both groups was similar to that seen for sentences. A repeated measures ANOVA showed significant main effects of number of training sessions [$F(2, 10) = 37.4$, $p < 0.001$] as well as processor [$F(4, 20) = 102$, $p < 0.001$], and a significant training sessions by processor interaction [$F(8, 40) = 89.6$, $p = 0.005$]. The lower bound of the 95%

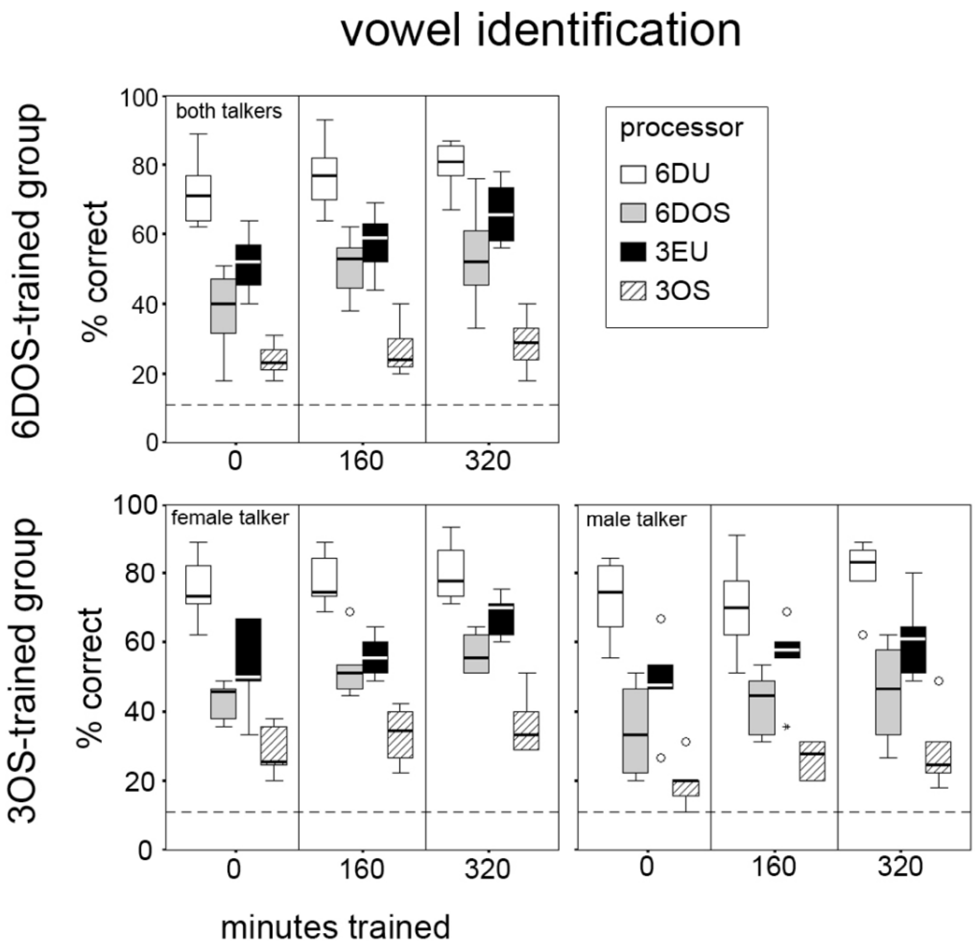

Figure 4-4. Experiment 1 vowel identification scores as a function of training time. Scores for the 6DOS-trained group are in the top panel, while scores for the 3OS-trained group are in the bottom two panels. The 3OS-trained group showed a significant talker effect, so plots for each of the two talkers are given separately. The vertical reference line separates the training time periods, and the horizontal reference line indicates chance performance, which was 11.1%.

confidence interval was above chance level performance (11.1%) at all test points and for all conditions, confirming that vowel recognition scores were significantly greater than chance.

Importantly, *post-hoc* testing showed significant improvements in all conditions between the pre-test and post-test sessions ($p < 0.05$), including 3OS, and a planned contrast of training sessions within the 3OS condition also showed significant improvement after training in this condition (p =0.01). While the significant interaction of training session with processor may indicate that improvement with the 3OS condition was smaller than for that in the other conditions, this is nonetheless evidence of learning in this condition. However, the outcome of *post-hoc* testing on the post-training vowel scores across conditions was very similar to that for sentences. Despite the improvement in the 3OS condition, there was no significant difference between 3EU and 6DOS after training, even though performance with 3EU was significantly worse than with 6DU (p = 0.02). Thus, while there was room for improvement from the addition of the odd channels (3EU < 6DU), and there was evidence of learning when these channels were shifted, subjects did not learn to integrate the mismatched frequency-place maps.

b. 3OS-trained group

The bottom two panels of Figure 4-4 show vowel identification scores for the 3OS-trained group. A repeated measures ANOVA showed a significant talker by processor [$F(5, 25) = 3.58$, (p = 0.014)] interaction, so the boxplots and subsequent analyses are given for each talker individually. In Bonferroni-corrected comparisons of talker within each condition, intelligibility with the female talker was significantly better than with the male only for conditions where spectral shifting was present, i.e. 6DOS, 6OS and 3OS. In unshifted conditions, there was no significant difference in intelligibility between the talkers. This may indicate adaptation to the shifted speech, since all experience with shifted speech outside of testing was also with a female talker. More evidence for this comes from a planned contrast of test sessions within the 3OS condition, which showed significant improvement with training in this condition only for the female talker ($p = 0.009$).

For both talkers, performance was significantly better than chance for all test conditions and at all test points, since the lower bound of the 95% confidence interval was always greater than 11%. The ANOVAs also showed significant main effects of session [*female*: $F(2, 10) = 21.2$, ($p < 0.001$); *male*: $F(2, 10) = 10.1$, ($p = 0.004$)] and processor [*female*: $F(5, 25) = 45.7$, ($p < 0.001$); *male*: $F(4, 19) = 33.7$, ($p < 0.001$)], but no significant session by processor interaction for either talker. Like the IEEE sentences for this group, then, this analysis suggests that there were significant and similar levels of learning for this group in all conditions, including 3OS.

Post hoc testing on the post-training scores revealed some differences with the 6DOS-trained group. While intelligibility with 3EU was significantly worse than 6DU for the male talker ($p = 0.006$), there was no significant difference between the 3EU and 6DU conditions for the female talker. This may indicate that for easier tasks such as closed-set vowel recognition, room for improvement from the additional shifted channels may be reduced. Secondly, post-training intelligibility with the male talker was significantly worse for 6OS than 3EU ($p = 0.045$). This is evidence that unlike the 6DOS-trained group, the 3OS-trained group may have continued to rely on cues to presentation ear after training. The 3OS-trained group received no explicit training with the 6DOS speech, and may have lacked the necessary experience to attune to the unshifted frequency components when these were not separated by ear, as in the 6OS processor.

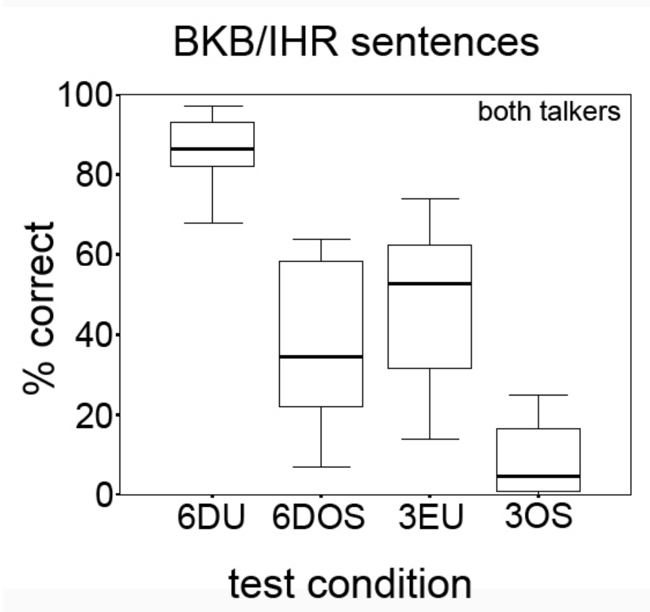

Figure 4-5. Experiment 1 (3OS-trained group) scores with high-context (BKB and IHR) sentences as a function of processor. BKB sentences were spoken by a male talker and IHR sentences by a female talker. Since there was no talker effect, these scores have been combined. These sentences were only tested after training.

4.2.2.3 Sentence perception – high context sentences

Figure 4-5 shows scores for the BKB and IHR sentences combined for the 3OS-trained group, which were only tested at the end of training. A repeated measures ANOVA showed a significant main effect of processor [$F(5, 23) = 60.0$, ($p < 0.001$)], but no significant interaction between talker and processor. *Post hoc* testing showed very similar results as with the IEEE sentences. Intelligibility with the 3EU condition was significantly worse than in the 6DU condition ($p = 0.006$). However, despite this indication of room for improvement, no evidence of integration of shifted and unshifted bands was found, since there was no significant differences between 3EU and 6DOS.

4.2.2.4 Different spectral information in odd and even channels

For the 3OS-trained group, *post-hoc* comparisons of post-training intelligibility with the IEEE sentences showed that the 3 odd unshifted channels (3OU) were significantly better than the 3 even unshifted channels (3EU) (*p* = 0.007). This is shown in the bottom right-hand panel of Figure 4-3. While data from the ANOVA showed that vowel recognition was equivalent between unshifted even (3EU) and odd (3OU) channels, a X^2 analysis comparing the vowel confusions from the 3EU and 3OU conditions showed error patterns that differed significantly ($p < 0.0001$). Thus, the vowels that subjects answered correctly differed with each of these processors, even though the overall accuracy was similar. Taken together with the finding that 6DU was always better than 3EU in this experiment, there is thus clear evidence that the spectral information in the even and odd bands was not redundant.

4.2.2.5 Comparison of 6DOS- and 3OS-trained groups

To test the hypothesis that training with the 3OS condition would better facilitate adaptation to the 6DOS processor, improvement scores for both groups from pre-training to post-training sessions were analyzed using a repeated measures ANOVA, with the training condition as a between-subjects factor. The 3OU condition was excluded from this analysis since it was only tested for the 3OS-trained group.

Boxplots showing improvement scores for the 6DOS and 3OS conditions by group are shown for IEEE sentences in Figure 4-6. For the IEEE sentences, the repeated-measures ANOVA showed a significant within-subjects effect of processor, a significant between-subjects effect of training condition, and a significant processor by training condition interaction [$F(4, 35) = 3.76$, $p = 0.015$]. *Post hoc* testing on the IEEE improvement scores showed that the 6DOS-trained group improved significantly more with training on that condition than the 3OS-trained group (p = 0.005). However, no significant group differences were found for vowel identification.

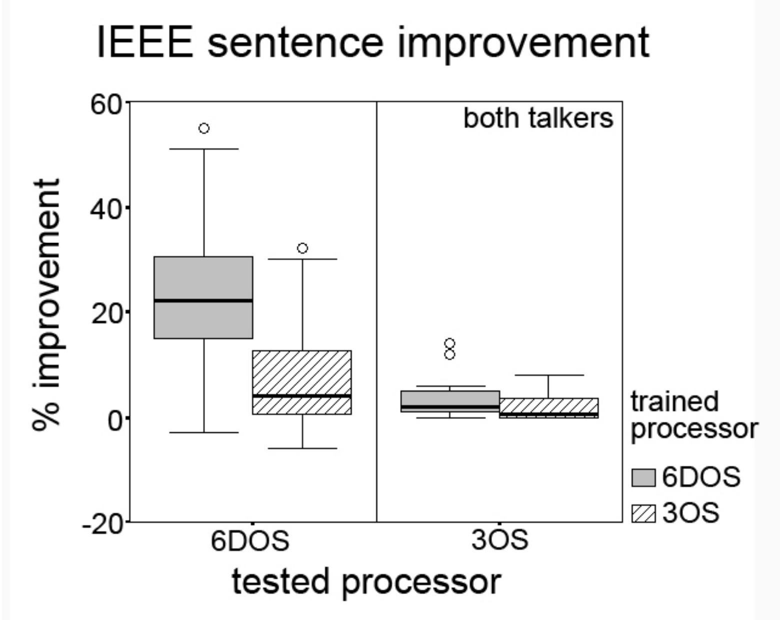

Figure 4-6. Increase in IEEE sentence scores with training for the 6DOS-trained group and the 3OS-trained group as a function of the processor used in testing.

4.2.3 Discussion

For both groups of subjects, performance with 6DU exceeded performance with 3EU in all but one subtest. Moreover, for the 3OS-trained group, the three odd unshifted channels (3OU) were *more* intelligible than the three even channels (3EU) after training with the IEEE sentence test, and vowel confusions differed significantly with these processors. This is clear evidence that the addition of the odd channels, when unshifted, provided useful cues for speech perception different to just the even unshifted bands. If subjects had adapted to the binaurally mismatched frequency-to-place map (6DOS), performance with the 6DOS condition would have also exceeded that with the 3EU condition after training. However, neither group of subjects showed significant post-training differences between the 6DOS and 3EU processor. This suggests that they relied on the sparser unshifted 3-channel input and did not take advantage of the improved spectral information available with the 6 channel binaurally mismatched processor.

The integration of mismatched frequency-place maps may require a longer period of adaptation than that needed for simple monaural basalward shifts. The improvements shown here were small and did not reach an asymptotic level after five hours twenty minutes of training, which may indicate subjects were still learning at the point of final testing. This was especially true for the 3OS condition, which proved particularly difficult for the listeners in this study in contrast to previous studies with basally shifted speech (Rosen et al., 1999). However, Tyler and Summerfield (1996) showed that cochlear implant patients are still adjusting to their speech processors six months and longer after implantation, and the training times considered here are very short by comparison.

It is also possible that the integration of mismatched frequency-place maps requires experience with both the shifted alone (3OS) and binaurally mismatched (6DOS) maps. For the 6DOS-trained group, training led to larger improvements overall, there was some evidence of adaptation specific to the shifted bands (3OS), and they showed significantly larger improvements with the 6DOS processor through training than the 3OS-trained group. However, the 6DOS-trained group also appeared to learn by relying on information contained in the unshifted frequencies, which when presented together with the shifted bands as in 6DOS, may have minimized the motivation for integration. While training with 3OS led to smaller overall improvements, for this group there was a similar level of learning for all conditions and for all test materials, since there were no training session with processor interactions, and there was also some evidence of learning specific to 3OS in the female talker vowel test. Thus while training with just the shifted components (3OS) may be insufficient to produce integration of binaurally-mismatched frequency-to-place maps, a period of training with just the shifted map may facilitate adaptation to the shifted components, which may only then lead to integration of the binaurally mismatched maps with sufficient further training.

At the same time, training with the 6DOS processor does appear necessary for frequency- rather than ear-based learning. The 6DOS-trained group showed no difference in intelligibility between dichotic (6DOS) and diotic (6OS) presentation. This indicates that they were not simply ignoring an ear, but rather they had learned by tuning into the unshifted bands or ignoring the shifted frequency bands, or possibly both. By contrast, the 3OS-trained group showed signs that they were relying on cues to presentation ear after training, since intelligibility was affected by dichotic versus diotic presentation mode for the male talker vowel test. Presumably, learning based on frequency rather than ear presentation cues is an indication of perceptual adaptation, so here 6DOS-training seems more beneficial.

Experiment 2 was designed to further explore the questions of type and time course of training in learning to accommodate mismatched frequency-place maps. The training period was doubled and listeners were trained with both the shifted bands alone (3OS) and with these bands in combination with unshifted bands (6DOS) to provide listeners with extensive experience in both conditions.

4.3 Experiment 2[6]

4.3.1 Method

4.3.1.1 Subjects

Six normally hearing speakers of British English took part, and each was paid for his or her participation. All had pure tone audiometric thresholds better than 20 dBHL at octave frequencies from 500 to 4000 Hz, and none had participated in Experiment 1. Ethical approval for this study was granted by the UCL Research Ethics Committee.

[6] Experiment 2 was carried out conjointly with K. Mair. The author of this thesis was responsible for designing the experiment, preparing stimuli and software and analyzing all data. The details are outlined in the Statement of Conjoint Work.

4.3.1.2 Signal processing

Signal processing was the same as for Experiment 1.

4.3.1.3 Training

Training for Experiment 2 was similar to that in Experiment 1, with the exception that subjects were trained in 30 sessions each of 20 minutes. Each training session was divided into 10 minute blocks which alternated between the 6DOS and 3OS processors. K. M., a speaker of Standard Southern British English, was the training talker for Experiment 2.

4.3.1.4 Procedure

Table 4-4 gives a description of the testing and training regime used for Experiment 2, which is similar to that of the previous experiment, with the following modifications. A familiarization test session was included – Session 1 – in which subjects were tested with the IEEE sentences, the BKB sentences and vowels. The purpose of this session was primarily for familiarization, so that any adaptation to the vocoder processing would occur before training commenced (Davis et al., 2005). In the first training session only (Session 3 in Table 4-4) the first five minutes with each processor was auditory-visual. All subsequent training for this experiment was auditory-only. Only IHR sentences were used for easier sentences, and these were tested throughout training. BKB sentences were used for easy material in only the very first pre-training session. Because there is some evidence to indicate that in dichotic listening, listeners give stronger weighting to the right ear (Kimura, 1961), the headphone orientation was counterbalanced across subjects. All other procedural considerations were the same as for Experiment 1.

Table 4-4. Sequence of training and testing conditions for Experiment 2

Session	Training	Processor	Testing
1	5 minutes AV 5 minutes AA	6DU	Familiarization: unprocessed vowels *Pre-test*: IEEE sentences (2 lists x 4 conditions x 2 talkers) BKB sentences (2 lists x 3 conditions x 2 talkers) bVd identification (5 tokens x 4 conditions x 2 talkers)
2	—	—	*Second Pre-test*: IEEE sentences (2 lists x 4 conditions x 2 talkers) IHR sentences (2 lists x 3 conditions x 2 talkers) bVd identification (5 tokens x 4 conditions x 2 talkers)
3	5 minutes AV 5 minutes AA x 2	6DOS + 3OS	none
4 – 17	10 minutes AA x 2	6DOS + 3OS	none
18	—	—	*Mid-test:* IEEE sentences (2 lists x 4 conditions x 2 talkers) IHR sentences (2 lists x 3 conditions x 2 talkers) bVd identification (5 tokens x 5 conditions x 2 talkers)
19 – 33	10 minutes AA x 2	6DOS + 3OS	none
33	—	—	*Post-test:* IEEE sentences (2 lists x 4 conditions x 2 talkers) IHR sentences (2 lists x 3 conditions x 2 talkers) bVd identification (5 tokens x 5 conditions x 2 talkers)

4.3.1.5 Conditions

Four test conditions were used in Experiment 2 – 6DU, 6DOS, 3EU and 3OS. Because we anticipated ceiling effects with the easy sentence material and the dichotic unshifted (6DU) processor, this condition was not tested with the IHR sentences.

4.3.2 Results

4.3.2.1 Test exposure learning

A repeated measures ANOVA was used to analyze the results from the first two test sessions, which factored out any early-stage adaptation to vocoding based on learning of the test material and task alone. No significant differences between these two test sessions were found for either the IEEE sentence test or the vowel test, which implies that any further improvements beyond the second test session were likely the result of training. All subsequent analyses are therefore based on the final three test sessions.

4.3.2.2 IEEE sentences

Scores for the IEEE sentences across training sessions are shown for the two talkers combined in the top left panel of Figure 4-7. Performance tended to improve with training in all conditions. A repeated measures ANOVA on data from the final three test sessions showed significant main effects of talker [$F(1, 5) = 28.6$, $p = 0.003$)], number of training sessions [$F(2, 10) = 29.0$, $p < 0.001$] and processor [$F(1, 5) = 72.2$, $p < 0.001$]. The lack of training session with processor interaction gives an indication that learning through training was similar in all conditions, including 3OS. This is a similar finding to the 3OS-trained group in Experiment 1, who had equivalent amount of training with the 3OS condition as the subjects in this experiment.

Here there was also direct evidence of learning of the shifted bands (3OS), since a planned contrast of training sessions in this condition showed intelligibility after training was significantly better than that before training ($p = 0.038$). Yet despite this learning, and the extended period of training in Experiment 2, there was still no evidence of integration of the mismatched maps. In post hoc tests, performance with

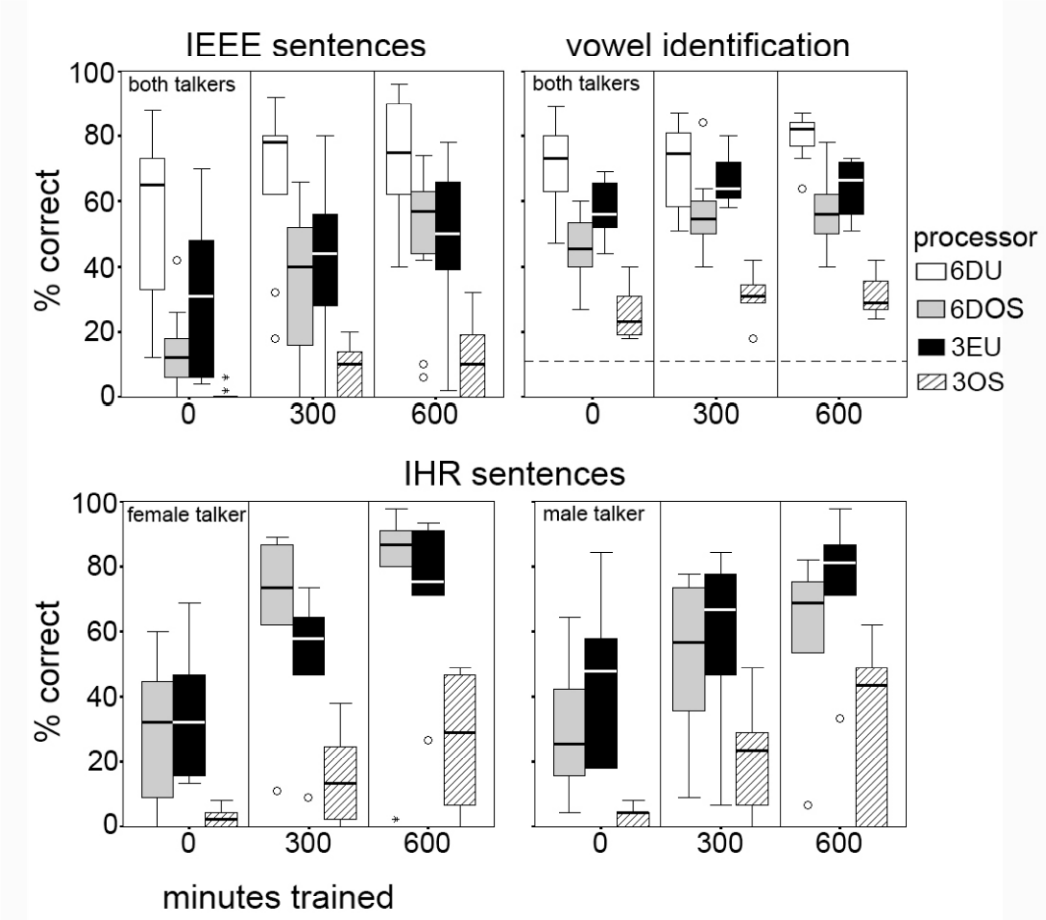

Figure 4-7. Performance as a function of amount of training for the final three test sessions of Experiment 2. The top left panel shows IEEE sentence scores, and the top right panel vowel identification. In the bottom two panels are scores for the IHR sentences, which have been split by talker because of a significant talker effect. Training time periods are separated by vertical lines, and for the vowel identification, the horizontal reference line indicates chance performance (11.1%)

6DOS did not differ from that with 3EU, even though there was room for improvement after 10 hours of training since 3EU was worse than 6DU (p = 0.02).

4.3.2.3 High-context (IHR) sentences

The bottom panel of Figure 4-7 shows scores for the IHR sentences for the female and male talkers individually. A repeated measures ANOVA showed a significant interaction between talker and processor, so scores for each talker were analyzed separately. These showed significant effects of training session and condition. For the female talker, there was also a significant session by condition interaction [$F(4, 20) = 3.62$, $p = 0.02$], which seems likely to be the result of ceiling effects with the 3EU processor after training. Indeed, the 6DU condition was not tested with easier sentences because ceiling effects were anticipated with the 6-channel unshifted processor. Easier sentences were nevertheless useful for examining learning in more difficult conditions of spectral shift.

The bottom panels of Figure 4-7 show that median intelligibility of the 3OS condition improved from close to floor before training to near 40% for both talkers after training, which is a much higher level of intelligibility for 3OS than that seen in Experiment 1. A planned contrast showed that improvement with training in this condition was significant ($p = 0.02$), as did post hoc testing, but only when the data was reanalyzed to exclude one listener who performed consistently worse than the rest on all conditions ($p = 0.04$). The lower level of intelligibility with the 3OS condition meant that this subject did not appear as an outlier.

Surprisingly, *post hoc* tests on the data from all six subjects showed that, for the male talker, post-training intelligibility with 3OS did not differ significantly from that of the 6DOS processor. This may have been the result of increased learning with the 3OS processor. However, performance for the 6DOS condition was significantly worse than the 3EU condition for the male talker, which may also indicate interference or distraction from the addition of the shifted channels in this condition. Despite clear evidence of 3OS learning, however, subjects did not demonstrate integration of the mismatched maps, since performance with 3EU and 6DOS did not differ significantly after training.

4.3.2.4 Vowel identification

Boxplots of vowel identification for the male and female talkers combined are given in the top right panel of Figure 4-7. A repeated measures ANOVA on the final three test sessions showed significant main effects of session [$F(2, 8) = 20.4$, $p = 0.001$], talker [$F(1, 5) = 20.0$, $p = 0.007$], and processor [$F(3, 15) = 77.8$, $p < 0.001$], but no significant interactions. The lower bound of the 95% confidence interval of the estimated marginal mean was always greater than 11%, indicating intelligibility was statistically greater than chance performance for all conditions and at all test points. A planned contrast showed evidence of learning of 3OS with training ($p = 0.046$). However, post hoc testing on the post-training vowel scores showed the same general trends as with the sentences: post-training scores with 3EU were significantly worse than 6DU ($p = 0.014$), yet there was no significant difference in intelligibility between the 6DOS and 3EU processors, so subjects did not learn to integrate the mismatched maps. Moreover, in a planned contrast in reference to 3EU, post-training vowel intelligibility was marginally worse with 6DOS than with 3EU, although this test just missed significance ($p = 0.051$). This is suggestive of distraction with the addition of the shifted channels, rather than integration.

4.3.3 Discussion

Subjects in Experiment 2 showed greater signs of adaptation to the shifted portion of the binaurally mismatched processor (3OS) than those in the previous experiment. First, interactions of session with processor were not found for the IEEE and vowel recognition, which indicates that learning through training was similar in all conditions. There was a significant session with processor interaction for the IHR sentences, but this was the result of ceiling effects with 3EU and not floor effects with 3OS. More significantly, there was direct evidence of learning of 3OS, since planned contrasts of sessions within the 3OS condition showed significantly better intelligibility after training with all test materials. Median intelligibility of easy sentences with just the three shifted bands (3OS) after 10 hours training improved from near floor to around 40%. In Experiment 1, by contrast, post-training intelligibility with 3OS was near floor with easy sentences for the 3OS-trained group.

It is interesting to note that this level of intelligibility with 3OS was only achieved with interspersed and/or concurrent training with unshifted vocoded speech: though the 3OS-trained group in Experiment 1 received the same amount of training with that condition as the subjects in Experiment 2, only the latter group of subjects showed such improvements. The increased improvement may have arisen from either more exposure time to the shifted speech (through exposure to both 6DOS and 3OS), or possibly the concurrent exposure to easier unshifted speech, or a combination of the two. This may also have been the result of training talker intelligibility. The talker for training in Experiment 2 was a native British English speaker and thus more similar in dialect to the talkers in the test material, whereas the talker in Experiment 1 was a native speaker of a northeastern American English dialect.

Yet despite this increased evidence of adaptation to the basally shifted speech, and the extended training period allowed for adaptation, it was again found that subjects did not learn to integrate the binaurally mismatched frequency-place maps. Performance with the 6DOS processor never exceeded that with the 3 unshifted channels alone (3EU), even though there was clear evidence of room for improvement from the addition of the odd channels, since 3EU was typically worse than 6DU. Moreover, there was some evidence of interference from the shifted bands when they were presented together with the unshifted bands, since post-training performance with 6DOS was worse than 3EU for the vowels and the male talker IHR sentences.

4.4 General discussion

4.4.1 Training condition effects

The training condition had a large effect on adaptation in both experiments. In Experiment 1, listeners trained with the binaurally mismatched processor (6DOS) adapted well to the unshifted portion of the signal, but sole exposure to this condition seemed to encourage subjects to ignore the information in the shifted ear, since intelligibility with 6DOS always matched that with 3EU. Training with just the shifted processor (3OS) led to poorer performance overall compared to training with the binaurally mismatched processor, but the group trained with this processor did show signs of adaptation specific to the shifted speech. These findings led to the hypothesis that for mismatched frequency-to-place maps, exposure to both the shifted signal and the binaurally mismatched signal may be important. The training condition was alternated over the extended time course in Experiment 2, and this appeared to have a large, positive effect on adaptation to all conditions, especially the shifted 3OS condition.

The findings on training condition effects imply that mixed and/or concurrent exposure to shifted and unshifted speech facilitates adaptation to both the unshifted and shifted speech. Recent work on simulations of monaural spectral shifts using similar sine-vocoded speech processing to that considered here is consistent with the present results. Li et al. (Li, Galvin III, and Fu, 2009) showed that passive vowel adaptation to a very large upward spectral shift – 8 mm – could be facilitated by interspersed, passive training with a smaller upward shift of 4 mm. Listeners exposed to both shift conditions showed improvements in the 8 mm condition with training, while listeners solely exposed to the 8 mm shift showed no signs of adaptation. While the complicating issue of binaural integration was not a factor in their study, the results are consistent with the present findings, and suggest that for more difficult spectral shift conditions, mixed exposure training may be important.

As more people are implanted bilaterally, the issue of training paradigm is likely to become even more important. The present findings lend support to Tyler et al.'s (2007) anecdotal evidence that training with both implants as well as each implant individually may be important in achieving optimal benefit from bilateral implantation.

4.4.2 Resistance to integration

Listeners in both experiments demonstrated evidence of adaptation following a period of training, yet intelligibility with the 6DOS processor never exceeded that with just the 3 unshifted channels (3EU). Moreover, in all tests of the 6DOS processor, post-training intelligibility closely matched that of the 3EU processor, indicating information derived from the two was equivalent, and listeners were relying on the "better ear". These findings thus indicate that at the very least, listeners are resistant to integrating cochlear frequency-to-place maps that differ greatly between the ears. The mechanism involved in the integration of binaurally mismatched frequency-place maps is likely to be different to that for adaptation to monaural basalward spectral shifts. Nevertheless, these results appear to contrast with earlier findings, which demonstrated adaptation to a monaural spectral shift in as little as 3 hours' training (Rosen et al., 1999).

4.4.2.1 Difficulty learning 3OS condition

In particular, learning of the 3OS condition proved more difficult than that found in previous studies. In Experiment 1, improvements in this condition were small and generally limited to improvements in vowel recognition. Experiment 2 showed significant improvements for sentence material as well as vowels, but these only appeared over the longer time course of training and with concurrent and interspersed training with unshifted channels (6DOS). This was somewhat surprising. The smaller number and non-contiguous output of the shifted channels, or differences in the parameters of the vocoding, such as carrier type and smoothing filter, may have influenced the outcome of training. Most simulation studies showing adaptation to upward spectral shifts have employed noise-vocoding (Faulkner, Rosen, and Wilkinson, 2001; Faulkner et al., 2006; Fu and Galvin III, 2003; Rosen et al., 1999; Smith and Faulkner, 2006), although there is also evidence of adaptation with sine-vocoding (Li and Fu, 2007; Li et al., 2009). While an earlier study indicated that these should be equivalent in intelligibility (Dorman et al., 1997), emerging evidence suggests that for small numbers of channels, there are significant differences in the intelligibility of unshifted noise and sine vocoders (Souza and Rosen, 2009), with noise vocoders being more intelligible when envelope smoothing frequency is low (30 Hz, similar to that used here). What may be a comparable resistance to learning of shifted spectral cues has been observed for frequency-lowered amplitude envelope signals when these were, as here, limited to fairly low rate (50 Hz) modulations and imposed on sinusoidal carriers (Grant, Braida, and Renn, 1994). However, that was a study of auditory-visual speech perception so listeners may have had less to gain from learning to reinterpret the frequency-shifted spectral information than would be the case here.

Even if this resistance to learning were in part due to the use of a low smoothing cut-off frequency with sine vocoding, it can be argued that low-frequency envelope smoothing should be preferred when simulating CI processors, because current cochlear implant systems do not deliver fine structure detail. Moreover, the use of noise-vocoders here would not have been suitable, since overlap of noise carriers between the ears may have led to confounding effects. Notwithstanding these considerations, there was clear evidence of adaptation to the 3OS processor, especially in Experiment 2, yet still no evidence of integration, so it seems unlikely that the more difficult 3OS condition is the sole explanation for this resistance to integration.

Another aspect of the processing used here that differed from previous studies was the sparsely sampled input in the monaural conditions. While the frequency range of analysis filters for binaural conditions was continuous, the monaural conditions comprised subsets of either even or odd channels only. When these were presented unshifted, as in the 3EU and 3OU conditions, subjects appeared to adapt readily. This adaptation to ‘spectral holes’ is consistent with previous studies of simulations of tonotopically matched, shallowly inserted electrodes, where subjects show some ability to adapt to missing spectral content, so long as the signal maintains other complementary cues useful for speech perception (Baskent and Shannon, 2005; Faulkner, Rosen, and Stanton, 2003). However, in the 3OS condition, the sparsely sampled channels were also shifted basally an equivalent of 6 mm. The large spectral shift in combination with the sparsely sampled input as in the 3OS condition proved difficult for subjects to learn, especially in Experiment 1. It appears that the spectral shift was the source of this difficulty, and not the frequency content in the odd channels *per se*, since the odd channels were more intelligible than the even channels after training. Applying a spectral shift to the sparsely sampled and frequency-shifted channels may have made the adaptation process more difficult, since the sparse sampling would have yielded a reduced access to the already limited redundant across-frequency speech cues that may facilitate the remapping between shifted and unshifted speech.

However, in Experiment 2, subjects did show adaptation to the 3OS condition to a level that is comparable with previous studies of monaural basally shifted vocoded speech, but there was still no sign of integration with the unshifted channels. This finding implies that it is the integration of shifted and unshifted channels, rather than adaptation to the sparse sampling and frequency shifting, that was the source of difficulty. Subjects were not able to integrate the redundant speech cues in the unshifted and shifted channels when presented together, even though they had presumably learned to make use of these across-frequency cues when listening to either of these subsets of channels (shifted or unshifted) on their own.

4.4.2.2 Altered frequency order

A novel aspect of the frequency-place mapping examined here, and a possible source of this resistance to integration, is the altered relative frequency order of the output bands in the binaurally mismatched processor. When combinations of shifted and unshifted bands were presented together, as in the 6DOS and 6OS processors, the output consisted of frequency information that had been shifted in both frequency and relative order. Ranking the output with reference to the relative frequency order of the input analysis bands yields, from low to high frequency, (2 – 1 – 4 – 3 – 6 – 5). Useful cues for speech perception, such as the first and second formants of vowels, may sound as if inverted. Evidence from the experiments here suggests that, despite the enriched frequency content, listeners do not integrate, or *spectrally fuse*, this altered output. In all conditions where the tonotopically matched three unshifted bands were present, listeners appeared to rely on these frequencies, perhaps treating the shifted channels as noise in the signal.

Further evidence for this comes from Experiment 1, where listeners trained with only the shifted components showed reduced intelligibility with the diotic 6OS processor than with the dichotic 6DOS for the female talker vowel test. The 3OS-trained group had no training with the 6DOS processor, and thus presumably relied somewhat on cues to ear presentation to attend to the three unshifted components of the 6DOS processor. When no such cues were available, intelligibility decreased. Listeners trained with the 6DOS processor showed equivalent intelligibility after training with the 6DOS and 6OS processors, indicating that learning was not based on cues to presentation ear.

The contrasting results shown here may be indicative of a constraint on plasticity for speech perception to cases where relative frequency order has been preserved. Or rather, adaptation to spectrally altered speech may occur more readily when relative frequency order is preserved. There is some evidence to the contrary: In his seminal study on adaptation to spectrally altered speech, Blesser (1972) trained listeners with speech that had been spectrally rotated around 1600 Hz, so that low frequencies became high and vice versa. Even under such drastic conditions, some listeners did eventually learn to converse, although this did not appear until much later in the study, which consisted of 20 45-minute sessions. Azadpour (Azadpour, 2008) also found that listeners could adapt to aspects of spectrally rotated speech over a similar time course to that in Experiment 1.

Even with spectrally rotated speech, however, some notion of relative ordering of frequency information is intact, even if it has been inverted. By contrast, the processing used here may selectively invert speech cues such as vowel formants which may result in a novel percept, rather than cues being presented to the wrong frequency region. In the present investigations, this appeared to hinder adaptation. Specifically, in Experiment 1, there was evidence of shift-specific adaptation to the 3OS condition for both groups for the vowel test, yet listeners still did not spectrally fuse this information with the unshifted channels as in the 6DOS condition. This was despite the fact that the frequency information contained in the unshifted even and odd channels was significantly different in terms of the resulting vowel confusions, which would imply that there was room for improvement by spectrally fusing these signals.

Since performance did not reach asymptote in either of the present experiments, it remains to be determined whether continued learning beyond that demonstrated here would be shown, or whether this is too difficult a mapping to learn fully. In this respect, it is difficult to interpret the implications of this resistance for cochlear implant patients, as with any simulation study, because the extent of their experience with any clinically fitted processor will be far greater than the 5-10 hours examined in the laboratory here. The group of patients investigated by Tyler and Summerfield (1996) reached asymptotic performance an average of 30-40 months post implantation. Nevertheless, an alternative interpretation of Blesser and Azadpour's findings that would be consistent with the present investigations is that altered frequency order hinders adaptation.[7]

4.4.3 Implications for bilateral cochlear implants

It is possible that listeners are able to adapt to frequency mismatches between ears, but that the mismatch examined here was too great. Positive results from speech perception with bilateral cochlear implants provide some evidence that this may be true. Several studies have shown that these patients experience a synergistic improvement when using both implants over either implant on its own, even for speech in a quiet laboratory (Dorman and Dahlstrom, 2004; Litovsky et al., 2006). This is the most similar clinical situation to the simulations studied here. If it is true that the Greenwood map exaggerates upward spectral shifts seen in modern CI users, as suggested by Stakhovskaya et al. (2007), then this may explain the discrepancy between the resistance to integrating mismatched maps shown here and positive reports with bilateral implants. The drastic spectral mismatch examined here may be uncommon in a clinical situation, and smaller mismatches may be tolerated. Nevertheless, the findings indicate that drastic mismatches between the

[7] It should be noted that this interpretation may be limited to cases of smaller numbers of channels, where formant cues are likely to fall within a single analysis/output filter. A more detailed discussion is given in Chapter 5.

ears should be avoided in cochlear implant fittings. Evidence from studies with a self-selecting tool for frequency-place mappings supports this conclusion. Fitzgerald et al. (2007) found that normal hearing subjects, when listening to simulations of CI processing, preferred frequency-place maps that minimized the mismatch between the ears, regardless of spectral shift.

4.5 Conclusion

Despite undergoing significantly longer training periods than in previous studies with monaural spectral shifts, subjects in the present investigations never learned to integrate the binaurally mismatched frequency-place map, and always relied on information from the ‘better ear’, if present. Post-training performance with the binaurally mismatched processor never exceeded that with the three unshifted channels, even after ten hours of training. These findings support the notion that, when fitting cochlear implants for bilateral devices, consideration should be given to keeping frequency-place maps similar in the two ears in order to optimize speech perception. This seems especially true for cases of drastic mismatch between the ears, such as might arise when one of two ears has a shallowly inserted electrode. The results are consistent with psychophysical studies in bilateral cochlear implant patients, which have also called for matching interaural electrode pairs in order to restore ITD (Blanks, Buss, Grose, Fitzpatrick, and Hall, 2008; Grose and Buss, 2007; Long, Eddington, Colburn, and Rabinowitz, 2003) and ILD sensitivity (Long et al., 2003).

Not all patients who receive bilateral cochlear implants achieve a binaural benefit, and the findings here indicate that a resistance to adapting to binaurally-mismatched frequency-to-place maps may underpin this lack of benefit. The evidence here also calls for careful attention to speech training procedures in bilateral cochlear implantation: training binaurally when there is a large difference in intelligibility between the signals at each ear may lead to adaptation by focusing attention on the "better ear". In instances where a single ear is providing a much less intelligible signal, training binaurally *and* monaurally may be crucial.

The results from the present investigation leave some questions unanswered, however. Many of the benefits of binaural hearing are realized for speech in noise, where differences in the sound signal at each ear can be used to obtain a better representation of what has been said. Indeed, of the two subjects tested by Dorman and Dahlstrom (2004), one showed binaural benefit over the "better ear" only in the noise condition. Better-ear effects may dominate in the absence of noise and thus prevent integration of the mismatched signals at each ear. In the present experiments, listeners adapted to the unshifted signal sufficiently well that the motivation for integration with channels with a large spectral shift may have been minimal, even if they showed some adaptation to the shifted channels when presented alone. If the shifted and unshifted signals here were presented in noise, on the other hand, listeners may have stronger motivation to rely on the shifted channels. When testing in noise, it may also be possible to use more spectral bands without achieving ceiling effects with the unshifted speech, and this increased spectral resolution may, in turn, facilitate learning.

In a related study, Long and colleagues (2004) used 12 channel binaurally interleaved vocoders similar to those described here but without a spectral shift. Either signal and noise were both presented at 0^{o} azimuth, or noise was presented a 0^{o} azimuth and the signal presented at 90^{o} with a corresponding ITD of 600 microseconds. Listeners in their experiment demonstrated a binaural advantage for consonant perception over either ear on its own, which, significantly, could not be attributed to binaural redundancy, since the interleaved channels, as here, ensured each ear was getting different information. However, the processor in their study was tonotopically matched in both ears. It merits exploration whether this advantage remains if a spectral shift is introduced to a single ear. The next chapter will re-examine the issue of adaptation to binaurally mismatched frequency-to-place maps for speech in noise.

Chapter 5. Perceptual Adaptation to Binaurally Mismatched Frequency-to-Place Maps II: Speech in Noise

The previous chapter presented evidence from simulations of binaurally-mismatched processors that showed that normal hearing listeners are resistant to integrating frequency-place map mismatches between the ears. However, there were aspects of the studies that made it difficult to draw clear conclusions. First, learning of the three odd shifted (3OS) condition was slower to appear than in previous studies with vocoder simulations of basally shifted speech. This raised the possibility that the more difficult 3OS condition was the bottleneck to integration in those experiments, rather than the mismatched frequency-place maps *per se*. However, over the extended time course of training in Experiment 2, subjects did adapt to the 3OS condition, but still showed no signs of integration. A more striking consideration in both experiments was the dominance of “better ear” effects. It appeared that providing unshifted speech in combination with speech that had been basally shifted quite drastically by 6 mm may have impeded integration, especially for speech in quiet.

In clinical bilateral cochlear implant situations, it is more likely that both ears will be subject to some degree of spectral shift, and the dominance of “better ear” effects may not be so prominent in the absence of an unshifted ear. For the purposes of maintaining continuity across experiments, and in the interests of avoiding the confounding effects outlined in 4.1, introducing asymmetrical spectral shifts was avoided in the present investigations. Another way to achieve a mitigation of “better ear” dominance, which also has strong grounding in the clinical literature on bilateral cochlear implantation, is to process the speech in noise. This chapter presents evidence from two experiments on simulations of binaurally mismatched frequency-to-place maps in noise. The aim was to more thoroughly test the hypothesis that listeners are resistant to integrating binaurally mismatched frequency-to-place maps.

The advantages for processing speech in noise were two-fold. First, for users of bilateral cochlear implants, it is not uncommon to see a binaural advantage only for speech in noise. For speech in quiet, binaural intelligibility is often similar to that with the “better ear” alone (Dorman and Dahlstrom, 2004; Tyler et al., 2007; Wackym et al., 2007). Similarly, the subjects in the experiments reported on in the previous chapter also demonstrated a “better ear” effect for the speech processed in quiet. As in the clinical literature, then, a binaural speech advantage may be most apparent in noise, even for binaurally mismatched frequency-to-place maps.

Also in contrast to speech in quiet, it is more common for users of bilateral CIs to show a binaural ‘superadditive’ advantage for speech perception in noise than in quiet (Dorman and Dahlstrom, 2004; Wackym et al., 2007). That is to say, intelligibility in the binaural condition is greater than the sum of the intelligibility at each ear. This advantage is even more apparent with electro-acoustic stimulation. Even though acoustic information in the low-frequency regions is insufficient to support robust speech recognition on its own, combining this information with electrical stimulation in the higher frequencies leads to a superadditive advantage for speech perception in noise. Li and Loizou (2008) posited that the low-frequency information allows the listener to “glimpse” redundant speech cues in the high, electrically-encoded frequencies, which in turn allows for stronger integration of both sets of cues compared to either electric or acoustic cues on their own. This advantage arises from the spectral properties of speech and noise which allow for a better SNR in the low-frequencies.

While CI speech processing may limit access to some of the low-frequency cues available through acoustic hearing, a sensitivity to the local SNR at each ear may also account for the binaural benefit for speech in noise shown in many users of bilateral CIs. As the signal-to-noise ratio (SNR) varies across the speech spectrum, users of bilateral CIs may be able to make use of momentary advantages in the SNR at each ear to achieve a binaural advantage for speech overall. This advantage would not be apparent for speech in quiet, since there would be no masking from noise, and the information at the “better ear” would always be maximal. Consider the speech (blue) and noise (pink) signals in the speech in speech-shaped noise utterance depicted in Figure 5-1. While the global SNR of the utterance is set at 10 dB, it is obvious from the figure that the speech and noise levels in relation to each other vary across the spectrum, and these levels will change with time. Considering the interleaved and monaurally shifted processing used in the previous experiments, it is feasible that important speech cues to be coded as unshifted sine carriers will be masked by the noise, but at the same time other speech cues to be coded as shifted sine carriers will not. It seems plausible that under these circumstances, a binaural advantage would arise from the

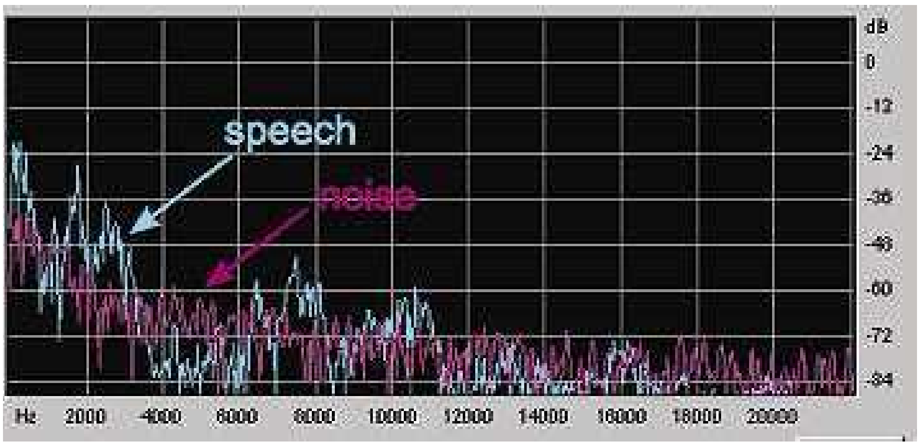

Figure 5-1. Spectra of speech combined with speech-shaped noise at an SNR of 10 dB. Note that while the *average* SNR is 10 dB, the *local* SNR across the frequencies varies. Here, the level of speech is greater than the level of noise up to about 1 kHz, but between 1 and 1.5 kHz, the levels of both signals are approximately equal.

additional information provided by the shifted channels at these instances that would not be apparent for speech in quiet.

A second reason for testing in noise was that it allowed for the use of more frequency channels. While the 3OS condition in the previous experiment proved difficult to learn, using more than three channels in each ear for speech in quiet would have yielded ceiling effects with the unshifted speech, even though this may have facilitated learning of the shifted speech. By processing the speech in noise, on the other hand, more channels can be used without achieving ceiling effects in the unshifted condition. Using more channels may also mitigate the adverse effects of sine vocoding with small numbers of channels and low-frequency smoothing filter (Souza and Rosen, 2009), sparse sampling of vocoders for monaural conditions, and the difficulty in adaptation to the shifted channels. If learning in the shifted condition is facilitated, then, this may also facilitate binaural integration.

Using more frequency channels has an additional advantage. As in the previous experiments, the introduction of the shift monaurally allowed for the odd channels to be upwardly shifted to the CF locations of the next odd-channel filter. When using more frequency channels over the same frequency bandwidth, then, the analysis channels are more densely spaced, so that an upward shift to the next odd CF would allow for a smaller spectral shift to be introduced overall. One limit on integration in the previous experiments may have been the large mismatch in frequency-place maps in the two ears. Testing in noise allowed for the examination of binaurally mismatched frequency-place maps with a more moderate spectral mismatch between the ears.

Speech processing in these experiments was similar to that presented in Chapter 4. However, the total number of channels was increased to 10, the upward spectral shift applied to the odd channels was reduced to 3.8 mm, and the speech was processed in speech-shaped noise with no directional cues. The experiments examined whether listeners could adapt to a more moderate spectral mismatch between the ears in the presence of background noise.

In Experiment 3, sixteen normally-hearing listeners were trained with the 10-channel binaurally mismatched processor, similarly to Experiments 1 and 2. It was presumed that the larger number of channels and the smaller degree of spectral shift would make the shifted condition easier to learn, even in the presence of background noise. Based on this assumption, there were three experimental hypotheses:

1) best local SNR hypothesis
2) adaptation hypothesis
3) better ear dominance hypothesis

The *best local SNR* hypothesis posits that listeners should adapt well to both the shifted and unshifted speech, and that listeners should show a binaural advantage for speech after training. This binaural advantage would arise from the ability to take momentary advantage of the better SNR at each ear in either the shifted or the unshifted channels, and the summation of this information over time would allow for a binaural advantage over just the unshifted channels alone. Note that this differs from the binaural squelch effect since the frequency information is different in each ear. This hypothesis differed from the *adaptation* hypothesis, which is what would be expected if listeners demonstrated complete or substantial adaptation to the shifted speech, so that shifted and unshifted channels could be continuously integrated to achieve binaural integration. The adaptation hypothesis would be consistent with findings of electrical pitch changes over time (Reiss et al., 2007). Finally, the "better ear" dominance hypothesis posits that listeners would adapt by attending to the "better ear" (in this case, unshifted channels) when presented with binaurally mismatched frequency-place maps in this way, so listeners would not integrate the shifted and unshifted channels. This hypothesis would be consistent with the results from Experiments 1 and 2.

While a failure to binaurally integrate might also lend support to the frequency order hypothesis proposed in those experiments, the present experiment does not explicitly test this hypothesis, since “better ear” dominance may or may not reflect relative frequency order effects. Furthermore, with the increased number of frequency channels used here, spectral information was more specifically allocated to sine carriers, so it was less likely that individual formants would be displaced as was possible in the previous experiments. Recall that in the six channel case, the range of the analysis filters spanned likely frequencies for the first and/or second formant frequencies of vowels. For example, an /ʌ/ vowel spoken by a female talker may have an F1 of 760 and an F2 of 1400. In the 6DOS binaurally mismatched processor, F1 falls within the third analysis filter (718 – 1208 Hz), and F2 the fourth (1208 – 1971 Hz). The unilateral spectral shift of the odd channels would have moved the F1 above F2, thus reversing relative frequency order. Here, however, ten analysis channels spanned the same frequency range (200 – 5000 Hz), so it was unlikely that individual formant frequencies would be displaced in this way: the proximity of F1 and F2 in relation to each other may be altered, but it was not likely that their relative order would change. Thus, a failure to integrate here may more strongly indicate “better ear” dominance than sensitivity to relative frequency order *per se*.

If a small binaural advantage was shown, then presumably this would result from a best local SNR account, since the additional information from the worse ear would contribute only occasionally. On the other hand, if a relatively large binaural advantage for the binaurally mismatched speech was found that approached that for the binaurally matched speech, then it would be hypothesized that integration was occurring by way of a more complete adaptation to each signal. Finally, if no binaural advantage was shown, then this would be consistent with the “better ear” dominance hypothesis, and the results from Experiments 1 and 2 in the previous chapter.

A secondary aim of Experiment 3 was to further explore the role of training condition in facilitating adaptation to the binaurally-mismatched processor. If adaptation to the binaural condition can occur only when adaptation to the shifted channels is sufficient, then it may be that longer training times with the monaural shifted speech are needed. This is what was anecdotally reported by the sequential bilateral implant subject in Tyler et al.’s (2007) study, who showed binaural advantage only after extensive training with the second implant alone. Furthermore, evidence from Experiment 2 suggested that training with both binaural and monaural conditions may be the best way to facilitate adaptation to the shifted condition, but because these conditions were alternated repeatedly in that study, it was unclear whether there were any order effects in terms of adaptation. Another consideration is that there was still no evidence of binaural integration after 10 hours of training. Too much exposure to the unshifted speech relative to the shifted speech may have prompted listeners to focus on the unshifted speech, rather than binaurally integrate. Experiment 3 used a crossover experimental design for the training procedure to further explore the effect of training condition and training condition order on adaptation and binaural integration.

Experiment 4 was designed to examine the issue of “better ear” effects and laterality in adaptation to binaurally mismatched frequency-to-place maps. In Experiment 1 of Chapter 4, there was evidence that subjects adapted to the binaurally mismatched processor by learning to attend to the information in the "better ear", but there was also mixed evidence about whether this learning was based on the frequency content of this information or rather on cues to ear presentation. Intelligibility was sometimes adversely affected when shifted and unshifted channels were not separated by ear. If listeners adapt to binaurally mismatched speech by learning to ignore the worse ear, then this has clear, adverse implications for fittings with bilateral cochlear implants. In this experiment, eight participants from Experiment 3 were invited to return for further testing and retraining. At the final test session, the headphone orientation was reversed and speech perception was re-tested, to examine whether speech perception was adversely affected if the learned cue to ear presentation was obstructed. The experiment also allowed for the examination of retention of any learning of the shifted and unshifted speech. If listeners demonstrated decreased intelligibility when the headphones were reversed, then this would point to the use of cues to ear presentation in adaptation to the binaurally mismatched processor. If this could also be linked to a decreased intelligibility with the diotic mismatched processor compared to the dichotic one, then there would be strong evidence that subjects had learned to ignore an ear, which would be consistent with the frequency order hypothesis.

5.1 Speech processor design

Speech processing, outlined in Figure 5-2, largely followed that presented in the previous chapter, with the exception that the number of analysis bands and output carriers was increased to ten and the unilateral spectral shift decreased to 3.8 mm, using a Greenwood map. The motivation for using interleaved processors was the same as for Experiments 1 and 2. To reiterate, the interleaved processing avoided providing the same frequency information to different cochlear locations, and different frequency information to the same cochlear location. In the case of speech in noise, the issue of delivering different frequency information to the same tonotopic

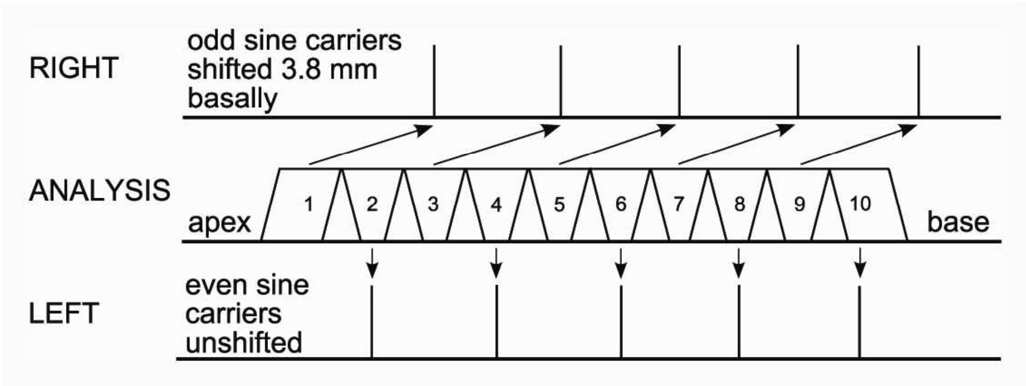

Figure 5-2. Speech processing showing a 10 channel binaurally mismatched frequency-to-place map.

location becomes even more important, since this may introduce a form of energetic masking in and of itself. Coupled with the additional energetic masking from the speech-shaped noise, this type of processing may have further distorted the signal and reduced intelligibility.

5.2 Selection of signal-to-noise ratio

As argued in Chapter 4, increasing the number of total channels beyond 6 may allow for ceiling effects with the unshifted speech. Therefore it was important to select a SNR that would be low enough to limit perception with the unshifted speech, but also high enough that learning of the shifted channels would not be obstructed. The most appropriate SNR is likely to vary depending on the difficulty of the speech task. The vowel recognition task involved closed-set identification of a single stimulus, which is likely to be a much easier task than open-set sentence recognition. It therefore seemed likely that the SNR for the vowel stimuli would need to be set lower than that for sentence stimuli in order to avoid ceiling effects.

A pilot study was conducted to determine the most appropriate SNR for the unshifted monaural condition for the vowel stimuli. Five normally hearing native British English speakers took part in the experiment. The pilot used an adaptive tracking procedure (Levitt, 1971) to determine the SNR that would allow for 50% vowel recognition accuracy of the vowel stimuli used in the previous experiments. Subjects listened to three repetitions of the entire set of vowel stimuli (135 instances of 9 bVd words, 5 tokens each, single female talker). The starting point and maximum SNR was 20 dB and the minimum allowable SNR was -6 dB. After two initial larger steps of 6 and 4 dB, the SNR was adapted in a one up, one down procedure in 2 dB steps until the entire set was played, and this lasted approximately 10 minutes. On completion, subjects' responses had converged on a mean SNR for the group of 0 dB. This was indicated by the 2 dB up-down steps of the final 25 responses, which typically ranged from either -2 – 0 dB or 0 – 2 dB.

Subjects also took part in an adaptive recognition procedure with the IEEE sentences. The aim was to find the noise level that would allow for 20% recognition of key words (1 keyword per sentence), which was the same level of accuracy achieved before training for the unshifted 3EU condition in Experiment 1. This level of accuracy seemed appropriate, since it would allow for room for improvement with training as in Experiment 1. However, the subjects failed to show convergence after twenty minutes of listening to sentences. Because the SNR gives only an indication of the average speech and noise spectra over a given sentence, the varying SNR over time may have resulted in widely varying recognition accuracy for speech. Hence, the SNR for the sentences was instead somewhat arbitrarily selected to be 10 dB higher than for vowel recognition, i.e. 10 dB. Friesen et al. (2001) showed that for 5 spectral channels (i.e. the number in the monaural unshifted condition), subjects were not at ceiling with an SNR of 10 dB, so this SNR had some grounding in the empirical literature. The 10 dB SNR was used for sentence tests and during speech training.

5.3 Experiment 3

5.3.1 Method

5.3.1.1 Subjects

Sixteen normally hearing speakers of British English took part, and each was paid for his or her participation. All had pure tone audiometric thresholds better than 20 dBHL at octave frequencies from 250 to 8000 Hz. Ethical approval for this study was granted by the UCL Research Ethics Committee.

5.3.1.2 Test conditions

Table 5-1 summarizes the five conditions tested in Experiment 3. These conditions were the monaural and binaural, 5- and 10-channel counterparts to those tested in Experiments 1 and 2. They were designed to assess what aspects of the signal were involved in any learning that took place. Test conditions were composed of monotic, dichotic and diotic combinations of the spectral components outlined in Figure 5-2 for the binaurally mismatched processor. Condition names were coded by number of channels, dichotic presentation, even- or odd-numbered channels, and shifted or unshifted presentation. When basalward shift was applied, it was always to the odd-numbered channels. Even-numbered channels were always unshifted.

Table 5-1. Conditions for Experiment 3

condition	***abbreviation***	***component bands and shift***	
		right	***left***
ten dichotic unshifted	10DU	odd	even
ten dichotic odd shifted	10DOS	odd → 3.8 mm	even
ten odd shifted	10OS	odd → 3.8 mm even	odd →3.8 mm even
five even unshifted	5EU		even
five odd shifted	5OS	odd → 3.8 mm	

The **ten dichotic unshifted (10DU)** condition served as a control to assess maximal intelligibility. Here, unshifted odd- and even-numbered channels were interleaved. The main experimental condition (Figure 5-2) was the **ten dichotic odd shifted (10DOS)** condition. Here, odd-numbered bands were shifted an equivalent of 3.8 mm basally and presented to the right ear, while even-numbered bands were presented to the left ear without a shift. This binaurally mismatched processor was one of the processors used in training. The **ten odd shifted (10OS)** condition contained the same information as the 10DOS condition, but was instead diotic: all bands were summed together and presented to both ears. In contrast to 10DOS, 10OS lacked any cue from ear of presentation to the carrier bands that were shifted. If performance here differed markedly from the 10DOS condition, this would suggest that listeners were learning to attune to information in an ear-specific manner. Conversely, if there was no difference, then we would infer that learning was occurring on the basis of the carrier frequencies. In the **five even unshifted (5EU)** condition, even-numbered channels alone were presented to the left ear unshifted. This was equivalent to the unshifted components of the 10DOS processor. In the **five odd shifted (5OS)** condition, odd-numbered channels were presented alone to the right ear with a 3.8 mm basalward shift. This comprised the shifted components of the 10DOS processor. This processor was also used in training.

5.3.1.3 Signal processing

Centre and crossover frequencies for the analysis and output filters were calculated using Greenwood’s equation and its inverse, relating position along the basilar membrane to characteristic frequency with an assumed cochlear length of 35 mm (Greenwood 1990). The equation describing the map is detailed in 4.2.1.3. The Greenwood map is the most widely used mapping for simulations of basalward spectral shift, so it was chosen so that results could be compared to earlier studies.

The amplitude envelope of each band was extracted with an analysis filter, full-wave rectification, and a smoothing filter. The envelope was then multiplied by a sinusoid with frequency matching the centre frequency (CF) of the band (or shifted equivalent). Finally, the modulated carriers were summed and presented to the left and/or right ears as determined by processor condition. Table 5-2 shows input and output CFs as well as filter cutoffs. All processor conditions used the same ten analysis filters and sine carriers at either the shifted or unshifted CFs of these analysis filters.

All vocoder processing was implemented offline in MATLAB. This ensured identical repetition of the test and training materials for each subject. Processing was executed at a 44.1 kHz sampling rate. Analysis bands were determined by a serial implementation of high-pass and low-pass third-order Butterworth IIR filters. Adjacent filter responses crossed at 3 dB down from the peak of the pass-band. Envelope smoothing used second-order low-pass Butterworth filters with a 32 Hz

Table 5-2. Analysis band cutoff and centre frequencies, and carrier frequencies for each band, in the unshifted (10DU, 5EU) and odd-band shifted (10DOS, 10OS, 5OS) conditions

Band	Analysis band cutoff (Hz)		Analysis band CF (Hz)	Carrier frequency (Hz)	
	lower	Upper		unshifted	odd-shifted
1	200	311	252	252	540
2	311	455	378	378	378
3	455	644	543	543	1032
4	644	889	758	758	758
5	889	1208	1038	1038	1869
6	1208	1625	1403	1403	1403
7	1625	2168	1879	1879	3290
8	2168	2876	2498	2498	2498
9	2876	3798	3306	3306	5703
10	3798	5000	4359	4359	4359

cutoff. Note that the use of a low-frequency smoothing filter with sine-vocoding meant that any cues to temporal fine structure were removed. This was essential for avoiding spectral overlap of the carriers, but it also better replicates the situation seen in real CIs.

An equal loudness correction was applied to each of the shifted bands to avoid substantial differences of relative loudness across the spectra of unshifted and shifted speech. The correction was set to half the difference (in dB) between the minimal audible field threshold of the analysis filter and that at the centre frequency of the shifted output filter. Minimal audible field values were taken from Robinson and Dadson (1956) and interpolated using a cubic spline fit to log frequency.

5.3.1.4 Training procedure (automated CDT)

A novel, computer-based administration of Connected Discourse Tracking (CDT) (De Filippo and Scott, 1978) was used in this experiment. The procedure, developed at UCL by Faulkner and Rosen (2008), was based on Stacey and Summerfield's (2007, 2008) automated sentence training procedure. The method aims to replicate live CDT as much as possible, and has been shown to be as effective as live CDT in eliciting adaptation to upwardly shifted vocoded speech. The computer plays out processed phrases from a story, and shows clickable buttons on the screen for the two to four correct keywords, along with buttons for corresponding foil words. Foil words were selected that differed from the keyword in their first or last phoneme or medial vowel. The subject had three attempts to click the correct keywords, which appeared highlighted with a green check when clicked correctly. With each incorrect word click, the word became highlighted with a red cross, the processed phrase was re-played, and the subject had another turn. After either all correct keywords were selected, or after three failed attempts, the phrase was repeated back a final time processed, and the correct phrase appeared as text on the screen. The computer then advanced to the next turn.

A key advantage of using this procedure was it ensured identical speech materials for each listener and training condition, which may not have been the case with the live procedure. Furthermore, the automated procedure saves much experimenter effort, which allowed for more subjects to be tested. Finally, this procedure allowed for CDT training with a talker of Standard Southern British English, of which the author of this thesis is not. The talker for this experiment was the same talker from Experiment 2, K. Mair. Recordings of four stories were made in an anechoic chamber, which were then processed offline to create the training materials. K. Mair was paid for her time for recording the stimuli. Texts for CDT were the same as those used in Experiments 1 and 2. They were chosen from the Heinemann Guided Readers series (Elementary Level), which are designed for learners of English as a second language and make use of controlled vocabulary and syntactic complexity.

5.3.1.5 Test materials

a. Sentence perception

The same recordings of the IEEE/Harvard sentence lists (Rothauser et al., 1969) used in Experiments 1 and 2 were used here. These have very little contextual information, but allow for better distinction between conditions because of their lower intelligibility in processed conditions compared to easier sentences such as the BKB or IHR lists. The 72 lists in the set each contained ten sentences, with five keywords in each sentence. The first 36 lists were designated for the female talker, and the remaining 36 lists were designated for the male. For each test session, two lists per condition were chosen from each talker set in a pseudo-random manner. No list appeared more than twice in the same condition across all of the subjects, and subjects never heard a list more than once. No feedback was given for the sentence material.

In order to better assess adaptation to the shifted 5OS condition, easier IHR (MacLeod and Summerfield, 1990) sentences were also tested at the final test session. The female and male talkers of these sentences were the same as for the BKB and IHR sentences of Experiment 2, though all sentences were IHR in this experiment. The female talker for the IHR lists was also the same as that for the IEEE sentences, but the male talker was different.

For sentence tests, the subject was asked to repeat back to the experimenter as many words as he or she could. Words were counted correct when the word root was repeated correctly.

b. Vowel identification

The vowel test was used to measure adaptation to the upward spectral shift, since this has a strong effect on the spectral content of the vowels. If learning takes place by remapping the novel shifted stimuli to existing vowel categories, then this should be reflected in the accuracy of vowel recognition. The same nine b-vowel-d words in the carrier sentence ‘Say bVd again’ used in Experiments 1 and 2 were used here. These consisted of five tokens of each bVd sentence, so that in an individual test of either a male or female talker in a given condition, there were 45 items. Vowels were restricted to monophthongs of similar duration, so that listeners would need to rely on primary, spectral cues for identification: /æ/ (bad); /ɑː/ (bard); /iː/ (bead); /e/ (bed); /ɪ/ (bid); /ɜː/ (bird); /ɔː/ (board); /ɑ/ (bod); /uː/ (booed); /ʌ/ (bud). A grid with all nine words appeared, and the subject clicked with the computer mouse on the button displaying the word they perceived. The vowels were represented on the buttons in the orthographic form given above. Before testing, the subject was given a practice session in which the vowel material was presented unprocessed, with a single token for each vowel and each talker. This enabled the subjects to familiarise themselves with the software and the task.

5.3.1.6 Procedure

The first session consisted of a ten minute familiarization with the CDT program with 6-channel *diotic* unshifted processing in quiet (i.e. 6U, similar to 6DU from Experiments 1 and 2).[8] This was to allow for early familiarization with the vocoded speech (Davis et al., 2005). Speech perception was then tested twice before training commenced, halfway through training, and at the end of training. As in Experiment 2, the very first test session served as familiarization, so that early, passive adaptation would be accounted for in the baseline measure. All subjects completed the entire cycle of training and testing within a maximum of two weeks, with no more than a two day gap between successive sessions. Sentence and vowel test presentation order was counterbalanced across the group. Within each test, stimuli were pseudo-randomized by block of condition and talker. Headphone orientation was also counterbalanced across subjects to control for any ear preferences in dichotic listening, which have previously been reported (Kimura, 1961). Table 5-3 describes the sequence of training and testing for Experiment 3.

The study employed a crossover design for the CDT training procedure, with eight subjects assigned to each training group. For the 10DOS/5OS – 5OS group, training in the first half of the experiment alternated in 10 minute blocks between the 10DOS and 5OS processors. The second half of the training comprised only the 5OS condition. The 5OS – 10DOS/5OS group received this same training, but in the

[8] The purpose of this procedure was familiarization with vocoding, and it was thought that 10 channels in quiet would be too easy.

Table 5-3. Sequence of testing and training for Experiment 3

Session	Training	Processor	Testing
1	10 minutes	6U	*Familiarization*: unprocessed vowels *Pre-test*: IEEE sentences (2 lists x 4 conditions x 2 talkers) bVd identification (5 tokens x 4 conditions x 2 talkers)
2			*Second pre-test*: IEEE sentences (2 lists x 4 conditions x 2 talkers) bVd identification (5 tokens x 4 conditions x 2 talkers)
4 – 6	40 minutes	10DOS + 5OS or 5OS	
7	10 minutes familiarization if not immediately following training session (see text)	10DOS + 5OS or 5OS	*Mid-test:* IEEE sentences (2 lists x 4 (5) conditions x 2 talkers) bVd identification (5 tokens x 4 (5) conditions x 2 talkers)
8 - 11	40 minutes	5OS or 10DOS + 5OS	
33	10 minutes familiarization if not immediately following training session	5OS or 10DOS + 5OS	*Post-test:* IEEE sentences (2 lists x 4 (5) conditions x 2 talkers) bVd identification (5 tokens x 5 conditions x 2 talkers) IHR sentences (2 lists x 4 conditions x 2 talkers)

reverse order. As in Experiment 1, training took place over eight 40-minute training sessions, with the half-way point speech perception test and subsequent training condition change occurring after four of these sessions. If the testing session did not take place immediately following a training session, then the subject heard a ten-minute CDT block with the 10DOS/5OS or 5OS processor (depending on subject group), which was not counted towards the total hours of training.

The 10OS condition served to measure laterality effects in adapting to the 10DOS processor, so it was only tested immediately after the subject group had trained with this processor. This was midway for the 10DOS/5OS – 5OS group, and at the endpoint for the 5OS – 10DOS/5OS group. Because it was expected that the smaller shift and the increased number of channels and consequent more dense spectral sampling in the 5OS condition would make this condition somewhat easier to learn than the 3OS condition in the previous experiments, all training was in auditory mode only.

Testing and training took place in a sound-treated room with presentation of the processed speech over Sennheiser HD280 headphones. The level was set by the experimenter to a comfortable listening level of approximately 65 dB. Subjects could request for the level to be adjusted, but none did, so this level was used by all participants.

5.3.2 Results

Test data was analyzed using repeated measures analysis of variance (ANOVA), with within-subject factors of test session, condition and talker, and a between-subjects factor of training group. Hyunh-Feldt epsilon corrections were applied to all F tests for factors with more than one degree of freedom. Hyunh-Feldt adjusted degrees of freedom have been rounded to the nearest integral value, and the significance criterion was $p = 0.05$. *A priori* hypotheses were tested using planned contrasts, and *post-hoc* testing was carried out using Bonferroni-adjusted paired comparisons.

5.3.2.1 Training group effects

The ANOVA revealed no significant training group effect for any of the test material, but the observed power was low in all instances, most likely due to the small number of participants in each group (eight). Because no significant effect could be shown, all subsequent analyses are based on the data from the sixteen subjects pooled.

5.3.2.2 Incidental learning

In an ANOVA of the first two baseline test sessions, *post hoc* testing revealed significant incidental learning from repeated test exposure without any training. For the sentence test, this was limited to conditions with unshifted speech (10DU, 10DOS and 5EU) ($p < 0.01$), but for the vowel test, subjects showed significant learning in all conditions ($p < 0.05$). Remarkably, at the unfavourable SNR of 0 dB for the vowel test, subjects were still able to passively adapt to the shifted speech. This incidental learning contrasts with Experiment 2 and was unanticipated. The larger number of channels available and smaller frequency shift may have facilitated passive adaptation compared to Experiment 2. Because training commenced after the second baseline test session, it is unknown whether passive adaptation would have continued beyond that shown in these first two sessions. Nevertheless, all subsequent analyses are based on the final three test sessions only, which at least partially discounts any passive adaptation.

5.3.2.3 IEEE sentence perception

Boxplots showing IEEE sentence intelligibility across training sessions for the four

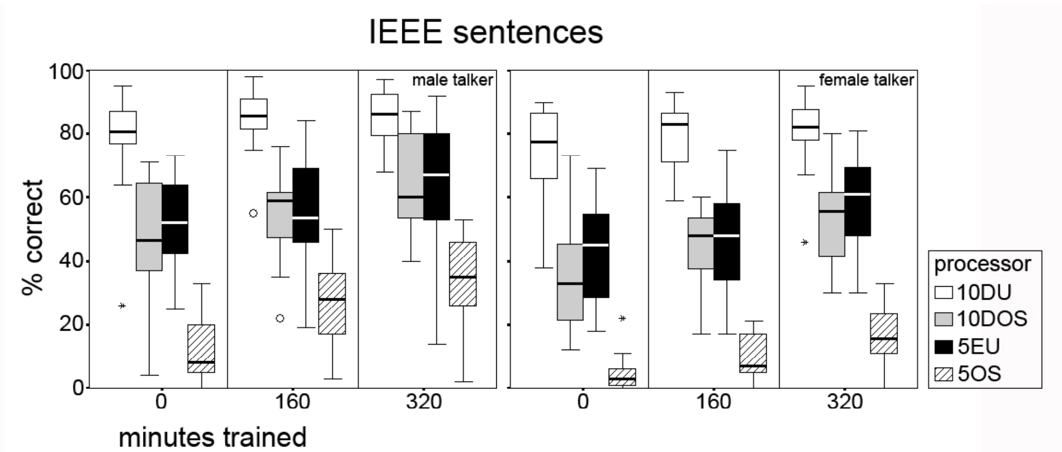

Figure 5-3. Boxplots of IEEE sentence recognition as a function of training time for Experiment 3. The scores have been broken down by talker because of a significant talker by processor interaction.

main conditions – 10DU, 10DOS, 5EU and 5OS – are given in Figure 5-3. A repeated measures ANOVA revealed a significant talker by processor interaction [$F(3,42) = 4.44$, $p = 0.008$], so the plots are shown for the two talkers individually. This interaction appears to reflect the higher intelligibility of the male talker for most conditions except the 10 channel unshifted condition (10DU), for which intelligibility with the two talkers was roughly equal.

Despite this talker effect, the analysis revealed three key findings common to both talkers. First, in planned contrasts, intelligibility with the 10DU condition always exceeded that with the 5EU condition ($p < 0.001$ for both talkers), which is an indication that there was room for improvement from the addition of the odd channels. Secondly, planned contrasts of sessions within the 5OS condition showed highly significant learning with training in this condition ($p < 0.001$ for both talkers). This level of learning was not found for the IEEE sentences in Experiments 1 and 2. Yet, despite this increased learning, intelligibility with the 10-channel binaurally mismatched processor (10DOS) never exceeded that with the 5-channel monaural unshifted processor (5EU), indicating subjects had not learned to binaurally integrate the mismatched maps over the five hours and twenty minutes examined here. For the most part, these conditions remained statistically equivalent. However, for the female talker, there was evidence that intelligibility of the 10DOS processor was significantly *worse* than that with the 5EU processor at the final test session [$F(1,15) = 4.64$, $p = 0.048$], suggesting interference from the additional channels rather than integration.

5.3.2.4 IHR sentence perception

Boxplots of easier IHR sentence intelligibility at the final test session are shown in Figure 5-4. Because the ANOVA showed a significant talker by processor interaction [$F(3,42) = 3.48$, $p = 0.02$], the data are presented for the two talkers separately. This

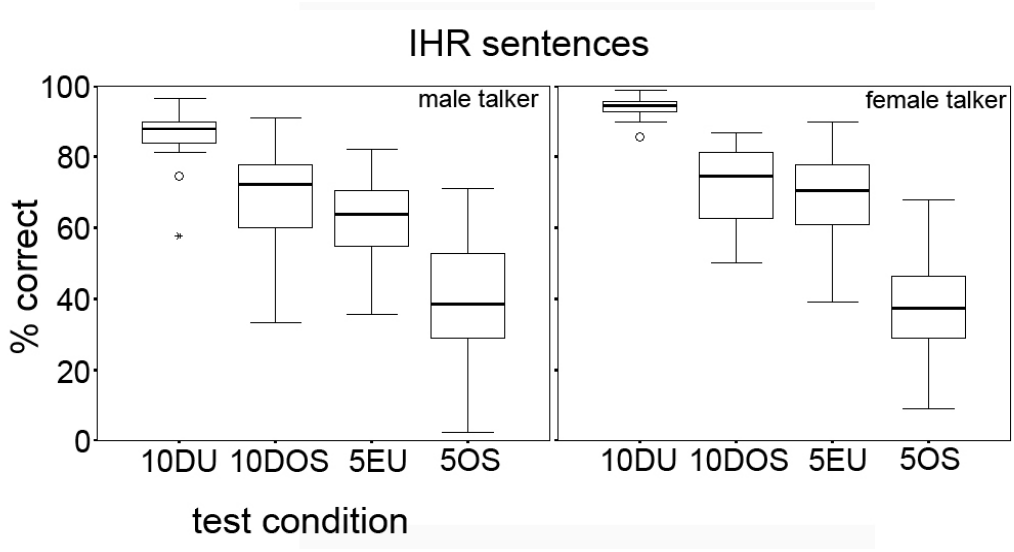

Figure 5-4. IHR sentence intelligibility for Experiment 3, which was measured at the final test session only. The data have been broken down by talker because of a significant talker effect.

interaction appears to be the result of better intelligibility with the female talker compared to the male talker for the unshifted, but not shifted, conditions.

The ANOVA revealed both similarities and differences with the IEEE sentences. First, while intelligibility with the 10DU unshifted condition was at or near ceiling for both talkers, intelligibility with the 5EU processor was significantly worse than the 10DU condition according to planned contrasts (both talkers $p < 0.001$), indicating there was room for improvement from the addition of the odd channels. Secondly, though there was no data from previous sessions to compare to, intelligibility with the 5OS condition was well above floor at around 40% for both talkers, which is a clear indication of adaptation to the 5OS condition. This is a similar level of intelligibility as that found for the 3OS condition in Experiment 2 after double the amount of training (10 hours), and far exceeds the near floor intelligibility of 3OS found in Experiment 1 with the same amount of training. Finally, a planned contrast comparing the 10DOS condition to the 5EU condition revealed no significant difference for the female talker. However, for the male talker, the same test revealed a small but significant advantage of around 6% for the binaurally mismatched processor (10DOS) compared to the monaural unshifted even channels (5EU) [$F(1,14) = 5.33$, $p = 0.036$]. In contrast to the IEEE sentences and the results from previous experiments, then, this somewhat surprising result seems to indicate a binaural advantage.

5.3.2.5 Vowel perception

Figure 5-5 shows boxplots of vowel perception across training sessions for the male and female talkers separately, since the ANOVA revealed a significant talker by processor interaction [$F(3,39) = 5.76$, $p = 0.003$]. Intelligibility with the male talker tended to be higher than that with this female talker. On inspection of the plots, the higher level of noise presentation for the vowel test – 0 dB SNR – appears to have affected the results significantly. First, there is a large variability in recognition accuracy for all conditions and with both talkers, probably reflecting variability in the listeners' ability to comprehend speech in noise. Secondly, and somewhat surprisingly, the decreased SNR seems to have suppressed intelligibility with the 10DU condition without suppressing that with the 5OS condition. Intelligibility of all

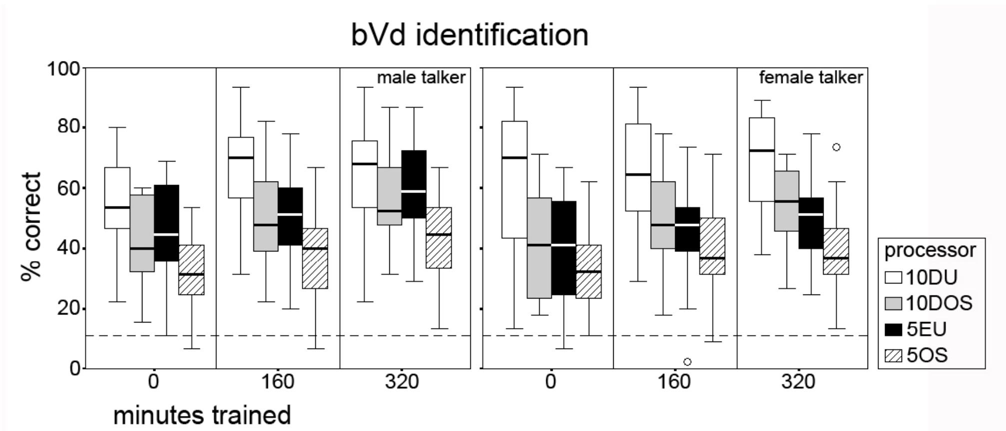

Figure 5-5. Boxplots of bVd word identification across training sessions. The scores have been broken down by talker because of a significant talker by processor interaction. The dashed reference line at the bottom denotes chance performance (11.1%).

conditions in this test was relatively clustered compared to Experiments 1 and 2, with 5OS remaining well above chance and close to that with the 10DOS and 5EU conditions. By contrast, median intelligibility with the 10DU condition remained low throughout the experiment, never exceeding 80%.

The ANOVA revealed similar findings to that from the IEEE sentences. According to planned contrasts, intelligibility with the 10DU condition remained significantly better than the 5EU condition, indicating room for improvement from the addition of the odd channels. At the final test session, this finding was more pronounced for the female talker ($p < 0.001$) than the male talker ($p = 0.041$), but this appears to be the result of the depressed intelligibility of 10DU in the presence of strong background noise, and not the result of ceiling intelligibility with 5EU. Secondly, planned contrasts also showed significant learning of the 5OS condition for both the male ($p = 0.002$) and female ($p = 0.036$) talkers. Despite this room for improvement and evidence of learning of the shifted channels (5OS), however, subjects did not learn to binaurally integrate in the 10DOS condition. For the male talker, there was evidence that intelligibility with 10DOS was significantly *worse* compared to the 5EU condition [$F(1,15) = 5.71$, $p = 0.03$], suggesting interference from the additional shifted channels rather than binaural integration. For the female talker, there was no significant difference between these conditions after training, also suggesting a failure to learn to integrate shifted and unshifted bands.

5.3.2.6 Laterality effects

The 10OS condition was tested after each group had trained with the 10DOS processor, so at the midway point for the 10DOS/5OS – 5OS group, and at the end point for the 5OS – 10DOS/5OS group. An ANOVA comparing the 10DOS condition and the 10OS condition was computed for each training group separately. This revealed no significant talker effects, so the data was pooled by talker. There was no difference between these conditions for bVd identification, but for IEEE

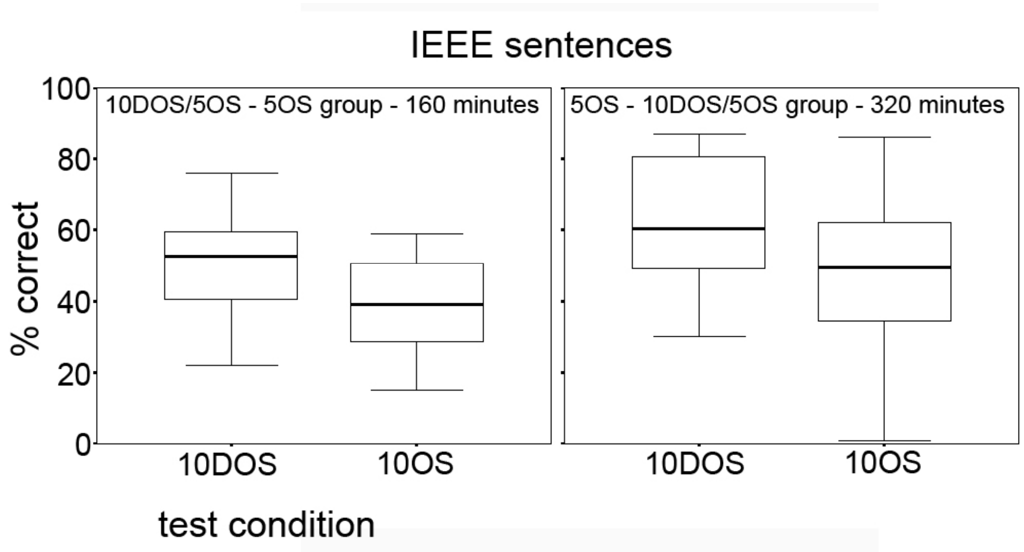

Figure 5-6. IEEE sentence intelligibility for the 10DOS and 5OS conditions broken down by training group. The group in the left-hand panel were tested with this condition after 160 minutes, while the group in the right hand panel after 320 minutes of training. Note that this is reflected in the lower intelligibility of these conditions for the group in the left-hand panel, but not the pattern between the conditions.

sentences, planned contrasts showed intelligibility with 10OS was significantly worse than 10DOS for both the group trained with 10DOS first [$F(1,7) = 25.8$, $p = 0.001$] and the group trained with 10DOS last [$F(1,7) = 18.6$, $p = 0.003$]. This finding, highlighted in Figure 5-6, is an indication that subjects were using cues to ear presentation to attend to the unshifted channels, since intelligibility was adversely affected when shifted and unshifted channels were not separated by ear.

5.3.3 Discussion

5.3.3.1 Training condition effects

There did not appear to be a training order effect in these experiments, but the observed power was low, so there may have been a small effect that was missed. However, the provision of explicit training time with both the binaurally mismatched processor (10DOS) and the shifted channels alone (5OS) had a similar strong and positive impact on adaptation to the 5OS condition to that shown for the 3OS condition in Experiment 2. This finding is consistent with emerging evidence about the benefits of exposure to an easier training condition when facilitating adaptation to a more difficult spectral shift condition (Li et al., 2009). While this advantage in facilitating adaptation to 5OS did not, typically, translate into binaural integration with the 10DOS processor, the results nevertheless suggest that mixed exposure may best facilitate adaptation to a second bilateral implant, as suggested by Tyler (Tyler et al., 2007).

5.3.3.2 Adaptation to shifted channels

Significant learning of the 5OS condition in this experiment was apparent at all test points and for all test materials. This finding is striking for several reasons. First, as in the 3OS condition, the 5OS condition represents a sparse sampling of a 10-channel vocoder, which may arguably be more difficult to learn when a spectral shift has been applied. Secondly, while the smaller 3.8 mm shift may have been easier to learn than the 6 mm shift examined previously, it is nevertheless outside the 3 mm range that has been suggested by Fu and Galvin to be the limit to passive, automatic adaptation to basalward shift (Fu and Galvin III, 2003). Third, the low-frequency smoothing filter used for the sine-vocoding may have further limited intelligibility (Souza and Rosen, 2009), though this admittedly may have been tempered by the increased number of spectral channels compared to the 3OS condition.

The most impressive consideration in this adaptation to 5OS, however, is that this occurred in the presence of background noise. For all speech materials, but particularly in the case of vowels where the SNR was at a very unfavourable 0 dB, subjects still showed significant adaptation to 5OS. This is the first such experiment, to the author’s knowledge, that has shown adaptation to basally shifted speech in the presence of background noise. At a minimum, this novel finding strongly bolsters support for the effectiveness of the computer-based CDT training used in this experiment in facilitating adaptation to simulations of cochlear implants.

5.3.3.3 Resistance to integration

Despite strong evidence of learning of the 5OS shifted condition for all test materials in these experiments, and evidence of room for improvement from the addition of the odd channels (10DU > 5EU at all test points), the bulk of the evidence suggests that subjects were resistant to integrating the binaurally mismatched frequency-to-place maps over the course of the training examined in this experiment. As in the previous experiments, intelligibility with the monaural unshifted channels (5EU) tended to match that with the binaurally mismatched processor (10DOS). Moreover, on two instances, there was evidence to suggest that the addition of the shifted channels to the unshifted channels served as a distraction, since post-training intelligibility with 10DOS was *worse* than 5EU. This occurred for the female talker for the IEEE sentences, and the male talker for the bVd identification.

However, this last finding contrasts with the result for the easier IHR sentences, where there was a small but significant advantage for the 10DOS condition for the male talker. Because this binaural advantage was small, appeared at only one test point and for only one talker, and contrasts with all other evidence from this experiment, it seems difficult to conclude that this was a strong or real effect. Though it must be conceded that the training times examined in this experiment were small compared to the omnipresent experience bilateral cochlear implant users will have with their implants, the evidence here is nevertheless broadly consistent with an account of "better ear" dominance, where subjects adapt to a binaurally mismatched processor by attending to the "better ear". It should be noted that this "better ear" dominance is not consistent with the right-ear dominance previously reported in dichotic listening experiments (Kimura, 1961). The headphone orientation was counterbalanced across the subjects, so if this had been the case, the effect would have been reflected in better speech scores at a particular ear regardless of condition, which was not found.[9] The dominance of a specific ear in dichotic listening and other binaural hearing tasks with speech stimuli has most commonly been shown when different speech signals are delivered to each ear (Beaman, Bridges, and Scott, 2007), so the lack of effect here may reflect that listeners perceived the signals as originating from one source, even if perception was typically based on the unshifted channels alone.

As proposed in Experiments 1 and 2, “better ear” dominance may reflect a sensitivity to relative frequency ordering of the output channels. However, the increased number of spectral channels used here – 10 – meant an increase in spectral resolution, making it less likely that individual formants would be displaced as in the 6 channel case of Experiments 1 and 2. The more narrow spectral bands of the analysis filters, and the smaller spectral shift, meant that even if an F1 was shifted upwards and an F2 was not, the F1 still remained below the F2 in the output. For example, an /i:/ vowel spoken by a female talker might have an F1 of 300 and an F2 of 2400. In the 10DOS processing here, the resulting vowel would have an F1 shifted upwards to 540 and an unshifted F2 of 2500. Since F1 and F2 values retained the same relative frequency ordering, it may have been easier to map the processed vowel onto an existing representation. The source of the “better ear” effect in these experiments could not be entirely attributable to sensitivity to relative frequency ordering, then, since relative frequency order was maintained more frequently here, yet “better ear” effects tended to dominate. Since the displacement of formants is likely to have a more drastic effect on the percept of a sound than a more fine-grained displacement of spectral cues as in the 10-channel case, it is

[9] A repeated measures ANOVA with a between-subjects factor of ear of shifted speech for the 10DOS condition showed no significant effect for IEEE sentences, IHR sentences or vowels.

difficult to determine whether sensitivity to relative frequency ordering contributed to the “better ear” dominance shown in this experiment, but the frequency-order hypothesis is still consistent with this account.

In considering the other experimental hypotheses, there does not appear to be any evidence in support of the adaptation hypothesis here. At no test point did intelligibility with the 10DOS processor approach that with the 10DU processor. While further training beyond the time examined in this experiment may prove differently, the finding of “better ear” dominance seems to suggest not. On the other hand, there does seem to be a small hint of support for the best local SNR hypothesis. For the IHR sentences and male talker, intelligibility with 10DOS exceeded that with 5EU by a small amount. This suggests that subjects had learned enough about the shifted condition that they may have learned to take momentary advantage of the SNR at the shifted ear when this proved more favourable than that at the unshifted ear.

A complicating factor with this interpretation, however, is the finding that intelligibility was adversely affected by diotic (10OS) versus dichotic (10DOS) presentation. When shifted and unshifted channels were not separated by ear, subjects showed decreased intelligibility. This suggests that subjects may have used cues to ear presentation in learning to attend to the unshifted speech only rather than binaurally integrating. After all, in the binaural integration of unshifted speech, there is no effect of dichotic versus diotic presentation – integration occurs readily. It could also be argued that this sensitivity to separation by ear was due to stronger learning of the 5OS condition. Recall that in the 6DOS-trained group of Experiment 1, where subjects were trained only with the binaurally mismatched processor, there was no effect of ear presentation. That group appeared to have learned by attending solely to the unshifted frequencies and/or ignoring the shifted frequencies, since intelligibility with 3EU always matched 6DOS, and intelligibility with 3OS was low overall. On the other hand, the increased amount of learning of 5OS shown here may have resulted in a sensitivity to *both shifted and unshifted speech* that was dependent on the different signals being separated by ear. When shifted and unshifted speech were presented diotically as in 10OS, intelligibility decreased. In essence, subjects may have adapted not by *ignoring* the shifted speech (as there was evidence for in Experiment 1), but rather by attending to both shifted and unshifted speech. If they had learned by ignoring the shifted speech, then intelligibility with 10OS should have matched that with both 10DOS and 5EU, with adaptation occurring on the basis of the unshifted channels (5EU) alone. However, because 10OS was significantly worse than both 10DOS *and* 5EU, it does seem likely that the reduced intelligibility with 10OS was a result of the decreased access to the unshifted speech when this was not separated by ear.

While Cutting also found reduced intelligibility for higher level linguistic fusions when these were presented diotically rather than dichotically (Cutting, 1976), the disruption was the result of energetic masking within an ear. This does not apply to the current investigation, which probed lower-level spectral fusion. Here there was no overlap of spectral information at the two ears, so no energetic masking could have taken place. Taking into account the evidence from the three experiments together, then, the most likely interpretation is that subjects learn to either attend to the unshifted ear, or ignore the shifted ear, in certain situations. This interpretation is further backed by the finding that intelligibility with 10DOS was significantly *worse* than 5EU on occasion, which suggested interference from the shifted bands in the 10DOS condition even in the presence of significant adaptation to the shifted bands in 5OS. If this were indeed the case, then this would have striking and adverse implications for bilateral cochlear implants. If listeners adapt to mismatched signals between the ears by learning to ignore an ear, then this would all but repudiate the purpose of providing a second implant.

Experiment 4 was designed to further probe the effect of ear presentation in adapting to the binaurally mismatched processor. Eight participants from Experiment 3 returned for further testing and training, and at the final test point, the orientation of the headphones was reversed. If this adversely affected perception of the 10DOS processor, then it would suggest that subjects were relying on cues to ear presentation to attend to the unshifted speech. A secondary aim of this experiment was to examine whether the small binaural advantage in the 10DOS condition for the IHR sentences shown here persisted or increased with further training. Finally, because the experiment took place weeks to months after the subjects' participation in Experiment 3, the study allowed for the examination of the retention of adaptation to the shifted, unshifted and combined speech.

5.4 Experiment 4

5.4.1 Method

5.4.1.1 Subjects

Subjects from Experiment 3 were contacted by email and asked to participate in a follow-up experiment, and eight accepted this offer. The time that had elapsed since their participation in Experiment 3 ranged from six to twenty-one weeks. This was a random sample of the previous sixteen subjects, so training group and headphone orientation were not counterbalanced across this group. All were paid for their participation.

5.4.1.2 Test materials

For sentence materials, the IEEE and IHR sentence sets were nearly exhausted in Experiment 3, so Experiment 4 used the easier BKB sentences (Bench et al., 1979). These were of a female talker from the same materials used in Experiments 1 and 2. This was also the same female talker from Experiment 3. A subset of fourteen of sixteen BKB sentences per list were used so that each sentence had three keywords. The BKB and IHR sentences are of similar complexity and construction, and the results from the two sets will be compared in this study. The four remaining lists of the IEEE sentences from a male talker were also used in the key condition (10DOS) so that results could be compared to Experiment 3. The vowel stimuli used here were from the female talker only of the same set used in the previous experiments.

5.4.1.3 Test conditions and signal-to-noise ratio

The purpose of this experiment was to monitor binaural advantage and/or integration and ear presentation, so only three conditions were tested here – the 10DOS, 5EU and 5OS conditions of Experiment 3 (see Table 5-1). Because the sentence materials used were easier sentences, and the subjects underwent further training, there was some concern that subjects may reach ceiling performance for the unshifted speech at the 10dB SNR used in the previous experiment. Therefore, sentence intelligibility was measured at two SNRs – 10 dB, in order to provide a direct comparison to Experiment 3, and 8 dB. Ceiling effects at an SNR of 0 dB were not a risk for the vowel identification, so only this SNR was tested for the vowels.

5.4.1.4 Procedure

The training and testing procedure for Experiment 4 is outlined in Table 5-4. Initially subjects were instructed to wear the headphones in the same orientation as they had for Experiment 3. The first session consisted of a baseline measure of speech

Table 5-4. Sequence of testing and training for Experiment 4

Session	Training	Processor	Testing
T3	————	————	*Post-test* from Experiment 3: IEEE sentences IHR sentences bVd identification
T4	————	————	*Baseline*: BKB sentences (1 list x 8 |10 dB x 5OS|5EU|10DOS) bVd identification (9 vowels x 5 tokens x 5OS|5EU|10DOS)
CDT	40 minutes	10DOS + 5OS	order counterbalanced across group
T5	————	————	*Post-test 1:* IEEE sentences (2 lists x 10DOS) BKB sentences (1 list x 8|10 dB x 5OS|5EU|10DOS) bVd identification (9 vowels x 5 tokens x 5OS|5EU|10DOS)
T6	————	————	*Post-test 2 *** HEADPHONES REVERSED* IEEE sentences (2 lists x 10DOS) BKB sentences (1 list x 8|10 dB x 5OS|5EU|10DOS) bVd identification (9 vowels x 5 tokens x 5OS|5EU|10DOS)

perception. This has been coded as T4, since it is the fourth such speech perception test these subjects had participated in, with T3 being the final test session of Experiment 3, and T0 the very first baseline measure of that experiment. Following this session, subjects underwent a further 40 minutes of computer-based CDT. The training condition alternated between the 10DOS and 5OS conditions in 10 minute blocks, and the order of these conditions was counter-balanced across the group. All subjects started with a new story, and the training talker was the same as in Experiments 2 and 3. Following training, speech perception was again measured with BKB sentences, bVd identification and the 10DOS condition only for the IEEE sentences. After this test session, subjects were instructed to reverse the orientation of the headphones. They then underwent a final speech perception test (T6) with these same speech materials. The entire experiment took approximately two hours per subject.

5.4.2 Results

Data was analyzed with repeated-measures ANOVA, and *a priori* hypotheses were tested with planned contrasts of sessions within a condition or conditions within a session. A linear regression was performed on the retention data to explore whether there were predictive factors in the retention of the shifted speech.

5.4.2.1 Retention and continued learning

Data from the last test session of Experiment 3 (T3) was compared to data from the baseline session of Experiment 4 (T4), and following 40 minutes of further training (T5), to examine the retention of adaptation and continued learning. For vowel identification, there were no significant changes in any of the conditions across sessions T3 – T5. However, for the IHR/BKB sentences, performance with the shifted speech (5OS) decreased significantly [$F(1,7) = 9.54$, $p = 0.018$] in the absence of experience, but performance with conditions containing unshifted speech (10DOS and 5EU) did not. Forty minutes of further training proved sufficient to restore intelligibility to the level achieved at T3, since there was no significant difference in this condition between T3 and T5. Further training did not improve

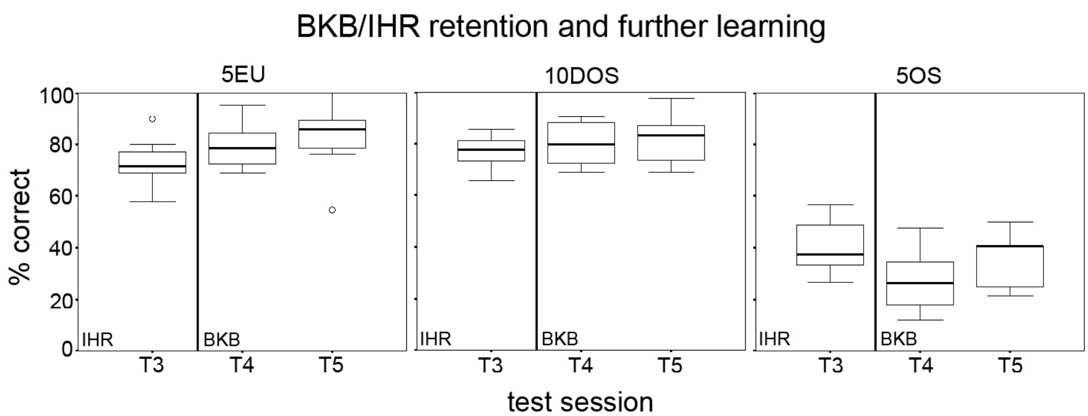

Figure 5-7. Boxplots of IHR and BKB sentence intelligibility as a function of test session and condition. Data for T3 are taken from the final test session of Experiment 3. The time elapsed between T3 and T4 varied across subjects from 3 to 21 weeks. Forty minutes of CDT was provided between T4 and T5. All data are for a female talker at a 10 dB SNR.

intelligibility with 10DOS or 5EU significantly, nor any of the conditions at the 8 dB SNR (T4 – T5 only). Boxplots highlighting intelligibility across sessions T3 – T5 for the BKB/IHR sentences at an SNR of 10dB are shown in Figure 5-7.

Analyses were performed to determine whether there were any predictive factors that could characterize the retention of adaptation to 5OS. The linear relationships are summarized in Figure 5-8 . The time elapsed between experimental sessions T3 and T4 varied widely across subjects from three to twenty-one weeks. However, a linear regression analysis showed no significant relationship between this decrease in intelligibility of 5OS between T3 and T4 and the time elapsed between these sessions. The subjects who returned for Experiment 4 were broadly classified into two groups: those who had participated more distantly and those more recently. This can be seen on the scatterplot, where subjects are clustered between weeks 5-10 and 17-21. This clustering is partially due to the random sampling of the subjects who returned but also to a break in testing during Experiment 3. It is interesting to

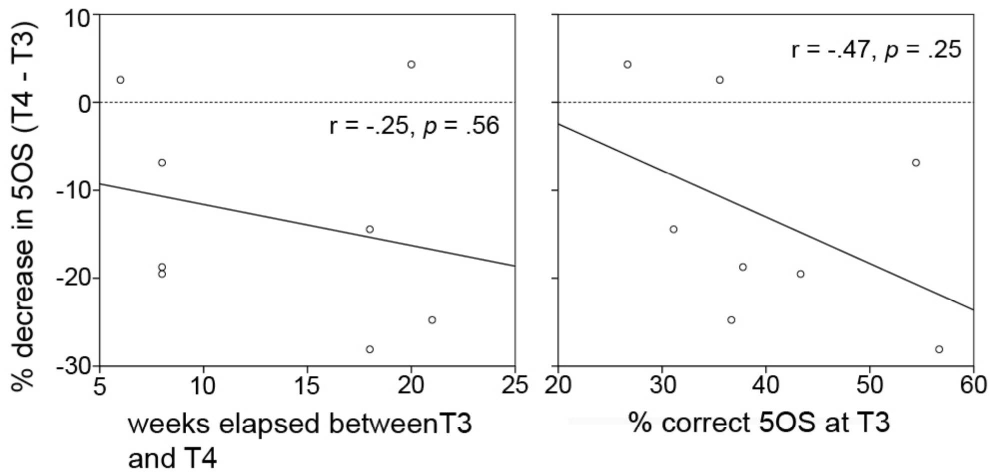

Figure 5-8. Figures showing regression lines for predictors of retention of adaptation to the 5OS shifted condition. Data is from the BKB sentences at an SNR of 10 dB. The dashed line indicates no change in performance between sessions T3 and T4. Neither time elapsed between test sessions nor intelligibility achieved on the previous experiment predicted retention. While only the linear relationships are shown, no other significant relationships were found.

note that despite this gap in testing, within each of these groups subjects seem to vary similarly in terms of their retention of 5OS. Though the sample size was small, this is an indication that retention of adaptation is related to the individuals' ability to cope with spectrally reduced speech rather than a gradual deterioration of adaptation to 5OS with increasingly distant exposure.

Along this line of reasoning, another potentially predictive variable of this decrease was the intelligibility achieved with 5OS at T3. Subjects who were better at perceiving 5OS after training may have been more likely to retain this ability. Performance at T3 versus the decrease in performance between T3 and T4 were also entered into a regression analysis, but again, no significant relationship was found. The decrease in performance may have also been related to the improvement shown over the time course in Experiment 3 rather than the raw post-training intelligibility score: subjects who learned more may have been better at retaining adaptation, or conversely subjects who learned more might have had more to forget. Data from IHR sentences were only collected at the final test session in that experiment, so this analysis could not be computed. However, based on the current analyses, the ability of subjects to retain adaptation to the shifted speech appears to be variable among subjects and not related to overall performance achieved with the shifted speech or time spent in absence of experience.

5.4.2.2 Cues to ear presentation

Repeated measures ANOVAs were computed on data from the BKB and IEEE sentences and bVd identification for the final two test sessions (T5 and T6) to determine whether reversing the orientation of the headphones had an effect on intelligibility. For the BKB sentences there was a significant effect of SNR [$F(1,7) = 15.1$, $p = 0.006$], but this did not interact with processor or session, so further analyses were based on pooled data from the two SNRs. Within each processor,

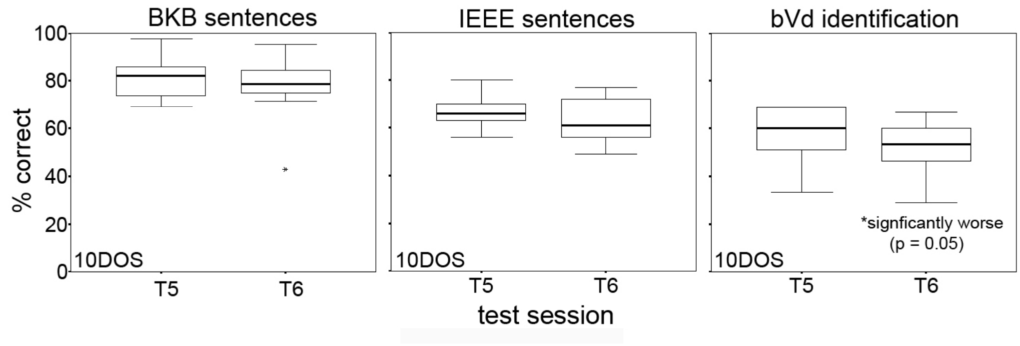

Figure 5-9. Boxplots comparing intelligibility of the 10DOS condition before and after headphone orientation reversal. The left panel shows BKB sentences, which were of a female talker at the combined 8 dB and 10 dB SNRs. The middle panel shows data from the IEEE sentences, which were from a male talker. The right panel shows data from the bVd identification, which were of a different female talker. Test sessions differed significantly only for vowel identification, which is denoted in the left panel.

planned contrasts of these test sessions revealed no significant difference. The same was found for the IEEE sentences (male talker, 10 dB SNR only), but for vowel identification, planned contrasts showed that intelligibility with the 10DOS processor decreased just significantly when the headphone orientation was reversed [$F(1,7) = 5.6$, $p = 0.05$]. There was no significant change in the 5EU or 5OS conditions. The findings for the 10DOS condition for each of the speech stimuli types are summarized in Figure 5-9.

While the analysis did not show statistical differences indicating ear sensitivity for the sentence tests, reversing the headphone orientation seems to have affected intelligibility here, too, at least for some subjects. For the BKB sentences, an outlier appeared at T6 who scored significantly lower than the median upon headphone orientation reversal, and for the IEEE sentences, intelligibility both decreased and became more variable at T6.

5.4.2.3 No binaural advantage

At the end of Experiment 3, a hint of an advantage for the binaural 10DOS condition over the 5EU condition appeared in the male talker IHR test. In order to test whether

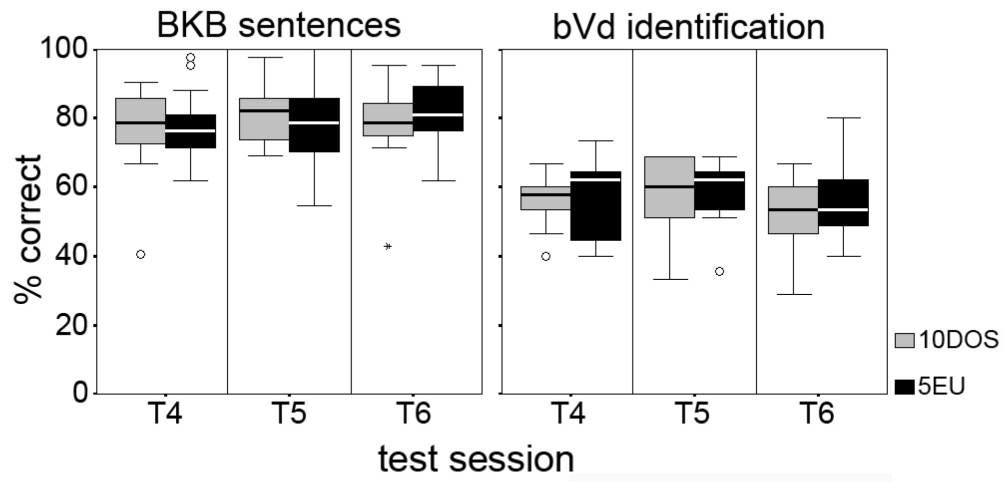

Figure 5-10. Boxplots comparing the 10DOS and 5EU conditions across the three test sessions of Experiment 4 for BKB sentences (left panel) and bVd identification (right panel). Data for the BKB sentences has been pooled over the 10 dB and 8 dB SNRs. The talkers were two different female talkers.

this finding remained or strengthened with the further training provided in this experiment, planned contrasts of the 10DOS and 5EU conditions at test sessions T4 – T6 were computed for the BKB and vowel stimuli. For the BKB sentences, scores were pooled for the two SNRs since there was no significant interaction of processor with SNR over these test sessions. No significant differences between these conditions were found at any test point for either sentences or vowels. Boxplots of these findings are shown in Figure 5-10. While the female talker in this experiment was not the same talker who provided the significant binaural advantage in the previous experiment, the data here suggest that this finding may have been fleeting and a statistical anomaly.

5.4.3 Discussion

The subjects in Experiment 4 demonstrated a high level of retention of adaptation to all conditions. While intelligibility with shifted speech (5OS) decreased in the absence of training, 40 minutes of further training was sufficient to restore intelligibility to the level achieved in the previous experiment. The ability of subjects to retain adaptation to the spectrally shifted speech could not be correlated with time spent in the absence of training or overall level of intelligibility in the previous experiment, so the factors governing this retention ability remain elusive but appear to be subject-specific.

The findings from Experiment 4 also suggest that, for most subjects, adaptation to the binaurally mismatched speech does not take place by attending to or ignoring a specific ear. While there was evidence that intelligibility was adversely affected when specific laterality cues were removed, this was not a strong effect and was only just significant for the vowel test. On closer examination of the spread of the data, it appears that this strategy may also be subject-specific, with most people adapting by attending to the frequency content of the signal rather than ignoring an ear. These findings were somewhat surprising, given the subjects' sensitivity to dichotic (10DOS) versus diotic (10OS) presentation of speech found in Experiment 3. While adaptation appears to take place on a frequency basis, cue to ear presentation may serve as one strategy in facilitating attending to these frequencies.

However, despite the fact that subjects do not appear to ignore an ear when adapting to the binaurally mismatched speech, even with the continued training provided in this experiment they still did not demonstrate binaural integration or a binaural advantage. Intelligibility with the binaurally mismatched processor (10DOS) closely mirrored that with the unshifted speech (5EU). Even though adaptation to the 5OS processor was robust and was retained in the absence of exposure, subjects did not appear to use this information when it was presented simultaneously with the unshifted speech. Findings from Experiment 4 are thus consistent with the findings of the previous three experiments, and suggest that listeners are resistant to integrating binaurally mismatched frequency-to-place maps.

5.5 General discussion

Experiments 3 and 4 aimed to further examine adaptation to binaurally mismatched frequency-to-place maps while addressing some of the shortcomings of Experiments 1 and 2. As was anticipated, the increased number of spectral channels and smaller degree of basalward spectral shift facilitated adaptation to the basally shifted speech (5OS) in these experiments. Adaptation may have also been further facilitated by the mixed exposure to the two training conditions, and increased training time with the shifted speech alone. This finding of significant adaptation is impressive, since the 5OS condition was composed of a spectrally sparse representation of the information that would be in a 10-channel vocoder, and was presented in the presence of background noise. Quite striking, too, was the subjects' retention of adaptation to the shifted speech over a period extending up to twenty-one weeks in absence of experience.

Yet, despite this increased learning and experience of the shifted condition (5OS), the bulk of the evidence suggests that subjects did not binaurally integrate this speech with the unshifted channels. While there was a hint of evidence of binaural advantage in Experiment 3, this finding did not remain with the further testing and training provided in Experiment 4. Furthermore, there was evidence in Experiment 3 that intelligibility with 10DOS was worse than 5EU, suggesting distraction from the addition of the shifted channels rather than integration. The findings from these experiments are thus consistent with an account of better-ear dominance, where subjects rely on information being provided by the "better ear" when listening to frequency-place maps that are mismatched between the ears.

Experiment 4 addressed the issue of whether this apparent "better ear" dominance reflected learning specific to an ear, or whether learning was based on adaptation to the frequency content providing the most reliably intelligible signal. The most likely interpretation is that adaptation to the 10DOS condition is based on attending to the unshifted speech, and that this is made easier when shifted and unshifted channels are separated by cues to ear presentation. However, most subjects do not appear to assign this cue weighting to a *specific* ear, so this attuning to the unshifted frequencies appears to occur at a more central stage of processing, and not by simply ignoring a specific ear. Still, at least one subject tested in Experiment 4 seemed acutely sensitive to specific ear cues, so this may be a strategy employed by some people.

This interpretation of “better ear” dominance based on higher-level adaptation does not exclude the possibility of binaural integration of mismatched frequency-place maps in the way that learning ear-specific cues does. If there had been a strong, negative effect from reversing the headphone orientation, then this would suggest that subjects had learned by ignoring an ear, which in turn would have negative implications for bilateral cochlear implant fittings. Nevertheless, the evidence from these experiments suggests that subjects are resistant to integrating binaurally mismatched frequency-to-place maps. As in Experiments 1 and 2, it could be argued that the level of intelligibility provided by the 5 channels of unshifted speech may have been sufficient to limit the motivation for integration with the shifted bands. Partially in order to limit the benefit from the increased number of unshifted channels, the speech in this experiment was presented in background noise. Yet, especially in the case of vowel identification, where the SNR was at an unfavourable 0 dB, subjects were far from achieving ceiling with the 5EU condition after training, but they still showed no signs of binaural integration. Even more striking was the level of adaptation to 5OS achieved with such a high level of background noise, which should have facilitated integration if it were possible.

The bulk of the evidence thus suggests that mismatches in frequency-place maps between the ears are difficult for subjects to learn, despite the fact that subjects may be able to adapt to the signals at the two ears individually quite well. Of course, the processing used in these experiments does not exactly mimic the situation seen in bilateral cochlear implants, and the extent of their experience will be far greater than the short-term training examined here. Nevertheless, the findings presented here have clear implications in the consideration of bilateral cochlear implant fittings. The next chapter will consolidate and critically assess the findings from the four experiments, discuss their implications for bilateral cochlear implant fittings and speech perception, and consider areas for future work.

Chapter 6. Resistance to Integration of Frequency-Place Mismatches: Implications for Bilateral Cochlear Implants

The preceding two chapters presented evidence from a series of experiments exploring normally hearing listeners' ability to adapt to frequency-place maps that were mismatched between the ears. The results from these experiments were broadly consistent and indicated that while listeners can adapt to spectrally reduced and basally shifted speech, they are resistant to integrating this with unshifted speech when both are presented simultaneously to different ears. The extent to which the processing used in these experiments reflects the situation in actual bilateral cochlear implant situations is uncertain. However, the persistent resistance to integration shown here would appear to contraindicate frequency-place mismatches between the ears in bilateral cochlear implant fittings. The findings thus have important implications for cochlear implants, where standard fitting practice does not typically account for any frequency-place mismatches between the ears (Muller et al., 2005; Gantz et al., 2002). Not all patients fitted bilaterally with cochlear implants demonstrate a binaural advantage, and this may reflect mismatches in frequency-place maps between the ears. This final chapter will provide a summary of the key findings from the four experiments and their limitations, discuss the implications for bilateral cochlear implant fittings and for theories of binaural hearing, and consider areas for future work.

6.1 Summary of results

6.1.1 Adaptation to shifted speech

Table 6-1 provides a summary of post-training intelligibility and improvement with training for the shifted conditions (3OS and 5OS). These results are also summarized in figures. Post-training results for all speech materials are shown in Figure 6-1, and improvement scores for IEEE sentence and bVd identification are shown in Figure 6-2. For simplification these have been pooled across talker, though in most cases there were significant talker effects, with the male talkers typically more intelligible than the female talkers. For comparisons between conditions or sessions, the significance value indicated on the figure comes from a planned

Table 6-1. Summary of post-training intelligibility and improvement with training for the monaural odd-channels shifted speech conditions (3OS and 5OS) in Experiments 1 – 3.

experiment	training condition	test condition	speech test	% correct after training		% improve with training	
				mean	s. d.	mean	s. d.
1	6DOS	3OS	IEEE	3.83	4.63	3.83	4.63
			bVd	28.6	6.37	4.81	6.42
	3OS	3OS	IEEE	2.25	2.86	2.00	2.89
			bVd	32.0	10.3	8.15	7.89
			IHR/BKB	8.50	9.73	—	—
2	6DOS/3OS	3OS	IEEE	11.5	10.5	10.8	10.4
			bVd	31.1	6.33	5.92	6.95
			IHR/BKB	29.8	22.8	26.6	20.0
3	10DOS/5OS	5OS	IEEE	24.9	15.4	16.4	11.8
			bVd	40.5	14.7	8.96	12.0
			IHR	38.6	15.9	—	—

6DOS: 6 dichotic odd shifted (binaural)
3OS: 3 odd shifted (monaural)
10DOS: 10 dichotic odd shifted (binaural)
5OS: 5 odd shifted (monaural)

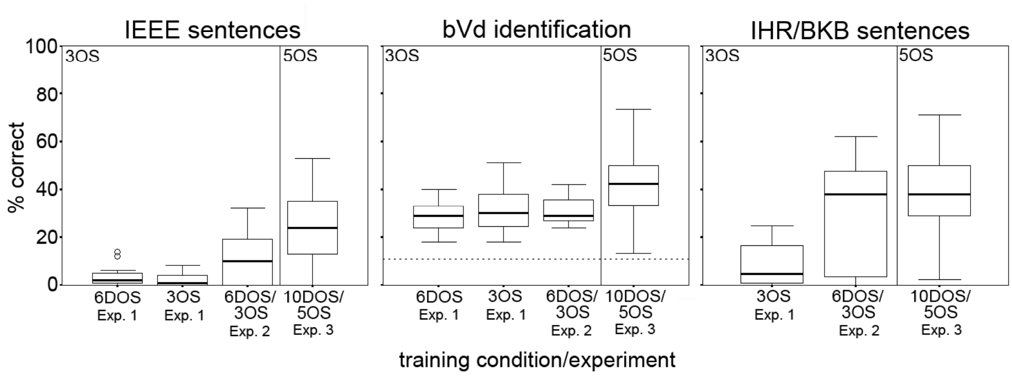

Figure 6-1. Post-training intelligibility of the monaural shifted conditions. Note that data from Experiment 2 is after 10 hours training, nearly double that of Experiments 1 and 3 (5 hours, twenty minutes). For the vowel identification, chance performance is indicated by a dashed line. The 5-channel condition of Experiment 3 is offset by a vertical line.

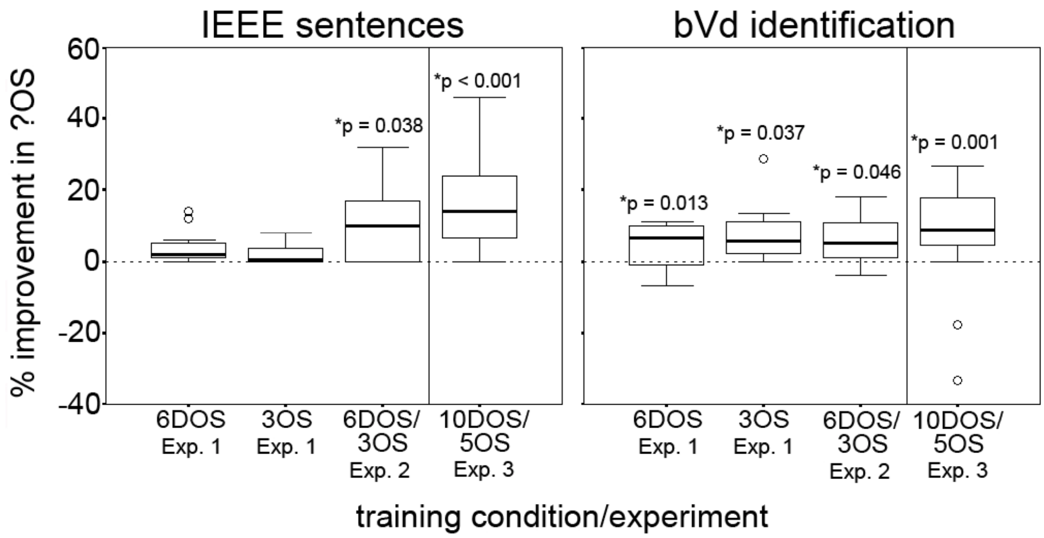

Figure 6-2. Boxplots showing improvement with training in the basally shifted conditions. The dashed line indicates no change, and the vertical line offsets the 3OS and 5OS conditions. Note that the training time in Experiment 2 was nearly double that in Experiments 1 and 3 (10 hours versus 5 hours, 20 minutes). Scores have been pooled over the male and female talkers. When an improvement between pre-test to post-test was significant, this is indicated with a p-value, but note that because the results are pooled, the p-value may differ from that presented in previous chapters.

contrast of conditions (3/5EU vs. 6/10DOS) or sessions within a condition (3/5OS: t1 vs. t3). Because results have been pooled, significance values reported here may differ from those presented in previous chapters. The aim here was to provide a general account of trends in the data across experiments, talkers and conditions.

Especially in the case of the first experiment, evidence of adaptation to the basally shifted speech was not as apparent as in previous studies of spectrally shifted, vocoded speech. While in Experiment 1 there was little evidence of adaptation to the shifted condition, in Experiment 2, intelligibility of the shifted condition for easier IHR and BKB sentences did approach that from previous studies (~40%), but only after 10 hours of training. By contrast, Rosen et al. found a similar level of intelligibility for BKB sentences after just three hours of training (Rosen et al., 1999). This difficulty with the shifted speech may have been the result of the sparse output of the sine carriers, sine-vocoding with low-frequency smoothing filter, and the training condition used. While it was anticipated that the sparse sampling of the sine carriers might make this processing more difficult to learn, the latter two effects were unknown at the time the speech processing was designed, but evidence emerging over the course of this research has clarified this difficulty to some extent (Souza and Rosen, 2009). This is discussed in the following section.

6.1.1.1 Vocoding parameters affecting adaptation

The design of the speech processing in these experiments differed from previous studies with basally shifted speech in the monaural case. Most studies demonstrating adaptation after short-term training used noise-vocoding (Faulkner et al., 2006; Faulkner et al., 2006; Rosen et al., 1999), although some have used sine vocoding (Li et al., 2009). In an earlier study Dorman showed that noise and sine vocoding were relatively equivalent in intelligibility for unshifted speech (Dorman et al., 1997b). However, Souza and Rosen (2009) have recently countered that for small numbers of channels (e.g. 3 as in Experiments 1 and 2), sine vocoding is less intelligible than noise vocoding when using low-frequency envelope smoothing filters. Neither of these studies addressed whether the same holds true for basally shifted speech, but it is possible that the processing used here may have been inherently more difficult to learn than the noise vocoding employed previously, since the processing here used sine vocoding with a low frequency (32 Hz) envelope smoothing filter.

At the time the speech processing was designed, this difficulty with sine-vocoding and low-frequency cutoff filters was not known, and if it had been then using longer training times may have been considered from the outset. While the use of more frequency channels may also have facilitated adaptation, a pilot study demonstrated that this would also have led to ceiling effects for the unshifted speech in quiet, so increasing the number of channels was ruled out at an early stage. The speech processing was designed to replicate CI processing as much as possible, and not to be artificially easy to learn, so the motivations for using the processing here remained despite this subsequently discovered difficulty in adaptation. It was necessary to use sine-vocoding in these experiments so that the signals at each ear were spectrally discrete and intelligibility scores could be interpretable. With this requirement in mind, using a higher-frequency envelope smoothing filter would have introduced side bands into the sine carrier signal, so that signals would no longer be composed of a small number of spectral components. These sidebands would have also cued fine structure, and since this information is not available to CI patients, it would not have been appropriate to include here even if it may have made the shifted signal easier to learn. In the end, stronger adaptation to the shifted speech was shown in Experiment 2 over the longer time course of training, and in Experiment 3 with the increased number of channels, so it seems clear that the sparse output of sine carriers and the sine-vocoding used here were not a barrier to adaptation to the shifted speech.

6.1.1.2 Adaptation and training condition

Another parameter of these experiments that had a strong effect on adaptation to the basally shifted speech was the training condition used. Whereas in previous experiments with adaptation to simple monaural shifts, the aim of the training was solely to facilitate adaptation to the shifted speech, it became increasingly apparent over the course of this research that the purpose of the training in these experiments was more complex. In retrospect, there appeared to be three distinct aspects of adaptation: adaptation to the sparsely sampled, basally shifted speech; adaptation to the unshifted speech, which, too, represented a sparse sampling of output channels; and binaural integration of shifted and unshifted channels. In attempting to facilitate adaptation to speech that was binaurally mismatched as here, there was no clear a priori hypothesis about the best way to facilitate both adaptation and integration. Should training be with the binaurally mismatched processor? Do subjects need experience with just the shifted speech? Are there any training order effects with shifted and unshifted speech? Each of these ideas was implemented in training, either alone or in succession, and since binaural integration was not typically shown in these experiments, the question of which training condition best facilitates integration is still unanswered. However, the data from these experiments provided clear evidence about training condition and adaptation to the shifted speech when listening to binaurally mismatched frequency-to-place maps, which also seems consistent with emerging evidence of adaptation to more difficult basalward shift conditions.

In the first of the three experiments, training was provided with either the binaurally mismatched (6DOS) or shifted channels only (3OS) processors, and neither of these training conditions facilitated adaptation to the shifted speech very well. When training with the binaurally mismatched processor alone, subjects appeared to learn by attending solely to the ear with the unshifted speech, and there was little evidence of adaptation specific to the shifted channels. Surprisingly, training with just the shifted channels alone also did not prove very beneficial either, since post-training intelligibility in all conditions was lower than that of the 6DOS group.

However, the vowel tests did show some evidence of learning specific to the shifted speech (3OS) which was not found for the 6DOS group. Taken together, these findings suggest that some experience with the shifted speech on its own is necessary for adaptation to the shifted speech, but for more difficult conditions of basalward shift such as the 3OS condition here, this alone may not be sufficient to provide adaptation, at least over shorter time periods.

In Experiment 2, the training time was doubled and training alternated between both the 6DOS and 3OS conditions, and this had a large, positive effect on adaptation specific to the shifted speech. Post-training intelligibility of the shifted speech was much stronger than that found in Experiment 1, and there were significant improvements for most test materials. The same type of alternating training was provided in Experiment 3, and a similar positive effect on adaptation to the shifted speech was found, though this may also have been facilitated by the increased number of channels and smaller basalward shift. While the idea that mixed exposure may facilitate adaptation to more difficult basally shifted speech conditions was unknown at the time the speech processing was designed, this idea is consistent with a new study on unsupervised adaptation to sine-vocoded, basally shifted speech. Li, Galvin and Fu (2009) showed that unsupervised adaptation to very difficult spectral shift conditions (e.g. an 8 mm basalward shift) is facilitated by mixed exposure to an easier basalward shift condition (4 mm). Only the subjects who were given experience with both degrees of basalward shift showed adaptation to the severely shifted speech. Though their study only considered monaural processing, the findings nevertheless seem consistent with the present investigations, and suggest mixed exposure training may be beneficial for speech training in cochlear implant users.

The findings here on the significance of training condition have important implications for bilateral cochlear implants, where exposure to different spectral shift conditions will presumably be mixed as well as concurrent. The negative results from Experiment 1 strongly imply that when there is a large difference in intelligibility of the signals provided binaurally, providing experience or training with the “better ear” may induce adaptation only to the “better ear”. While training with a very difficult spectral shift condition alone, such as the 3OS condition used here, may also be insufficient to produce adaptation *and* integration, it is clearly necessary at some stage. This is what was suggested by Tyler et al. in their study on patients receiving bilateral cochlear implants sequentially, where a patient reported achieving a benefit from the second implant only after explicit and extensive training with the second implant in the absence of the first (Tyler et al., 2007).

Despite some of the difficulties encountered with adaptation to the shifted speech in these experiments, Experiments 3 and 4 demonstrated novel evidence that listeners can adapt to basally shifted speech in the presence of background noise. This was even more impressive given the sparse spectral representation and lack of temporal modulation cues above 32 Hz.

6.1.1.3 Adaptation and vowel identification

In previous studies the vowel identification task has been used as a measure of adaptation to basally shifted or CI-processed speech. Because vowels can be primarily described by their first and second formant frequencies, changes in their classification over time are easy to map and allow for a measure of whether identification migrates from the spectrally shifted formant frequencies towards the unshifted formant frequencies of the expected target vowel after a period of training. As is highlighted in the right hand panel of Figure 6-2, vowel improvements in all three experiments were small, even though identification significantly improved for all groups for at least one talker. Though the small improvements may seem suggestive of a lack of adaptation to the basalward shift on the one hand, this finding may also reflect the speech training procedure used in these experiments. Subjects were trained with connected discourse texts, which have been consistently shown, as here, to improve the recognition of words in sentences for this type of speech (Faulkner, Rosen, Watt, and Gedgaudaite, 2008; Rosen et al., 1999). The very nature of the training material may thus encourage a remapping of the novel acoustic representation onto existing lexical representations, rather than a phonetic-based acoustic remapping.

It is also possible that lexically-based remapping governs the adaptation to basally shifted speech de facto, regardless of the training materials used. This is what Stacey and Summerfield (2008) concluded when they compared phonetic, sentence and word-based training for improving the recognition of basally shifted (6 mm), noise-vocoded speech. They found similarly small improvements in vowel recognition for all three types of training, but in contrast to their sentence and word-based training, which was effective at improving sentence and word recognition, their phonetic training procedure was ineffective for all speech materials.

Most users of CIs do not undergo extensive speech training, so any adaptation is based on unsupervised connected discourse. To the extent that such experience is comparable to the speech training provided here, this lexical adaptation hypothesis is also consistent with the slow remapping of perceptual vowel spaces seen in some CI users. If improvements seen in word recognition are based on a remapping of lexical representations, then more fine-grained phonetic awareness based on this remapping may be slow to appear. For example, the subjects reported in Svirsky et al.'s (2004) longitudinal study of perceptual vowel spaces took between one month and two years to achieve near-normal target vowel spaces. This is also consistent with the small improvements in vowel identification seen here, since presumably the short-term training was insufficient to initiate fine-grained phonetic awareness of the spectrally shifted signal in this way. Nevertheless, an inability to correctly identify vowels does not seem to preclude moderate levels of speech recognition, so the findings here as in previous studies suggest that training with lexical stimuli is most effective at eliciting adaptation to lexically meaningful speech (Davis et al., 2005).

6.1.2 Resistance to integration

Figure 6-3 shows boxplots summarizing the comparison of the binaurally mismatched processors (6DOS and 10DOS) with the unshifted monaural conditions (3EU and 5EU) for Experiments 1 – 3. The results have been pooled by talker to provide a simplified account of the findings, though there were occasionally talker effects. The findings are also summarized in Table 6-2. Planned contrasts were used to determine whether there was a significant difference (in either direction) between the binaural and monaural conditions, and significant differences between these conditions are marked with an asterisk and *p* value above or below the better condition in the pair.

At each test point in the series of experiments, intelligibility with the 6- or 10-channel binaural unshifted conditions (6DU and 10DU) was always significantly better than that with the 3- or 5-channel monaural unshifted conditions (3EU and 5EU), which indicated that in all instances there was room for improvement from the addition of the odd channels (note that these data are not shown here). If listeners had been able to integrate the shifted and unshifted bands, and the addition of the shifted bands to the unshifted bands did not cause a distraction, then intelligibility with the binaurally mismatched processor should also have exceeded that with the unshifted channels. However, the overarching finding of these comparisons was that intelligibility with the binaurally mismatched conditions either closely mirrored or was

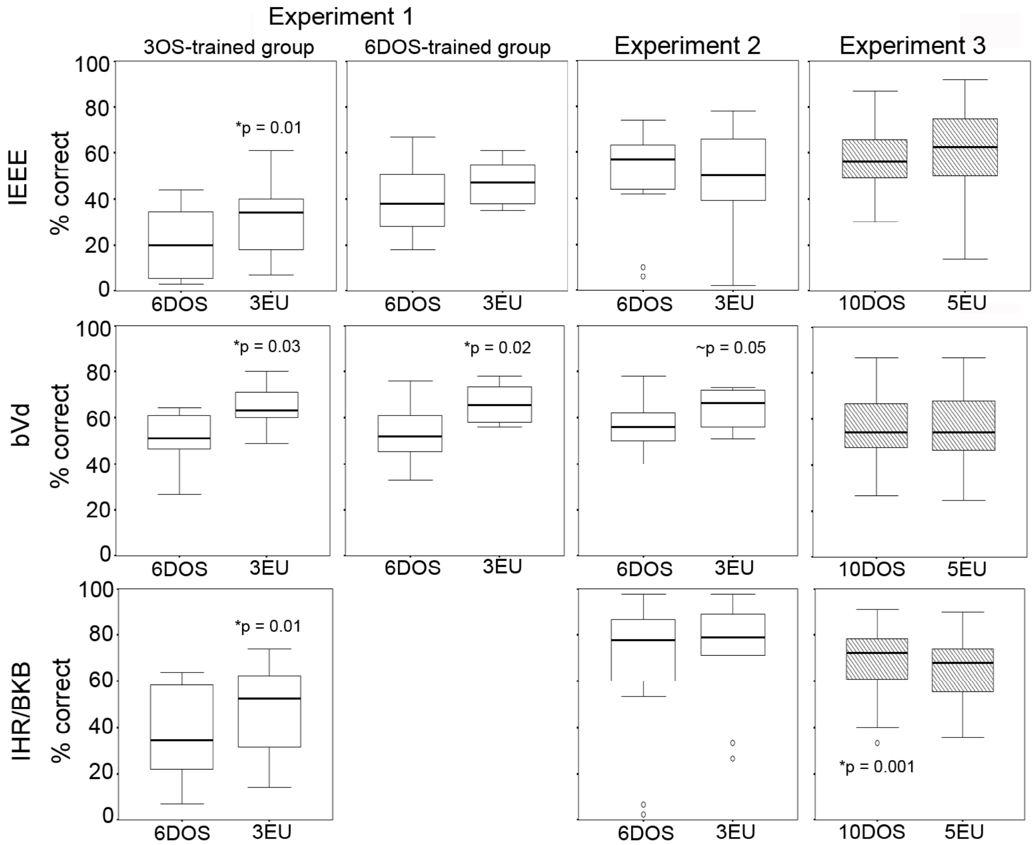

Figure 6-3. Boxplots showing a comparison of post-training intelligibility with binaurally mismatched (6/10DOS) and monaural even unshifted (3/5EU) conditions. The results are pooled by talker. From left to right, panels are arranged in order of experiment and/or training condition. From top to bottom, the results indicate IEEE sentences, vowel identification and easier IHR/BKB sentences. Where a significant advantage for a condition was found, this is indicated with a *p*-value above or below the advantageous condition. Note that this value may differ from that presented in previous chapters, because the results have been pooled over talkers. Results from Experiment 3 have been shaded a different colour to distinguish the 5- and 10-channel conditions and speech in noise from the other experiments.

significantly *worse* than the monaural, even unshifted channel conditions, suggesting that subjects did not learn to binaurally integrate the shifted and unshifted bands.

The only exception to this was for the easier IHR sentences in Experiment 3, where a small (5%) but significant advantage was found for the binaurally mismatched processor.[10] Because this advantage only appeared for the easier sentences and did not appear after the further testing and training in Experiment 4, it seems difficult to conclude that this was a strong or reliable effect. The training times in these experiments were clearly short compared to the experience of users of bilateral cochlear implants, but to the extent that short-term training can be seen as a window into the adaptive processes governing basally shifted speech, then the results

Table 6-2. Summary of post-training intelligibility for the binaurally mismatched (6/10DOS) and even unshifted (3/5EU) speech conditions in Experiments 1 – 3.

experiment	training condition	speech test	% correct binaural 6/10DOS		% correct monaural 3/5EU	
			mean	s. d.	mean	s. d.
1	6DOS	IEEE	40.1	15.6	47.1	9.43
		bVd	52.9	12.2	66.1	8.13
	3OS	IEEE	20.3	15.7	31.5	15.8
		bVd	51.1	11.7	64.6	9.26
		IHR/BKB	37.3	21.1	47.4	19.1
2	6DOS/3OS	IEEE	50.3	21.7	49.8	21.0
		bVd	56.9	10.8	64.0	8.59
		IHR/BKB	66.7	31.4	73.7	22.1
3	10DOS/5OS	IEEE	58.6	15.1	62.1	16.9
		bVd	55.1	13.8	55.3	16.3
		IHR	70.2	14.0	65.2	13.0

[10] The significance value reported in the figure is lower than that reported in the previous chapter. The results here have been pooled by talker for simplification of exposition, even though there was a significant talker with processor interaction when analysing the entire IHR sentence set [$F(3,45)=160$, $p = .019$]. Though the binaural advantage was small (5%), the increased number of data points from the pooled talkers allowed for the detection of the small advantage.

appear to reflect a resistance to integrating frequency-place maps that are mismatched between the ears, even though adaptation to the monaural shifted and unshifted signals at each ear is apparent.

For speech in quiet and for larger spectral shifts, it seems clear that the binaurally mismatched frequency-to-place map was difficult for subjects to integrate, and adaptation was based on learning to attend to the signal provided by the “better ear”. As for reasons described in 6.1.1.1 above, the shifted 3OS signal in particular proved difficult for subjects to learn, though, and may have influenced the outcome of integration. Yet while adaptation to the shifted signal (5OS) was stronger and occurred over a shorter time course in Experiments 3 and 4, there was still little evidence of binaural integration. The findings appear to support a resistance to integration that may be indicative of “better ear” dominance.

As demonstrated in Experiment 3, however, the situation appears to be complicated for speech in noise, where a small advantage provided by the binaurally mismatched processor in the case of the easier IHR sentences seems to conflict with this resistance to integration shown at all other test points. In constructing a theory that accounts for both this resistance and apparent integration seen in the data from the first three experiments, the most consistent explanation for the conflicting results comes from a combination of the “better ear” dominance and best local SNR hypotheses proposed in Chapter 5. Under this account, subjects adapt to both the unshifted and shifted speech signals with training. When listening to the binaurally mismatched processor, they attend to the ear with the signal that is most intelligible. In the case of speech in quiet, this will always be the unshifted ear (e.g. “better ear” dominance). For speech in noise, this will also typically be the unshifted ear unless the SNR at the shifted ear is more favourable. Such an account does not necessarily posit the *integration* of unshifted and shifted signals, but rather a selective and instantaneous attending to the ear with the most intelligible signal. While the experiments in this thesis did not directly address this question of continuous integration versus selective attention, indirect evidence in favour of such an account comes from the fact that the binaural advantage, when shown, was small and usually fell well short of the binaural integration provided by the 6- or 10-channel binaural unshifted conditions (6DU and 10DU). This finding appears more in line with a small and occasional additive advantage of the shifted signal than with a continuous integration of shifted and unshifted speech.

It is possible that this selective attention is only feasible when cognitive demands are reduced, as in the case of the easier, more predictable IHR sentences. The low-predictability IEEE sentences may have placed additional cognitive demands that distracted attentional resources away from attending to the ear with the best SNR, and under such conditions subjects may default to continuously attending to the easier ear, as in the theory of “better ear” dominance. In the case of vowel identification, the higher level of noise would have further increased cognitive demand, thus also making the best local SNR strategy more difficult to put in place. Gallun et al. found that subjects did not always attend to the ear with the best SNR when listening to dichotic stimuli, even when it would be more beneficial to do so, and they also attributed this to task-based demands (Gallun, Mason, and Kidd, Jr., 2007).

The extent to which this instance of binaural advantage was also influenced by the increased number of channels and smaller spectral shift in the 5OS condition of Experiment 3 is unknown. An alternative account of the findings from Experiment 3 may be that the easier spectral shift condition would ultimately lead to strong binaural integration but that the training times explored in these experiments were too short for these effects to appear. However, such an account cannot explain the discrepancy with the more difficult sentence data and the vowel identification, nor the findings from the first two experiments, where intelligibility in the monaural condition often exceeded that in the binaural condition. At a minimum, the data point to a clear *resistance* to integrating binaurally mismatched frequency-to-place maps, consistent with an account of “better ear” dominance, that poses the more important question of whether this type of processing would prove advantageous in real bilateral cochlear implant situations.

6.1.3 Laterality effects

Boxplots showing post-training intelligibility in *dichotic* (6DOS and 10DOS) and *diotic* (6OS and 10OS) combinations of shifted and unshifted bands are summarized in Figure 6-4 and Table 6-3. This comparison was tested in order to determine whether subjects were learning to attend to the unshifted speech by ignoring an ear. Presumably, if this were happening, then intelligibility would decrease when unshifted and shifted channels were presented diotically rather than dichotically. The outcome of this test depended on the training condition and test material. When

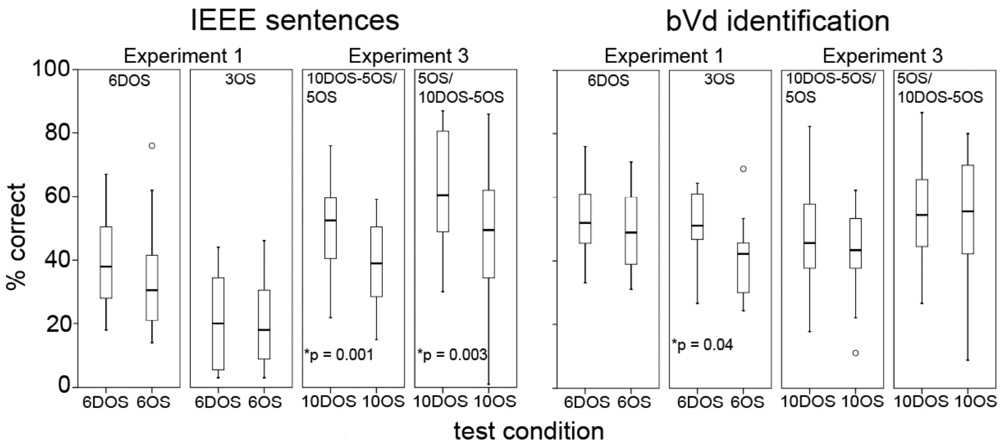

Figure 6-4. Boxplots showing post-training intelligibility in dichotic (DOS) versus diotic (OS) combinations of shifted and unshifted bands in Experiments 1 and 3. The data have been pooled by talker, and any significant advantage for the dichotic condition is denoted underneath with a *p* value. Note that in Experiment 3, the 10OS condition was tested at different time periods for each training group, so these data have been split into groups.

Table 6-3. Summary of post-training intelligibility for dichotic (?DOS) and diotic (?OS) speech conditions in Experiments 1 and 3. Data have been pooled by talker.

experiment	training condition	speech test	% correct dichotic ?DOS		% correct diotic ?OS	
			mean	s. d.	mean	s. d.
1	6DOS	IEEE	40.1	15.6	34.6	18.7
		bVd	52.9	12.2	49.1	12.7
	3OS	IEEE	20.3	15.7	19.7	13.1
		bVd	51.1	11.7	40.7	12.5
3	10DOS/5OS – 5OS (t2)	IEEE	49.5	14.4	37.8	14.2
		bVd	47.5	16.1	42.2	14.1
	5OS – 10DOS/5OS (t3)	IEEE	62.6	17.5	46.1	21.6
		bVd	54.2	16.4	52.9	21.4

subjects were given explicit training time with the shifted speech alone, there appeared to be a significant effect of diotic versus dichotic presentation. For example, the 3OS-trained group in Experiment 1 showed significantly better dichotic versus diotic presentation for the vowels, and in Experiment 3 IEEE sentence intelligibility was significantly better for dichotic than diotic presentation. By contrast, there were no significant differences in dichotic versus diotic presentation for the 6DOS-trained group in Experiment 1, though IEEE sentence intelligibility was somewhat better with dichotic presentation.

The cause of this outcome and this apparent effect of training condition was uncertain, but its implications for bilateral cochlear implants potentially very significant. If subjects were adapting to the binaurally mismatched processor by learning to ignore the worse ear, then this would suggest that the addition of a second cochlear implant could be without purpose if the signal were spectrally mismatched compared to the original implant. Experiment 4 probed this issue further by testing subjects with the headphone orientation reversed, to see if subjects had adapted by learning to ignore the ear with the worse (shifted) signal. For the most part, there was no effect of headphone reversal, although intelligibility did decrease just significantly when headphone orientation was reversed for the vowel test. One subject did appear to adapt to the binaurally mismatched processor by attending to a specific ear, since BKB intelligibility decreased markedly when headphone orientation was reversed. This is shown as the outlier in the left-hand panel of Figure 5-9.

Considering these findings in the context of the previously outlined results and proposed theories, they are also consistent with a combined theory of better-ear dominance and a best local SNR account of the small binaural advantage. While subjects do use cues to ear presentation when listening to the binaurally mismatched processor, the key insight from Experiment 4 is that these cues are not ear-specific. The advantage for dichotic presentation found in Experiments 1 and 3 seems to arise from the *separation* of shifted and unshifted signals, which allows for selective attention to be drawn to the ear providing the more intelligible signal, whether this contains shifted or unshifted speech. For speech in quiet, separation always focuses attention on the unshifted ear, but for speech in noise attention may be drawn to either ear, depending on the local SNR at a given point in an utterance. If the SNR of the unshifted speech is lower than the SNR of the shifted speech, then the listener may momentarily attend to the shifted ear because the signal it is providing is more intelligible. When the ear separation cue is taken away as in diotic presentation, intelligibility decreases, presumably because it is more difficult to access shifted and unshifted channels separately when they are mixed at each ear. This account resolves the apparent discrepancy between the diotic disadvantage for mismatched speech found in Experiments 1 and 3, and the binaural advantage of

10DOS over 5EU found for the easier sentences in Experiment 3. After all, if subjects had learned by ignoring an ear, as was initially hypothesized, then we would not expect to see such a binaural advantage. The binaural advantage for dichotic mismatched speech found in Experiment 3, and the diotic disadvantage found in Experiments 1 and 3, arise from a single mechanism in the proposed theory.

While a sensitivity to ear presentation further supports both the theory of “better ear” dominance and the best local SNR account of the binaural advantage, these findings are inconsistent with an interpretation of continuous binaural integration, which is presumably what underpins intelligibility with the 6- and 10-channel unshifted speech conditions (6DU and 10DU). Though in pilot studies no significant difference was found between dichotic and diotic presentation of the 6-channel unshifted processor, presumably any difference would be in the opposite direction, with dichotic presentation being *less* intelligible than diotic presentation. Further, the binaural advantage seen in the unshifted conditions (6DU) was far greater than the 5% for the odd-shifted condition (10DOS) seen in Experiment 3, which suggests a more robust integration process.

Finally, it is significant to note that, again, the training condition had a strong impact on adaptation. It was initially hypothesized that the lack of sensitivity to dichotic presentation found for the 6DOS-trained group was positive evidence of adaptation, and that the dichotic advantage demonstrated by the 3OS-trained group in Experiment 1 indicated a general lack of adaptation to both the shifted and unshifted signals and was thus a negative finding. Recall that the 6DOS-trained group demonstrated higher intelligibility scores than the 3OS-trained group overall in all but the shifted (3OS) condition. However, results from later experiments seem to indicate the opposite. When no explicit training with the shifted speech was given as in the 6DOS-trained group of Experiment 1, then subjects adapted by attending solely to the unshifted frequencies or ignoring the shifted frequencies. Intelligibility was not affected by dichotic versus diotic presentation, so subjects were equally good at picking out the unshifted frequencies whether or not these were separated by ear. By contrast, subjects given explicit training time with the shifted speech showed sensitivity to dichotic versus diotic presentation. It seems plausible that this very experience with the shifted speech served to cue that the shifted speech was meaningful, which in turn allowed for the eventual slight binaural advantage shown in Experiment 3. If all training had taken place with the combination of shifted and unshifted bands, then subjects may have defaulted to attending to the unshifted speech as in the 6DOS-trained group of Experiment 1.

The processing used in these experiments departs from what may be seen in a clinical situation, and the implications of this are discussed in a forthcoming section. Leaving aside those issues for the moment, the findings lend further support to the importance of explicit training time with the more difficult condition when providing signals that differ greatly in their intelligibility between the ears.

6.2 Better ear dominance hypothesis

The theory that best accounts for the data presented in these experiments is one of “better ear” dominance, where listeners rely on information provided by the “better ear” when they are presented with binaurally mismatched frequency-to-place maps. The theory can be summarized as follows:

1) When listening to binaurally mismatched frequency-to-place maps where the signal at one ear is substantially more intelligible than that at the other, subjects attend to the ear providing the more intelligible signal.

2) If explicit training is provided with the less intelligible (here, basally shifted) signal alone, then subjects will adapt to this signal.

3) Despite this adaptation, attention will still be drawn to the better (i.e. unshifted) ear, unless certain conditions make this signal difficult to attend to and cognitive demands are simultaneously reduced:

4) When listening to speech in noise, spectral properties important for the recognition of speech may be selectively masked more at the better (unshifted) ear than at the worse ear at a given point in time. When this happens, attention may be momentarily drawn to the worse (shifted) ear. This allows for the recognition of information that would have otherwise been masked, and thus in turn a slight advantage over simply attending to the better, unshifted ear.

5) When the worse and better signals are not separated by ear, it is more difficult for subjects to attend to the most intelligible (unshifted) frequencies.

6) Under difficult listening conditions, cognitive demands may be sufficiently high that the better (unshifted) ear will dominate, regardless of a momentary advantage in the signal-to-noise ratio at the shifted (worse) ear.

The best local SNR account for the slight and fleeting binaural advantage for the binaurally mismatched frequency-to-place map can be included in this framework, since it, too, involves selectively attending to the (momentarily) “better ear”. Such a theory also allows for the possibility of a stronger binaural advantage effect after further speech training, but it is clearly distinct from an account of *binaural integration*. The finding of reliance on the unshifted speech was so pervasive in these experiments that it seems to reflect an inherent resistance to such an integration. This was in marked contrast to the case of the dichotically presented, interleaved unshifted speech conditions (6DU and 10DU), which were almost always substantially, and always significantly, more intelligible than their monaural even unshifted components (3EU and 5EU).

While the finding of a slight binaural advantage suggests the possibility that further training may indeed lead to an increase in and/or more consistent binaural advantage under certain conditions, the more pervasive reliance on the “better ear” also indicates that there may be a limit to any binaural advantage that may appear with further experience in the long term. The extent of any binaural advantage is presumably governed by both the degree of long-term adaptation to the shifted speech and the ease with which attention can be drawn away from the ear with the more intelligible, tonotopically matched signal. But, given this distinction between binaural *advantage* arising from binaurally mismatched processors on the one hand, and binaural *integration* arising from binaurally matched frequency-to-place maps on the other, might a better type of speech processing provide binaural integration through aligned frequency-place maps in each ear?

In Chapter 4, the frequency-order hypothesis was proposed to explain why subjects were so resistant to adapting to binaurally mismatched frequency-to-place maps as they were implemented here. Though the further experiments presented in Chapter 5 did not directly test this hypothesis, such an account does provide an explanation for why listeners were resistant to integrating the mismatched frequency-place maps. In particular, the hypothesis can account for this distinction between *integration* when frequency-place maps are aligned, and “better ear” dominance and/or binaural *advantage* when maps are mismatched. When the maps are aligned, integration is facilitated because relative frequency order is preserved. When maps are mismatched as in the processing here, listeners attend to one or the other of the frequency-order-preserved maps (either shifted or unshifted), but not both simultaneously.

In accordance with this distinction, alternative processing strategies that take account of this sensitivity to relative frequency order may prove more beneficial. Consider the speech processing used here and the locations of the output sine carriers as an abstract model for fixed electrode locations of bilateral cochlear implants. Using these same output sine carriers, rather than introducing a large spectral shift to one ear only, the speech could be processed to introduce a more slight but uniform basalward shift to both ears. A schematic representation of this type of processing for the six channel case is shown in Figure 6-5. Note that this fitting approach is based on the *same stimulation places*.

Speech processed in this way would require the listener to adapt to basally shifted speech in both ears, but the shift in each ear would be half of that of the monaural shift in the binaurally mismatched processors, and thus presumably much easier to adapt to. One can imagine that this type of processing would also be much easier for subjects to binaurally integrate for two reasons. First, the processing would eliminate the “better ear” effects seen in the binaurally mismatched processors, since the same degree of basalward shift would be applied to both ears. Secondly, this type of processing would preserve relative frequency ordering: the output channels represent a dichotic version of a slightly upwardly shifted sine-excited vocoder. Whether interleaved, dichotic presentation of spectrally shifted speech can

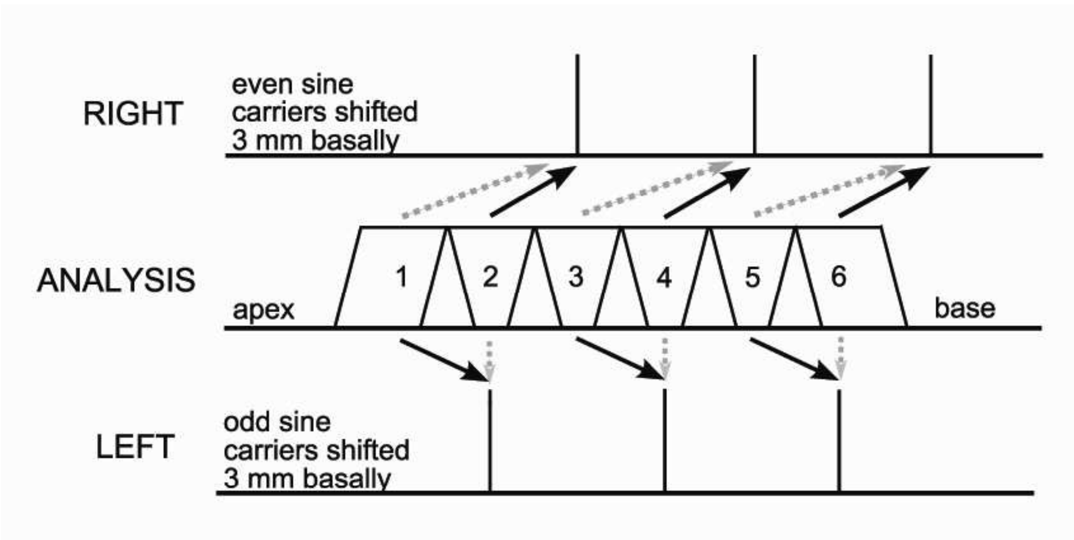

Figure 6-5. Alternative speech processing scheme. The analysis filters and ‘electrode’ (e.g. sine carriers) are in the same positions as for the binaurally mismatched processor in Experiments 1 and 2, but the frequency allocation to these carriers has been modified. The grey dashed arrows indicate the output processing from the experimental processor in Experiments 1 and 2. The black solid arrows indicate the proposed alternative processing. By introducing a smaller basalward shift to both ears, relative frequency order is maintained, and the degree of basalward shift is halved.

be readily adapted to is yet another question for future research, but evidence from Lawson et al.'s experimental zipper processors suggests that cochlear implant users are able to (Lawson et al., 2000). That issue aside, the results from the experiments in this thesis indicate that such an alternative to processing for bilateral cochlear implants would likely produce binaural integration approaching that in the unshifted dichotic conditions (6DU and 10DU), and not simply a binaural advantage.[11]

6.3 Limitations of the findings

The experiments produced consistent and reliable results across a range of conditions showing resistance to integration of binaurally mismatched frequency-to-place maps. However, their implications for bilateral cochlear implant fittings must be interpreted in the context of how well the experimental conditions reflect a clinical cochlear implant situation. Limitations to the present investigations may include a poverty of experience, the extent to which the interleaved processing departs from that with actual bilateral cochlear implants, and the large degree of spectral shift (at least in the first two experiments). However, these caveats are balanced by the substantial benefit of allowing for testing under highly-controlled conditions which can produce more reliable results than when testing real implant users.

6.3.1 Short-term training

While much of the theoretical groundwork on cochlear implants has been conducted in the laboratory using simulations in normally hearing listeners, the poverty of experience in these conditions compared to the omnipresent experience of CI users will always be a limitation of simulation based studies. In addition, normally hearing listeners return to normal, unprocessed speech when not being tested, which may

[11] Consideration was given to training and testing with this type of processing, too, but in the end it was decided that subjects would likely reach ceiling relatively quickly. Moreover, the primary objective of this thesis was to determine whether subjects could adapt to very different frequency-place maps in each ear, rather than in the optimization of frequency-place maps for binaural integration.

also further limit any adaptation to the processed speech. Notwithstanding these shortcomings, many of the findings from CI simulation studies have been approximately replicated in CI patients (Baskent and Shannon, 2004; Baskent and Shannon, 2005).

With this caveat in mind, an indirect method for assessing whether the resistance to integration shown in these experiments is indicative of a cognitive process and not simply to insufficient exposure is to interpret the findings within the context of previous studies of adaptation to basalward shift. Here, adaptation to the shifted speech was comparable to previous studies, but this resistance to integration was somewhat anomalous and may well prove stubborn in the longer term. For example, Rosen et al. (1999) showed adaptation to 4-channel noise-vocoded speech with a 6.5 mm basalward shift in just one hour and twenty minutes, and further studies have shown adaptation in comparably short time periods (Faulkner et al., 2006; Faulkner et al., 2008; Smith and Faulkner, 2006).

Though the processing in the present experiments was markedly different to that typically considered in previous adaptation studies (e.g. sparsely sampled output at each ear, binaurally mismatched, sine-excited vocoding), the training times were comparably longer to compensate for the anticipated increased difficulty. Yet despite the extended course of training in these experiments (five to ten hours), subjects showed very little evidence of a binaural advantage or binaural integration for the binaurally mismatched processing conditions, even though there was substantial evidence of adaptation to the monaural shifted speech.

A further counter-argument against a deficiency in the training time and procedure is the high level of retention of adaptation shown in Experiment 4. For unshifted conditions (5EU and 10DOS), intelligibility remained unchanged in the absence of experience, and for the shifted condition (5OS), forty minutes of further training was sufficient to restore the small but significant decrease in intelligibility that arose from an absence of training. Given this high level of adaptation and retention, it seems that the short-term training procedure used here was relatively efficient at facilitating adaptation to shifted speech. By contrast, in Fu et al.'s study of perceptual adaptation in CI patients (2002), equivalent levels of improvement in unpredictable sentences to those shown in Experiments 3 and 4 (20%) appeared only after three months of passive experience with similarly shifted (2 – 4 mm) experimental CI processors. If adaptation through active speech training shown in the laboratory is more rapid to appear than that from passive experience in actual CI users, as is suggested by this comparison with Fu et al.'s study, then the resistance to integration shown in these experiments may indeed reflect a limitation on learning such processing in the longer term.

Results from normal-hearing children suggest that the normal speech acquisition process can take upwards of seven years to establish (Eisenberg, Shannon, Martinez, Wygonski, and Boothroyd, 2000), and it is possible that the acquisition of novel speech through a cochlear implant may take a similar amount of time to be completed if the signal delivered by the CI is markedly different from the speech representations acquired through normal hearing. For CI users with deeper insertions and substantial nerve survival, adaptation to a CI may be better likened to the acquisition of a second language. In either case, the goal of the restoration of hearing through cochlear implants should be to approximate already-established speech representations as much as possible, since adaptation to markedly different speech processing may be very slow to appear. However, this aim must be balanced with the goal of providing the most useful speech information. For shallowly inserted CI electrode arrays, preserving the natural tonotopic alignment is at the expense of providing low-frequency information important for speech perception. The general consensus is that there is a tradeoff between these two aims, and a frequency-place map that provides some useful low-frequency information but avoids extreme basalward shifts that are difficult to learn is probably the most useful (Baskent and Shannon, 2004; Baskent and Shannon, 2005; Baskent and Shannon, 2007). Any persisting resistance to learning shown in short-term training studies may thus reflect novelties in the speech processing that would prove difficult to learn in the longer term, and if possible should be avoided. It is plausible that the resistance to integrating binaurally mismatched frequency-to-place maps shown here may reflect such a novelty and may ultimately underpin the lack of binaural advantage for bilateral CIs seen in some patients. However, longer-term studies in bilateral CI users are necessary to confirm this hypothesis.

6.3.2 Interleaved speech processing

Perhaps the most significant departure from actual bilateral cochlear implants in these experiments is the interleaved speech processing of the binaurally mismatched speech used in these experiments. In an abstract sense, the sparse sampling of the sine carriers at each ear can be thought of as mimicking sparsely populated nerve survival in a person with bilateral cochlear implant. While the orderly nature of the interleaved channels here is unlikely to arise without the precise determination of pitch ranking of the electrodes in each ear (Lawson et al., 2000), the limited nature of the signal at each ear here had some basis in clinical findings (Long, Deeks, Leow, and Carlyon, 2004). More significantly, in experiments using dichotic presentation of noise-vocoded speech, processing that had been interleaved, as here, was more intelligible than processing that had been split at each ear into low and high frequency bands (Loizou et al., 2003). In context of the resistance to integration shown here, this finding suggests that the interleaving of channels between the ears was not a likely source of difficulty.

Notwithstanding these issues, the motivation for the speech processor design arose primarily from the requirement that the information carried by the signal in each ear was not redundant for the reasons outlined in 4.1. To recap, presenting the same information to each ear but with a unilateral basalward shift could introduce regions where either the same frequency information was delivered to different tonotopic locations in each ear and/or different frequency information was delivered to the same tonotopic location in each ear. Either of these effects could adversely affect speech intelligibility, and would thus mask any improvements or difficulties that were the result of the binaural mismatching.

Yet, typical bilateral cochlear implant systems do deliver roughly the same signal to each ear, but the varying tonotopic locations of this delivery and patterns of nerve survival at each ear are likely to substantially alter the way this information is perceived. A systematic exploration of these additional and potentially confounding effects has not yet been performed in bilateral cochlear implant patients. However, a simulation study that looked at normal-hearing listeners' preferences for interaural frequency-place maps suggested that frequency-place maps that are aligned between the ears are preferable, even if this introduces a large spectral shift to both ears (Fitzgerald, Tan, and Svirsky, 2007). Further evidence from electro-acoustic stimulation patients and simulations of the same in normally hearing listeners suggest that regions of conflict between the ears, where different information is delivered to the same cochlear place, may yield speech intelligibility that is worse than if the conflicting region were omitted (Smith and Faulkner, 2007; Vermeire et al., 2008).

It could be counter-argued that providing the same information to each ear should allow for stronger integration facilitated by enhanced signal redundancy, and the resistance shown in these experiments may reflect a lack of such redundancy cues introduced by the interleaved processing. However, a large frequency-place mismatch between the ears may counteract any benefit from redundancy, if redundant speech cues are delivered to different cochlear places that are no longer compared interaurally. For example, while the restoration of binaural squelch would presumably enhance the detection of speech in noise, binaural squelch is dependent on binaural interaction (Buss, Pillsbury, Buchman, Pillsbury, Clark, Haynes, Labadie, Amberg, Roland, Kruger, Novak, Wirth, Black, Peters, Lake, Wackym, Firszt, Wilson, Lawson, Schatzer, D'Haese, and Barco, 2008), which in turn is thought to be at least somewhat dependent on interaural comparisons within a given frequency region, so mismatches in frequency-place maps between the ears might eliminate any potential benefit, even if redundant information were available. Clinical evidence here is conflicting, with evidence both for (Blanks, Roberts, Buss, Hall, and Fitzpatrick, 2007) and against (Loizou, Hu, Litovsky, Yu, Peters, Lake, and Roland, 2009) the restoration of binaural squelch in bilateral CI users with likely frequency-place mismatches between the ears. Further testing of adaptation to mismatched frequency-to-place maps under more natural conditions is clearly warranted, but the findings in these experiments nevertheless point to a cognitive resistance to integrating such mismatches.

6.3.3 Degree of spectral shift

Evidence emerging over the course of the work presented here has suggested that the Greenwood map that estimates the CFs of CI electrodes is not a realistic model for more modern cochlear implants that place CI electrodes near the modiolus (Sridhar et al., 2006; Stakhovskaya et al., 2007). Here, a spiral ganglion map seems more appropriate which assigns substantially lower CFs to a given electrode contact position than the Greenwood map, especially for more apical electrode locations. If it is true that the Greenwood map exaggerates upward spectral shifts seen in modern CI users, then it is possible that the basalward shifts introduced here overestimated the CFs of the shifted sine carriers, which in turn may have influenced the outcome of the training studies. There is also clinical evidence from electro-acoustic patients suggesting pitch matches at apical electrodes much lower than would be predicted by a Greenwood map, but these findings are tempered by the low reliability of the pitch ranking procedures used to obtain these rankings (Blamey et al., 1996; Reiss et al., 2007; Boex et al. 2006).

The Greenwood map is used frequently in studies of CI simulations, and it was chosen here so that the results could be directly compared to similar previous research. Perhaps the time has come to more critically assess its use in simulation studies of this kind. This issue notwithstanding, the concern about shifted frequency overestimation seems most relevant to the 6 mm shift applied in Experiments 1 and 2, which subjects did find difficult to learn. It could be argued that the drastic mismatch between ears in these experiments is unlikely in a clinical situation, and an overestimation of shifted frequency CFs was the source of resistance to integration. However, this does not negate the finding that such a mismatch is difficult to learn. Furthermore, there was also little evidence of integration per se with the more moderate 3.8 mm spectral shift applied in Experiments 3 and 4, so even if the shifted frequencies were overestimated, the findings still indicate a resistance to integrating frequency-place mismatches between the ears, and such processing may have adverse consequences for bilateral CI fittings.

6.3.4 Directional cues for speech in noise

The binaural advantage of bilateral cochlear implants has been most consistently shown for speech in noise with directional cues, where large differences in the SNR at each ear allow the listener to focus attention on the ear with the best SNR (Wackym et al., 2007; Litovsky et al., 2006; van Hoesel et al., 2002). Presumably, some level of frequency-local redundancy of speech cues must be accessible for listeners to demonstrate a binaural advantage beyond that of the “better ear”, since the cognitive processes underlying such a binaural benefit are thought to involve frequency-sensitive comparisons of the signals at each ear. The interleaved processing used in the present experiments may have eliminated the possibility of binaural squelch and/or summation in this way, but it is still possible that adding directional cues to the noise may have further motivated integration. If the noise presented in Experiments 3 and 4 had instead been processed to sound as if originating from the same side of the head as the unshifted speech, so that the SNR at the unshifted ear was consistently lower across frequencies than that at the shifted ear, then the more favourable SNR at the shifted ear may have changed the way subjects adapted to both the shifted and unshifted speech, and perhaps a stronger binaural advantage may have resulted.

6.4 Implications for bilateral cochlear implants

Notwithstanding the limitations of the research presented in these experiments, the findings have important implications for bilateral cochlear implant fittings. Most current bilateral cochlear implant systems function as two cochlear implants independently, and very little attention is given to aligning frequency-place maps between the ears during implant fittings (Litovsky et al., 2006; Gantz et al., 2002). More importantly, the method for doing so – pitch ranking of interaural electrodes – is particularly unreliable, since it relies on the subjective ability of patients to determine which pitch is higher or lower, a task that many find difficult. In Reiss et al.’s study of pitch ranking in acoustic and electric hearing (Reiss et al., 2007), they also found that subjects’ pitch rankings were particularly unstable over time. The situation is likely to be different for bilateral cochlear implants, since differential adaptation to the signals at both ears is likely, whereas the acoustic signal in electro-acoustic hearing is likely to remain relatively stable (and adaptation thus only occurring monaurally). Frequency-place mappings of bilateral implants are not adjusted to reflect any migration of interaural pitch rankings that may occur with experience, which may complicate the situation further.

The resistance to learning binaurally mismatched frequency-to-place maps shown in these experiments suggests the need for careful consideration of interaural cochlear frequency-place mappings in bilateral cochlear implant fittings. While there was some evidence of a binaural advantage shown here in Experiment 3, this was a very small advantage which did not remain stable over time. These findings are further supported by evidence from simulation studies and electro acoustic hearing that have shown that regions of conflict in the two ears lead to decrements in speech perception compared to when regions of conflict are avoided (Smith and Faulkner, 2007; Vermeire et al., 2008). This growing body of evidence thus calls for the consideration of alternative frequency-place mapping strategies that do not introduce mismatches between the ears. This is not to say that such mismatches could never be learned in the long term, but rather this type of processing proves difficult for listeners to accommodate, whereas alternative processing strategies such as that described in 6.2 could prove more beneficial.

Recent data from a study on speech perception in electro-acoustic hearing in newly implanted patients may or may not be consistent with this contention. Simpson et al. (2009) provided CI+HA stimulation to newly implanted patients with either a conventional CI map, or a map that had been "pitch-matched" to exclude low-frequency information provided through acoustic hearing. Newly-implanted patients wore the pitch-matched map for twelve weeks, then the conventional map for twelve weeks, and then returned to the pitch-matched map for a further four weeks. Interestingly, listeners showed a binaural benefit from the CI+HA stimulation whether there was a region of conflict or not. Although CI+HA intelligibility with the pitch-matched, non-conflicting map at 28 weeks did exceed that with the CI+HA conventional map at 24 weeks by about 10%, this difference was not significant. The within-subjects design makes the findings somewhat difficult to interpret, since comparisons between the conditions could not be made for the same point in time. For example, it is not clear whether the increased performance with the non-conflicting map at 28 weeks was the result of four more weeks' experience with the implant (regardless of the map), or whether the lack of conflict made the map genuinely more intelligible. Further research in this regard is warranted, but it is possible that this advantage for the non-conflicting map would become more pronounced in the long term. Moreover, it is possible that regions of conflict may be better tolerated when the stimulating signals at each ear are very different, as in electro-acoustic hearing.

Evidence from psychophysics also suggests that interaural frequency-place mapping is important for restoring binaural hearing phenomena in bilateral implants. Current cochlear implant systems do not contribute the fine structure information thought to underpin many of these binaural hearing phenomena in normal hearing listeners. However, Long and colleagues showed at least one bilateral cochlear implantee to be sensitive to the interaural time difference (ITD) and interaural level difference (ILD) cues provided solely through envelope information. Crucially, though, this was limited to cases where interaural electrode pairs were matched in frequency (Long et al., 2003). Grose and Buss (2007) showed bilateral CI listeners were also sensitive to mismatches in interaural cochlear places in a gap detection task. Normal hearing listeners also demonstrate best ITD (Nuetzel and Hafter, 1981; Blanks et al., 2008) and ILD (Francart and Wouters, 2007) detection for similar interaural cochlear place regions. The sensitivity to interaural cochlear place does not appear to be as strong in CI users as in normally hearing listeners (Blanks et al., 2008), but this limit is likely due to the very broad place coding and gross envelope signal provided by the implant compared to the fine structure-rich, frequency-tuned signal provided in normal hearing. As CI systems are improved to include more temporal fine structure information, it is likely that this sensitivity to interaural cochlear place will become more pronounced in CI listeners as well, which may further indicate the need for matched interaural frequency-place maps.

Though perhaps through a different underlying mechanism than has been considered here, then, evidence from psychophysics is consistent with the present research. The converging evidence indicates that frequency-place maps should be kept similar in the two ears in order to optimize bilateral cochlear implant devices. To the author's knowledge, no systematic study of bilateral CI users not achieving benefit from an additional implant has been conducted. It would be interesting to explore whether mismatches in frequency-place maps between the ears underpins this lack of benefit. The findings from the present investigations indicate this possibility.

6.5 Future research

Bilateral cochlear implantation has met with many successes, with many people who receive them achieving a binaural benefit for speech compared to the use of a single implant (Litovsky et al., 2006; Offeciers et al., 2005; Tyler et al., 2005; Tyler et al., 2007; Wackym et al., 2007). They have been particularly useful in restoring the ability to recognise speech in noise and the ability to localize sounds in space. While there have been many clear successes, not all people receiving bilateral implants achieve a binaural advantage beyond that of their best CI ear. Much more research needs to be carried out to determine why this might be the case, but the results from the present investigations indicate that a mismatch in frequency-place maps between the ears may underpin this lack of benefit. There is clear room to extend the present investigations to include bilateral cochlear implant patients.

However, a major obstacle to conducting similar research in the clinical population is the lack of a reliable method for determining interaural cochlear frequency-to-place maps. For example, an alternative method to processing that would avoid frequency-place mismatches between the ears was proposed in 6.2, but this type of processing would require a precise determination of the perceived CF at each electrode contact in order to be implemented. The unreliability of the currently used pitch-ranking procedure is highlighted in Reiss et al.'s (2007) study of electro-acoustic patients, where subjects demonstrated huge (up to 2 octaves) variability in their pitch rankings over time that were unlikely to be accounted for by perceptual adaptation alone. Considering the additional significance of aligning frequency-place maps between the ears in restoring binaural hearing phenomena, it seems clear that more robust methods for matching frequency-place maps between the ears for humans need to be developed.

Earlier studies with bilateral CI patients, also acknowledging the importance of interaural matching of frequency-place maps, looked at alternative subjective procedures to the more common interaural pitch ranking task (Eddington, Tierney, Noel, Herrmann, Whearty, and Finley, 2002). One such technique was to determine the frequency of a signal in each ear that elicited the perception of a fused signal. However, sequentially implanted bilateral CI users found this task difficult. The extended experience in adapting to the initial CI meant that the timbres of the signals from each of the implants were difficult to fuse. In the end, new research into more objective techniques, such as the binaural interaction component of the evoked auditory brainstem response (EABR), as described by Smith and Delgutte (2007), could prove particularly promising. The technique has the benefit of being highly reliable, but as it is currently implemented is too lengthy for clinical utility.

In the absence of novel and more reliable procedures for estimating frequency-to-place maps in bilateral CI users, alternative methods for frequency-place mapping based on currently available techniques may also prove interesting. For example, the finding that conflicting sites of stimulation has a negative impact on speech perception is proving to be consistent. Alternative strategies for mapping between the ears that could avoid such conflicts might stimulate one ear with lower frequencies and one ear with higher frequencies. It has already been demonstrated that normal hearing listeners can adapt to spectral information that has been warped around a "hole" (Smith and Faulkner, 2006), so as long as this type of mapping would not introduce a region of conflict in the two ears, it may prove more beneficial than providing the same frequency information to both ears that could be both mismatched and conflicting.

The research on conflicts of stimulation at each ear has been weighted towards delivering different frequency information to the same cochlear place. While interesting, the finding may be simply the result of energetic masking. The opposite effect of delivering the same frequency information to different cochlear places also merits exploration, and may provide more insight into how frequency-place mismatches are accommodated in the absence of simple masking effects. This research would more easily be conducted using simulations in normally hearing listeners.

Finally, though no substantial amount of binaural integration was shown in these experiments, the results nevertheless highlighted the important role of the training condition in facilitating adaptation to both spectrally shifted and unshifted speech. Specifically, the findings in the present research suggest that under circumstances where there is a large difference in intelligibility of the signals at each ear (cf. Experiments 1 and 2), specific training time with the worse ear is essential to prevent adaptation based on ignoring the information provided by the worse ear. These are the first studies to explore this issue when presenting shifted and unshifted speech simultaneously, and the findings indicate a clear role for targeted speech training with the more difficult ear on its own. Research needs to explore this effect in users of bilateral cochlear implants. Anecdotal evidence reported in Tyler et al.'s study (2007) has already called for similar training, but widespread uptake in the clinical population is not currently evident.

6.6 Conclusion

Despite undergoing significantly longer training periods than in previous studies with monaural spectral shifts, subjects in the present investigations never learned to integrate the binaurally mismatched frequency-place maps, and almost always relied on information from the better, unshifted ear when present. Post-training performance with the binaurally mismatched processor typically mirrored that with the monaural unshifted channels, even after ten hours of training. While there was some evidence of a binaural advantage for the binaurally mismatched processor for speech in noise, this was a very small advantage that did not persist with further training and testing. The hypothesis based on the findings from these experiments is one of "better ear" dominance, where listeners rely on information from the "better ear" when listening to mismatched frequency-to-place maps. Included in this hypothesis is the occasionally employed strategy of best local SNR for speech in noise, where listeners will rely on the 'worse' ear (in terms of spectral shift) if it momentarily provides a better SNR, provided that all other cognitive demands are sufficiently low. Through this latter strategy, listeners may occasionally demonstrate a binaural advantage for speech in noise, but this falls short of the binaural integration shown when relative frequency order is preserved, as in the dichotic unshifted conditions. It is this preference for frequency-place maps that maintain relative frequency order that is hypothesized to underpin the resistance to adaptation to the mismatched maps shown in these experiments.

Limitations on the applicability of these findings primarily lie in the short-term training used and the abstract nature of the speech processing. To the extent that these parameters reflect the situation in a bilateral cochlear implant setting, the findings suggest that mismatches in frequency-place maps between the ears are difficult to accommodate and should be avoided in cochlear implant fittings. This is especially true for cases of drastic mismatch between the ears, such as might arise when one of two ears has a shallowly inserted electrode. Although this outcome contrasts somewhat with earlier findings for monaural spectral shifts, where adaptation has been shown in comparably short time periods, the results are consistent with psychophysical studies in bilateral cochlear implant patients, which have also called for matching interaural electrode pairs in order to restore ITD and ILD sensitivity. To this end, more robust techniques for aligning frequency-place maps between the ears need to be developed.

References

Advanced Bionics (2008). How a cochlear implant works. http://www.bionicear.com/Your_Journey_to_Hearing/Adults/Learning/How_a_CochlearImplant_Works.cfm?langid=1 [On-line].

Advanced Bionics (2004). Hearing with two ears: Technical advances for bilateral cochlear implantation. *Advanced Bionics Corporation*.

Assmann, P. F. and Nearey, T. M. (2003). Frequency shifts and vowel identification. In *Proceedings of the 15th International Congress of the Phonetic Sciences*. Barcelona, Spain.

Azadpour, M. (2008). *Perceptual learning in speech processing: Comprehending spectrally-rotated speech.* Unpublished PhD Dissertation, University of Trieste.

Baskent, D. and Shannon, R. V. (2007). Combined effects of frequency compression-expansion and shift on speech recognition. *Ear and Hearing, 28,* 277-289.

Baskent, D. and Shannon, R. V. (2005). Interactions between cochlear implant electrode insertion depth and frequency-place mapping. *Journal of the Acoustical Society of America, 117,* 1405-1416.

Baskent, D. and Shannon, R. V. (2004). Frequency-place compression and expansion in cochlear implant listeners. *Journal of the Acoustical Society of America, 116,* 3130-3140.

Beaman, C. P., Bridges, A. M., and Scott, S. (2007). From dichotic listening to the irrelevant sound effect: A behavioural and neuroimaging analysis of the processing of unattended speech. *Cortex, 43,* 124-134.

Bench, J., Kowal, A., and Bamford, J. M. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially hearing children. *British Journal of Audiology, 13,* 108-112.

Blamey, P. J., Dooley, G. J., Parisi, E. S., and Clark, G. M. (1996). Pitch comparisons of acoustically and electrically evoked auditory sensations. *Hearing Research, 99,* 139-150.

Blanks, D. A., Buss, E., Grose, J. H., Fitzpatrick, D. C., and Hall, J. W. (2008). Interaural time discrimination of envelopes carried on high-frequency tones as a function of level and interaural carrier mismatch. *Ear and Hearing, 29,* 674-683.

Blesser, B. (1972). Speech perception under conditions of spectral transformation I. Phonetic characteristics. *Journal of Speech and Hearing Research, 15,* 5-41.

Boex, C., Baud, L., Cosendai, G., Sigrist, A., Kos, M.-I., and Pelizzone, M. (2006). Acoustic to electric pitch comparisons in cochlear implant subjects with residual hearing. *Journal of the Association for Research in Otolaryngology, 7,* 110-124.

Broadbent, D. E. and Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America, 29,* 708-710.

Buss, E., Pillsbury, H. C., Buchman, C. A., Pillsbury, C. A., Clark, M., Haynes, D. S., Labadie, R. F., Amberg, S., Roland, P. S., Kruger, P., Novak, M. A., Wirth, J. A., Black, J. M., Peters, R., Lake, J., Wackym, P. A., Firszt, J. B., Wilson, B., Lawson, D., Schatzer, R., D'Haese, P. S. C., and Barco, A. (2008). Multicenter U.S. bilateral MED-EL cochlear implantation study: Speech perception over the first year of use. *Ear and Hearing, 29,* 20-32.

Ching, T. Y. C., Incerti, P., Hill, M., and van Wanrooy, E. (2006). An overview of binaural advantages for children and adults who use binaural/bimodal hearing devices. *Audiology and Neurotology, 11,* 6-11.

Ching, T. Y. C., Psarros, C., Hill, M., Dillon, H., and Incerti, P. (2005). Should children who use cochlear implants wear hearing aids in the opposite ear? *Ear and Hearing, 22,* 365-380.

Ching, T. Y. C. (2005). The evidence calls for making binaural-bimodal fittings routine. *The Hearing Journal, 58,* 32-41.

Ching, T. Y. C., Incerti, P., and Hill, M. (2004). Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear and Hearing, 25,* 9-21.

Clarke, C. M. and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116,* 3647-3658.

Colburn, H. S. and Durlach, N. I. (1978). Models of Binaural Interaction. In E.C.Caterette and M. P. Friedman (Eds.), *Handbook of Perception* (San Diego, CA: Academic).

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83,* 114-140.

Davis, M. H., Johnsrude, I. S., and Hervais-Adelman, A. Taylor, K. and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134,* 222-241.

De Filippo, C. L. and Scott, B. L. (1978). A method for training and evaluating the reception of ongoing speech. *Journal of the Acoustical Society of America, 63,* 1186-1192.

Dinse, H. R., Godde, B., Reuter, G., Cords, S. M., and Hilger, T. (2003). Auditory cortical plasticity under operation: Reorganization of auditory cortex induced by electric cochlear stimulation reveals adaptation to altered sensory input statistics. *Speech Communication, 41,* 201-219.

Dorman, M. F., Spahr, A. J., Gifford, R., Loiselle, L., McKarns, S., Holden, T., Skinner, M., and Finley, C. C. (2007). An electric frequency-to-place map for a cochlear implant patient with hearing in the nonimplanted ear. *Journal of the Association for Research in Otolaryngology, 8,* 234-240.

Dorman, M. F. and Dahlstrom, L. (2004). Speech understanding by cochlear-implant patients with different left- and right-ear electrode arrays. *Ear and Hearing, 25,* 191-194.

Dorman, M. F., Loizou, P. C., and Rainey, D. (1997a). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *Journal of the Acoustical Society of America, 102,* 2993-2996.

Dorman, M. F., Loizou, P. C., and Rainey, D. (1997b). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America, 102,* 2403-2411.

Doucet, M. E., Bergeron, F., Lassonde, M., Ferron, P., and Lepore, F. (2006). Cross-modal reorganization and speech perception in cochlear implant users. *Brain, 129,* 3376-3383.

Dunn, C. C. and Ou, H. (2008). New methodologies for determining if two cochlear implants are necessary. Audiology Online [On-line].

Dunn, C. C., Tyler, R. S., Oakley, S., Gantz, B. J., and Noble, W. G. (2008). Comparison of speech recognition and localization performance in bilateral and unilateral cochlear implant users matched on duration of deafness and age at implantation. *Ear and Hearing, 29(3),* 352-359.

Durlach, N. I. and Colburn, H. S. (1978). Binaural Phenomena. In E.C.Caterette and M. P. Friedman (Eds.), *Handbook of Perception*. San Diego, CA: Academic.

Eddington, D. K., Tierney, J., Noel, V., Herrmann, B., Whearty, M., and Finley, C. C. (2002). *Speech processors for auditory prostheses: Third quarterly progress report*.

Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, M. G., and Parkin, J. L. (1978). Auditory prosthesis research with multiple channel intracochlear stimulation in man. *Annals of Otology, Rhinology and Laryngology, 87,* 1-39.

Eisenberg, L., Shannon, R. V., Martinez, A. S., Wygonski, J., and Boothroyd, A. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America, 107,* 2704-2710.

Evans, B. G. and Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America, 115,* 352-361.

Fant, G. (1976). Nonuniform normalization of formant data. *Journal of the Acoustical Society of America, 59*(S1), S71.

Faulkner, A., Rosen, S., Watt, C., and Gedgaudaite, K. (2008). Comparison of live-voice and computer-based training for upward-shifted vocoded speech simulating a cochlear implant. In *British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness*.

Faulkner, A., Rosen, S., and Norman, C. (2006). The right information may matter more than frequency-place alignment: Simulations of frequency-aligned and upward shifting cochlear implant processors for a shallow electrode array insertion. *Ear and Hearing, 27,* 139-152.

Faulkner, A. (2006b). Adaptation to distorted frequency-to-place maps: Implications of simulations in normal listeners for cochlear implants and electroacoustic stimulation. *Audiology and Neurotology, 11,* 21-26.

Faulkner, A., Rosen, S., and Stanton, D. (2003). Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth. *Journal of the Acoustical Society of America, 113,* 1073-1080.

Faulkner, A., Rosen, S., and Wilkinson, L. (2001). Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant processor. *Ear and Hearing, 22,* 431-438.

Fishman, K. E., Shannon, R. V., and Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *Journal of Speech, Language and Hearing Research, 40,* 1201-1215.

Fitzgerald, M. B., Tan, C.-T., and Svirsky, M. A. Customized selection of frequency maps in an acoustic simulation of a cochlear implant. In *Conference on Implantable Auditory Prostheses*. Tahoe City, CA, 2007.

Francart, T. and Wouters, J. (2007). Perception of across-frequency interaural level differences. *Journal of the Acoustical Society of America, 122,* 2826-2831.

French, N. R. and Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America, 19,* 90-119.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America, 110,* 1150-1163.

Fu, Q. J. and Galvin III, J. J. (2007). Computer-assisted speech training for cochlear implant patients: Feasibility, outcomes, and future directions. *Seminars in Hearing, 28,* 142-150.

Fu, Q.-J. and Galvin III, J. J. (2007a). Perceptual learning and auditory training in cochlear implant recipients. *Trends in Amplification, 11,* 193-205.

Fu, Q. J. and Galvin III, J. J. (2007b). Maximizing cochlear implant patients' performance with advanced speech training procedures. *Hearing Research, 242,* 198-208.

Fu, Q.-J., Galvin III, J. J., Wang, X., and Nogaki, G. (2005a). Moderate auditory training can improve speech performance of adult cochlear implant patients. *Acoustics Research Letters Online, 6,* 106-111.

Fu, Q.-J., Nogaki, G., and Galvin III, J. J. (2005b). Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation. *Journal of the Association for Research in Otolaryngology, 6,* 180-189.

Fu, Q.-J. and Galvin III, J. J. (2003). The effects of short-term training for spectrally mismatched noise-band speech. *Journal of the Acoustical Society of America, 113,* 1065-1072.

Fu, Q.-J., Shannon, R. V., and Galvin III, J. J. (2002). Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. *Journal of the Acoustical Society of America, 112,* 1664-1674.

Fu, Q.-J. and Shannon, R. V. (1999). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *Journal of the Acoustical Society of America, 105,* 1889-1900.

Furst, M., Bresloff, I., Levine, R. A., Merlob, P. L., and Attias, J. J. (2004). Interaural time coincidence detectors are present at birth: Evidence from binaural interaction. *Hearing Research, 187,* 63-72.

Gallun, F. J., Mason, C. R., and Kidd, Jr., G. (2007). The ability to listen with independent ears. *Journal of the Acoustical Society of America, 122,* 2814-2825.

Gantz, B. J., Tyler, R. S., Rubinstein, J. T., Wolaver, A., Lowder, M. W., Abbas, P. J., Brown, C., Hughes, M. L., and Preece, J. P. (2002). Binaural cochlear implants placed during the same operation. *Otology and Neurotology, 23,* 169-180.

Grant, K. W., Braida, L. D., and Renn, R. J. (1994). Auditory supplements to speechreading: Combining amplitude envelope cues from different spectral regions of speech. *Journal of the Acoustical Society of America, 95,* 1065-1073.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species -- 29 years later. *Journal of the Acoustical Society of America, 87,* 2592-2605.

Grose, J. H. and Buss, E. (2007). Within- and across-channel gap detection in cochlear implant listeners. *Journal of the Acoustical Society of America, 122,* 3651-3658.

Hamzavi, J., Pok, S. M., Gstoettner, W., and Baumgartner, W.-D. (2004). Speech perception with a cochlear implant used in conjunction with a hearing aid in the opposite ear. *International Journal of Audiology, 43,* 61-65.

Harnsberger, J. D., Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., and Meyer, T. (2001). Perceptual "vowel spaces" of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. *Journal of the Acoustical Society of America, 109,* 2135-2145.

Hofman, P. M., van Riswick, J. G. A., and van Opstral, A. J. (1998). Relearning sound localization with new ears. *Nature Neuroscience, 1,* 417-421.

Iverson, P., Smith, C. A., and Evans, B. G. (2006). Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement and duration. *Journal of the Acoustical Society of America, 120,* 3998-4006.

Iwaki, T., Matsushiro, N., Mah, S.-R., Sato, T., Yasuoka, E., Yamamoto, K.-I., and Kubo, T. (2004). Comparison of speech perception between monaural and binaural hearing in cochlear implant patients. *Acta Otolaryngologica, 124,* 358-362.

James, C., Blamey, P. J., Shallop, J., and Incerti, P. (2001). Contralateral masking in cochlear implant users with residual hearing in the non-implanted ear. *Audiology and Neuro-Otology, 6,* 87-97.

Kent, R. D. and Read, C. (1999). *The Acoustic Analysis of Speech*. San Diego, CA: Singular.

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., and Neely, J. G. (1998). In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays. *Annals of Otology, Rhinology and Laryngology, 107,* 1-16.

Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. *Canadian Journal of Psychology, 15,* 166-171.

Kuhn-Inacker, H., Shehata-Dieter, W., Muller, J., and Helms, J. (2004). Bilateral cochlear implants: A way to optimize auditory perception abilities in deaf children? *International Journal of Pediatric Otorhinolaryngology, 68,* 1257-1266.

Lawson, D., Brill, S., Wolford, R., Wilson, B., and Schatzer, R. (2000). *Speech Processors for Auditory Prostheses: Ninth quarterly progress report*.

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

Li, N. and Loizou, P. C. (2008). A glimpsing account for the benefit of simulated combined acoustic and electric hearing. *Journal of the Acoustical Society of America, 123,* 2287-2294.

Li, T., Galvin III, J. J., and Fu, Q.-J. (2009). Interactions between unsupervised learning and the degree of spectral mismatch on short-term perceptual adaptation to spectrally shifted speech. *Ear and Hearing, 30,* 238-249.

Li, T. and Fu, Q.-J. (2007). Perceptual adaptation to spectrally shifted vowels: Training with nonlexical labels. *Journal of the Association for Research in Otolaryngology, 8,* 32-41.

Libby, E. R. (1981). *Binaural Hearing and Amplification*. (vols. 1) Chicago: Zenetron.

Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. (2006). Simultaneous bilateral cochlear implantation in adults: a multicenter clinical study. *Ear and Hearing, 27,* 714-731.

Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., and Roland, P. S. (2009). Speech recognition by bilateral cochlear implant users in a cocktail party setting. *Journal of the Acoustical Society of America, 125,* 372-383.

Loizou, P. C., Mani, A., and Dorman, M. F. (2003). Dichotic speech recognition in noise using reduced spectral cues. *Journal of the Acoustical Society of America, 114,* 475-483.

Loizou, P. C., Dorman, M. F., and Tu, Z. (1999). On the number of channels needed to understand speech. *Journal of the Acoustical Society of America, 106,* 2097-2103.

Loizou, P. C. (1998). An overview of signal processing strategies for converting sound into electrical signals in cochlear implants. *IEEE Signal Processing Magazine,* 101-130.

Long, C. J., Deeks, J. M., Leow, M.-C., and Carlyon, R. P. (2004). Simulations of bilateral cochlear implant speech reception in noise. In *Association for Research in Otolaryngology Midwinter Meeting*. Daytona Beach, Florida.

Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (2003). Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user. *Journal of the Acoustical Society of America, 114,* 1565-1574.

MacLeod, A. and Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation and recommendations for use. *British Journal of Audiology, 24,* 29-43.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

Merzenich, M. M. and Reid, M. D. (1974). Representation of the cochlea within the inferior colliculus of the cat. *Brain Research,* 397-415.

Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology, 9,* 399-406.

Moore, B. C. J. (2003). *The Psychology of Hearing*. (5 ed.) San Diego, CA: Academic.

Moore, B. C. J. (1998). Physiological Aspects of Cochlear Hearing Loss. In *Cochlear Hearing Loss* (pp. 1-39). London: Whurr.

Mosnier, I., Sterkers, O., Bebear, J. P., Godey, B., Robier, A., Deguine, O., Fraysse, B., Bordure, P., Mondain, M., Bouccara, D., Bozorg-Grayeli, A., Borel, S., mbert-Dahan, E., and Ferrary, E. (2009). Speech Performance and Sound Localization in a Complex Noisy Environment in Bilaterally Implanted Adult Patients. *Audiology and Neuro-Otology, 14,* 106-114.

Muller, J., Schon, F., and Helms, J. (2005). Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system. *Ear and Hearing, 23,* 198-206.

Nuetzel, J. M. and Hafter, E. R. (1981). Discrimination of interaural delays in complex waveforms: Spectral effects. *Journal of the Acoustical Society of America, 69,* 1112-1118.

Offeciers, E., Morera, C., Muller, J., Huarte, A., Shallop, J., and Cavalle, L. (2005). International consensus on bilateral cochlear implants and bimodal stimulation. *Acta Otolaryngologica, 125,* 918-919.

Potter, R. and Steinberg, J. C. (1950). Toward the specification of speech. *Journal of the Acoustical Society of America, 22,* 807-820.

Reiss, L. A. J., Turner, C. W., Erenberg, S. R., and Gantz, B. J. (2007). Changes in pitch with a cochlear implant over time. *Journal of the Association for Research in Otolaryngology, 8,* 241-257.

Remez, R. E., Rubin, P. E., Pisoni, D., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947-949.

Robinson, D. and Dadson, R. S. (1956). A redetermination of the equal-loudness relations for pure tones. *British Journal of Applied Physics, 7,* 166-181.

Rosen, S., Faulkner, A., and Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 106,* 3629-3636.

Rothauser, E. H., Chapman, N. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17,* 227-246.

Rubin, B. A., Uchanski, R. M., and Braida, L. D. (1992). Integration of acoustic cues for consonants across frequency bands (A). *Journal of the Acoustical Society of America, 91,* 2473-2474.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270,* 303-304.

Simpson, A., McDermott, H. J., Dowell, R. C., Sucher, C., and Briggs, R. J. S. (2009). Comparison of two frequency-to-electrode maps for acoustic-electric stimulation. *International Journal of Audiology, 48,* 63-73.

Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., Neely, J. G., Kletzker, G. R., Brunsden, B., and Blocker, B. (2002). CT-Derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients. *Journal of the Association for Research in Otolaryngology, 3,* 332-350.

Smith, M. W. and Faulkner, A. (2007). Adaptation to conflicting frequency-to-place maps in simulations of bimodal listening. In *Conference on Implantable Auditory Prostheses*. Tahoe City, CA.

Smith, M. W. and Faulkner, A. (2006). Perceptual adaptation by normally hearing listeners to a simulated "hole" in hearing. *Journal of the Acoustical Society of America, 120,* 4019-4030.

Smith, Z. M. and Delgutte, B. (2007). Using evoked potentials to match interaural electrode pairs with bilateral cochlear implants. *Journal of the Association for Research in Otolaryngology, 8,* 134-151.

Souza, P. and Rosen, S. (2009). Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech. *Journal of the Acoustical Society of America, 126,* 792 - 805.

Sridhar, D., Stakhovskaya, O., and Leake, P. (2006). A frequency-position function for the human cochlear spiral ganglion. *Audiology and Neurotology, 11,* 16-20.

Stacey, P., Raine, C., O'Donoghue, G. M., Tapper, L., Twomey, T., and Summerfield, A. Q. (*Submitted*). Effectiveness of computer-based auditory training for adult users of cochlear implants. *International Journal of Audiology*.

Stacey, P. and Summerfield, Q. (2008). Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally distorted speech. *Journal of Speech, Language and Hearing Research, 51,* 526-538.

Stacey, P. and Summerfield.Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *Journal of the Acoustical Society of America, 121,* 2923-2935.

Stakhovskaya, O., Sridhar, D., Bonham, B., and Leake, P. (2007). Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants. *Journal of the Association for Research in Otolaryngology, 8,* 220-223.

Stern, R. M., Jr. and Trahiotis, C. (1995). Models of binaural interaction. In B.C.J. Moore (Ed.), *Handbook of Perception and Cognition* (pp. 347-386). New York: Academic Press.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research, 28,* 455-462.

Stuhlman, Jr., O. (1943). *An Introduction to Biophysics*. New York: John Wiley and Sons, Inc.

Svirsky, M. A., Silveira, A., Neuburger, H., Teoh, S.-W., Helms, J., and Suarez, H. (2004). Long-term auditory adaptation to a modified peripheral frequency map. *Acta Otolaryngologica, 124,* 381-386.

Svirsky, M. A., Silveira, A., Suarez, H., Neuburger, H., Lai, L. T., and Simmons, P. M. (2001). Auditory learning and adaptation after cochlear implantation: A preliminary study of discrimination and labeling of vowel sounds by cochlear implant users. *Acta Otolaryngologica, 121,* 262-265.

Tonndorf, J. (1962). Time/frequency analysis along the partition of cochlear models: A modified place code. *Journal of the Acoustical Society of America, 34,* 1337-1350.

Tyler, R. S., Dunn, C. C., Witt, S. A., and Noble, W. G. (2007). Speech perception and localization with adults with bilateral sequential cochlear implants. *Ear and Hearing, 28,* 86S-90S.

Tyler, R. S., Gantz, B. J., Rubinstein, J. T., Wilson, B. S., Parkinson, A. J., Wolaver, A., Preece, J. P., Witt, S. A., and Lowder, M. W. (2005). Three-month results with bilateral cochlear implants. *Ear and Hearing, 23,* 80S-89S.

Tyler, R. S. and Summerfield, Q. (1996). Cochlear implantation: Relationships with research on auditory deprivation and acclimatization. *Ear and Hearing, 17,* 38S-50S.

van Hoesel, R., Ramsden, R., and O'Driscoll, M. (2002). Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user. *Ear and Hearing, 23,* 137-149.

Vermeire, K., Anderson, I., Flynn, M., and Van de Heyning, P. (2008). The influence of different speech processor and hearing aid settings on speech perception outcomes in electric acoustic stimulation patients. *Ear and Hearing, 29,* 76-86.

Verschuur, C. A., Lutman, M. E., Ramsden, R., Greenham, P., and O'Driscoll, M. (2005). Auditory localization abilities in bilateral cochlear implant recipients. *Otology and Neurotology, 26,* 965-971.

von Bekesy, G. (1960). *Experiments in Hearing*. New York: McGraw-Hill.

Wackym, P. A., Runge-Samuelson, C. L., Firszt, J. B., Alkaf, F. M., and Burg, L. S. (2007). More challenging speech-perception tasks demonstrate binaural benefit in bilateral cochlear implant users. *Ear and Hearing, 28,* 80S-85S.

Wilson, B. (2005). Signal processing for binaural devices. In *International Binaural Symposium*. Manchester, UK.

Wilson, B., Lawson, D., Muller, J., Tyler, R. S., and Kiefer, J. (2003). Cochlear implants: Some likely next steps. *Annual Review of Biomedical Engineering, 5,* 207-249.

Wilson, B., Finley, C. C., Lawson, D., Wolford, R., Eddington, D. K., and Rabinowitz, W. M. (1991). Better speech recognition with cochlear implants. *Nature, 352,* 236-238.

Yost, W. A. (2000). *Fundamentals of Hearing (Fourth Edition)*. San Diego, CA: Academic.

Zeng, F.-G. (2004). Auditory Prostheses: Past, Fresent and Future. In F.-G.Zeng, A. N. Popper, and R. F. Fay (Eds.), *Cochlear Implants: Auditory Prostheses and Electric Hearing*. New York: Springer.