FP6–004171 HEARCOM

Hearing in the Communication Society

INTEGRATED PROJECT

Information Society Technologies

D-6-3: Simulation studies of bimodal CI and hearing aid fittings

Contractual Date of Delivery:	28 February 2007(+45 days)
Actual Date of Submission:	13 April 2007
Editor:	A Faulkner
Sub-Project/Work-Package:	SP3/WP6
Version:	1.0
Total number of pages:	11

Dissemination Level		
PU	Public (subject to copyright restrictions only)	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Deliverable D-6-3

VERSION DETAILS	
Version:	1.0
Date:	12 April 2007
Status:	Draft/ Under review/ **Submitted**

CONTRIBUTOR(S) to DELIVERABLE	
Partner	**Name**
UK-UCL	A Faulkner
UK-UCL	C Siciliano
UK-UCL	S Rosen
UK-UCL	K Mair

DOCUMENT HISTORY			
Version	**Date**	**Responsible**	**Description**
0.1	01 April 2007	A Faulkner	Draft 1
1.0	12 April 2007	A Faulkner	Revised following internal peer review

DELIVERABLE REVIEW			
Version	**Date**	**Reviewed by**	**Conclusion***
0.1	01 April 2007	M. Lutman,	Minor revision
0.1	01 April 2007	B. van Dijk	Minor revision

* e.g. Accept, Develop, Modify, Rework, Update

Table of Contents

Acknowledgement

Supported by grants from the European Union FP6, Project 004171 HEARCOM. Additional support was received in the form of a Wellcome Trust Summer Vacation Scholarship to Katharine Mair. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Pre-Amble

The combined use of hearing aids and cochlear implants is a topic of growing importance as was noted in the overview of cochlear implant (CI) fitting methods included in D-6-1. One major challenge of this development for fitting is to ensure that the implant user can effectively combine and integrate information from the two prostheses. One important aspect of this is to develop a more complete understanding of the effects of any mismatching of acoustic frequency to cochlear place mapping between electrical and acoustic stimulation.

This is an interim report of work carried out as part of task 3 of HearCom WP6. Work on this task is planned to continue through to month 42, and in this period will include studies in users of cochlear implants.

1 Executive Summary

This deliverable is presented in three parts. The second part (Appendix 1) is a paper recently published in Audiology and Neurotology (Faulkner, 2006). The third part (Appendix 2) is the draft of an MS intended for submission to the Journal of the Acoustical Society of America. The first part provides a short overview of the research and our conclusions.

The work described uses speech processed through cochlear implant simulations and presented to normal hearing adults to examine the effects of conflicting mappings of acoustic frequency to cochlear place that may arise when a cochlear implant (CI) is used in conjunction with an acoustic hearing aid. While listeners seem well able to learn to accommodate to simple shifts of frequency to place mapping, combinations of CI and hearing aid are likely to result in conflicts of frequency to place mapping between the electrical and acoustic input modalities. Our simulation studies indicate that when such conflicts exist between the two ears across the complete frequency range of speech, while listeners show some ability to adapt to the conflict, this adaptation does not allow listeners to make effective use of all of the presented information, rather they learn to ignore the information carried in one of the two conflicting frequency to place representations.

The implication for the fitting of cochlear implants combined with hearing aids is that such conflicts of frequency to place mapping may need to be avoided.

2 Introduction

There is increasing interest in the combined use of acoustic hearing aids with cochlear implants (Offeciers et al., 2005). One major challenge of this development for fitting is to ensure that the implant user can effectively combine and integrate information from the two prostheses. One promising approach to a more optimal combination of information from two prostheses is to develop a more complete understanding of the effects of any mismatching of acoustic frequency to cochlear place mapping between the two. Even with a single cochlear implant it is clear that acoustic frequency to cochlear place mapping can have major effects on benefit (Fu & Shannon, 1999; Shannon, Zeng, & Wygonski, 1998). However these effects appear to be substantially reduced with experience (Rosen, Faulkner, & Wilkinson, 1999). Several considerations suggest that the importance of mismatches between frequency-to-place maps, whether with monaural electro-acoustic stimulation or in the use of contralateral hearing aids together with a cochlear implant or indeed with two contralateral implants, may be more significant than in the use of a CI alone.

One form of mismatch of frequency to place mapping will arise when the upper frequency limit of residual hearing lies well below the effective frequency of the most apical CI electrode (see for example Figure 1). There will in consequence be a medial range of cochlear places that cannot be stimulated either acoustically or electrically. This could arise with the use of a short CI electrode implanted in an ear with residual low-frequency acoustic function (Gantz & Turner, 2004) but could also arise with a shallow CI insertion to one ear and a contralateral hearing aid. Mid-frequency information is often of crucial importance in speech recognition, yet the preservation of frequency-place alignment would require that this information be discarded. The preservation of mid-frequency information in this case requires a warping of the frequency-to-place map to distribute this information around the places that cannot be stimulated. As a recent study shows, mismatches of this sort are readily learnable for normal hearing listeners, even with only a few hours of experience (Smith & Faulkner, 2006; Smith et al., 2006); this study is also briefly overviewed in Appendix 1 of this deliverable).

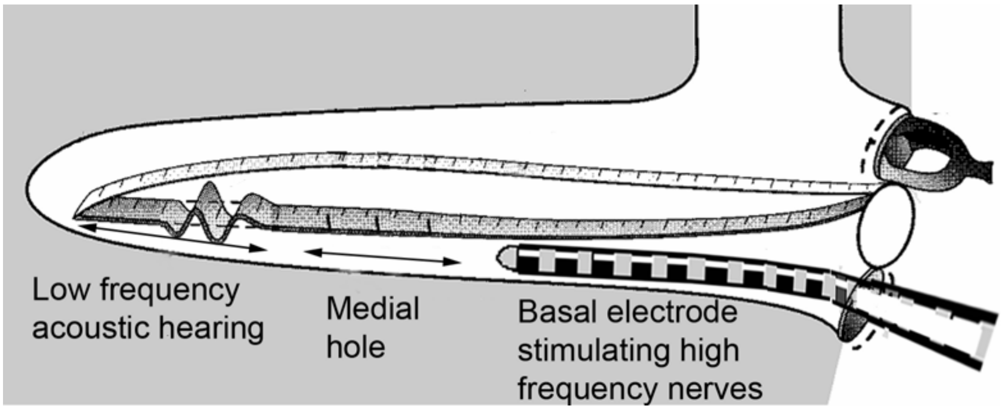

Figure 1 Illustration of a monaural hybrid electrical / acoustic prosthesis in a configuration leading to a significant medial cochlear region that cannot respond to electrical or acoustic stimulation.

A second issue arises with binaural fittings. Consider a CI electrode array inserted such that the most apical contact has an effective CF of 1000 Hz or more, used in combination with contralateral amplification to acoustic hearing. Acoustic hearing will necessarily give rise to a frequency-to-place mapping that is determined by the basilar membrane, while the cochlear implant would often be fitted with a basally shifted frequency-to-place mapping. In this instance, there is a mismatch of frequency-to-place mapping between the two ears. What might be the effect of such a mismatch? Can listeners adapt to such mismatches after training such that they can combine spectral information from the two ears? These questions are also important in the case of binaural cochlear implants, where the two electrode arrays may be inserted to significantly different depths, so that here again there may be a mismatch of the frequency mapping between the two ears unless the different insertion depths are taken into account in the fitting of the speech processors.

Studies already performed within HearCom and presented in detail in part 3 of this deliverable address what will be termed a “global” mismatch of frequency to place mapping, where spectral information across the whole frequency range important to speech is presented to different cochlear places in each ear. As the objective is to assess the extent to which listeners are able, after training, to make combined use of information presented to both ears, the spectral detail presented to each ear alone is restricted. The processing used to achieve this in a simulation for listeners with normal hearing is outlined in Figure 2.

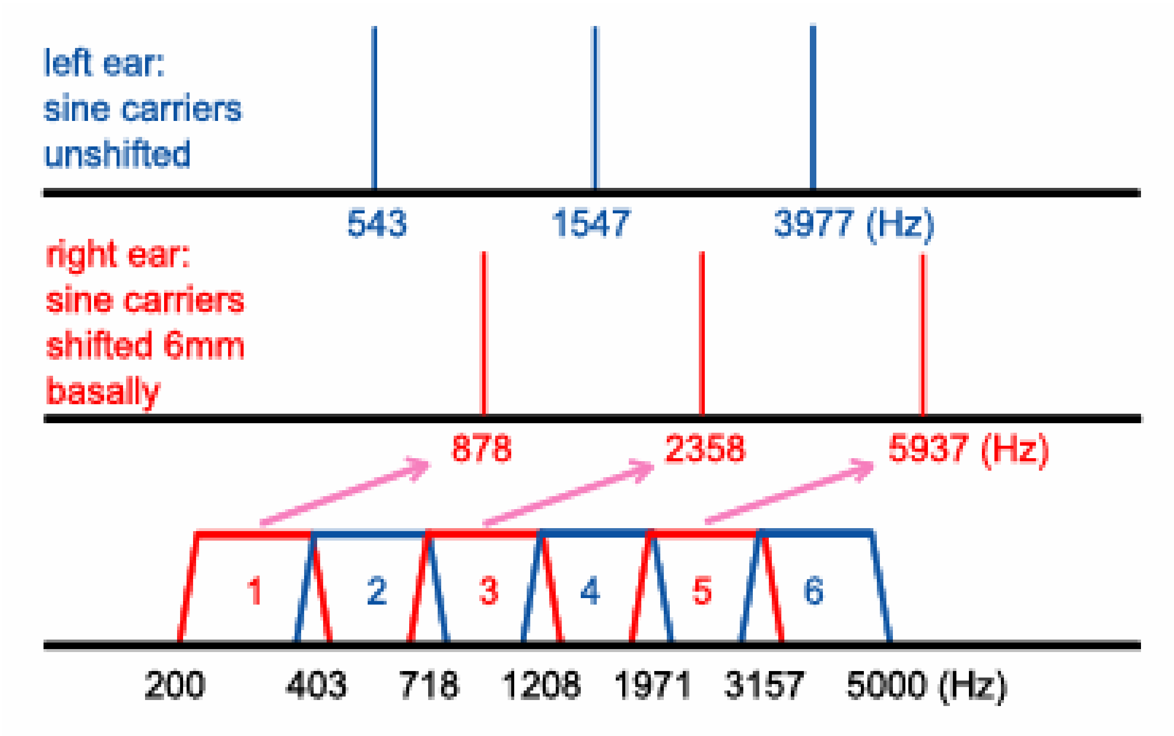

Figure 2: scheme of speech processing used to induce "global" mismatch of frequency to place mapping between two ears. The lowest panel represents 6 spectral analysis channels covering the frequency range 200-5000 Hz. Temporal envelope information from channels 1, 3 and 5 is presented to the right ear (see mid panel), imposed on sinusoidal carriers whose frequencies are shifted upwards compared to the analysis channels to an extent equivalent to a 6mm basalward shift on the basilar membrane. Temporal envelope information from channels 2, 4 and 6 is presented to the left ear, (top panel) imposed on sinusoidal carriers that match the centre frequencies of these three analysis bands.

The effect of this global mismatch has been measured by comparing speech perceptual performance through this processing scheme with that for a processor that is identical except that the carrier signals presented to the right ear are not shifted upwards. Hence the identical information is presented, with and without a global mismatch of frequency to place mapping.

The effect of this mismatch prior to the opportunity for leaning is substantial. Before subjects are exposed to training with the mismatched processor, median scores for the recognition of words from IEEE sentences are reduced from 62% to 10% by the introduction of the mismatch (compare conditions DU and DOS for session 0 in Figure 3). After 10 hours of training, the effect of the mismatch is much reduced, but still present. Here the effect of the mismatch is to reduce median scores from around 70% to around 56% (compare conditions DU and DOS for session 3 in Figure 3). Normal listeners can, therefore learn to be less affected by this global mismatch than they are at first exposure. However, the results across several studies indicate that these listeners do not learn to combine spectral information across the two ears in the presence of this global mismatch. Rather, their performance never exceeds that with the unshifted ear alone (compare conditions DU and 3US in Figure 3). The training effects that are observed for the mismatched processor seem to have two components. One is learning to make better use of the 3 unshifted bands presented to the left ear. A second effect that seems to be present is that listeners also learn to ignore the information from the shifted ear rather than to integrate this with information from the unshifted ear.

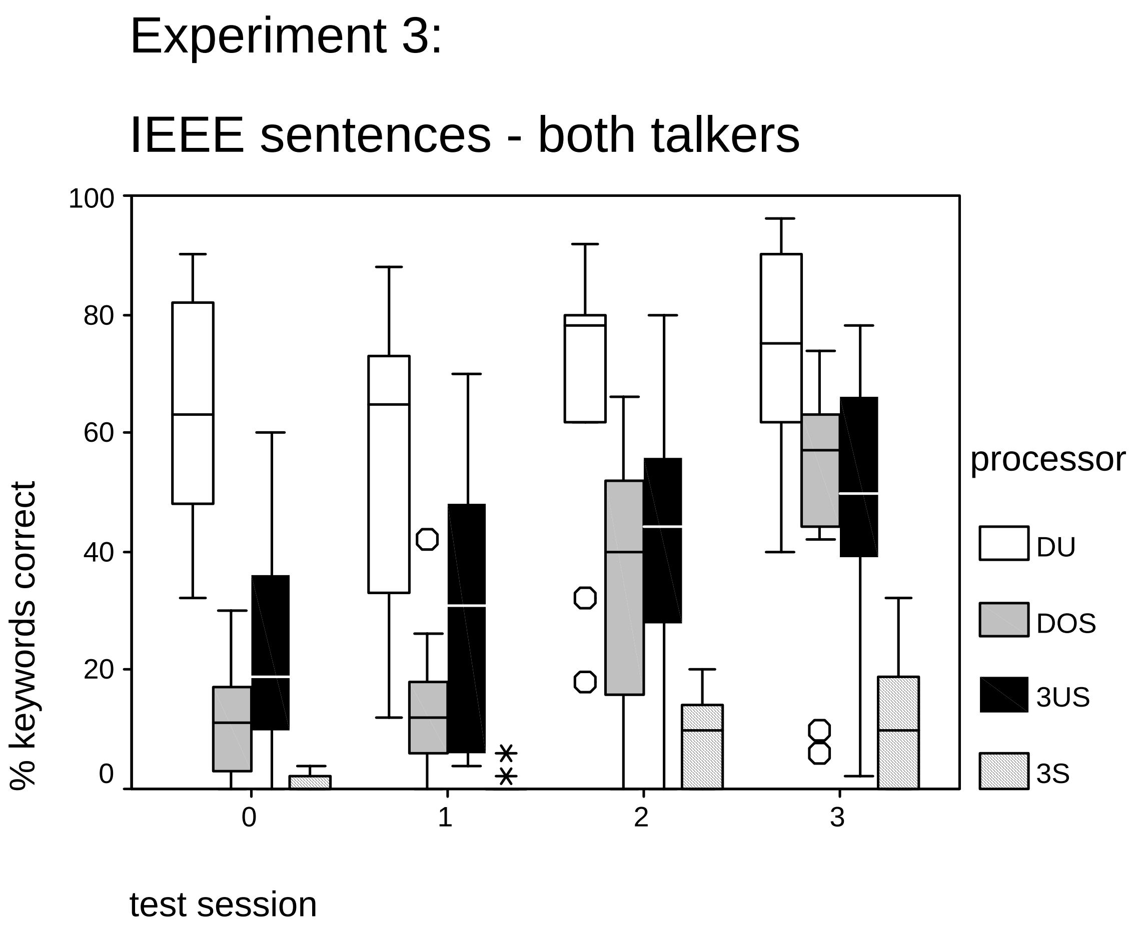

Figure 3. Illustrative results before, during, and after 10 hours or training (see appendix 2 for details). Key to processor conditions: DU – 3 bands to each ear with no spectral shift; DOS – 3 unshifted bands presented to left ear and 3 upward shifted bands presented to right ear; 3US – 3 unshifted bands to left ear; 3S – 3 upward shifted bands to right ear.

3 Dissemination and Exploitation

The work reported in this deliverable has been disseminated as follows.

3.1 Conference/Lecture presentations

Siciliano, C., Faulkner, A., Rosen, S. and Saul, L (2006) "Perceptual Adaptation to a Binaurally-Mismatched Frequency-to-Place Map" Poster 98, Assoc. Res. Otolaryngol. MidWinter meeting 2006 (Baltimore, Feb 2006).

Invited lecture by A Faulkner at "1st International Electro-Acoustic Workshop", Toulouse, December 2005

As part of material presented by A Faulkner for the Oticon Research Seminar Series, University of Manchester. 20 February 2007

3.2 Journal articles

Faulkner, A. (2006) "Adaptation to distorted frequency-to-place maps: implications for cochlear implants and electro-acoustic stimulation", Audiology and Neurotology, 11, Suppl. 1, 21-26. (Appendix 1)

Siciiliano, C., Faulkner, A., Rosen, S. & Mair K (in preparation) perceptual adaptation to a binaurally mismatched frequency-to-place map: implications for binaural rehabilitation with cochlear implants. MS for J Acoust. Soc. Am (Appendix 2)

4 Ethical issues

The work reported here has received approval from the UCL/UCLH Hospitals Committee Alpha for the Ethics of Human Research.

5 Conclusions

This research is consistent with the view that global mismatches of frequency-to-place representation will reduce the benefit that can be achieved in the combined used of a cochlear implant with a hearing aid. Further research is underway to explore the effects of mismatches restricted to the low frequencies that are likely to be accessible to residual hearing, and future work is planned to develop new psychophysical methods to assess the mismatch of acoustic to electrical frequency to place maps in users of cochlear implants. The outcome of these studies should make it possible to give guidance for the clinical fitting of a CI combined with a hearing aid.

References

Faulkner, A. (2006). Adaptation to distorted frequency-to-place maps: implications of simulations in normal listeners for cochlear implants and electro-acoustic stimulation. *Audiology and Neuro-Otology, 11,* 21-26.

Fu, Q. J. & Shannon, R. V. (1999). Effects of electrode location and spacing on phoneme recognition with the nucleus-22 cochlear implant. *Ear and Hearing, 20,* 321-331.

Gantz, B. J. & Turner, C. (2004). Combining acoustic and electrical speech processing: Iowa/nucleus hybrid implant. *Acta Oto-Laryngologica, 124,* 344-347.

Offeciers, E., Morera, C., Muller, J., Huarte, A., Shallop, J., & Cavalle, L. (2005). International consensus on bilateral cochlear implants and bimodal stimulation - Second Meeting Consensus on Auditory Implants, 19-21 February 2004, Valencia, Spain. *Acta Oto-Laryngologica, 125,* 918-919.

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Perceptual adaptation by normal listeners to upward shifts of spectral information in speech and its relevance for users of cochlear implants. *Journal of the Acoustical Society of America, 106,* 3629-3636.

Shannon, R. V., Zeng, F.-G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America, 104,* 2467-2476.

Smith, M. & 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.

Audiology & Neurotology

Audiol Neurotol 2006;11(suppl 1):21–26
DOI: 10.1159/000095610

Received: January 20, 2006
Accepted after revision: June 12, 2006
Published online: October 6, 2006

Adaptation to Distorted Frequency-to-Place Maps: Implications of Simulations in Normal Listeners for Cochlear Implants and Electroacoustic Stimulation

Andrew Faulkner

Department of Phonetics and Linguistics, UCL, London, UK

Key Words
Electroacoustic stimulation · Frequency mapping · Speech perception

Abstract
The ideal cochlear implant electrode array positioning enables stimulation over a range of cochlear positions whose characteristic frequencies cover the frequency range of speech and match the speech processor filter frequencies. However, the electrode positions achieved in practice may not meet this specification. Users of conventional monaural cochlear implants seem able to perceptually adapt to a mismatch of speech processor filters to electrode positions. In electroacoustic stimulation, it is important to consider possible inconsistencies between acoustic and electrical frequency-to-place mapping. Two simulation studies are outlined that address normal listeners' ability to perceive speech presented through distorted frequency maps. The first presented a map that is spectrally warped around a 10-mm medial cochlear area. Listeners were able to adapt to this map after a few hours of training. The second study presented a binaural mapping in which one ear was subject to a 6-mm basalward shift. Here listeners were unable to learn to integrate speech information across the two mismatched ears, rather they seem to learn to ignore the shifted information. Frequency-to-place mapping is likely to be an important factor in the successful use of a combination of electrical and acoustic hearing.

Introduction

The normal ear exhibits a mapping of frequency to cochlear place that is determined by the properties of the basilar membrane, and allows the association of a characteristic frequency (CF) to an auditory nerve fibre. It has been argued that the frequency alignment of cochlear implant (CI) speech processor filters to auditory nerve CFs is an important factor in determining speech perceptual benefit, at least in acquired hearing loss. This claim is based on data both from implant users and from acoustic simulations in normal hearing listeners, in which a shift of the frequency-to-place map equivalent to a displacement of 3 mm or more along the basilar membrane leads to a substantial decrement in speech perception [Dorman et al., 1997; Fu and Shannon, 1999]. This review examines the claimed importance of frequency alignment in the light of studies of perceptual learning for speech subjected to a shift of frequency-to-place mapping, with particular reference to shifts of frequency mapping that may arise with the combined use of a CI and amplification to residual hearing.

The frequency-to-place mapping of a CI depends on several factors. These include the frequency ranges covered by the speech processor analysis filters, the position of the electrode array within the cochlea, and the topography of the neural elements that are stimulated by each electrode. We can very readily and exactly characterise the frequency ranges of the speech processor filters. For

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1420–3030/06/0117–0021$23.50/0

Accessible online at:
www.karger.com/aud

Andrew Faulkner
Department of Phonetics and Linguistics, UCL
Wolfson House, 4 Stephenson Way
London NW1 2HE (UK)
Tel. +44 20 7679 7408, Fax +44 20 7679 5107, E-Mail andyf@phon.ucl.ac.uk

example, a typical fitting of a Nucleus speech processor could use analysis filter bands covering the frequency range 150–10823 Hz. The electrode position is harder to determine, although it can be revealed through CT [Ketten et al., 1998]. The topography of the stimulated neural elements is not at present completely understood. The standard approach for estimating the effective CF at each electrode contact has been based on the frequency-to-place mapping of the organ of Corti established by Greenwood [1990]. As recently shown by Sridhar et al. [this issue, pp. 16–20], the Greenwood map is not a realistic model for the CFs of CI electrodes placed near to the modiolus, for which a spiral ganglion map seems appropriate. A spiral ganglion frequency-to-place map assigns substantially lower CFs to a given electrode contact position than does Greenwood's 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. An additional source of uncertainty is that the degree of neural degeneration in an individual implant recipient cannot be known in vivo, so that an estimation of the location of the neural elements that are stimulated by a contact in a pathological cochlea is always subject to some degree of error.

Notwithstanding our incomplete knowledge of the effective CFs along an electrode array, shifts of frequency-to-place mapping have demonstrable and sizable effects, so it remains the case that it is important to understand the significance of this aspect of mapping in making the best use of a CI, whether monaurally or bilaterally, and also when considering the combined use of a CI with residual hearing, whether ipsi- or contralateral.

In vivo CT data from human CI recipients suggest that typical array insertion depths may be around 20 mm from the cochlear base, with some deeper insertions up to 25 mm [Ketten et al., 1998; Skinner et al., 2002]. On the assumption that Greenwood's organ of Corti map is appropriate for estimating electrode CF, a 20-mm insertion corresponds to a CF for the most apical contact of around 1000 Hz, while the most basal contact will have a CF of above 10 kHz. Given this range of CFs over the electrode array, setting a speech processor to avoid a misalignment of speech processor filters to electrode CFs would entail delivering speech information extracted from the acoustic frequency range encompassed by the electrode CFs. Both the classic articulation index data [French and Steinberg, 1947], which assign relative importance values to frequency bands of speech, and recent data from simulations of CIs [Faulkner et al., 2003; Shannon et al., 2002] indicate that such a fitting would lead to reduced speech intelligibility as a result of the loss of significant low-frequency speech cues that lie below 1000 Hz. Clearly the preservation of frequencies that convey significant speech information is an important factor in choosing a frequency-to-place map for a CI and this may mitigate the disadvantage of a frequency-to-place map that does not match that of normal hearing.

A second reason to suppose that a natural frequency-to-place map may not be essential comes from studies showing perceptual adaptation to shifted frequency maps. Several studies have used acoustic simulations to study the ability of normal listeners to adapt to frequency-to-place maps entailing a substantial basalward shift of cochlear stimulation of around 6 mm [e.g. Rosen et al., 1999; Fu and Galvin, 2003]. In each case, the basalward shift had a large immediate impact on the intelligibility of speech, but after several hours of training, the effects of the shift were very much reduced. Direct evidence of adaptation to a change of frequency-to-place mapping in CI users has been reported by Fu et al. [2002]. In this study speech processor filters were shifted downwards in frequency by 0.68 or 1 octave. The 3 subjects all showed a marked initial drop in consonant and vowel recognition followed by a partial recovery of performance after 2 weeks of use of the lowered filters, while there was little further change in performance after 3 months of use.

Hence, both normal hearing listeners and CI users demonstrate the ability to at least partially adapt to changes of frequency-to-place mapping. When such changes are accompanied by an increase in the information provided by the shifted map, it may well be the case that, after a period of adaptation, a misaligned map which provides the most appropriate information is more beneficial than an aligned map that, as a result of a less than ideal electrode array position, must entail the omission of important frequency information [Faulkner et al., 2006].

Most studies of frequency-to-place mapping have focused on basalward shifts of approximately constant distance along the BM, which are close to shifts on a logarithmic frequency scale. Such shifts relate well to variations of insertion depth and are useful analogies to the frequency-to-place mapping in conventional monaural implantation. There is now increasing interest in the combined use of acoustic hearing aids with CIs [e.g. Offeciers et al., 2005]. In considering both monaural electroacoustic stimulation (EAS) and the use of contralateral hearing aids together with a CI, other forms of distortion of frequency-to-place mapping are likely to arise.

One such distortion will arise when the upper frequency limit of residual hearing lies well below the effective CF of the most apical electrode contact. There will in consequence be a medial range of cochlear places that cannot be stimulated either acoustically or electrically. Mid-frequency information is often of crucial importance in speech recognition, yet the preservation of frequency-place alignment would require that this information be discarded. The preservation of mid-frequency information in this case requires a warping of the frequency-to-place map to distribute this information around the places that cannot be stimulated. The first study reviewed here is a simulation in normal listeners that investigates their ability to adapt to maps that employ such frequency warping.

A second issue arises with binaural fittings. Consider a CI electrode array inserted such that the most apical contact has an effective CF of 1000 Hz or more, used in combination with contralateral amplification to acoustic hearing. Acoustic hearing will necessarily give rise to a frequency-to-place mapping that is determined by the basilar membrane, and hence, that accords to the Greenwood organ of Corti map, while this CI would typically be fitted with a basally shifted frequency-to-place mapping. In this instance, there is a mismatch of frequency-to-place mapping between the two ears. The second study reviewed here asks what the effect of such a mismatch might be and whether listeners can adapt to such mismatches after training. These questions are also important in the case of binaural CIs, where the two electrode arrays may be inserted to significantly different depths, so that here again there may be a mismatch of the frequency mapping between the two ears unless the different insertion depths are taken into account in the fitting of the speech processors.

Frequency-to-Place Mapping in EAS and Binaural Implants

Mapping around Medial Non-Responsive Regions of the Cochlea

In monaural EAS there may very well be a medial non-responsive cochlear region that is beyond the reach of the electrode array yet is lacking functional inner hair cells and hence cannot be stimulated acoustically. This may arise, for example, from the use of a short CI electrode [Gantz and Turner, 2004]. The preservation of important mid-frequency speech information can in such cases be achieved by warping the frequency-to-place map around the medial region. Acute studies of such frequency maps suggest that warping the frequency-to-place map in this way does not lead to improved speech recognition compared to maps that discard mid-frequency information [Shannon et al., 2002]. However, these were acute studies and listeners had no opportunity for extended experience of the frequency-warped maps. Given that listeners can learn to use distorted frequency-to-place maps that involve a simple basalward shift [Rosen et al., 1999], it may be premature to dismiss frequency-warped maps on the basis of acute experiments. For this reason, a recent study by Smith and Faulkner [2006] investigated whether normal listeners could learn frequency maps in which mid-frequency information was warped around a simulated non-responsive 10-mm medial region of the cochlea.

Method

One processor, called 'dropped', used a matched mapping that omitted frequency bands matching the 'missing' medial region (fig. 1). An 'adjacent-warp' condition shifted mid-frequency information to places just adjacent to the medial region, while the two lowest and two highest bands preserved a matched mapping (fig. 1). A 'spread-warp' condition encoded 6 bands covering the speech range, all being spectrally warped onto 3 apical and 3 basal bands (fig. 1). A final unwarped control condition mapped the 6 input bands of the spread-warp processor onto 6 carriers matching the frequencies of the input bands. All of the processors employed 6th order Butterworth analysis filters, while envelope extraction in each band used half-wave rectification and a 3rd order Butterworth 400-Hz low-pass filter.

Each of the 8 normal-hearing young adult participants was trained for a total of 3 h in each of the two warped conditions. Training employed live-voice connected discourse tracking, requiring the listener to repeat processed speech phrase-by-phrase. The subjects were split into two subgroups so that the order of the training conditions was counterbalanced over subjects.

Results

As figure 1 shows, the recognition of words in IEEE sentences increased significantly with training in both warped conditions. Subjects learned the 'spread-warp' mapping more readily, and here performance after training was statistically indistinguishable from that in the control condition in which the same 6 input bands were presented without frequency warping. Similar conditions were explored by Shannon et al. [2002] but without any opportunity for training. They found that performance was equivalent in 2 conditions similar to our adjacent-warp and dropped conditions, and concluded that listeners could not use warped spectral information. However, it is clear that this conclusion does not hold for listeners with just a few hours of listening experience.

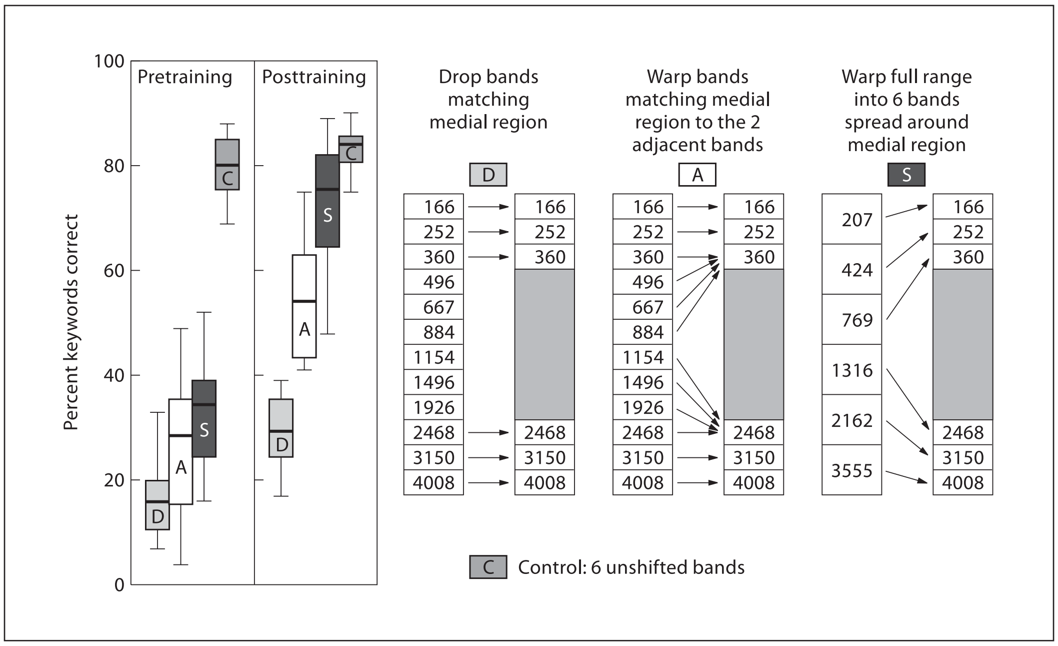

Fig. 1. Performance with IEEE sentences before and after training [Smith and Faulkner, 2006]. The main vocoder conditions are represented to the right. The characters overlaid on the graph indicate the condition. D: Mid-frequency information is dropped; A: mid-frequency information is warped into the adjacent bands; S: the 'spread-warp' condition in which 6 analysis bands covering the speech frequency range are warped to both apical and basal regions; C: control condition in which 6 bands are presented without warping. The box and whisker plots show the median score (bar) the interquartile range (box) and the complete range of scores (whiskers).

Binaural Mismatches of Frequency-to-Place Mapping

The frequency-to-place map distortions considered so far have all been monaural. When an implant is used in conjunction with a contralateral hearing aid, or with bilateral implants, it is readily conceivable that the two ears would be operating with different frequency-to-place maps. We have examined such a situation in a further simulation study [Faulkner et al., 2005]. Here we simulate two ears in which one is subjected to a basalward place shift. The issue of interest is whether normal listeners can learn to combine cues from this binaurally mismatched mapping.

Method

In order for the successful integration of cues to be readily observed, each ear received limited spectral information. A total of 6 analysis bands covering a frequency range of 200–5000 Hz were interleaved between the two ears so that each ear had access only to 3 spectral bands, with the carriers to one ear shifted upwards to represent a 6-mm basalward basilar membrane displacement. Envelopes were extracted from each analysis band using half-wave rectification and a 32-Hz 3rd order low-pass filter and imposed on sinusoidal carriers. Carrier frequencies were set at the logarithmic center of each unshifted band, and at frequencies shifted upwards to represent a 6-mm basalward basilar membrane shift from the center of each shifted band. Six normal listeners received 5 h of training with live-voice connected discourse tracking with speech presented through this binaurally mismatched processor.

Results

Intelligibility scores for words in simple BKB sentences are shown in figure 2. Although subjects did show statistically significant improvements in their performance with the binaurally mismatched processor over 5 h of training, and also showed signs of improvement when hearing only the 3 shifted bands, performance with the binaurally mismatched processor never rose above the performance seen with only the 3 unshifted bands. At the final post-training session, BKB sentence scores with these two processors were statistically indistinguishable. Additional tests with more difficult IEEE sentences showed the same outcome, and since with these materials maximum performance never exceeded 70% correct, this finding cannot simply be attributed to a ceiling effect. The main result suggests that what is learned in this instance is that the shifted information should be ignored. The training period in this study is plainly very short compared to the duration of exposure received by an implant user, but we find no indication that our subjects had learned to combine spectral cues to speech between the two mismatched ears. This finding suggests that if a substantial mismatch of frequency-to-place map arises between the two ears in the combined fitting of a CI and a hearing aid, or with a binaural CI fitting, this mismatch may prevent the most effective use of the spectral information conveyed by the two devices.

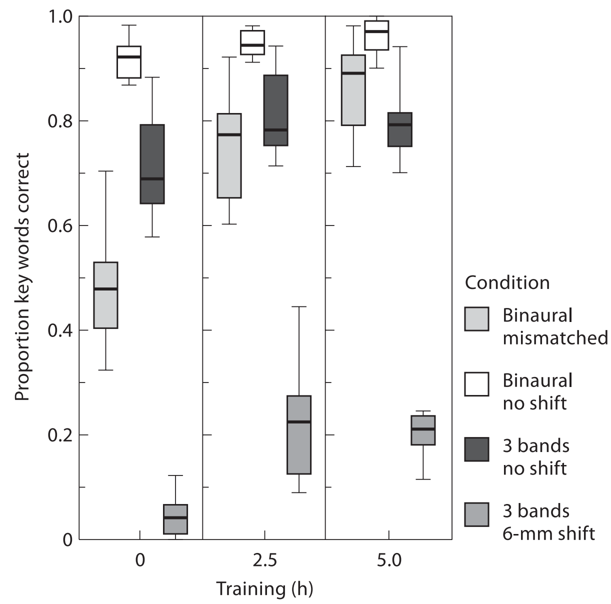

Fig. 2. Box and whisker plot showing performance in the recognition of words in BKB sentences over the course of training. Apart from the binaurally mismatched condition described in the main text, data are shown for 3 other conditions. 'Binaural – no shift' presents 6 interleaved bands to the two ears with no frequency shift. The two 3-band conditions represent performance with the individual ear components of the binaurally shifted processor.

Discussion

Smith and Faulkner's [2006] finding of perceptual adaptation to a map warped around a medial dead region reinforces a gathering body of evidence that the human perceptual system is remarkably flexible in its adaptability to distortions of frequency-to-place maps, at least within the realm of speech perception. It also underlines the need for considerable caution in the interpretation of any effects of acute manipulations of frequency-to-place mapping. Additionally, this study makes clear that listeners can adapt to frequency-to-place distortions that go beyond the constant shifts of place along the basilar membrane that have been considered in previous work.

With specific reference to EAS, these data suggest that CI users with a medial non-responsive region need not be fit by speech processors that match filters to electrode positions, but rather that the CI speech processor should provide information over the full range of frequencies important for speech and that is not accessible to residual hearing. Gantz and Turner [2004] found poorer speech performance in 3 monaural EAS subjects implanted with a short 6-mm electrode compared to those with a more deeply inserted 10-mm electrode. They suggest that the poor performance of their 6-mm electrode subjects can be attributed to a shifted frequency-to-place mapping. Our results might seem to be in conflict with that interpretation. This conflict could be resolved by the assumption that there is a limit to plasticity for information received in extremely basal neural elements. The short hybrid electrode matches CFs ranging from around 8 kHz upwards on an organ of Corti map, and there are no published data suggesting plasticity to frequency-place misalignment at such high frequencies. There may also be interactions between the responses to acoustic and electrical stimulation involved in monaural EAS that are not represented in our simulations which could impact upon the ability of the system to adapt to the extreme frequency-to-place map distortion presented by the short electrode.

Our simulations of binaurally mismatched maps seem relevant to frequency-to-place maps that overlap and are in conflict by representing similar frequencies in quite different cochlear places and/or very different frequencies in similar places. One important question that arises in this context is whether perceptual adaptation to frequency-shifted speech is a central cortical process, or whether it might have a lateralised component specific to each ear. It seems very unlikely that there is a lateralised adaptation during a few hours of experience, but this appears considerably more plausible with the many months of exposure experienced by implant users.

More research is clearly needed to test more strongly the finding that binaural mismatches of place map cannot be learned. If this is indeed the case, then there is a clear implication that such mismatches would be best avoided in the clinical fitting of both binaural CIs and of a CI with a contralateral hearing aid. This in turn leads to a need for clinically applicable methods to assess and correct any mismatch. If, on the other hand, such mismatches can be learned, but the learning process is a relatively difficult one compared to the learning of a novel monaural frequency-to-place map, there is a need to develop training techniques that facilitate the learning required to make full use of fittings with binaurally mismatched frequency-to-place maps.

Acknowledgements

Work partially supported by the RNID and the European Commission (FP6-004171 HEARCOM). Donations of research interface hardware and software tools from Cochlear Europe and Advanced Bionics Europe have been received. Financial support for attendance at workshops has also been received from both these sources.

References

Dorman MF, Loizou PC, Rainey D: Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. J Acoust Soc Am 1997;102:2993–2996.

Faulkner A, Rosen S, Norman C: 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 Hear 2006;27:139–152.

Faulkner A, Rosen S, Saul L: Perceptual adaptation to a binaurally-mismatched frequency-to-place map: what is learned? Conference on Implantable Auditory Prostheses, Asilomar, August 2005.

Faulkner A, Rosen S, Stanton D: Simulations of tonotopically-mapped speech processors for cochlear implant electrodes varying in insertion depth. J Acoust Soc Am 2003;13: 1073–1080.

French NR, Steinberg JC: Factors governing the intelligibility of speech. J Acoust Soc Am 1947;19:90–119.

Fu QJ, Galvin JJ: The effects of short-term training for spectrally mismatched noise-band speech. J Acoust Soc Am 2003;113:1065–1072.

Fu QJ, Shannon RV: Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 cochlear implant. Ear Hear 1999;20:332–344.

Fu QJ, Shannon RV, Galvin JJ: Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant. J Acoust Soc Am 2002;112: 1664–1674.

Gantz BJ, Turner C: Combining acoustic and electrical speech processing: Iowa/nucleus hybrid implant. Acta Otolaryngol 2004;124: 344–347.

Greenwood DD: A cochlear frequency-position function for several species – 29 years later. J Acoust Soc Am 1990;87:2592–2605.

Ketten DR, Vannier MW, Skinner MW, Gates GA, Wang G, Neely JG: In vivo measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays. Ann Otol Rhinol Laryngol 1998;175:1–16.

Offeciers E, Morera C, Muller J, Huarte A, Shallop J, Cavalle L: International consensus on bilateral cochlear implants and bimodal stimulation – Second Meeting Consensus on Auditory Implants, 19–21 February 2004, Valencia, Spain. Acta Otolaryngol 2005;125: 918–919.

Rosen S, Faulkner A, Wilkinson L: Perceptual adaptation by normal listeners to upward shifts of spectral information in speech and its relevance for users of cochlear implants. J Acoust Soc Am 1999;106:3629–3636.

Shannon RV, Galvin JJ, Baskent D: Holes in hearing. J Assoc Res Otolaryngol 2002;3: 185–199.

Skinner MW, Ketten DR, Holden LK, Harding GW, Smith PG, Gates GA, Neely JG, Kletzker GR, Brunsden B, Blocker B: CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients. J Assoc Res Otolaryngol 2002;3:332–350.

Smith M, Faulkner A: Perceptual adaptation by normally hearing listeners to a simulated 'hole' in hearing. J Acoust Soc Am 2006, in press.

Sridhar D, Stakhovskaya O, Leake PA: A frequency-position function for the human cochlear spiral ganglion. Audiol Neurotol 2006;11(suppl 1):16–20.

PERCEPTUAL ADAPTATION TO A BINAURALLY MISMATCHED FREQUENCY-TO-PLACE MAP: IMPLICATIONS FOR BINAURAL REHABILITATION WITH COCHLEAR IMPLANTS

Catherine Siciliano, Andrew Faulkner, Stuart Rosen and Katharine Mair

Department of Phonetics and Linguistics, UCL, Wolfson House, 4 Stephenson Way,

LONDON NW1 2HE

Draft MS prepared for submission to J. Acoust. Soc. Am.

12 April 2007

I. INTRODUCTION

The advantages of binaural hearing have been well established in normal hearing listeners, yet the possibility of binaural rehabilitation for those with profound hearing loss has only been realised in recent years. Because of the great success reported with the cochlear implant (CI), criteria of candidacy for the procedure have been relaxed to include listeners who have some residual hearing. In this case, it is possible for the patient to use an acoustic hearing aid (HA) in the ear contralateral to the implant in order to make maximal use of any residual hearing. In a typical case, the HA is used to amplify low frequencies, while the cochlear implant is used to electrically stimulate higher frequencies. Another possibility for binaural rehabilitation is the use of two (bilateral) cochlear implants (Offeciers et al., 2005).

For those with a cochlear implant, the restoration of binaural hearing is especially important because many CI users perform well in quiet, but their speech reception drops dramatically when listening in noise. Normal hearing listeners show a binaural advantage for speech in noise that arises from a combination of binaural squelch, the head-shadow effect and redundancy. Limitations in cochlear implant technology and/or the listener's auditory processing capacities may preclude the re-establishing of all of these benefits in impaired listeners. However, a binaural advantage for speech perception has been reported in patients with bilateral cochlear implants (Litovsky, Parkinson, Arcaroli, & Sammeth, 2006; Dorman & Dahlstrom, 2004; Tyler et al., 2005), and implants used in conjunction with acoustic hearing aids (Ching, Incerti, & Hill, 2004; Ching, Incerti, Hill, & van Wanrooy, 2006; Iwaki et al., 2004; Ching, 2005). Significantly, this has not been limited to speech in noise, as studies have also shown binaural advantage for quiet speech tested in a laboratory for both bilateral cochlear implants (Dorman et al., 2004) and cochlear implants in conjunction with hearing aids (Hamzavi, Pok, Gstoettner, & Baumgartner, 2004).

While the potential benefits for bilateral rehabilitation are manifold, the use of bilateral devices poses new problems in the consideration of CI frequency-place mapping. The average length of the human cochlea is about 35mm, but cochlear implants are inserted to depths of 25mm at best, and in many cases much more shallowly than this. 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 et al., 1998; Skinner et al., 2002). The result of the shallow insertion is a tonotopic mismatch between the analysis filter of the implant speech processor and the characteristic frequency (CF) of the basilar membrane at a given electrode: the most apical electrode of a cochlear implant will stimulate nerves normally “tuned” to higher frequencies (i.e. a basalward shift). 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 been based on the frequency-to-place mapping of the organ of Corti established by Greenwood (1990), which at the median insertion depth of 20 mm found by Ketten et al. leads to an estimated CF of 1000 Hz. As recently shown by Leake and colleagues, the Greenwood map is not a realistic model for the CFs of CI electrodes placed near to the modiolus, for which a spiral ganglion map seems appropriate (Sridhar, Stakhovskaya, & Leake, 2006). A spiral ganglion frequency-to-place map assigns substantially lower CFs to a given electrode contact position than does Greenwood’s 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, shifts of frequency-to-place mapping have demonstrable and sizable effects, and since there are considerable variations between implant recipients of electrode insertion depth, this aspect of mapping is likely to be of importance when considering the combined use of a CI with residual hearing, whether ipsi- or contra-lateral.

Frequency-place mismatch has been associated with a decrement in speech intelligibility in cochlear implant patients, especially in the case of shallow electrode insertions. This has led to an ongoing debate about whether it is best to preserve tonotopic matching at the expense of a more spectrally rich but frequency-shifted map when fitting CI speech processors. Using simulations of cochlear implant speech processing with noise-vocoded speech processors in normal hearing listeners, Dorman and colleagues (Dorman, Loizou, & Rainey, 1997a) claimed that while basalward frequency shifts up to the equivalent of 3mm along the basilar membrane (Greenwood, 1990) could be tolerated, shifts larger than 3mm lead to large decreases in speech intelligibility. For a sentence test, performance (measured as keywords correct) declined from near 100% for a simulated full insertion (i.e. tonotopically matched) to just below 50% for an upward shift of 4 mm. A similar result with vowel recognition was shown for cochlear implant patients (Fu & Shannon, 1999).

A caveat with these studies is that the listeners in these experiments were not given time to adapt to the distorted signal. In studies using noise-vocoded simulations as well as with cochlear implant patients, listeners given time to adapt to altered frequency-place maps do show improved performance after a period of training, suggesting some level of plasticity for learning altered frequency-place maps (Rosen, Faulkner, & Wilkinson, 1999; Fu & Shannon, 2002; Fu, Nogaki, & Galvin III, 2005b; Fu, Galvin III, Wang, & Nogaki, 2005a; Faulkner, Rosen, & Norman, 2006; Svirsky et al., 2005).

Fu et al. (2002) examined whether cochlear implant listeners could 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. However, all three subjects showed continuing improvements in speech recognition throughout the three month period of wearing the experimental processors.

Rosen et al. (1997) were able to show, through the use of noise-vocoded simulations of cochlear implant speech processing, a similar process in normal hearing listeners on a much shorter time scale. They used a 4-channel noise vocoder to simulate the reduced spectral resolution that is typical for cochlear implant users, and then shifted the entire spectrum upwards in frequency to an equivalent of 6.5 mm along the cochlea according to Greenwood's formula, or the equivalent of an incomplete insertion of a cochlear implant electrode. When these listeners were tested prior to any experience with the spectrally-shifted speech, their performance was at floor level with easy sentence material. With only 3 hours training, though, these listeners improved recognition of words to 30%. This study demonstrated very clearly the importance of adaptation in examining speech perception with altered frequency-place maps.

In two later studies, Fu and colleagues demonstrated that targeted auditory training can improve speech perceptual performance with altered frequency-place maps in both normal hearing listeners (with simulations of spectrally shifted speech) and cochlear implant patients (Fu et al., 2005b; Fu et al., 2005a). Normal hearing listeners were trained with spectrally shifted speech with an 8-band sine vocoder simulation of CI speech processing. The best results were achieved with a computer-based targeted 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 short training sessions. Implant patients were trained for an hour a day, 5 days per week, for an entire month or longer. A self-directed computer based training program was used to test minimal phonetic contrasts within a multitalker speech database. All 10 CI patients demonstrated significantly increased speech perception performance following a month of training.

Growing evidence from simulation studies indicates that when given time to adapt, subjects will perform better at listening tasks with a shifted frequency-place map than one that is tonotopically matched but with gaps in frequency regions important for speech. This is what might be predicted from the articulation index, which assigns a high importance value to frequency bands of speech below 1000 Hz and thus below the most apical point of stimulation in a cochlear implant (French & Steinberg, 1947).

Faulkner, Rosen & Norman (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 6mm 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. Post-training performance on sentences and vowels was, in some conditions, better with the shifted than the matched processor.

In a second study, Smith and Faulkner (2006) examined frequency-place mapping around a simulated “hole” in hearing (i.e. a dead region in the cochlea). They found that after 3 hours of training, listeners could adapt to the frequency-place map in which spectral regions were warped around a simulated 10mm hole in the mid-frequency region of the cochlea. Significantly, performance in the warped conditions was higher than for that in the tonotopically matched conditions in which frequency information in the ‘hole’ was simply dropped. This study is particularly relevant in the consideration of bilateral/bimodal stimulation, where there is likely to be a gap in stimulation between the highest perceived frequency in the HA ear and the lowest perceived pitch percept at the most apical electrode of the implant ear.

These studies converge on the idea that speech that is warped with respect to frequency-place mapping in the cochlea, when presented to postlingually deafened adults, is initially difficult to perceive, but may be learned to varying degrees of success when the listener is given time to adapt. Crucially, however, these studies have examined the monaural and/or monotic case, when listeners have been implanted with a single cochlear implant, or when the listener is presented with the same signal to both ears.

However, in the case of bilateral implants and traditional long-electrode cochlear implants in combination with a contralateral HA, the signals and thus frequency-place maps in the two ears may be very different. For bilateral implants, 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. In the case of CI+HA, the HA ear will retain the natural frequency place map (albeit band limited in all likelihood), while the CI ear will experience some frequency shifting. To what extent are listeners able to adapt when presented with frequency-place maps that differ between the two ears?

Evidence from dichotic listening experiments suggests that information presented in complement across the two ears will be easily integrated. Broadbent and Ladefoged (1957) showed that listeners presented with the F1 and F2 of a /da/ syllable to contralateral ears perceived the syllable as /da/. This process, later coined *spectral fusion* by Cutting (1976), is robust to 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, & Braida, 1992). An experimental approach to bilateral CI processing, uses "*zipper processors*", which excite electrodes at each ear in an interleaved fashion, thus increasing the number of effective electrodes and decreasing channel interaction. These rely on the listeners' ability to spectrally fuse the incoming signal into a single percept (Lawson, Brill, Wolford, Wilson, & Schatzer, 2000).

Dorman and Dahlstrom (2004) reported a binaural advantage for speech perception in two bilateral implant patients who had frequency-place maps that differed between the two ears, supporting the hypothesis that spectral fusion of mismatched frequency maps is indeed possible in a clinical situation.

This research uses sine-excited vocoder simulations to examine the extent to which normally hearing listeners can adapt to a binaurally mismatched (and unilaterally shifted) cochlear frequency-to-place map after a period of training. In Experiment 1, we trained listeners with a binaurally mismatched processor over eight 40-minute training sessions, testing the extent of adaptation before, during and after training. In Experiment 2, we looked at whether adaptation could be facilitated by training with just the shifted frequency map. In Experiment 3, we doubled the length of training and alternated training material from that of Experiments 1 and 2, to monitor how listeners' performance changed over time, and whether adaptation to binaurally mismatched frequency-place maps required a longer period of training.

If listeners cannot adapt to a spectral mismatch between ears, then this must be taken into account when optimizing the presentation of information to binaural devices. If, however, spectral mismatch can be learned, then specialised training techniques to accommodate this mismatch should be developed.

II. EXPERIMENT 1

A. Method

1. Subjects

Six normally hearing speakers of British English took part, and each was paid for his or her participation. They all had normal hearing by self report.

2. Signal Processing

Speech processing used interleaved six-band sine-excited vocoders similar to those used by Dorman and colleagues (Dorman, Loizou, & Rainey, 1997b; Loizou, Mani, & Dorman, 2003). From low frequency (apical end) to high frequency (basal end), bands 1, 3 and 5 were presented to the right ear with the equivalent of a 6mm basalward shift (assuming 35mm cochlea) (test conditions and training), while bands 2, 4 and 6 were presented to the left ear without a shift. This interleaved binaural configuration was designed to encourage integration of the signals at each ear – a six channel vocoder being more spectrally rich than a three channel one. The use of sine excited vocoders was preferred because this eliminated any interaction between channels, thus ensuring a clearer interpretation of the intelligibility scores between conditions. Six channels were used to ensure that intelligibility of either ear on its own would not be so great as to mask any potential improvements gained through integration of the two ears. With a six channel vocoder, intelligibility of either ear would be for a 3 channel vocoder, while integration would be for a 6 channel vocoder. Four channels to each ear (i.e. an 8 channel vocoder), or anything greater, would probably yield sufficient intelligibility such that any improvements with integration may be masked. Conversely 2 channels (for an effective 4 channel vocoder) to each ear would not be enough to facilitate adaptation with the shifted information. See Dorman, Loizou, & Rainey (1997c); Loizou, Dorman, & Tu, (1999) for further discussion of spectral resolution with CI simulations. While six is a significantly fewer number of channels compared to the number of electrodes in current CIs, studies indicate that most CI users only have the use of 6 to 8 effective channels (Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001).

Centre and crossover frequencies for the analysis filters 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 35mm (Greenwood 1990).

$$frequency = 165.4(10^{0.06x} - 1)$$

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

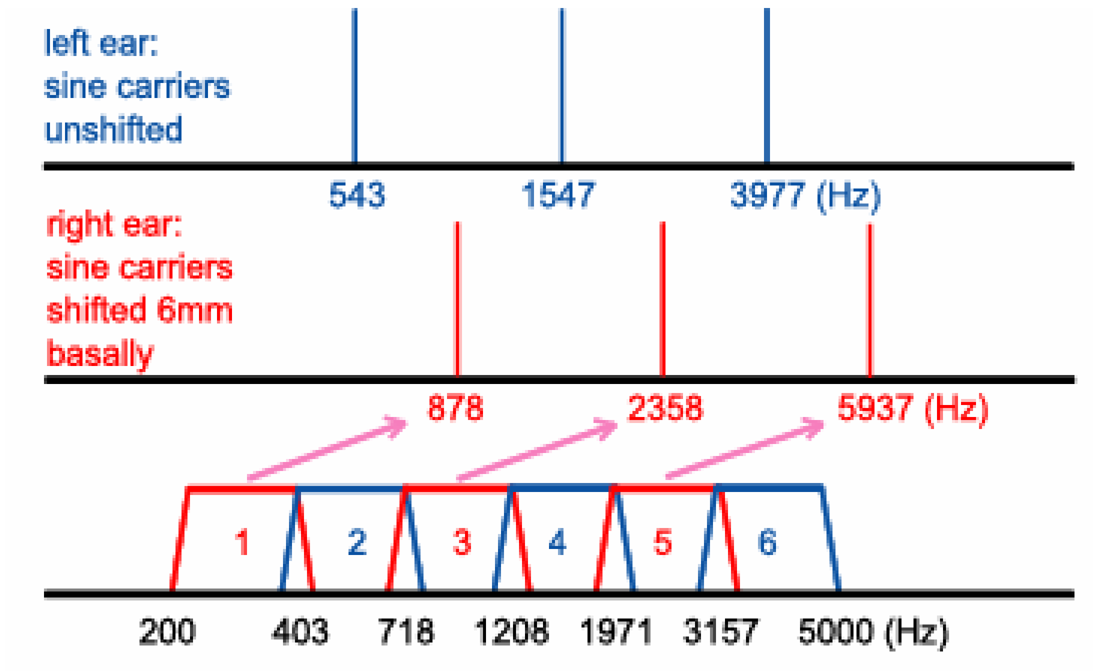

Figure 1: Speech processing producing a binaurally-mismatched frequency-to-place map. The lowest panel represents 6 spectral analysis channels covering the frequency range 200-5000 Hz. Temporal envelope information from channels 1, 3 and 5 is presented to the right ear (see mid panel), imposed on sinusoidal carriers whose frequencies are shifted upwards compared to the analysis channels to an extent equivalent to a 6mm basalward shift on the basilar membrane. Temporal envelope information from channels 2, 4 and 6 is presented to the left ear, (top panel) imposed on sinusoidal carriers that match the centre frequencies of these three analysis bands.

Each band was processed with an analysis filter, full-wave rectification, envelope smoothing, and multiplication of a sinusoid with frequency matching the CF of the band by the time averaged absolute amplitude of the band. Finally, bands one, three and five were summed and presented to the right ear, while bands two, four and six were summed and presented to the left ear. Each channel received speech with no pre-emphasis as its input. Input and output spectra are shown in Figure 1, while Table I shows input and output CFs as well as filter cutoffs.

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

Table I: analysis band cutoff and centre frequencies, and carrier frequencies for each band in the unshifted and odd-band shifted conditions.

An online 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 filters were Butterworth IIR designs with three orders per upper and lower side. 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.

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 loudness of each of the shifted bands was corrected to reflect the loudness of the input stimulation in the unshifted band at that CF. 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.[1] An overall level correction was applied to the shifted processed speech to ensure that processed speech had similar SPLs to the given speech input.

3. Training

Subjects were trained with Connected Discourse Tracking (CDT) (De Filippo & Scott, 1978). In this method, the experimenter reads from a text to the subject, who then repeats back what he or she heard. This allows the listener to acclimate to the spectrally shifted speech while engaging in a communication task that is similar to a conversation. A metric of communication rate can be calculated based on the number of words repeated back correctly per minute, allowing a way to monitor progress throughout training. CDT has been shown to be an effective training method for spectrally shifted speech (Rosen et al. 1999, Faulkner et al. 2006).

The talker in this experiment was the first author, CS, who is a native speaker of a north-eastern dialect of American English, although 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. The subject was given three chances to repeat back the phrase; if the listener's response matched what the talker had said, the talker would move on to the next phrase. If after the third attempt, the listener could still not reproduce the phrase, the talker repeated this back to the subject as unprocessed speech (to the left ear only) by pressing a switch. The talker used a digital timer to measure five minute blocks so that the number of words correctly repeated could be

[1] Minimal audible field values were taken from Robinson & Dadson (1956) and interpolated to exact frequencies using a cubic spline fit to log frequency.

counted for each five minute block of training. When each five minute block finished, a mark was made in the text at the last correctly repeated word in the block.

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. They have 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 could be used to communicate through during blocks of auditory-visual training. During auditory-only training, the window was blinded. A constant speech-shaped (pink) masking noise at 45 dBA was played in the listener's room to mask any unprocessed speech information that could be transmitted through the wall and window.

condition	***abbreviation***	***component bands and shift***	
		Right	***left***
dichotic unshifted	DU	1, 3, 5	2, 4, 6
dichotic odd channels shifted	DOS	1, 3, 5 → 6mm	2, 4, 6
odd channels shifted	OS	1, 3, 5 → 6mm 2, 4, 6	1, 3, 5 → 6mm 2, 4, 6
three unshifted channels	3US		2, 4, 6
three shifted channels	3S	1, 3, 5 → 6mm	

Table II: Conditions for Experiment I

4. Test Conditions

Table 2 summarizes the five conditions tested in Experiment 1. These conditions were designed to assess what aspects of the signal subjects were learning, if any, and at what stage of processing this learning occurred. The **dichotic unshifted (DU)** served as a control condition to assess baseline performance. This processor was interleaved, although none of the channels were shifted. A previous study indicated that performance with the 6-channel dichotic interleaved processor, where all output channels were matched to input analysis filters, yielded intelligibility equivalent to a 6-channel non-interleaved monotic processor (Faulkner, 2006). The test condition, which subjects were also trained with, was the **dichotic odd-shifted (DOS)** condition. Here, bands 1, 3 and 5 were shifted an equivalent of 6mm and presented to the right ear, while bands 2, 4 and 6 were presented to the left ear without a shift. In the **odd-shifted (OS)** condition, bands 1, 3 and 5 were shifted but bands 2, 4 and 6 were matched; here, however, all bands were summed together and presented to both ears. This condition was designed to tease apart any effects of laterality and binaural integration. If performance here differed markedly from the DOS condition, then we would guess that listeners were learning to attune to information at a single ear. Conversely, if there was no difference here, then we would guess that learning was occurring at a higher stage of processing. In the **three unshifted (3US)** condition, listeners were tested with channels 2, 4 and 6, which had input and output filters matched in frequency. In the **three shifted (3S)** condition, channels 1, 3 and 5 were presented with a 6mm basalward shift.

5. Test materials

Sentence Perception

In an earlier study (Faulkner, 2006), listeners showed near-ceiling effects for some processors after a period of training with high context sentences. For this reason, the IEEE/Harvard sentence lists were used, which have very little contextual information. Recordings of the sentences were produced by one male and one female talker of British English. The female talker's speech was extracted from audio-video recordings of the sentences. The male sentences were from an anechoic digital recording of the talker. For both talkers the material was from a 16-bit 48 kHz digital audio recording. The 72 lists in the set each contained ten sentences with five keywords in each sentence. The first half of the lists was designated for the female talker, and the second half for the male. For each test session, two lists per condition were chosen from each subset in a pseudo-random manner. No list appeared more than twice in the same condition across all of the subjects. No visual feedback was given for the sentence material.

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. The male talker was the same as for the sentence tests, while a different female talker was used here. Recordings were anechoic digital recordings sampled at 48 kHz. Five tokens of each bVd word 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. The bVd words were: /æ/ (bad); /ɑː/ (bard); /iː/ (bead); /e/ (bed); /ɪ/ (bid); /ɜː/ (bird); /ɔː/ (board); /ɑ/ (bod); /uː/ (booed); /ʌ/ (bud). The text shown here is the text that appeared on the computer program during testing. A grid with all nine words appeared, and the subject clicked on the button displaying the bVd word they perceived with the computer mouse. 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.

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 either training or testing sessions. A crossover design was used for testing, so that in each test session, half the subjects were tested first with sentences and then with vowels, while the other half of the subjects were tested first with vowels and then consonants. Table 3 describes the sequence of training and testing for Experiment 1.

Session	**Training**	**Processor**	**Testing**
1	5 minutes audio visual (AV) 5 minutes auditory alone (AA)	DU	*Familiarization:* unprocessed vowels *Pre-test*: IEEE sentences (2 lists x 5 conditions x 2 talkers) bVd identification (5 tokens x 5 conditions x 2 talkers)
2	5 minutes AV 35 minutes AA	DOS	None
3	5 minutes AV 35 minutes AA	DOS	None
4	5 minutes AV 35 minutes AA	DOS	none
5	5 minutes AV 35 minutes AA	DOS	none
6	none or familiarization		*Mid-test:* IEEE sentences (2 lists x 5 conditions x 2 talkers) bVd identification (5 tokens x 5 conditions x 2 talkers)
7	5 minutes AV 35 minutes AA	DOS	None
8	5 minutes AV 35 minutes AA	DOS	None
9	5 minutes AV 35 minutes AA	DOS	None
10	5 minutes AV 35 minutes AA	DOS	None
11	none or familiarization		*Post-test:* IEEE sentences (2 lists x 5 conditions x 2 talkers) bVd identification (5 tokens x 5 conditions x 2 talkers)

Table III. Sequence of training and testing conditions 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, followed by the pre-test. Previous experiments with (unshifted) vocoded speech have demonstrated that listeners require a period of rapid adaptation before they can reliably perceive the distorted speech (Davis, Johnsrude, & Hervais-Adelman, 2005). While there is great variability among listeners, performance jumps from near floor levels to near ceiling levels after just a few minutes experience. This familiarization block ensured this rapid adaptation process was achieved before testing commenced. Similar to the procedure carried out through the remaining training sessions, the first five minutes of the familiarization block were auditory-visual (AV), so the subject could see the experimenter speaking. After five minutes, the experimenter blocked the window so that the remaining five minutes were auditory alone (AA). It was thought that visual feedback would facilitate the learning process. However, the primary concern was improvements in auditory conditions, so we wanted training to reflect this.

Following the pretest, subjects were trained with CDT in four 40-minute training sessions with the dichotic odd-shifted (DOS) speech. For the first five minutes of each session, visual feedback was also given. During the remainder of the session, the experimenter employed the rule that training would be auditory-only so long as the subject was tracking at a minimum of 10 words per minute. No subject ever fell below this minimum level of performance, so this rule never needed to be invoked.

Subjects were tested again, in the same order as designated at the first session, 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) refresher block, which was not counted towards the total hours of training. Following this second test, the subjects underwent four more 40-minute training sessions, and again completed a final round of testing according to the same protocol.

Testing took place in a sound-proofed room with diotic[2] presentation of the processed speech over Sennheiser HD280 headphones. Stimuli were presented at a comfortable listening level. The Euraud ASTEC software suite was used to control the experiment. For sentence tests, the subject was seated facing away from the computer screen, and the experimenter faced the computer. When each stimulus was played, the subject was asked to repeat back to the experimenter as many words as he or she could. Using the loose keyword scoring paradigm, where words were counted correct when the word root was

[2] For monotic test conditions, presentation was monotic, although this was determined by the stimulus and not by the headphones or experimental control suite.

repeated correctly (Bench & Bamford, 1979), the experimenter clicked on the buttons indicating the words answered correctly. The computer stored the information for subsequent retrieval and scoring. For the vowel test, the subject was able to run the experiment on their own, clicking the mouse to indicate the bVd word they had heard.

B. Results

Test data was analyzed using repeated measures analysis of variance (ANOVA), with within subject factors of test session, condition and talker. Hyunh-Feldt epsilon corrections were applied to all F tests for factors with more than one degree of freedom.[3]

Connected Discourse Tracking

CDT rates over training sessions for auditory-visual (AV) and auditory-alone (AA) presentation modes are shown in Figure 2. For AV presentation, the mean number of words tracked per minute across all training sessions was 73.6 (S.D. 14.0). For AA presentation, mean number of words tracked per minute was 59.8 (S.D. 13.1). As is expected with CDT, performance increased across training sessions. This improvement can be partially attributed to increasing familiarity with the talker and the text material. For each subject, a decrease in tracking rate was seen at the point of introduction of new text material. A repeated measures ANOVA showed significant main effects of presentation mode [$F(1, 5) =121$, $p < 0.001$] and training session [$F(5, 25) = 21.1$, $p < 0.001$], as well as a significant training session by presentation mode interaction [$F(7, 34) = 3.86$, $p = 0.004$], confirming that performance improves over time and is affected by presentation mode. The session by condition interaction is likely to be attributed to the greater increase in words tracked per minute over time in the AA mode than the AV mode.

[3] Hyunh-Feldt adjusted degrees of freedom have been rounded to the nearest integral value.

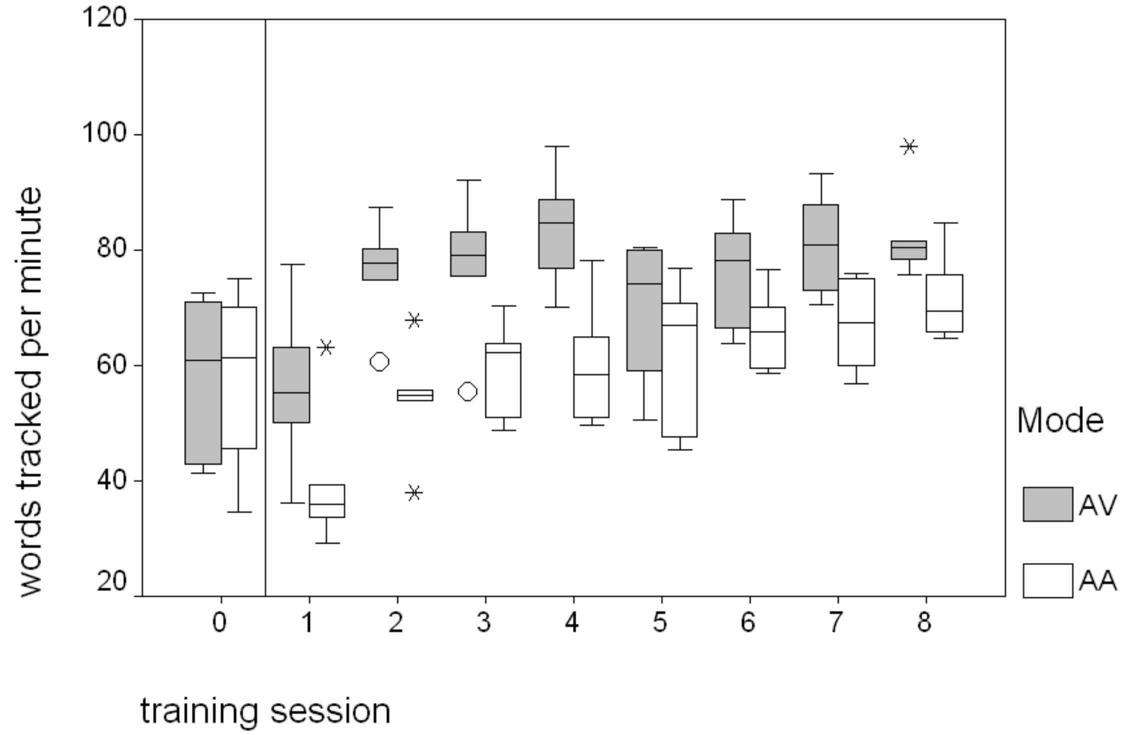

Figure 2: Box and whisker plot showing connected discourse tracking rate by session and presentation mode in experiment 1. In session 0 the DU (unshifted) speech processor was used. From sessions 1 onwards, the processor was always the dichotic odd shifted (DOS) condition. Boxes show the interquartile range over subjects. The bar shows the median and the whiskers show the complete range excluding any outlying values (shown as open circles or asterisks).

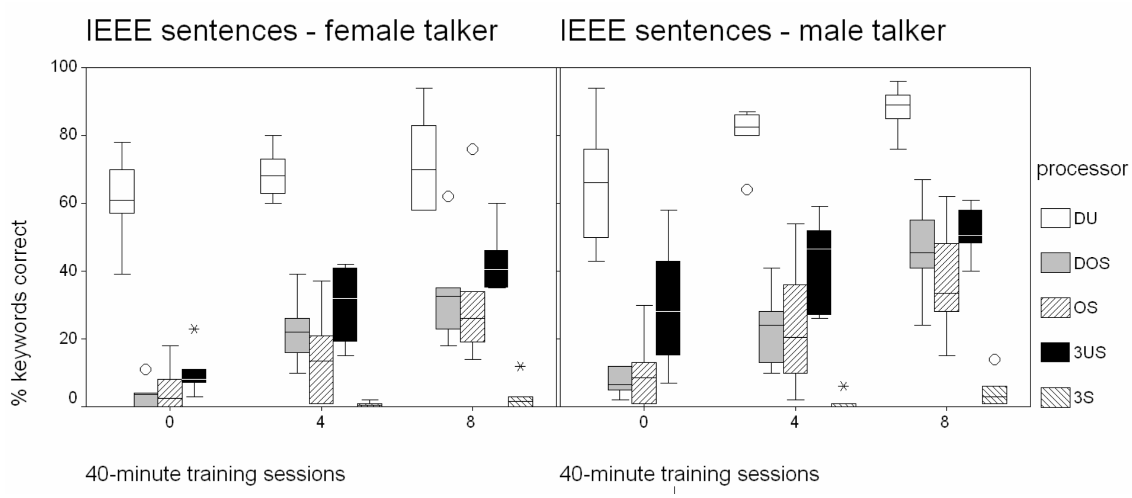

Figure 3 IEEE sentence scores from experiment 1 as a function of training and processor for female and male test talkers.

Sentence Perception

Keywords correct for the sentence tests across training sessions is shown for each talker in Figure 3. A repeated measures ANOVA showed significant main effects of talker [$F(1, 5) = 28.2$, $p = 0.003$)], number of training sessions [$F(2, 10) = 61.9$, $p < 0.001$] and processor [$F(3, 17) = 136$, $p < 0.001$]. There was also a significant talker by processor interaction [$F(3, 17) = 6.57$, $p = 0.003$]. Because of this interaction, separate ANOVAs for each talker were performed. These showed main effects of number of training sessions [*female:* $F(1, 5) = 31.7$, $p = 0.002$; *male*: $F(2, 10) = 50.7$, $p < 0.001$] and processor [*female*: $F(3, 16) = 119$, $p < 0.001$; *male*: $F(3, 18) = 113$, $p < 0.001$] for both talkers. For both talkers, there was also a highly significant session by condition interaction ($p \leq 0.001$), which can be attributed to floor effects in the 3S conditions across all sessions.

Performance in the DOS and OS conditions never exceeded that in the 3US condition. This is an indication that subjects did not adapt to the binaural mismatch, since they did not improve performance with the additional shifted channels. However, for both talkers, Bonferroni-adjusted paired comparisons revealed significant differences between post-test and pre-test performance in all conditions except the 3S condition, indicating some extent of adaptation after the training.

While ANOVA values showed significant differences between the patterns of performance with male and female talkers, Bonferroni-adjusted paired comparisons revealed that this could be accounted for largely by differences in the DOS and 3US conditions between the two talkers, as well as better overall performance with the male talker in the DU condition.

Bonferroni-adjusted paired comparisons on the post-training sentence scores showed performance with the 3 unshifted channels (3US) to be significantly worse than with the dichotic unshifted (DU) condition ($p = 0.001$), but significantly better than the 3 shifted channels (3S) condition ($p < 0.001$). No significant difference between the 3US, DOS and OS conditions was found.

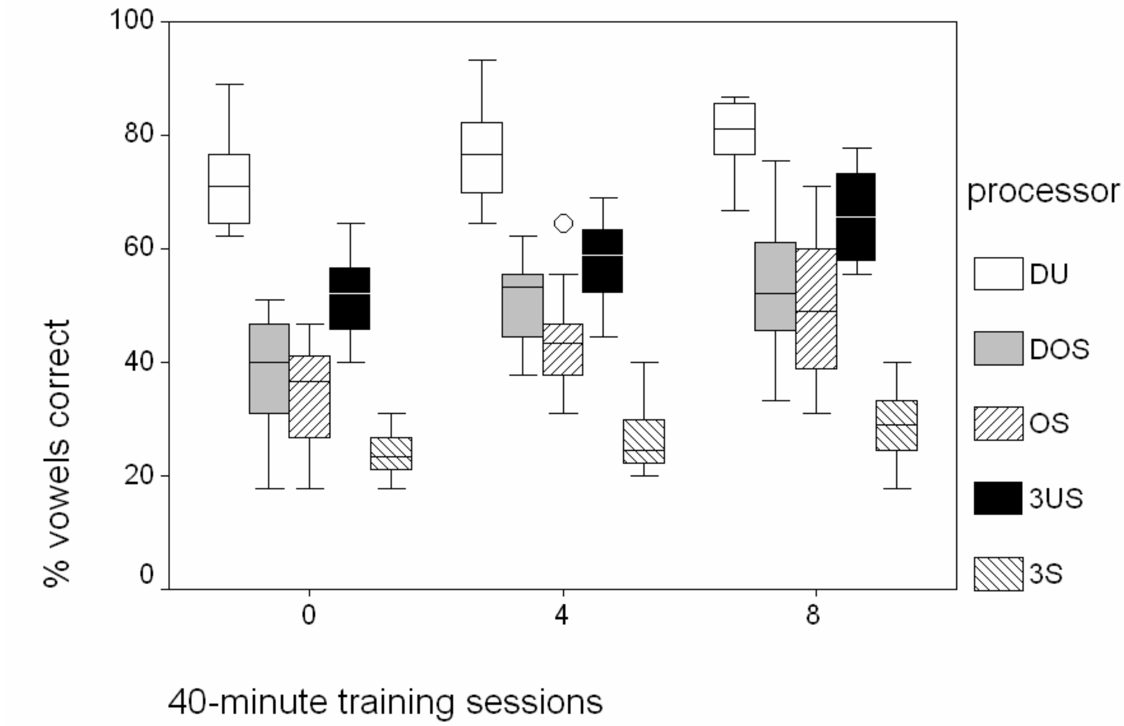

Figure 4: Vowel identification scores from experiment 1 combined over talkers as a function of training and processor.

Vowel Identification

Figure 4 gives shows vowel identification across the training sessions. As can be expected from a closed-set identification task, performance in all conditions was higher for the vowel identification than for the sentence perception test. However, the overall pattern of results across training sessions and conditions is similar. 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 [$F(8, 40) = 89.6$, $p = 0.005$]. There was no significant effect of talker, or talker by processor interaction. Therefore, the data plotted in Figure 4 are for the two talkers combined, and separate analyses for the individual talkers were not performed.

Bonferroni-adjusted comparisons showed significant improvements in identification scores in all conditions between the pre-test and post-test sessions ($p < 0.05$). As found with the sentence perception tests, Bonferroni-adjusted paired comparisons on the post-training scores revealed no significant difference between the 3US condition and the DOS, and OS conditions. Performance with the 3US condition was significantly worse than for the DU condition ($p = 0.02$) and significantly better than the 3S condition ($p < 0.001$). There was no talker by processor interaction.

C. Discussion

The data show clear general trends over all of the processor conditions for both sentence and vowel speech materials. While subjects showed some degree of learning in all conditions after five hours, twenty minutes of training, performance with the DOS, OS and 3S conditions never exceeded that with the 3US condition. This suggests that listeners are relying on the sparser 3-channel tonotopic map at the expense of improved spectral detail with the 6-channel mismatched map. This learning appears to take place at a stage of processing subsequent to binaural integration, since performance was not significantly affected by diotic (OS) versus dichotic presentation (DOS). Because subjects did not show any significant improvement with the 3S condition, it appears that they may also be learning to ignore this information. It is plausible that listeners need to be forced to adapt to the mismatched frequency map. Listeners appear to be able to recognize speech sufficiently enough with the three matched channels that adaptation to the shifted frequency-place map is not occurring.

Studies of adaptation with vocoded and spectrally shifted speech have traditionally examined either effects of the noise vocoding itself (Dorman, Loizou, & Rainey, 1997d; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), or effects of spectral shifting (Shannon, Zeng, & Wygonski, 1998; Rosen et al., 1999), but not the interaction of these effects. Thus, in the case where subjects have been trained with spectrally shifted speech, they have not been given tonotopically matched speech simultaneously. A period of training with the spectrally shifted speech in the absence of any tonotopically matched speech may be necessary before integration of these mismatched maps can occur.

Fu, Shannon, & Galvin III (2002) fitted Nucleus-22 implant patients with an experimental processor that shifted the analysis filters downwards by 1 or 0.68 octaves, and tested them over a 3 month period. While performance with the experimental processor was initially far below that with their clinical processor, throughout the three month period, they gradually improved. At the end of the 3 months, their performance on sentence tests approached that with their clinical processor, although not for vowels. When they returned to their clinically fitted processor, they immediately performed as well as they had before the period with the experimental processor. This provides some evidence that listeners can adapt to shifted frequency-place maps while retaining their natural tonotopic maps.

Another aspect to consider is the duration of the training in this experiment. While previous studies have demonstrated significant improvements in perception of spectrally shifted speech in as little as three hours (Rosen et al., 1999), these were looking at the case of a frequency map with four contiguous channels that preserved the relative order of spectral bands. The shifted channels here consist of three non-contiguous spectral channels with gaps in spectral information, interleaved with 3 channels that are tonotopically matched. It is possible that the adaptation to this mismatched map, in which information is interleaved and partially shifted in frequency as well as spectral order, is significantly more difficult to learn and thus requires a longer period of adaptation. Performance did not reach asymptotic level with any of the mismatched or shifted processors, which may indicate subjects are still learning at the point of final testing. Tyler and Summerfield (1996) showed that cochlear implant patients are still adjusting to their speech processors six months and longer after implantation, so this is probably quite possible with simulations as well.

Experiments II and III were designed to explore these questions of type and time course of training in learning to accommodate mismatched frequency maps. In Experiment II, we look at whether training subjects with just the shifted frequency map yields any synergistic improvements in speech perception scores with the mismatched frequency map. In Experiment III, we examine the possibility that long-term training is necessary to elicit adaptation. We also look at the interaction of training paradigm and training duration here.

II. EXPERIMENT II

The aim of Experiment II was to examine whether subjects were better able to adapt to the binaurally mismatched speech after training with the spectrally mismatched speech in the absence of tonotopically matched speech. In Experiment I we showed that subjects are resistant to integrating a frequency-place map that is not consistent between the two ears, although it was not clear whether the training they received was the best method of eliciting the binaural integration. In this experiment, we maintained the same experimental design as in Experiment I, with the exception that training was with the 3S processor. We expected that if subjects were able to learn to adapt to the 3 spectrally-shifted bands through a period of training, then binaural integration with the additional 3 matched bands would be facilitated. We also included an additional control test condition, 3NS, which was the 3 odd channels unshifted. Finally, because we anticipated that this training would be much more difficult for subjects, we added an additional post-training session with easier sentence materials to more readily show learning in the 3S condition. While a pilot study had shown these materials to elicit ceiling effects in subjects who had been trained with the DOS condition, subjects in Experiment II would have had no experience with matched vocoded speech apart from that gained through listening to the test material.

A. Method

1. Subjects

Six normally hearing speakers of British English took part, and each was paid for his or her participation. They all had normal (< 20 dBHL) pure-tone thresholds at .5, 1, 2 and 4 kHz.

2. Signal Processing

Signal processing was identical to that for Experiment I with the exception that training for this experiment was with a 3-channel sine vocoder with a shift equivalent to 6mm along the cochlea. This was exactly the same vocoder that was presented to the right ear during training in Experiment I, with the left ear receiving no input for this experiment.

Furthermore, an additional test processor was used that consisted of the 3 odd channels - 3NS, which had analysis and output filters matched in frequency.

3. Training

The training procedure for Experiment II was identical to that for Experiment I with the exception that the training processor was the 3S processor – the three odd spectrally shifted bands.

4. Testing

Sentence Perception

As for Experiment I, the IEEE sentence materials were used to monitor performance across training sessions. Because the additional 3NS condition was being tested in this experiment, the test was lengthened by four sentences at each session – two more lists for the new condition with the male and female talkers.

Because the 3S training processor was thought to be much more difficult to learn than the DOS processor, we also tested subjects with the easier BKB and IHR sentences. There were not enough lists in either set to allow the use of two talkers in all conditions, so we designated the BKB sentences for the male talker, and the IHR sentences for the female talker. These sentence lists have similar syntactic constructions and are essentially equivalent in intelligibility. The male talker recordings were 16-bit 48 kHz digital audio recordings of the BKB sentences made simultaneously with an audio-visual recording (see reference note), from a different male talker of British English from that used in Experiment I. The female recordings of the IHR Adaptive Sentence Lists (MacLeod & Summerfield, 1990) were from an anechoic digital recording (16-bit, 44.1 kHz) of the same talker as Experiment I.

The BKB sentence lists consisted of 16 sentences per set with either 3 or 4 keywords per sentence, so that there was always 50 keywords per sentence set. IHR sentences consisted of 15 sentences of 3 keywords each, with 45 keywords per sentence set. To maintain consistency, only the first 12 sentence sets from each list type were used.

Vowel Identification

The vowel identification task was the same as that for Experiment I, with the addition of a male talker and female talker test for the 3NS condition.

5. Conditions

The same five conditions outlined in Table 2 were used in Experiment II as in Experiment I. Additionally, a 3NS condition was tested which consisted of the 3 odd frequency bands, tonotopically matched, which were presented to the right ear. The purpose of this condition was to monitor whether the potential intelligibility of the 3 shifted channels was similar to that with the 3 even unshifted channels.

6. Procedure

The procedure for Experiment II was the same as for Experiment I, with the addition of a twelfth test session in which the easier IHR and BKB sentences were tested.

As in Experiment I, we had an a priori rule to alternate training with AV and AA presentation should the subjects not achieve a baseline tracking rate of 10 words per minute, although it was never necessary to invoke this rule.

B. Results

Connected Discourse Tracking

CDT rates over training sessions for auditory-visual (AV) and auditory-alone (AA) presentation modes are shown in Figure 5. For AV presentation, the mean number of words tracked per minute across all training sessions was 55.6 (S.D. 18.0). For AA presentation, mean number of words tracked per minute was 27.8 (S.D. 13.8). As seen in Experiment I, performance increased across training sessions, although for each subject, a decrease in tracking rate was seen at the point of introduction of new text material. A repeated measures ANOVA showed significant main effects of presentation mode [$F(1, 5) = 28.8$, $p = 0.003$] and training session [$F(7, 35) = 6.64$, $p < 0.001$]. The interaction between training session and presentation mode was nearly significant [$F(7, 35) = 2.24$, $p = 0.054$].

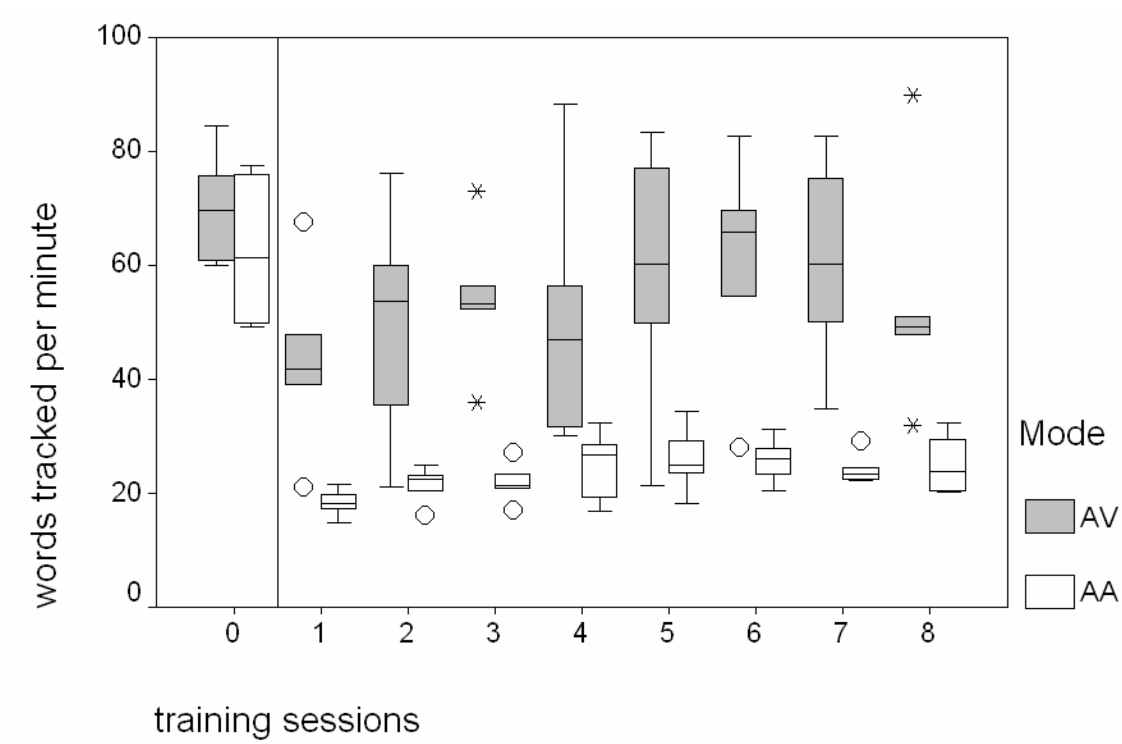

Figure 5: Box and whisker plot showing connected discourse tracking rate by session and presentation mode in experiment 1. In session 0 the DU (unshifted) speech processor was used. From sessions 1 onwards, the processor was always the 3 shifted bands only (3S) condition.

Sentence Perception – IEEE sentences

Keywords correct scores for the sentence tests across training sessions are shown for the two talkers combined in Figure 6. 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 [$F(3, 16) = 85.9$, $p < 0.001$]. There was no significant talker by processor interaction, so analyses shown here are for the two talkers combined. There was also no significant session by condition interaction.

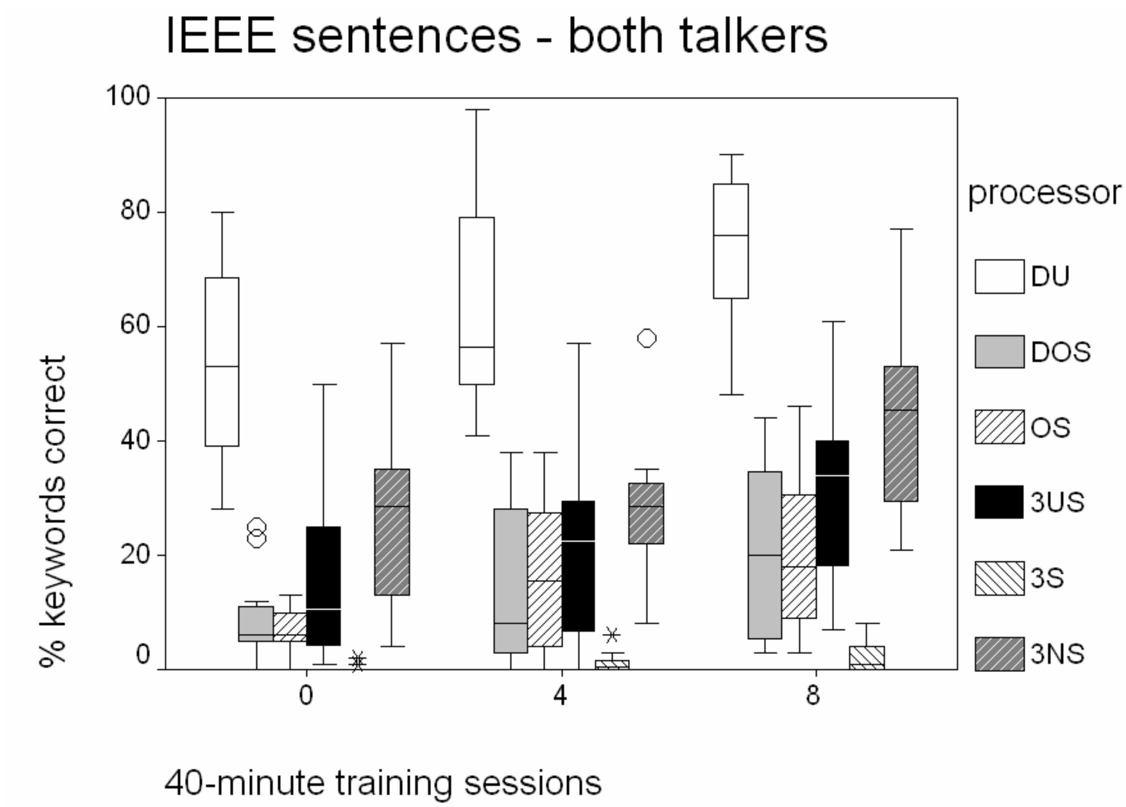

Figure 6: IEEE sentence scores from experiment 2 as a function of training and processor.

Performance in the DOS and OS conditions never exceeded that in the 3US condition. Bonferroni-adjusted paired comparisons of scores within each condition showed that subjects improved significantly across training only with the DU and 3NS processors. They did not significantly improve across training sessions in any of the other conditions, although performance never decreased across training sessions in any of these conditions. Bonferroni-adjusted paired comparisons on the post-training sentence scores revealed no significant difference between the 3US condition and the DOS and OS conditions. Intelligibility with the 3US condition was significantly worse than the DU ($p < 0.001$) and 3NS ($p = 0.007$) conditions, but significantly better than the 3S condition ($p = 0.04$). This indicates a similar pattern of performance as in Experiment I, with the exception that this group of listeners may still be relying on cues of laterality in tuning into the 3 matched even bands, since performance was better with the dichotic presentation. The data also indicate that information in the 3 odd bands, when presented unshifted, may be more rich in speech cues than that in the 3 even bands, since subjects showed significantly better performance with the 3NS condition compared to the 3US condition at each test session.

An analysis of covariance was performed to test for a possible interaction between the amount of training and the processor condition. The analysis examined the DOS and 3US conditions, and showed a highly significant ($p < 0.001$) effect of training, but no effect of condition, and no significant condition by training interaction, nor condition by training by talker interaction. This indicates that for both talkers, the pattern of learning was the same, and that this was also similar for these two conditions.

Sentence Perception – Post-training High Context Sentences

Figure 7 shows scores for the easier BKB and IHR sentences at the end of training. A repeated measures ANOVA showed a significant main effect of processor [$F(5, 23) = 60.0$, ($p < 0.001$)]. There was no significant interaction between talker and processor, so analysis is for the two talkers combined.

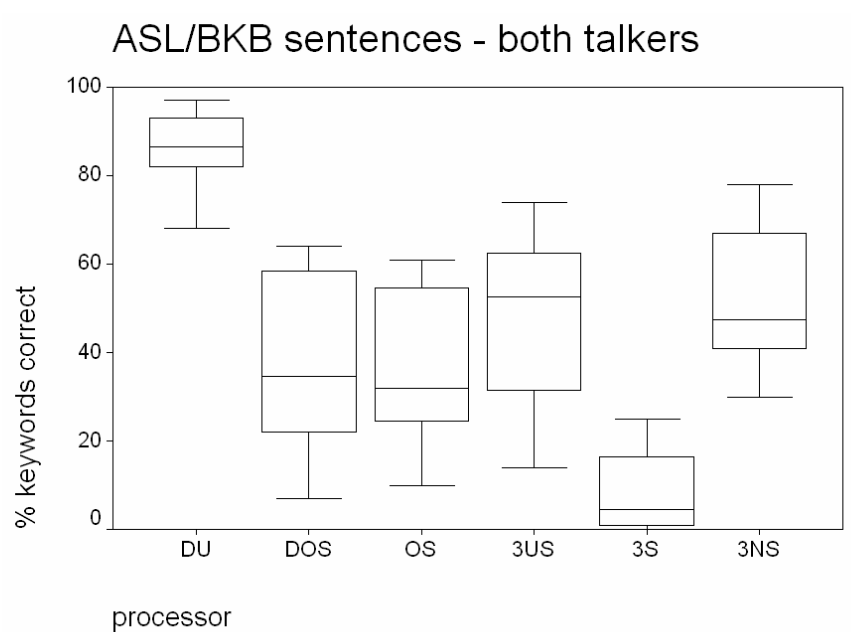

Figure 7: Scores with ASL and BKB sentences after training in experiment 2 as a function of processor.

Bonferroni-adjusted paired comparisons showed that intelligibility with the 3US condition was significantly worse only from the DU condition (p = 0.006), and significantly better only than the 3S (p = 0.008) conditions. No significant difference was found between the 3US processor and the DOS, OS and 3NS processors. This is very similar to the pattern seen with the IEEE sentences, although for these sentences, diotic versus dichotic presentation did not affect intelligibility. Also, for these talkers and this material, there was no difference in intelligibility between the 3 odd (3NS) and 3 even unshifted bands (3US).

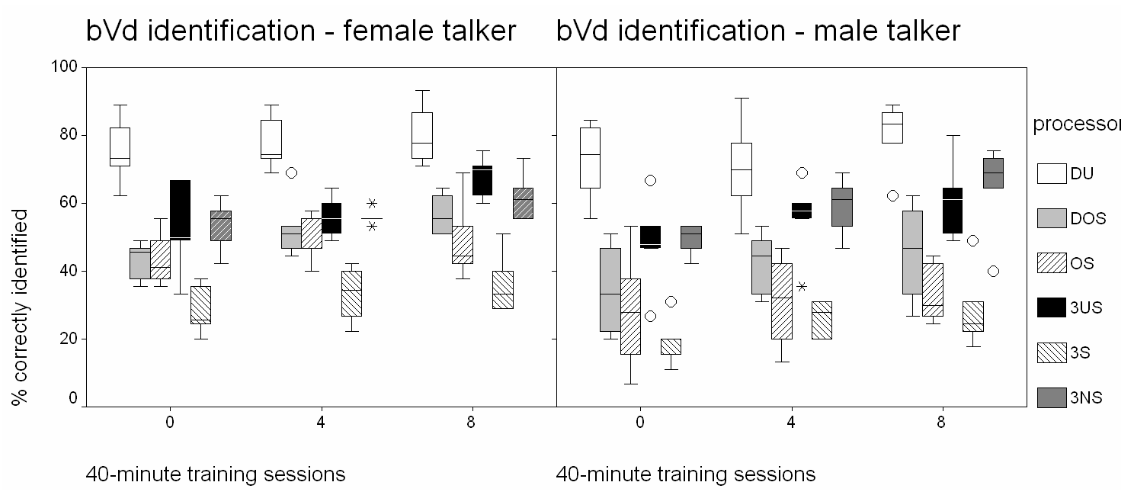

Figure 8: vowel identification in experiment 2 for female and male talkers as a function of training and processor.

Vowel Identification

Figure 8 shows vowel identification for the two talkers separately. A repeated measures ANOVA showed significant main effects of session [$F(2, 10) = 23.4$, ($p < 0.001$)], talker [$F(1, 5) = 8.25$, ($p = 0.035$)] and processor [$F(4, 19) = 54.4$, ($p < 0.001$)]. The ANOVA also showed significant session by processor [$F(10, 50) = 60.9$), ($p = 0.019$)] and talker by processor [$F(5, 25) = 3.58$, ($p = 0.014$)] interactions. The session by condition interaction can be largely accounted for because of the lack of improvement across sessions in some of the conditions (i.e. OS, 3S and 3NS conditions), whereas the remaining conditions did show significant improvements.

Because of the talker by processor interaction, data from the two talkers were analyzed in separate repeated measures ANOVAs. These again showed significant main effects for both 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*: $F(4, 19) = 33.7$, ($p < 0.001$)]. When analyzed separately by talker, the session by processor interaction was no longer significant.

For all conditions, scores for the male talker were somewhat lower than those for the female talker. For the female talker, Bonferroni-adjusted paired comparisons on the post training scores revealed no significant difference between the 3US processor and any of the processors except the 3S processor, for which performance was significantly better (p = 0.006).

With the female talker, Bonferroni-adjusted paired comparisons of sessions within each condition showed significant improvements across training in the DU (p = 0.019), DOS (p =0.004), and 3S (p = 0.027) processors, but not with the OS, 3US and 3NS processors. For the male talker, the same analysis showed significant improvements in the DU (p = 0.038) and DOS (p = 0.020) processors, but not for the 3US, OS, 3S and NS processors.

A univariate analysis of covariance (ANCOVA) was performed on the DOS and 3US conditions to determine any interactions between amount of training and performance with each processor. While a highly significant effect of training was shown, there was no significant effect of processor, or training by processor, processor by talker, or processor by talker by training interactions.

C. Discussion

Despite undergoing five hours and twenty minutes of training with the 3 odd spectrally shifted frequency bands, subjects did not seem to show any adaptation to this material as indicated in the results from all test materials. This is despite the fact that from the outset, intelligibility with the unshifted odd 3 frequency bands was significantly better than that with the 3 even unshifted bands, indicating these 3 bands contain spectrally rich information sufficient for speech perception. While we hypothesized that allowing listeners time to adapt to the more difficult 3S processor might better equip them to integrate this information to improve performance with the DOS test processor, the data does not support this hypothesis. Subjects in this group never achieved performance better than the 3US reference processor except for the 6-channel unshifted DU processor. This is somewhat surprising, given the evidence from previous studies with spectrally shifted speech. While acute studies show listeners are resistant to learning spectrally shifted speech, accumulating evidence has demonstrated this can be overcome by the kind of training used in this experiment (Faulkner et al., 2006; Faulkner, Rosen, & Stanton, 2003; Faulkner, Rosen, & Wilkinson, 2006; Faulkner, 2006; Rosen et al., 1999; Smith et al., 2006).

In this experiment, the training the subjects received, when considered on its own, was with a 3-channel vocoder, albeit with non-contiguous analysis filters. This is not so different from the 4-channel vocoder described in Rosen, Faulkner and Wilkinson (1999). While their vocoder used more channels, the first channel was so low in frequency that its contribution to speech perception was questionable. Moreover, they only trained subjects for three hours, whereas subjects in this experiment were trained for nearly double that amount of time.

It is possible that improvements shown were a result of experience with the test materials rather than experience received through the training. While subjects did not show significant improvements across training with the 3S condition, despite being trained with this condition, they did show improvements in the other conditions. The only experience with unshifted vocoded speech these subjects had was during testing.

III. COMPARISON OF DATA FROM EXPERIMENT I AND EXPERIMENT II

While the analyses of the subject groups from Experiments I and II revealed interesting patterns of adaptation on their own, the primary aim of the second experiment was to determine whether training with the 3S processor would facilitate adaptation more than training with the DOS processor. In order to examine this question, then, a statistical comparison of the two subject groups was also performed. Data from Experiments I and II were analyzed using a repeated measures analysis of variance, this time entering training paradigm as a between-subjects factor. The 3NS processor tested in Experiment II was excluded from these analyses. This condition was primarily a reference condition for Experiment II only, so is not pertinent to this particular analysis.

The between groups analysis was computed on the improvement scores across training rather than on raw intelligibility scores. The question in this comparison is which training method of the two is better, so improvement scores are a better measure of adaptation.

A. IEEE Sentence Perception

For each condition, improvement scores across training were computed by subtracting mean keywords correct before training from the mean keywords correct score after 8 training sessions. Boxplots showing improvement scores by condition for each group of subjects are given in Figure 9. Improvement scores were entered into a repeated-measures ANOVA, with within subjects factors of talker and processor, and a between subjects factor of training paradigm. This showed a significant main effect of processor [$F(4, 35) = 8.83$, $p < 0.001$] and a significant processor by training paradigm interaction [$F(4, 35) = 3.76$, $p = 0.015$]. A between-groups Bonferroni-adjusted paired comparisons on the improvement scores showed that the only condition for which training paradigm was a statistically significant factor was for the DOS condition ($p = 0.005$).

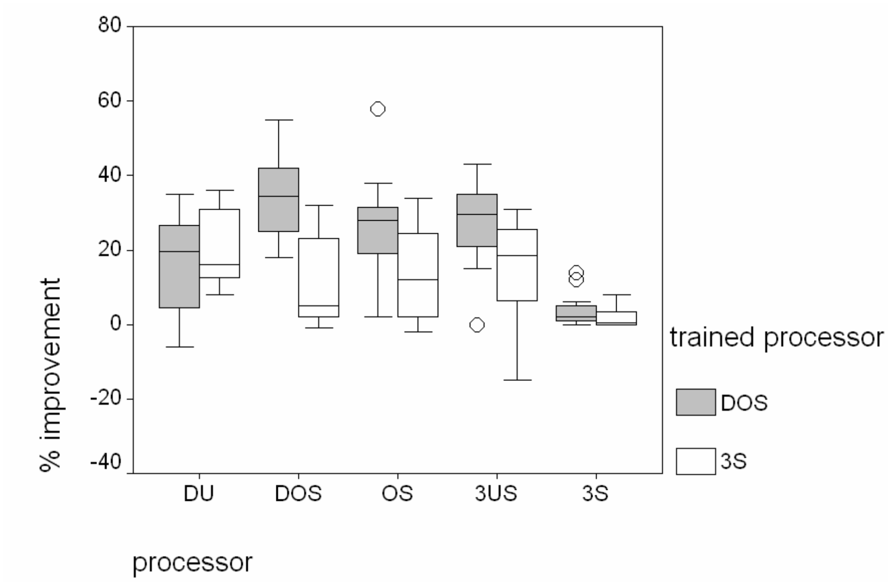

Figure 9: Increase in IEEE sentence scores with training for experiment 1 (training with DOS processor) and experiment 2 (training with 3S processor) as a function of the processor used in testing.

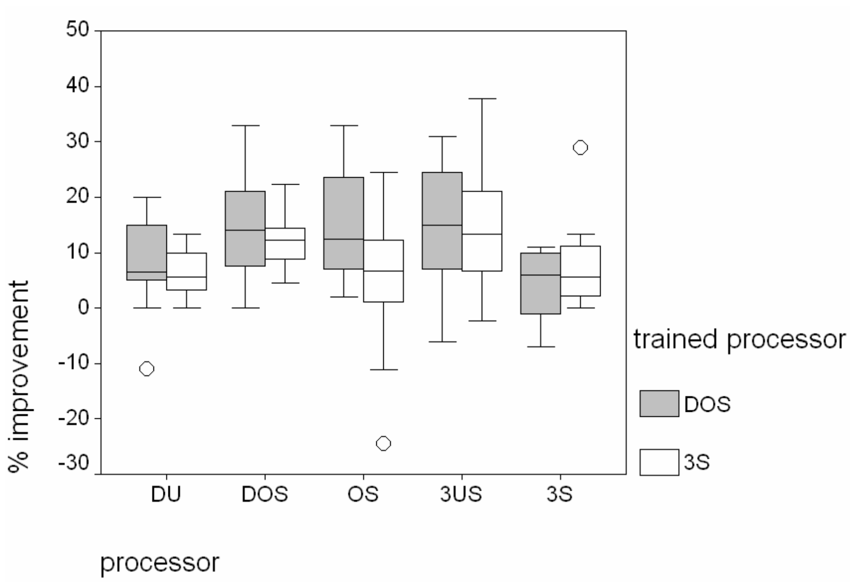

Figure 10: Increase in vowel identification scores with training for experiment 1 (training with DOS processor) and experiment 2 (training with 3S processor) as a function of the processor used in testing.

B. Vowel Identification

As for the IEEE sentence test, improvement scores across training were computed by subtracting the mean number of vowels identified correctly before training from the mean correctly identified after the 8 training sessions. The improvement scores by condition for each subject group are shown in the box-plots in Figure 10. Improvement scores were analysed with a repeated-measures ANOVA, with within subjects factors of talker and processor, and a between-subjects factor of training paradigm. This showed a significant main effect of processor [$F(4, 40) = 4.78$, $p = 0.003$]. There were no other significant main effects or interactions. The between-groups Bonferroni-adjusted paired-comparisons on these improvement scores revealed no significant differences between the two groups of subjects in any of the processing conditions.

To further examine the effects of training paradigm on vowel identification, an analysis of the vowel confusions was performed using a chi-square test. This analysis was restricted to the DOS, 3US and 3S processing conditions. For each group of subjects, a confusion matrix was computed. A reference matrix was computed by calculating the mean between each subject group for each cell. This was used for a chi-square analysis, with 80 degrees of freedom. For each of these three processing conditions, the analysis revealed no significant difference in the patterns of confusions of these two subject groups.

Despite the lack of difference between the subject groups, the confusion matrices for these three conditions display quite distinct patterns. Bubble plots for the two subject groups combined are given in Figures 11 a-c, where the stimulus is plotted along the ordinate, and the response on the abscissa. In these plots, strong diagonals denote high levels of identification, whereas strong verticals may denote a response bias. The bubble plot for the 3US condition shows quite a strong diagonal. This is also true, albeit to a lesser extent, for the DOS condition. However, in the 3S condition, the pattern of data appears to be randomly distributed, suggesting that subjects did not demonstrate perceptual adaptation, since they appear to be employing a guessing strategy for this task in this condition.

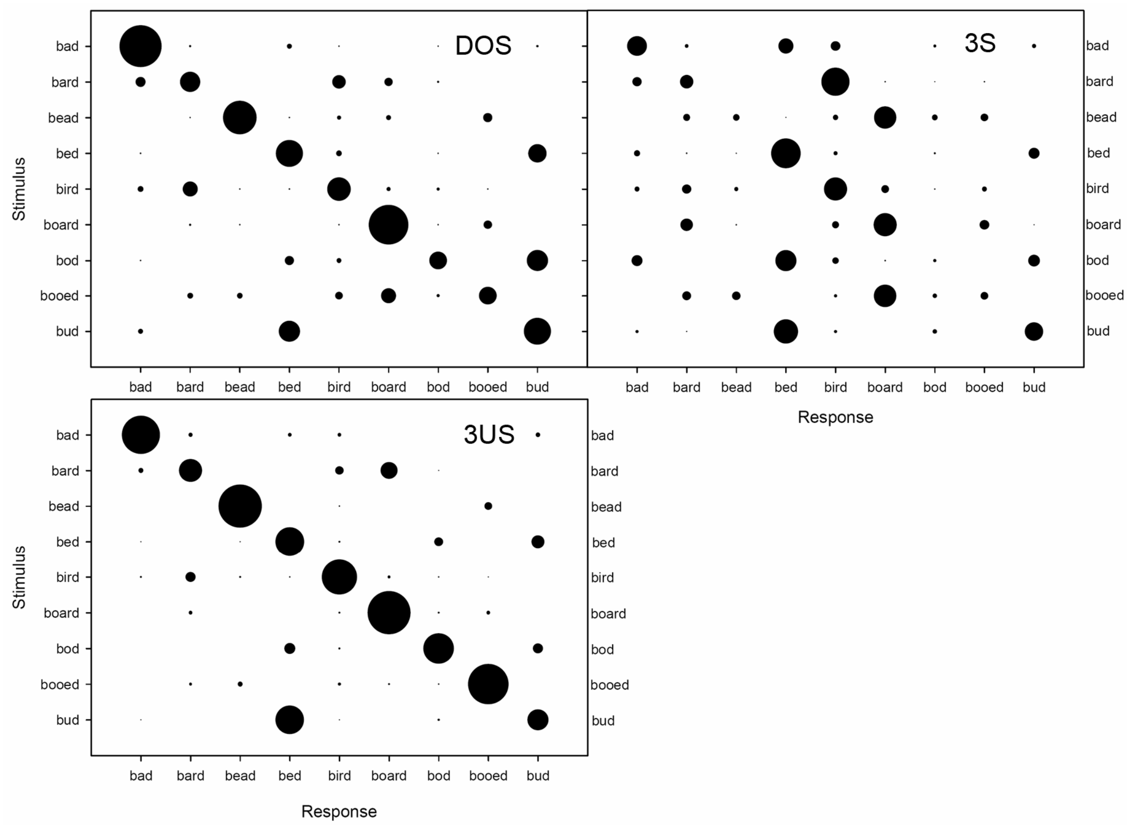

Figure 11: Bubble plots showing post-training vowel confusions. Data from experiments 1 and 2 are combined. The 3 panels show confusions for the DOS processor (upper left), the 3S processor (upper right) and the 3US processor (lower left).

C. Discussion

From the individual analyses of each group of subjects, and the comparison of the two groups, it seems clear that training with just the three shifted conditions does not yield any benefit over training with the DOS condition. Despite some evidence of perceptual adaptation, subjects trained with either processor never achieve a performance greater than that with the 3 unshifted channels (3US). Moreover, they do not demonstrate anywhere near the level of improvement in intelligibility with the 3 shifted channels (3S) after 5 hours, 20 minutes of training, that previous studies with spectrally shifted speech have demonstrated in just three hours.

Because of this discrepancy, it seems possible that the task of integrating mismatched frequency place maps may require a time course for adaptation much longer than previously demonstrated for these other processors. While improvements shown in these first two experiments were small, they did not reach an asymptotic level, so it seems plausible that subjects are still adapting to the processors after training in these experiments was complete. Accumulating evidence from cochlear implant patients is consistent with the hypothesis of long-term adaptation (Svirsky et al., 2005; Tyler & Summerfield, 1996).

There are several aspects of the dichotic shifted processor that are distinct from the spectral shifting examined in previous studies, and these may prescribe a longer time course for training. While the analysis filters cover the frequency range of 200 – 5000 Hz continuously, the channels of the output processor are divided between the ears and interleaved. This creates two effective 3-channel vocoders comprised of non-contiguous channels. When presented dichotically and unshifted, this does not introduce any obstacles to speech perception (Loizou et al., 2003; Faulkner, 2006). However, when a spectral shift is introduced to the untrained ear, the signals at the two ears are no longer perceived as a single contiguous representation of the signal, but rather two distinct signals. From the patterns of intelligibility throughout training for both sub-groups, it appears that subjects adapt relatively quickly to the tonotopically matched 3US processor. However, data from Experiments I and II indicate that subjects do not use information from the 3S processor. The introduction of the spectral shift means that frequency information has been shifted in both frequency and relative order (i.e. 2 – 1 – 4 – 3 – 6 – 5), and listeners seem resistant to learning this type of spectral warping. In all conditions, they learn to pick out the tonotopically matched frequencies, treating the shifted frequencies as noise in the signal. It is not clear, however, whether this is because the time course of training has been too short, or whether specific aspects of the processor are too difficult to learn, or whether there is some interaction among all of these things. We therefore conducted a third experiment to examine a combination of these factors. If the limit on perceptual adaptation lies in the time course of the adaptation, then we should start to see an acceleration in intelligibility with the DOS processor compared to the 3US processor over long-term training. However, if there is something about the spectral mismatch introduced in the signal that subjects cannot adapt to, then we should see results similar to Experiments I and II.

IV. EXPERIMENT III

The aim of Experiment III was to explore perceptual adaptation to the DOS processor after an extended period of training – double that examined in Experiments I and II. Training for this experiment employed both the DOS and 3S processors, so that subjects would have extensive experience with both the shifted (3S) processor and the dichotic odd shifted (DOS) processors. Performance with both easy and difficult sentence sets as well as the vowel material was monitored throughout training to examine the progression of adaptation.

A. Method

1. Subjects

Six normally hearing speakers of British English took part, and each was paid for his or her participation. They all had normal (< 20 dBHL) pure-tone thresholds at .5, 1, 2 and 4 kHz.

2. Signal Processing

Signal processing was the same as for Experiments I and II, using both experimental processors during training and most of the same test materials.

3. Training

Training for Experiment III was similar to that in Experiments I and II, with the exception that subjects were trained in 30 20-minute sessions (10 hours total). Each session was divided into 10 minute blocks which alternated between the DOS and 3S processors. Author K. M., a speaker of Standard Southern British English, was the speaker for Experiment III.

Session	Training	Processor	Testing
0	5 minutes AV 5 minutes AA	DU	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)
1	—	—	*Second 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)
3 - 17	10 minutes x 2	DOS + 3S	None
18	—	—	*Mid-test:* IEEE sentences (2 lists x 4 conditions x 2 talkers) ASL sentences (2 lists x 3 conditions x 2 talkers) bVd identification (5 tokens x 5 conditions x 2 talkers)
19 - 33	10 minutes x 2	DOS + 3S	none
34	—	—	*Post-test:* IEEE sentences (2 lists x 4 conditions x 2 talkers) ASL sentences (2 lists x 3 conditions x 2 talkers) bVd identification (5 tokens x 5 conditions x 2 talkers)

Table IV: sequence of training and testing conditions for experiment 3

4. Testing

Experiment III tested both easy (ASL) and difficult (IEEE) sentence material throughout training, as well as the 9 vowel identification test. All test materials were the same as for Experiments I and II. The exception to this is that the full set of 18 ASL lists was used rather than a subset of the BKB and ASL sentences. However, the same talkers were used for this set as for Experiment II. As for Experiments I and II, subjects were given a familiarization block for the vowel test with unprocessed speech during the first pretest session.

5. Conditions

In Experiment III, we limited the testing to four conditions – the DOS, 3S, 3US and DU conditions. Because we anticipated ceiling effects with the easy sentence material and the unshifted processor, this condition was not tested for the ASL sentence set.

6. Procedure

Table 4 gives a description of the testing and training regime used for Experiment III. A 10 minute familiarization training block with the DU processor was given prior to beginning the experiment. This was broken down into 5 minutes of AV presentation and five minutes of AA presentation. A familiarization test session was also included - Session 1 – in which subjects were tested with the IEEE sentences, the vowels and the BKB sentences. 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)). Session 2 was largely a repeat of the first session, however the ASL sentences were tested rather than the BKB sentences. In Sessions 3 – 17, subjects were trained alternately with the DOS and 3S processors in 10 minute blocks. Training was auditory-only in presentation. The rule for switching to AV presentation if performance fell below a 10 words per minute threshold was in place, but it was never necessary to invoke this rule. In Session 18, subjects were tested again with the IEEE sentences, the ASL sentences and the vowel identification. Sessions 19 – 33 were further training sessions. Finally, in Session 34, subjects were tested again with the IEEE sentences, ASL sentences and vowels. The order of testing was counterbalanced across subjects. All other procedural considerations were the same as for Experiments I and II.

B. Results

Connected Discourse Tracking

CDT rates over training sessions for the dichotic odd-shifted channels (DOS) and 3 shifted channels (3S) processors are shown in Figure 12. These represent auditory-only scores. For

DOS presentation, the mean number of words tracked per minute across all training sessions was 63.2 (S.D. 16.6). For 3S presentation, mean number of words tracked per minute was 34.8 (S.D. 12.1). Tracking with both processors continues to increase with training. Midpoint CDT rates are similar to Experiments I (for DOS blocks) and II (for 3S blocks), while endpoint CDT rates are slightly higher.

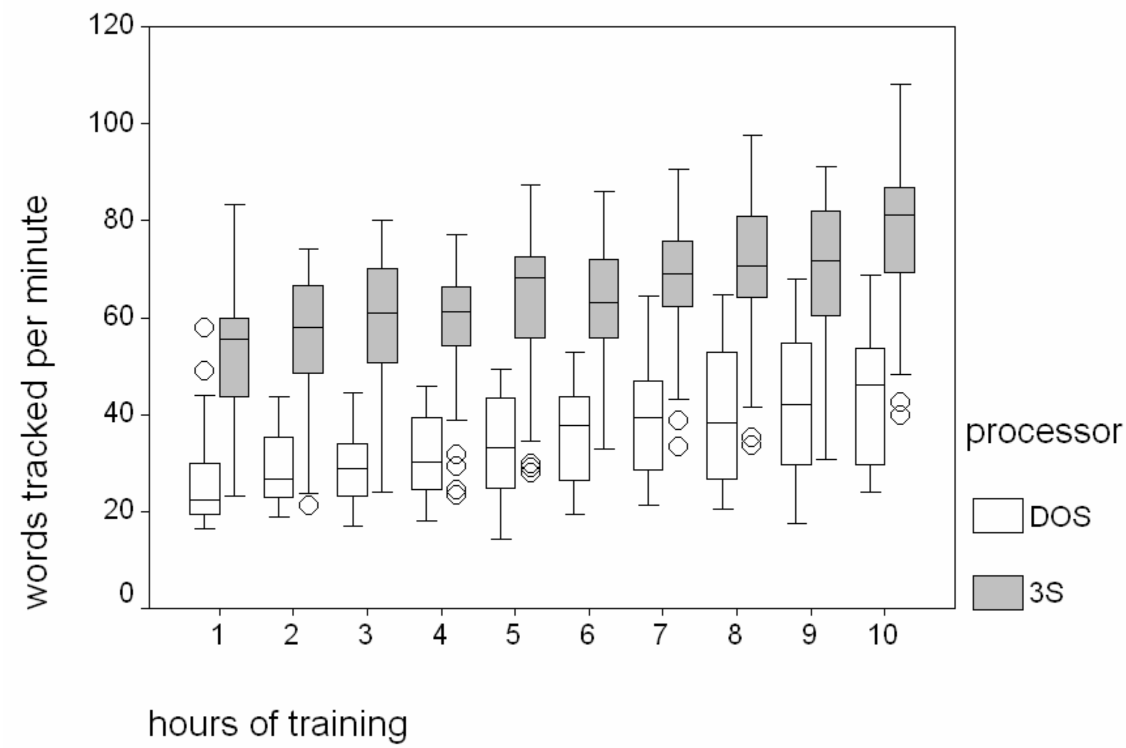

Figure 12: CDT rates in experiment 3 over training. Only auditory performance is shown in the figure.

Sentence Perception – IEEE sentences

Keywords correct scores for the sentence tests across training sessions are shown for the two talkers combined in Figure 13. A repeated measures ANOVA on the first two sessions was computed. This showed significant effects of talker [$F(1, 5) = 18.1$, $p = 0.008$] and processor [$F(1, 7) = 41.9$, $p < 0.001$], and a significant session by condition interaction [$F(2, 11) = 6.43$, $p = 0.012$]. However, a Bonferroni-adjusted paired comparison of the intelligibility scores by session within each condition) revealed no significant difference in intelligibility between these sessions in any of the conditions. A second repeated measures ANOVA was computed on the final three sessions. This 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 [$F(1, 5) = 72.2$, $p < 0.001$]. There was a significant session by talker interaction [$F(2, 10) = 6.85$, $p = 0.013$], although no significant talker by processor interaction, so analyses shown here are for the two talkers combined. There was also no significant session by condition interaction.

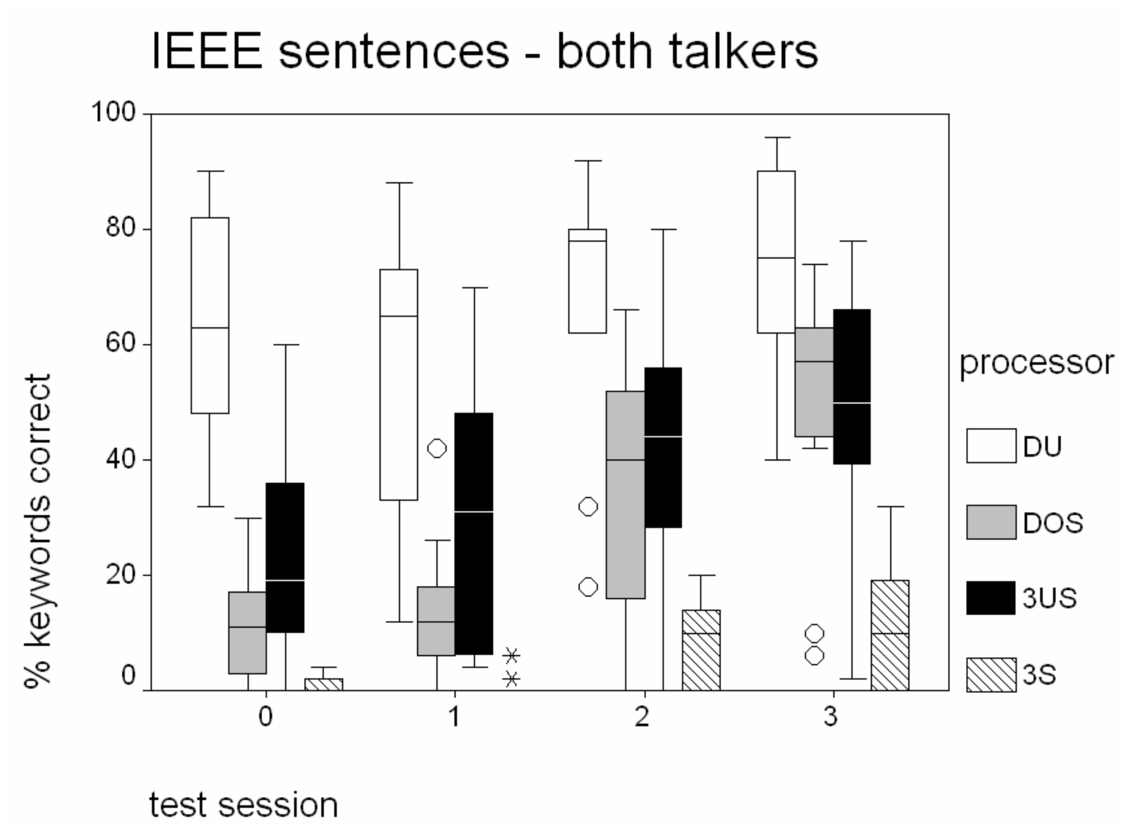

Figure 13: IEEE sentence scores from experiment 3 over session for each processor. Sessions 0 and 1 are both prior to training.

As in Experiments I and II, performance in the DOS condition never exceeded that in the 3US condition. Bonferroni-adjusted paired comparisons on the session by condition interaction (with respect to condition) showed that subjects improved significantly across training only with the DU ($p = 0.015$) and DOS processors ($p = 0.009$). They did not show significant improvements across training sessions in the 3US and 3S conditions, although performance did increase.

Bonferroni-adjusted paired comparisons on the post-training sentence scores revealed that performance with the 3US processor was significantly worse than with the DU processor (*p* = 0.02) and significantly better than the 3S processor (*p* = 0.02). There was, however, no significant difference in intelligibility with the 3US processor and the DOS processor after training.

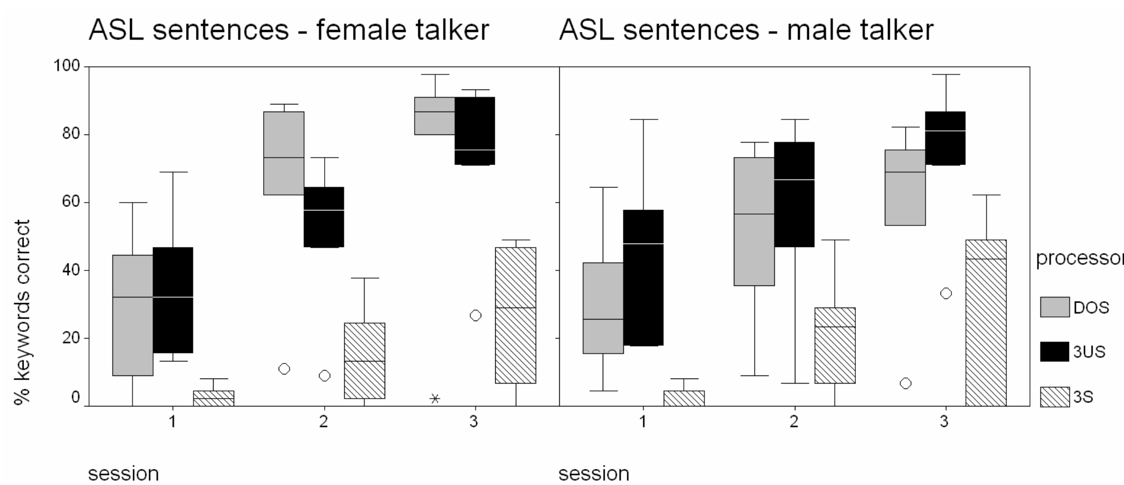

Figure 14: ASL sentence scores in experiment 3 as a function of session and processor.

Sentence Perception – High Context Sentences

Figure 14 shows scores for the IHR-ASL sentence sets for the male and female talkers individually. A repeated measures ANOVA showed a significant main effect of session [F(1, 6) = , p = 0.002] and processor [F(2, 8) = 31.4, p < 0.001]. There was also a significant interaction between talker and processor [F(2, 8) = 6.58, p = 0.022]. Because of this interaction, scores for each talker were entered into two separate ANOVAs. These showed significant effects of training [*female:* F(1, 6) = 23.9, p = 0.003; *male:* F(1, 6) = 21.7, p = 0.002] and processor [*female:* F(1, 7) = 23.6, p = 0.001; *male:* F(2, 10) = 33.4, p < 0.001].

For the female talker, there was also a significant session by condition interaction [$F(4, 20) = 3.62$, $p = 0.023$]. Data from the male talker did not show this interaction.

For the female talker, Bonferroni-adjusted paired comparisons showed significant improvements across training in the DOS ($p = 0.018$) and 3US ($p = 0.017$). Improvement across training in the 3S condition was also near significance ($p = 0.063$). Data from the male talker showed a slightly different pattern. Bonferroni-adjusted paired comparisons across training sessions with respect to condition showed significant improvement in only the DOS condition ($p = 0.042$). Improvement in the 3US condition was near significance ($p = 0.06$. These differences account for the interaction of session and talker.

For both talkers, Bonferroni-adjusted paired comparisons on the post-training scores showed that there was no significant difference in intelligibility between the 3US condition and the DOS condition ($p > 0.05$). However, intelligibility with the 3US processor was significantly higher than the 3S processor for both talkers [*female*: $p = 0.005$; *male:* $p = 0.004$] after training. This was also true of the DOS processor for the female talker ($p = 0.022$).

Vowel Identification

Boxplots of vowel identification for the male and female talkers combined are given in Figure 15. A repeated measures ANOVA on the first two sessions was computed. This showed significant main effects of talker [$F(1, 5) = 13.5$, $p = 0.014$] and processor [$F(3, 15) = 59.8$, $p < 0.001$]. However, Bonferroni-adjusted paired comparisons with respect to condition did not reveal any significant differences in any of the conditions between these two training sessions.

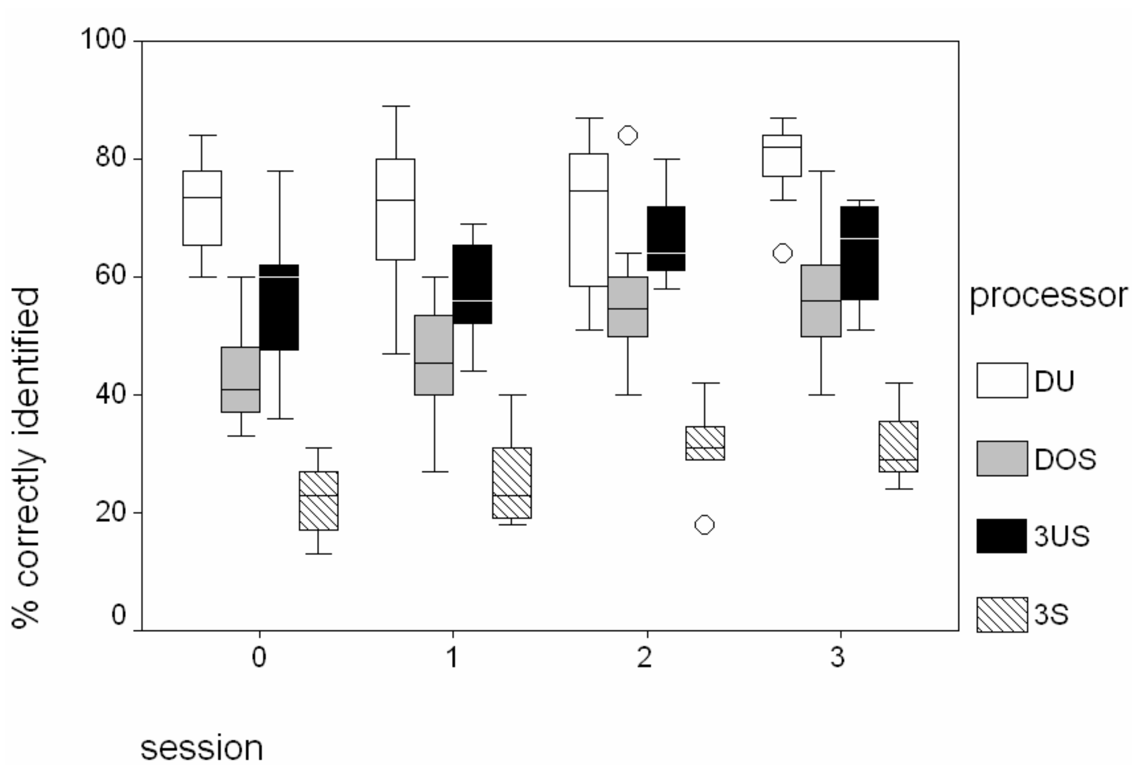

Figure 15: vowel identification in experiment 3 as a function of test session and processor. Sessions 0 and 1 were both prior to training

A second repeated measures ANOVA was computed on the final three test sessions. This 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 [$F(3, 15) = 77.8$, $p < 0.001$]. There were no significant interactions. A Bonferroni-adjusted paired comparison of session with respect to condition showed that intelligibility with the DOS condition significantly improved across training ($p = 0.002$). However, intelligibility with the other processors did not differ significantly across training. Bonferroni-adjusted paired comparison on the post-training scores revealed no significant difference in intelligibility with the 3US processor and the DOS processor. However, intelligibility with this processor remained significantly less than with the DU processor ($p = 0.014$), and significantly greater than the 3S processor ($p < 0.001$).

C. Discussion

Despite the extensive period allowed for training in Experiment III, subjects still never performed better with the dichotic mismatched processor (DOS) than with the 3 matched channels alone (3US), suggesting that during the initial period of learning with this stimuli, they are ignoring the information from the 3 shifted channels. This is despite having an equal amount of explicit training time with this processor as with the dichotic mismatched processor. However, even after 10 hours of training, subjects' performance with these two processors had still not reached a plateau, suggesting that they may still be learning more about this 3 non-contiguous channel speech processor.

Data from the easier sentences in this experiment show greater signs of adaptation than in the previous two experiments: performance with just the three shifted channels (3S) after 10 hours training is reaching the level of the 3 unshifted channels (3US) before training. Post-training scores for this subject group were also much higher in the 3S condition than for the group in Experiment II, although this depended somewhat on the talker. While some improvement in performance is expected due to familiarization with the task, this extent of improvement was not found in Experiment II, when subjects would have experienced a similar amount of testing and thus similar familiarization. In addition to the extended amount of training for Experiment III, subjects also received training with both the 3US and DOS processors. This type of training regimen may have been better suited to eliciting the desired adaptation to the dichotic-mismatched processor, which may not have been the case when trained with either of these processors on their own. Finally, the talker for training in Experiment III was a native British English speaker, whereas the talker in Experiments I and II was a native speaker of a Northeastern American dialect. It is possible that the talker from Experiment III was easier to understand, for reasons of dialect or for some unknown reason.

IV. GENERAL DISCUSSION

While results from Experiment III demonstrate some evidence of adaptation from the extended period of training, the data in these three experiments together indicate that at the very least, listeners are resistant to learning cochlear frequency-to-place maps that differ greatly between the ears. Because performance did not reach asymptote in any of the experiments, it remains to be determined whether training and testing for these experiments were conducted in the midst of the acute phase for this type of processing, 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, 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 in Tyler and Summerfield (1997) reached asymptotic performance an average of 30-40 months post implantation. Svirsky et al. (2005) used a method-of-adjustment procedure to monitor over time the perceptual vowel spaces of cochlear implant patients. They found that patients showed significantly altered vowel spaces immediately following implantation, but by the end of the study implantees' vowel spaces were near normal. The duration of the study, however, was 24 months, suggesting a very long time course for perceptual adaptation in CI listeners.

One distinguishing factor between the frequency-mapping examined here, and frequency mapping examined previously, is the non-contiguous presentation of the output frequency bands. Simulation studies that have examined tonotopicity, spectral shifting and frequency matching, as well as spectral holes in hearing, have all had output filters representing a continuous representation of the signal, although sometimes this has been warped, as in the case of EAS simulations. The processors examined here, when considered on their own, have outputs that are non-contiguous channels. When these are presented unshifted, listeners seem to adapt readily. However, the introduction of a spectral shift to the non-contiguous channels may introduce a level of abstraction and discontinuity that is much more difficult to learn, and thus may require more explicit feedback in learning. Because subjects are so resistant to adapting to the 3S processor, it is not surprising that they have not learned to integrate this to synergistic effect in the DOS processor over the time course of this study.

Despite up to 5 hours of training with the 3S processor, however, evidence of adaptation to the 3S processor begins to appear at the end of Experiment III. This Experiment was the only to combine the two processors in training, so it is possible that this method of training is necessary to facilitate adaptation.

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 et al., 2004; Litovsky et al., 2006). This is the most similar clinical situation to the simulations studied here.

Another explanation for resistant adaptation is that frequency shifts based on Greenwood's formula have been exaggerated. Newer cochlear implants position the electrodes near the modiolus where they stimulate cells in the spiral ganglion rather than the Organ of Corti, as is assumed in Greenwood's formula. If this is the case, then predictions of electrode CFs should be based on a map of the spiral ganglion (Sridhar et al., 2006).

A final point to consider is that the simulations used here were processed in quiet. 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. If the portion of the signal in these processors that was unshifted was presented in noise, listeners may have stronger motivation to rely on the 3 shifted channels. One example of this is from a recent study using simulations of electric-acoustic stimulation (EAS). The listeners demonstrated a greater super-additive effect with speech processed in noise for acoustic (low-pass filtered speech) and electric (tonotopically matched frequency map varying in degree of simulated insertion depths) hearing together over either condition on its own (Dorman, Spahr, Loizou, Dana, & Schmidt, 2005). The effects of binaurally mismatched speech in noise remain to be examined.

Acknowledgments

Work supported by a grant from the European Commission [FP6 – 004171 HEARCOM] as well as Study Assistance from the Staff Development and Training Unit, Department of Human Resources, UCL and a Wellcome Trust Summer Vacation Scholarship to K Mair..

ReferencesReferences

Bench, J. & Bamford, J. (1979). *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children*. London: Academic Press.

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

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

Ching, T. Y. C., Incerti, P., Hill, M., & 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.

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., & Hervais-Adelman, A. T. K. M. 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. & 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.

Dorman, M. F. & 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., & 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., & Rainey, D. (1997d). 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.

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997c). 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.

Dorman, M. F., Loizou, P. C., & 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.

Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J., & Schmidt, J. S. (2005). Acoustic simulations of combined electric and acoustic hearing (EAS). *Ear and Hearing, 26,* 371-380.

Faulkner, A. (2006). 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., & 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., Rosen, S., & 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., & Wilkinson, L. (2006). Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cohclear implant processor. *Ear and Hearing, 22,* 431-438.

Fishman, K. E., Shannon, R. V., & 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.

French, N. R. & 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., & 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., Galvin III, J. J., Wang, X., & Nogaki, G. (2005a). Moderate auditory training can improve speech performance of adult cochlear implant patients. Acoustics Research Letters Online 6[3], 106-111.
Ref Type: Journal (Full)

Fu, Q.-J., Nogaki, G., & 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. & 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.

Fu, Q.-J. & Shannon, R. V. (2002). Frequency mapping in cochlear implants. *Ear and Hearing, 23,* 339-348.

Fu, Q.-J., Shannon, R. V., & 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.

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.

Hamzavi, J., Pok, S. M., Gstoettner, W., & 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.

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

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., & 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.

Lawson, D., Brill, S., Wolford, R., Wilson, B., & Schatzer, R. (2000). *Speech Processors for Auditory Prostheses: Ninth quarterly progress report* (Rep. No. 9).

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

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

MacLeod, A. & 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.

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

Rosen, S., Faulkner, A., & 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.

Rubin, B. A., Uchanski, R. M., & Braida, L. D. (1992). Integration of acoustic cues for consonants across frequency bands (A). Journal of the Acoustical Society of America 91[4], 2473-2474.
Ref Type: Abstract

Shannon, R. V., Zeng, F.-G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America, 104,* 2467-2476.

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

Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A. et al. (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. & 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.

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

Svirsky, M. A., Silveira, A., Neuburger, H., Teoh, S.-W., Helms, J., & Suarez, H. (2005). Long-term auditory adaptation to a modified peripheral frequency map. *Acta Otolaryngologica, 124,* 381-386.

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

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