Resistance to learning binaurally mismatched frequency-to-place maps: Implications for bilateral stimulation with cochlear implants[a]

Catherine M. Siciliano,[b] Andrew Faulkner, Stuart Rosen, and Katharine Mair
Speech, Hearing and Phonetic Sciences, Division of Psychology and Language Sciences, UCL, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom

(Received 14 December 2007; revised 8 December 2009; accepted 22 December 2009)

Simulations of monaural cochlear implants in normal hearing listeners have shown that the deleterious effects of upward spectral shifting on speech perception can be overcome with training. This study simulates bilateral stimulation with a unilateral spectral shift to investigate whether listeners can adapt to upward-shifted speech information presented together with contralateral unshifted information. A six-channel, dichotic, interleaved sine-carrier vocoder simulated a binaurally mismatched frequency-to-place map. Odd channels were presented to one ear with an upward frequency shift equivalent to 6 mm on the basilar membrane, while even channels were presented to the contralateral ear unshifted. In Experiment 1, listeners were trained for 5.3 h with either the binaurally mismatched processor or with just the shifted monaural bands. In Experiment 2, the duration of training was 10 h, and the trained condition alternated between those of Experiment 1. While listeners showed learning in both experiments, intelligibility with the binaurally mismatched processor never exceeded, intelligibility with just the three unshifted bands, suggesting that listeners did not benefit from combining the mismatched maps, even though there was clear scope to do so. Frequency-place map alignment may thus be of importance when optimizing bilateral devices of the type studied here.
[DOI: 10.1121/1.3293002]

PACS number(s): 43.71.Ky, 43.71.Es, 43.66.Sr, 43.66.Rq [KWG] Pages: 1645–1660

I. INTRODUCTION

For users of a cochlear implant (CI), the restoration of binaural hearing could provide a range of advantages including binaural squelch, the head-shadow effect, and increased stimulus redundancy, all of which may lead to better understanding of speech in noisy listening environments. Even though current CIs do not provide the fine structure information that contributes to binaural processing in normal hearing, a binaural advantage for speech perception has been reported in patients with bilateral cochlear implants (Dorman and Dahlstrom, 2004; Tyler *et al.*, 2005; Litovsky *et al.*, 2006; Wackym *et al.*, 2007; Tyler *et al.*, 2007) and for implants used in conjunction with acoustic hearing aids (HAs) (Iwaki *et al.*, 2004; Ching *et al.*, 2004; Hamzavi *et al.*, 2004; Ching *et al.*, 2005; Ching, 2005; Ching *et al.*, 2006).

While the potential benefits are manifold, the bilateral use of auditory devices raises new questions in CI frequency-place mapping. Cochlear implant electrode arrays are usually designed for an insertion depth of 25 mm into the typically 35 mm long cochlea. In many cases, however, the insertion achieved is shallower than this. Estimates based on *in vivo* computed tomography from 26 Nucleus-22 implant recipients showed insertion depths ranging from 11.9 to 25.9 mm, as well as considerable variation in cochlear length from 29.1 to 37.4 mm (Ketten *et al.*, 1998; Skinner *et al.*, 2002). The topography of the neural elements stimulated by the electrode contacts is not at present completely understood. The standard approach for estimating the effective characteristic frequency (CF) at each electrode contact has followed the frequency-to-place mapping of the organ of Corti established by Greenwood (1990). At the median insertion depth of 20 mm found by Ketten *et al.* (1998), the Greenwood map leads to an estimated CF of 1000 Hz. However, this map is not a realistic model for the CFs of CI electrodes placed near the modiolus, for which a spiral ganglion map seems more appropriate (Stakhovskaya *et al.*, 2007). 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, which would require imaging of the cochlea and the electrode array for each subject.

Notwithstanding our incomplete knowledge of the effective CFs along an electrode array, the altered frequency-to-place mapping resulting from incomplete insertion of the implant electrode array and band-limited speech processing has demonstrable and sizable effects for monaural CIs (Skinner *et al.*, 2002). This has led to an ongoing debate about whether it is best to preserve tonotopic matching at the ex-

---

[a] Portions of this work were presented at the 2006 Midwinter Meeting of the Association for Research in Otolaryngology, the 2007 Conference on Implantable Auditory Prostheses, and the 2007 British Society of Audiologists Short Papers Meeting on Experimental Studies of Hearing and Deafness.

[b] Author to whom correspondence should be addressed. Present address: Audiology and Deafness Research Group, School of Psychological Sciences, University of Manchester, Manchester M13 9PL, United Kingdom. Electronic mail: sicilian@post.harvard.edu

pense of a frequency-shifted map that is maximally informative for speech. Earlier acute studies using vocoder-based simulations of cochlear implant speech processing suggested that a basalward frequency shift larger than 3 mm, using an organ of Corti map, leads to large decreases in speech intelligibility (Dorman *et al.*, 1997a; Shannon *et al.*, 1998). Analogous downward shifts of CI processor analysis filter frequencies have been shown to have a similar effect on vowel recognition by CI patients (Fu and Shannon, 1999). Yet, more recent studies with both noise-vocoded simulations (Rosen *et al.*, 1999; Fu *et al.*, 2005b; Faulkner *et al.*, 2006; Smith and Faulkner, 2006) and cochlear implant patients (Fu *et al.*, 2002; Svirsky *et al.*, 2004; Fu *et al.*, 2005a) have allowed listeners time to adapt to altered frequency-place maps and have shown improved performance after a period of training. For example, Rosen *et al.* (1999) trained normal hearing listeners with noise-vocoded simulations of a 6.5 mm basalward shift and found that word recognition had improved from floor to 30% after 3 h of training.

However, with bilateral implants or a cochlear implant in combination with a contralateral HA, the 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 with a limited frequency range), while the CI ear is likely to be subject to basalward place shifting. To what extent are listeners able to adapt when presented with frequency-place maps that differ between the two ears?

Dorman and Dahlstrom (2004) reported a binaural advantage for speech perception in two bilateral implant patients who had different cochlear implants in each ear, which may support the hypothesis that information from mismatched frequency-to-place maps can be combined. Patients showed improvements of 32%–34% on HINT sentences with the addition of the second implant over performance with the better ear alone. However, the study only included two subjects, and the method for determining mismatch between the ears—pitch-ranking of electrodes—was inexact. The authors conceded that the degree of mismatch between the ears was unknown, and thus the question of adaptation to such a mapping remains open.

Evidence from dichotic listening experiments suggests that speech information presented in complement across the two ears is easily integrated. Broadbent and Ladefoged (1957) showed that listeners presented with the F1 and F2 of a /da/ syllable separately to opposite ears perceived the syllable as /da/. This process, later termed *spectral fusion* by Cutting (1976), is robust to 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 *et al.*, 1992).

To the extent that the information delivered to each ear with bilateral CIs is in complement, a binaural advantage for speech perception should be expected. However, this might be achieved through various possible mechanisms. For example, the binaural combination of information is likely to be relatively easy if there are similar basalward shifts to each ear. Alternatively, if listeners were able to adapt to different degrees of basalward shift in each ear, then they might still be able to combine spectral cues from the two ears to show a binaural speech advantage: information that may have been perceived as conflicting initially may eventually be perceived as complementary after experience with the implants. However, many bilateral CI users do not show a binaural advantage for speech over the best ear alone for speech in quiet (Tyler *et al.*, 2007) or for speech in noise in the absence of directional differences (van Hoesel *et al.*, 2002; Wackym *et al.*, 2007). A recent study suggests that better-ear listening is especially apparent for bilateral CI users who show asymmetrical monaural speech scores in each ear (Mosnier *et al.*, 2009). It thus seems plausible that misaligned frequency-place maps between each ear may underpin a lack of binaural advantage for speech shown in some bilateral CI users.

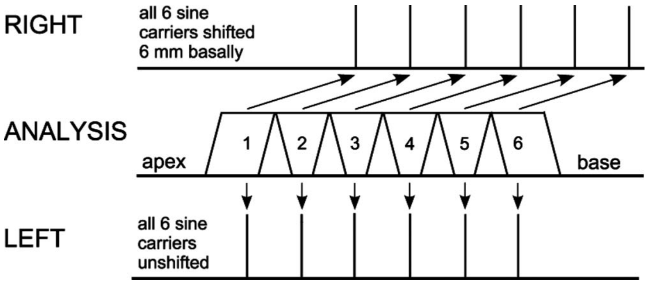

FIG. 1. Speech processing depicting sine-vocoded simulations of bilateral cochlear implants. Here, all analysis channels are presented to both ears, with the sine carriers at the right ear being upwardly shifted by an equivalent of 6 mm basilar membrane difference. In this configuration, the spectral information in each analysis band is delivered to different places in each ear. Further, the carrier frequencies for bands 1–4 in the right ear are the same as the carrier frequencies for bands 3–6 in the left ear so that there is an inter-aural conflict at each of these places.

A. Considerations for simulation methods sensitive to binaural advantage

Compared to monaural stimulation with cochlear implants, bilateral stimulation introduces complicating factors that may lead to decrements in speech perception. Consider the simplest case where contiguous frequency bands are vocoded and presented to each ear with an upward shift to one ear only, as in the bilateral six-channel sine vocoder illustrated in Fig. 1. Two distinct aspects of mapping conflict are evident. First, for each analysis band, the same spectral content is delivered to different tonotopic locations in each ear. Second, the carrier frequencies for bands 1–4 in the shifted ear are tonotopically aligned to the carriers for bands 3–6 in the unshifted ear so that the spectral information carried at each of these corresponding places is different in the two ears. The second of these conflicts may in addition lead to central masking (Mills *et al.*, 1996). A related but conceptually distinct global consequence of a mismatch of frequency mapping between the ears is that the central processing of spectral speech features cannot integrate information from the two ears without reference to an ear-specific transformation of the peripheral excitation pattern which needs to be learned.

A bilateral implant simulation of the type shown in Fig. 1, while representative of the information presented to an implant user, is not ideal for the study of adaptation to inter-aural mismatches of frequency mapping in normally hearing listeners. The primary difficulty with such a simulation is the redundancy of information between the shifted and unshifted ears. For tonotopically aligned speech in quiet, a binaural advantage can only be expected when there is some non-redundant information in each ear that can be exploited. The binaural advantage shown in some users of bilateral cochlear implants may arise from inter-aural variations in nerve survival or differences in spatial specificity at each ear that result in a sparcity of redundant information in each ear. However, for the normal hearing listener the signals will be completely redundant, which is especially problematic in a study where the intent is to encourage adaptation to a spectral shift to one ear only, because the information that can be gained by such adaptation is minimal. A second limitation of a bilateral implant simulation that employes the same contiguous analysis bands for both ears is that the local conflicts noted above are confounded with the global conflict that arises from the interpretation of two mismatched excitation patterns.

This redundancy can be eliminated through a configuration related to so-called *zipper processing*, which was initially proposed as an alternative processing strategy for bilateral CIs. Here only alternate electrodes at each ear were excited in an interleaved fashion with the aim of reducing interaction between electrodes (Lawson *et al.*, 2000). Data from pitch-ranking of electrodes across the ears were used to ensure, as far as possible, that the two ears were fitted without a significant mismatch of frequency-to-place mapping. The technique did not lead to improved speech scores compared to monaural fittings with the same total number of channels, but nor did it lead to decrements in speech intelligibility. While "zipper processing" has not been implemented clinically and is thus not representative of typical bilateral CI processing, the approach does provide a configuration that reduces redundancy and local conflicts of mapping between the two ears. It was thus ideal for the consideration of mismatched maps in normally hearing listeners in the absence of the confounding effects outlined above.

All speech processing was based on sine-excited vocoding (Loizou *et al.*, 2003). Sine carriers were chosen primarily to constrain the excitation patterns to avoid overlap of excitation in the two ears, which would be likely at the edges of noise-band carriers. The zipper-like (Lawson *et al.*, 2000) interleaved processing was chosen because it maximizes the potential advantage from the binaural combination of information between the two ears. In tests of dichotic speech perception in noise, Loizou *et al.* (2003) showed that for dichotically presented (unshifted) noise-vocoded speech, intelligibility was higher (and thus presumably binaural information more effectively combined) when channels were interleaved as here, rather than split according to low and high frequencies. Similar results were found for quiet speech in users of bilateral cochlear implants (Mani *et al.*, 2004).

The extent of mismatch between the ears was large at 6 mm to simulate the effect of a shallowly implanted electrode in one of two stimulated ears. Six channels were used, interleaved as three channels to each ear, to minimize the possibility of ceiling performance in unshifted conditions and at the same time to provide a shifted signal with sufficient information to be learnable. In relation to ceiling limits, it has been widely found that an unshifted three-channel vocoder leads to substantially poorer scores for most speech materials than an unshifted six-channel vocoder (Dorman *et al.*, 1997b; Fishman *et al.*, 1997; Loizou *et al.*, 1999). Thus, any improvements gained through adaptation to a mismatch between the two ears should be easily detectable. A pilot study showed that unshifted conditions were prone to ceiling effects with some easier speech materials, so the use of more than six channels in quiet was not explored (Faulkner, 2006). In relation to learnability, a 6.5 mm shifted four-band processor has been shown to allow fairly rapid adaptation given training (Rosen *et al.*, 1999).

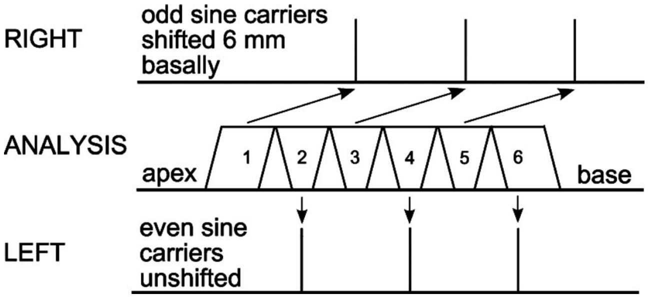

FIG. 2. Speech processing producing a binaurally mismatched frequency-to-place map. The middle panel represents the six spectral analysis bands covering the frequency range 200–5000 Hz. Temporal envelope information from channels 1, 3, and 5 is presented to the right ear (see top panel) imposed on sinusoidal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

Figure 2 shows a schematic representation of the binaurally mismatched processor that was employed. From low frequency (apical end) to high frequency (basal end), odd-numbered bands 1, 3, and 5 were presented to the right ear with the equivalent of a 6 mm basalward basilar membrane shift (assuming a 35 mm cochlea). Even-numbered bands 2, 4, and 6 were presented to the left ear without a shift.

B. Overview of research questions

In Experiment 1, we trained one group of listeners with a binaurally mismatched processor over eight 40 min training sessions, testing the extent of adaptation before, during, and after training. It was hypothesized that the presence of moderately intelligible unshifted frequency components may hinder adaptation to the shifted frequency map and thus to the binaurally mismatched processor. Consequently we also trained and tested a second group of listeners with just the shifted components to examine whether adaptation to the binaurally mismatched map could be facilitated by training with the shifted processor alone in the absence of the conflict between the two maps. In Experiment 2, we looked at how listeners adapt to the binaurally mismatched map over an

TABLE I. Conditions for Experiment 1.

Condition	Abbreviation	Component bands and shift Right	 Left
Six dichotic unshifted	6DU	1, 3, 5	2, 4, 6
Six dichotic odd shifted	6DOS	1, 3, 5 → 6 mm	2, 4, 6
Six odd shifted	6OS	1, 3, 5 → 6 mm 2, 4, 6	1, 3, 5 → 6 mm 2, 4, 6
Three even unshifted	3EU		2, 4, 6
Three odd shifted	3OS	1, 3, 5 → 6 mm	
Three odd unshifted	3OU	1, 3, 5	

extended course of training. We doubled the length of training and alternated training conditions from that of the two groups from Experiment 1 to see whether adaptation to binaurally mismatched frequency-place maps required a longer period and/or different type of training.

Perceptual adaptation to binaurally mismatched frequency-place maps has not been previously explored. If listeners do not learn to benefit from the binaurally mismatched maps, then this may reflect a constraint on plasticity for speech perception that could have implications for bilateral cochlear implant fittings. Alternative frequency-place mappings that avoid such binaural mismatches might be more optimal for speech recognition. If, however, a spectral mismatch can be learned, then the circumstances which facilitate this adaptation, such as specialized training techniques, need to be clarified.

II. EXPERIMENT 1

A. Method

1. Subjects

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

2. Test conditions

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

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

3. Signal processing

Center and crossover frequencies for the analysis and output filters were calculated using Greenwood's equation and its inverse, relating distance x (in mm) from the apex along the basilar membrane to characteristic frequency (in Hz). The assumed cochlear length was 35 mm (Greenwood, 1990).

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

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

The amplitude envelope of each band was extracted with an analysis filter, full-wave rectification, and a smoothing filter. The envelope was then multiplied by a sinusoid with frequency matching the center frequency of the band (or shifted equivalent). Finally, the requisite bands were summed and presented to the left and/or right ears as determined by pro-

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

Band	Analysis band cutoff (Hz) Lower	Analysis band cutoff (Hz) Upper	Analysis band center frequency (Hz)	Carrier frequency (Hz) Unshifted	Carrier frequency (Hz) 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

cessor condition. Table II shows input and output center frequencies as well as filter cutoffs. All processor conditions used the same six analysis filters and sine carriers at either the shifted or unshifted center frequencies of these analysis filters.

A real-time implementation of the vocoder processor was used for live training, while offline processing of the test material was implemented in MATLAB. This ensured identical repetition of test materials. Offline processing was executed at a 44.1 kHz sampling rate. Analysis bands were determined by a serial implementation of high-pass and low-pass third-order Butterworth IIR filters. 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, Sverige) and ran on a DSP card (Loughborough Sound Images TMSC31, Loughborough, UK). The computational power of the DSP was limited so the sampling rate was restricted to 16 kHz, and elliptical rather than Butterworth filter designs were used with the same 3 dB crossover frequencies as for the offline processing. Analysis filters consisted of fourth-order band-pass designs, while third-order low-pass filters were used for envelope smoothing.

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

4. Training

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

The talker for the CDT portion of this experiment was the first author, C.S. Although she is a native speaker of a north-eastern dialect of American English, she had been living in the U.K. for 5 years at the time of testing and has been judged to have an accent similar to Standard Southern British English. CS's speech was not used for any of the testing. The talker read from the text in short phrases, and the listener repeated back what he or she heard. If the listener's response matched what the talker had said, the talker would move on to the next phrase. Otherwise the phrase was repeated. If after the third presentation, the listener could still not reproduce the phrase, the listener was presented the phrase as unprocessed speech (to the left ear only). Texts for CDT were chosen from the Heinemann Guided Readers series (elementary level). These texts are designed for learners of English as a second language and make use of controlled vocabulary and syntactic complexity. During training, the talker and subject were situated in adjacent sound-treated rooms. The room had a double-glazed window that enabled auditory-visual (AV) training. During auditory-only training, the window was blinded. A constant pink masking noise at 45 dBA was played in the listener's room to mask any speech from the talker that might be transmitted through the wall and window. The talker heard the listeners' responses over an intercom, and no attempt was made to prevent the listener hearing their own unprocessed voice when responding. Of the 12 subjects, 6 were trained with the 6DOS processor (6DOS-trained group), and the remaining 6 were trained with the 3OS processor (3OS-trained group).

5. Test materials

a. Sentence perception The IEEE/Harvard sentence lists (Rothauser *et al.*, 1969) were used, which have very little contextual information. Digital recordings of the sentences were from one male and one female talker of British English (16 bit, 48 kHz downsampled to 44.1 kHz). The 72 lists in the set each contained ten sentences with five keywords in each sentence. The first 36 lists were designated for the female talker, and the remaining 36 lists were designated for the male. A subset of 32 lists from each talker was used for the 6DOS-trained group, who were tested with fewer conditions. For each test session, two lists per condition were chosen from each talker set in a pseudo-random manner. No list appeared more than twice in the same condition across all of the subjects, and subjects never heard a list more than once. The subject was asked to repeat back to the experimenter as many words as he or she could, and no feedback was given. Words were counted correct when the word root was repeated correctly.

b. Vowel identification Vowel identification was included as a measure of the contribution of spectral cues in each condition and also as a source of both confusion and recognition accuracy data. If vowel confusions in the presence of spectral shift become more similar to those for unshifted processors after training, then this would reflect adaptation to the shifted speech. The task also allowed for the

TABLE III. Sequence of training and testing conditions for Experiment 1.

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

comparison of the intelligibility of the information contained in the unshifted odd and even subsets of channels, which was important for demonstrating that the information in these subsets of channels was not entirely redundant, and that there was room for improvement beyond the three channel unshifted speech. Further, since vowels can be described primarily by their first and second formants, vowel confusions can be easily mapped in terms of the relationship between the expected vowel and the output at the sine-carrier frequency of a given formant.

Nine b-vowel-d words in the carrier sentence "Say bVd again" were recorded by a male and female speaker of British English in anechoic conditions at a 48 kHz sampling rate and subsequently downsampled to 44.1 kHz. The male but not the female talker was the same as for the sentence test. Five tokens of each bVd word were recorded from each talker so that in an individual test of one talker in a given condition, there were 45 items. Vowels were restricted to monophthongs of similar duration so that listeners would need to rely on spectral cues for identification: /æ/ (bad), /ɑː/ (bard), /iː/ (bead), /e/ (bed), /ɜː/ (bird), /ɔː/ (board), /ɒ/ (bod), /uː/ (booed), and /ʌ/ (bud). A grid with all nine words appeared, and the subject clicked with the computer mouse on the button displaying the word they perceived. The vowels were represented on the buttons in the orthographic form given above. Before testing, the subject was given a practice session in which the vowel material was presented unprocessed, with a single token for each vowel and each talker. This enabled the subjects to familiarize themselves with the software and the task.

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 2 weeks, with no more than a 2 day gap between successive sessions. Sentence and vowel test presentation was counterbalanced across the group. Within each test, stimuli were pseudo-randomized by block of condition and talker. Table III presents the sequence of training and testing for Experiment 1.

In the first session, subjects were acclimatized to unshifted sine-vocoding with a 10 min block of CDT with the DU processor prior to the pre-training test session. Previous experiments have demonstrated that listeners require a short period of acclimatization before they can reliably perceive speech through unshifted vocoders with a limited number of channels (Davis *et al.*, 2005). As in subsequent CDT training sessions, the first 5 min of the familiarization block were AV, while the remaining 5 min were auditory alone (AA). Following the pretest, subjects were trained with CDT in four 40 min training sessions with either the dichotic odd shifted speech (6DOS-trained group) or the monaural three shifted channel speech (3OS-trained group). Subjects were tested after the fourth training session. If the testing session did not take place immediately following a training session, then the experimenter administered a 10 min (five AV, five AA) CDT block with the 6DOS or 3OS processor, which was not counted toward the total hours of training. Following this mid-training test session, subjects underwent four more 40 min training sessions with the same processor as in the first four sessions, and then completed the post-training testing.

Testing took place in a sound-treated room with presentation of the processed speech over Sennheiser HD280 headphones. The level was set by the experimenter to a comfortable listening level, and this level was used by all participants.

B. Results

Test data were analyzed using repeated measures analysis of variance (ANOVA), with within subject factors of test session, condition, and talker, and where relevant, a between subjects factor of processor used in training. Hyunh–Feldt epsilon corrections were applied to all F tests for factors with more than one degree of freedom. Hyunh–Feldt adjusted degrees of freedom have been rounded to the nearest integral value, and the significance criterion was $p=0.05$. *A priori* hypotheses were tested using planned contrasts, and *post-hoc* testing was carried out using Bonferroni-adjusted paired comparisons. Data were typically pooled by talker, unless there was a significant interaction between talker and processor that could not be attributed to floor effects with the shifted (3OS) condition.

1. IEEE sentence perception

a. 6DOS-trained group Keywords correct for the IEEE sentence test across training sessions for the two talkers combined are shown in the left-hand panel of Fig. 3. For the

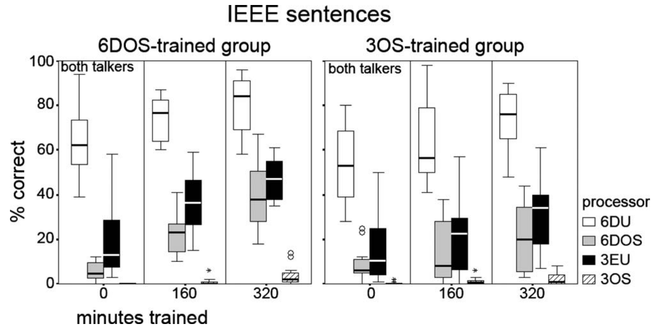

FIG. 3. Experiment 1 box-and-whisker plots of IEEE sentence scores as a function of training time. The box shows interquartile range over six subjects, the bar shows the median, and the whiskers show the complete range excluding any outlying values (shown as open circles or asterisks). Plots are shown for the two talkers combined. A vertical reference line separates each training time period. Scores for the 6DOS-trained group are in the left-hand panel, while scores for the 3OS-trained group are in the right panel. The plots show the four main conditions: 6DU, 6DOS, 3EU, and 3OS.

6DOS-trained group, performance with the 3OS processor remained close to floor throughout training, while intelligibility with all other processors tended to increase. This was indicated by a significant training session with processor interaction $[F(8,40)=445,\ p<0.001]$. Scores have been pooled by talker, although they were slightly better with the male talker in all but the 3OS condition, which is reflected by the significant interaction of talker with processor $[F(3,17)=6.57,\ p=0.003]$. When the data were reanalyzed excluding the 3OS condition, these interactions were no longer significant, and significant main effects were evident for number of training sessions $[F(2,10)=71.5,\ p<0.001]$ and processor $[F(2,9)=146,\ p<0.001]$.

Post-hoc testing on the post-training sentence scores revealed three key findings. First, performance with the three unshifted channels (3EU) was significantly worse than with the dichotic unshifted (6DU) condition ($p=0.001$), which is a clear indication that there was room for improvement with the additional channels in the shifted ear. Second, there was no significant difference between the 6DOS condition and the 3EU condition, which indicates that subjects did not show a binaural advantage. If they had, we would expect to see performance with this processor exceeding that in the 3EU condition, as was found for the 6DU condition.

b. 3OS-trained group Scores for IEEE sentence recognition for the 3OS-trained group are shown in the right-hand panel of Fig. 3. A repeated measures ANOVA showed significant main effects of number of training sessions $[F(2,10)=36.0,\ p<0.001]$ and processor $[F(3,16)=85.9,\ p<0.001]$, but there was no significant training session with processor interaction, indicating significant improvement with training that was relatively similar in all conditions, including 3OS.

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

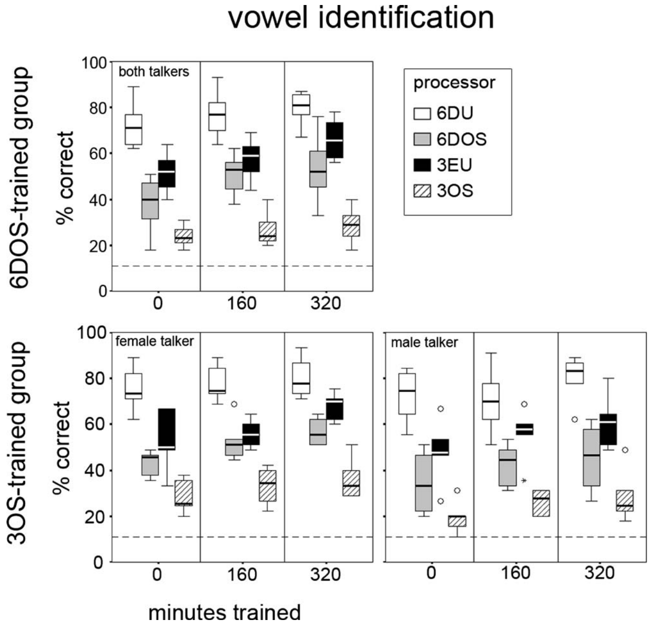

FIG. 4. Experiment 1 vowel identification scores as a function of training time. Scores for the 6DOS-trained group are in the top panel, while scores for the 3OS-trained group are in the bottom two panels. The 3OS-trained group showed a significant talker by processor interaction, so plots for each of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

2. Vowel identification

a. 6DOS-trained group Vowel identification scores for the two talkers combined are summarized in the top panel of Fig. 4. The overall pattern of results is similar to that seen for sentences. There were significant main effects of number of training sessions $[F(2,10)=37.4,\ p<0.001]$ and processor $[F(4,20)=102,\ p<0.001]$ and a significant training sessions with processor interaction $[F(8,40)=89.6,\ p=0.005]$. The lower bound of the 95% confidence interval was above chance level performance (11%) at all test points and for all conditions, confirming that vowel recognition scores were significantly greater than chance.

Importantly, *post-hoc* testing showed significant improvements in all conditions between the pre-test and post-test sessions ($p<0.05$), including 3OS, and a planned contrast of training sessions within the 3OS condition also showed significant improvement after training in this condition ($p=0.01$). While the significant interaction of training session with processor may indicate that improvement with the 3OS condition was smaller than the improvement in the other conditions, this is nonetheless evidence of learning in this condition. Despite the improvement in the 3OS condition, again there was no significant difference between 3EU and 6DOS after training, even though performance with 3EU was significantly worse than with 6DU ($p=0.02$). Thus, while there was room for improvement by the addition of the odd channels (3EU < 6DU), and there was evidence of learning

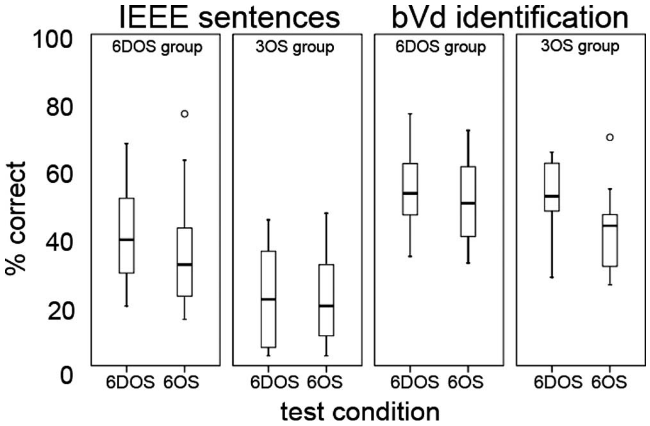

FIG. 5. Comparison of dichotic (6DOS) versus diotic (6OS) presentation of mismatched frequency-place maps within each stimulus type and training group from Experiment 1. The IEEE scores are in the left two panels (6DOS and 3OS groups, respectively), while the bVd identification scores are in the two panels on the right (6DOS and 3OS groups, respectively).

when these channels were shifted, subjects did not show a binaural advantage with the mismatched frequency-place maps.

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

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

3. Comparison of dichotic versus diotic presentation

Mismatched maps were compared under dichotic and diotic conditions to test the hypothesis that subjects had learned to ignore information from the shifted ear. Boxplots comparing post-training intelligibility with the dichotic and diotic presentations of the mismatched maps are presented in Fig. 5. The left two panels give scores for IEEE sentences, and the right two panels for bVd identification. The results are presented for the two talkers combined. The scores were entered into a repeated measures ANOVA. For the 6DOS-trained group, there was no significant difference between dichotic (6DOS) versus diotic (6OS) presentation for sentences or vowels. This suggests that if this group had adapted by suppressing or ignoring information, this could only have been on the basis of carrier frequency rather than ear of input. By contrast in the 3OS-trained group, Bonferroni-corrected comparisons showed that post-training vowel identification was significantly worse for the diotic condition than the dichotic condition ($p=0.04$). This suggests that subjects in the 3OS-trained group were still relying on cues to ear presentation when listening to the mismatched maps since intelligibility was decreased when this cue was removed.

4. Comparison of unshifted odd and even channels

Intelligibility with unshifted even and odd channels (3EU and 3OU) was compared for the 3OS-trained group to confirm that the speech information in each subset of channels was not redundant. This is important for demonstrating that there was speech information that could be gained through the binaural combination of odd and even channels. The post-training results are summarized in the left (IEEE sentences) and right (vowel identification) panels of Fig. 6.

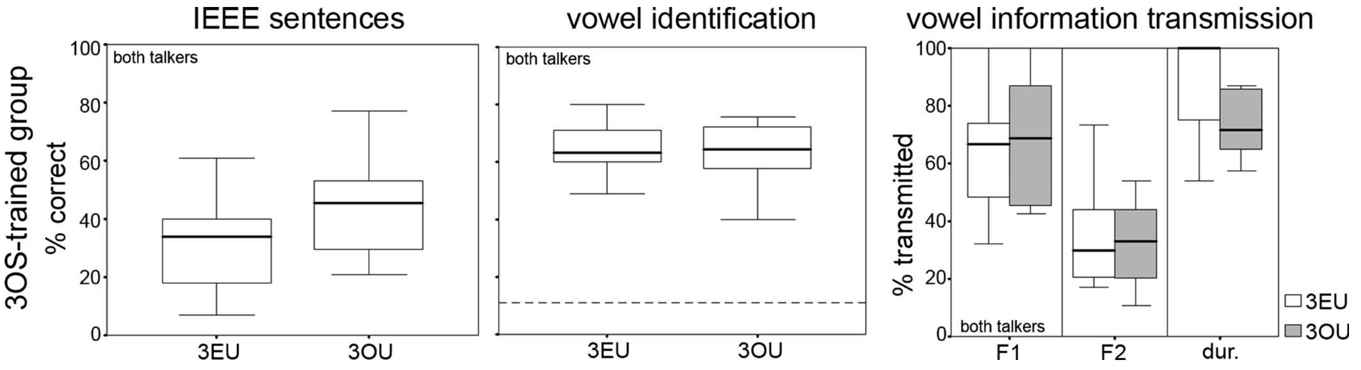

FIG. 6. Comparison of post-training scores for the unshifted even (3EU) and odd (3OU) channels for the 3OS-trained group of Experiment 1. The left panel shows IEEE sentence scores, the middle panel shows bVd identification, and the right panel shows percent of vowel feature information (F1, F2, and duration) transmitted. The results are shown for the two talkers combined.

TABLE IV. Feature definitions for the F1, F2, and duration features for each of the nine vowels. F1 was binary (open, close); F2 was tertiary (front, back, central); and duration was binary (long, short).

Feature	æ	ɑː	iː	e	ɜː	ɔː	ɒ	uː	ʌ
F1	o	o	c	o	c	o	o	c	o
F2	f	b	f	f	c	b	b	b	b
Duration	l	l	l	s	l	l	s	l	s

Surprisingly, intelligibility with the IEEE sentences was significantly better with the 3OU channels than the 3EU channels after training ($p=0.007$).

For vowel recognition, the ANOVA revealed no significant difference between these conditions for overall vowel recognition accuracy. The post-training vowel confusions for 3EU and 3OU were also analyzed in terms of information transfer (Miller and Nicely, 1955). Features of F1, F2, and duration were analyzed according to the feature definitions set out in Table IV. The results for the two talkers combined are summarized in the right-hand panel of Fig. 6. A repeated measures ANOVA revealed no significant difference between 3EU and 3OU in terms of information transfer for any of the three features. However, an χ^2 analysis showed that vowel confusions differed significantly between the two processors ($p<0.0001$). Hence there are clearly differences of detail between the vowel features conveyed by odd compared to the even bands, but no evidence of a difference across the set of vowels in the level of feature information provided by these two sets of bands.

These detailed differences in vowel confusions probably reflect the complementary representation of spectral information in each of the subsets of channels and also the sparse spectral coding which presents vowel formant energy only at the discrete carrier frequencies of the sine vocoder. For example, the /iː/ vowel was almost always incorrectly recognized as /uː/ in the odd unshifted condition (3OU) (87% of the time), but was recognized fairly accurately in the even unshifted condition (3EU) (77% correct). The F2 of this vowel for the female talker was around 2850 Hz, but the sine carrier for the analysis filter encompassing F2 was 2498 Hz, so F2 energy was lowered in frequency which could account for /iː/ being recognized as /uː/. For the /uː/ vowel, both F1 (319 Hz) and F2 (1980 Hz) frequencies fell within the analysis filters of odd-numbered bands. As would be expected, this vowel was recognized accurately in the odd channel condition (3OU) (85% correct), but in the even channel condition (3EU), accuracy was low (28%), and there was no consistent pattern in confusions since energy from formant frequencies was largely absent in this condition for this vowel. Taken together with the finding that 6DU was always better than 3EU in this experiment, there is thus clear evidence that the lack of binaural advantage shown with the binaurally mismatched processor (6DOS) compared to the even unshifted channels (3EU) is not attributable to a redundancy of speech information in the even and odd bands.

5. Vowel information transmission analysis

A further information transmission analysis of the post-training vowel confusions was performed to examine in more detail the information provided in the different processor conditions. The results for Experiment 1 are summarized in the left and middle panels of Fig. 7. Repeated measures ANOVAs for each training group were computed on the percentage of information transferred, with within-subjects factors of processor (6DOS, 3EU, and 3OS only), feature (F1, F2, and duration), and talker. For each group, there was a significant interaction of processor with feature [6DOS-trained group: $F(4,20)=3.17$, $p=0.036$; 3OS-trained group: $F(4,20)=12$, $p<0.001$]. Despite the equivalent post-training intelligibility scores for vowel identification, Bonferroni-corrected comparisons of processors within each feature revealed that the information transmitted with the 6DOS processor was significantly less than that with the unshifted 3EU processor, though this depended on feature and training group. For the F1 feature, this reduction in information was only marginally significant for the 6DOS-trained group ($p=0.048$), but was more significant for the 3OS-trained group ($p=0.008$). In fact, for the 3OS-trained group, there was no significant difference in F1 information for the 6DOS and 3OS conditions. Less F2 information was transmitted for 6DOS than 3EU for the 6DOS-trained group only ($p=0.03$), while less duration information was transmitted for the 3OS-trained group only ($p=0.025$). Not surprisingly, the information transmitted with the 3OS processor was always significantly worse than that with 3EU. This pattern of results suggests that subjects were not simply ignoring the shifted information in the 6DOS condition, and that training condition impacted the binaural combination of information.

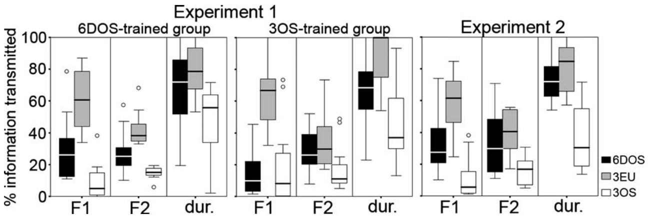

FIG. 7. Percent of vowel feature information (F1, F2, and duration) transmitted for the 6DOS, 3EU, and 3OS conditions. The results are shown for the two talkers combined. Scores are broken down by experimental group: the left panel shows Experiment 1 6DOS-trained group, the middle panel Experiment 1 3OS-trained group, and the right panel scores from Experiment 2.

C. Discussion

For both groups of subjects, performance with 6DU exceeded performance with 3EU in all but one subtest. Moreover, for the 3OS-trained group, the 3OU channels showed higher IEEE scores than the 3EU channels after training, and vowel confusions differed significantly with these processors. This is clear evidence that the information conveyed by the odd channels was not redundant in the presence of the even channels, which suggests that there was room for improvement if listeners could learn to use unshifted and shifted channels together for binaural advantage. However, neither group of subjects showed a binaural advantage for the binaurally mismatched processor since post-training intelligibility with the 6DOS processor never exceeded that with the 3EU processor for any of the speech materials.

Post-training intelligibility was similar in the 6DOS and 3EU conditions, which is consistent with the view that subjects had adapted to 6DOS by attending to the "better ear"—i.e., the unshifted frequencies. However, the information transfer analysis revealed that for both groups of subjects, vowel information transferred with the binaural 6DOS condition was significantly worse for some features than that for the monaural 3EU condition, which suggests a more complicated listening strategy. The binaural combination of shifted and unshifted channels may have led to a fused but "incorrect" percept—a process referred to as "psychoacoustic fusion" by Cutting (1976). It is also possible that the combination of shifted and unshifted channels was somehow distracting and thus led to more errors. The different pattern of results in the two groups suggests that the training condition affected the binaural combination of information. For the F1 feature, explicit training with the 6DOS condition seems to have mitigated the deficit with 6DOS compared to 3EU since for this group the difference was only marginally significant. By contrast, explicit training with 3OS seems to have been more beneficial for F2 information since only the 6DOS-trained group showed significantly worse F2 information transmission with 6DOS than 3EU.

Notwithstanding these apparent training condition effects, it seems clear that neither group of subjects learned to combine the shifted and unshifted channels in a way that led to a binaural advantage for speech over the time course in Experiment 1. It is possible that the period of learning required to show a binaural advantage for mismatched frequency-place maps may be longer than that needed for simple monaural basalward shifts. The improvements shown here were small and did not reach an asymptotic level after 5 h, 20 min of training, which may indicate that subjects were still learning at the point of final testing. This was especially true for the 3OS condition, which proved particularly difficult for the listeners in this study. While the difficulty in learning 3OS appears to contrast with previous studies with monaural basally shifted speech (Rosen *et al.*, 1999), the sparse spectral representation in the processing is a novel feature that has not been previously studied and may have been a source of difficulty. Tyler and Summerfield (1996) showed that cochlear implant patients are still adjusting to their speech processors 6 months and longer after implantation, and the training times considered here were very short by comparison.

It is also possible that adaptation to mismatched frequency-place maps requires experience with both the shifted alone (3OS) and binaurally mismatched (6DOS) maps. For the 6DOS-trained group, training led to larger improvements overall, and there was some evidence of adaptation specific to the shifted bands (3OS). While training with 3OS led to smaller overall improvements, for this group there was evidence of shift-specific learning and better transmission of F2 information in the 6DOS condition. While training with just the shifted components (3OS) alone may be insufficient to allow for a binaural advantage from the mismatched maps, a period of training with just the shifted map may facilitate adaptation to the shifted components, which may only then lead to a binaural advantage with sufficient further training.

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

III. EXPERIMENT 2

A. Method

1. Subjects

Six normally hearing speakers of British English took part and were paid to participate. They all had normal (<20 dB HL) pure-tone thresholds at 0.5, 1, 2, and 4 kHz. None had participated in Experiment 1.

2. Signal processing

Signal processing was the same as for Experiment 1.

3. Training

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

4. Test materials

Test materials for Experiment 2 included the same IEEE sentence and vowel identification tests from Experiment 1. Because of floor effects with the 3OS condition in Experiment 1, we also used easier high-context BKB (Bench*et al.*, 1979) and IHR (Institute of Hearing Research, MacLeod and Summerfield, 1990) sentences in order to more clearly demonstrate learning with 3OS. These two sentence sets have similar syntactic constructions and are essentially equivalent in intelligibility. The female talker was the same as that for the IEEE sentences, but the male talker was different. The BKB sentences consist of 16 sentences per list with either 3 or 4 keywords per sentence, making 50 keywords per list. The IHR sentences consist of 15 sentences of 3 keywords each, hence 45 keywords per list. BKB sentences were used for familiarization testing, with IHR sentences used for all subsequent testing.

5. Procedure

Table V sets out the testing and training regime used for Experiment 2, which is similar to that of the previous experiment, with the following modifications. A familiarization test session was included at session 1 in which subjects were tested with the IEEE sentences, the Bamford–Kowal–Bench (BKB) sentences, and vowels. This was primarily for familiarization so that any rapid adaptation to the unshifted vocoder processing would occur before training commenced

TABLE V. Sequence of training and testing conditions for Experiment 2.

Session	Training	Processor	Testing
1	5 min AV, 5 min AA	6DU	Familiarization: Unprocessed vowels; *Pre-test*: IEEE sentences (2 lists × 4 conditions × 2 talkers), BKB sentences (2 lists × 3 conditions × 2 talkers), bVd identification (5 tokens × 4 conditions × 2 talkers)
2	⋯	⋯	*Second Pre-test*: IEEE sentences (2 lists × 4 conditions × 2 talkers), IHR sentences (2 lists × 3 conditions × 2 talkers), bVd identification (5 tokens × 4 conditions × 2 talkers)
3	5 min AV × 2, 5 min AA × 2	6DOS+3OS	None
4–17	10 min AA × 2	6DOS+3OS	None
18	⋯	⋯	*Mid-test*: IEEE sentences (2 lists × 4 conditions × 2 talkers), IHR sentences (2 lists × 3 conditions × 2 talkers), bVd identification (5 tokens × 5 conditions × 2 talkers)
19–33	10 min AA × 2	6DOS+3OS	None
33	⋯	⋯	*Post-test*: IEEE sentences (2 lists × 4 conditions × 2 talkers), IHR sentences (2 lists × 3 conditions × 2 talkers), bVd identification (5 tokens × 5 conditions × 2 talkers)

(Davis *et al.*, 2005). BKB sentences were used for easy material in only the very first pre-training session. After session 1, only IHR sentences were used for easy sentence material. Also, in the first training session only (session 3 in Table V) the first 5 min with each processor was auditory-visual. All subsequent training for this experiment was auditory only. All other procedural considerations were the same as for Experiment 1.

6. Conditions

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

B. Results

1. Test exposure learning

A repeated measures ANOVA was used to compare results between the first two pre-training test sessions. No significant differences between these two test sessions were found for either the IEEE sentence test or the vowel test, which implies that any further improvements beyond the second test session were likely the result of training rather than incidental learning. All subsequent analyses are therefore based on the final three test sessions.

2. IEEE sentences

Scores for the IEEE sentences across training sessions are shown for the two talkers combined in the top left panel of Fig. 8. Performance tended to improve with training in all conditions. An ANOVA on data from the final three test sessions showed significant main effects of number of training sessions $[F(2,10)=29.0,\ p<0.001]$ and processor $[F(1,5)=72.2,\ p<0.001]$. While there was a significant session with talker interaction $[F(2,10)=6.85,\ p=0.013]$, there was no training session with processor interaction, which indicates that learning through training was similar in all conditions, including 3OS. This is a similar finding to the 3OS-trained group in Experiment 1, who had an equivalent amount of training with the 3OS condition as the subjects in this experiment.

Here there was also direct evidence of learning of the shifted bands (3OS) since a planned contrast of training sessions in this condition showed intelligibility after training was significantly better than before training ($p=0.038$). Yet despite this learning, and the extended period of training in Experiment 2, there was still no evidence of an advantage for the mismatched maps. In *post-hoc* tests, performance with 6DOS and 3EU did not differ, even though there was room for improvement after 10 h of training since 3EU was worse than 6DU ($p=0.02$).

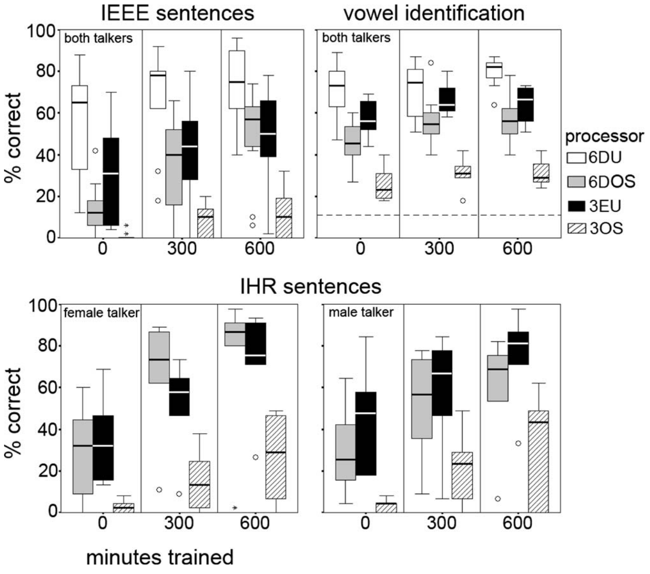

FIG. 8. Performance as a function of amount of training for the final three test sessions of Experiment 2. The top left panel shows IEEE sentence scores, and the top right panel vowel identification. In the bottom two panels are scores for the IHR sentences, which have been split by talker because of a significant talker effect. Training time periods are separated by vertical lines, and for the vowel identification, the horizontal reference line indicates chance performance (11.1%).

3. High-context (IHR) sentences

The bottom panels of Fig. 8 show scores for the IHR sentences for the female and male talkers individually. ANOVA showed a significant interaction between talker and processor [$F(2,8)=6.58$, $p=0.022$], so scores for each talker were analyzed separately. These showed significant effects of training session and condition. The bottom panels of Fig. 8 show that median intelligibility of the 3OS condition improved from close to floor before training to near 40% for both talkers after training, which is a much higher level of intelligibility for this processor than that seen in Experiment 1. A planned contrast showed that improvement with training in this condition was significant ($p=0.02$).

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

4. Vowel identification

Boxplots of vowel identification for the two talkers combined appear in the top right panel of Fig. 8. ANOVA on the final three test sessions showed significant main effects of session [$F(2,8)=20.4$, $p=0.001$] and processor [$F(3,15)=77.8$, $p<0.001$], but no significant interactions. The lower bound of the 95% confidence interval of the estimated marginal mean was always greater than 11%, indicating that intelligibility was statistically greater than chance for all conditions and at all test points. While a planned contrast showed significant learning of the shifted channels (3OS) with training ($p=0.046$), *post-hoc* testing on the post-training vowel scores revealed no significant difference in intelligibility between the 6DOS and 3EU processors, so subjects did not show a binaural advantage for the mismatched maps, despite having room for improvement to do so ($\text{6DU}>\text{3EU}$, $p=0.014$). Moreover, a planned contrast showed that post-training vowel intelligibility was marginally worse with the binaurally mismatched processor (6DOS) than with the monaural unshifted bands (3EU). Although this test just missed significance ($p=0.051$), the finding is suggestive of distraction with the addition of the shifted channels, rather than binaural benefit.

As for Experiment 1, the post-training vowel confusion matrices were entered into an information transfer analysis to explore the differences in information transmission with the 6DOS, 3EU, and 3OS processors. The findings are shown in the right panel of Fig. 7. The percent information transmitted was then entered into a repeated measures ANOVA, with within-subjects factors of talker, processor (6DOS, 3EU, and 3OS only), and feature (F1, F2, and duration). The ANOVA revealed no significant effect of talker, but there was a significant interaction of processor with feature [$F(4,20)=4.16$, $p=0.013$]. The F1 information transmitted was significantly lower for 6DOS than for 3EU ($p=0.036$), but there was no difference between these conditions for F2 and duration information. Information transmission with 3OS was always significantly worse than 6DOS and 3EU.

C. Discussion

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

Intelligibility with 3OS was only clearly above floor levels with interspersed and/or concurrent training with unshifted vocoded speech: although the 3OS-trained group in Experiment 1 received the same amount of training with that condition as the subjects in Experiment 2, only the latter group of subjects showed such improvements. The greater improvement may have arisen from either more exposure time to the shifted speech (through exposure to both 6DOS and 3OS), from the concurrent and/or interspersed exposure to easier unshifted speech, or a combination of the two. Recently Li *et al.* (2009) showed that unsupervised adaptation to a large basalward shift (8 mm) only occurred with interspersed exposure to a more moderate basalward shift (4 mm). It thus seems plausible that exposure to an easier condition may facilitate the learning of more difficult speech conditions. However, the increased learning of 3OS in Experiment 2 may also have been the result of a change in training talker.

Despite this increased evidence of adaptation to the basally shifted speech and the extended training period allowed for adaptation, again no binaural advantage was found for the binaurally mismatched frequency-place maps. Performance with the 6DOS processor never exceeded that with the three unshifted channels alone (3EU), even though there was clear room for improvement from the addition of the odd channels, since 3EU was typically worse than 6DU. Moreover, there was some evidence of interference from the shifted bands when they were presented together with the unshifted bands since post-training performance with 6DOS was worse than 3EU for the vowels, the male talker IHR sentences, and F1 information transmission. This latter finding is similar to the 3OS-trained group in Experiment 1 and suggests that explicitly training with the 3OS condition may make the shifted F1 information difficult to ignore when presented with unshifted information in 6DOS.

IV. GENERAL DISCUSSION

A. Resistance to learning

Listeners in both experiments demonstrated evidence of adaptation following a period of training, yet intelligibility with the 6DOS processor never exceeded that with just the three unshifted channels (3EU). The findings thus indicate that listeners are resistant to learning cochlear frequency-to-place maps that differ greatly between the ears. The mechanism involved in adaptation to binaurally mismatched frequency-place maps is likely to be different from that for adaptation to monaural basalward spectral shifts, where significant if partial adaptation is evident after as little as 3 h training (Rosen *et al.*, 1999).

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

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

It is possible that aspects of the training procedure, such as the listener hearing their own voice unprocessed during training, or the listener returning to everyday speech conditions outside of the experiment, may have hindered adaptation. However, the training procedure was identical to that used in previous studies. Rosen *et al.* (1999) showed adaptation to four-channel noise-vocoded speech with a 6.5 mm basalward shift in just 1 h and 20 min, and further studies have shown adaptation in comparably short time periods (Faulkner *et al.*, 2006; Smith and Faulkner, 2006). Though the processing in the present experiments was markedly different from that typically considered in previous adaptation studies (e.g., sparsely sampled output at each ear and binaurally mismatched, sine-excited vocoding), the training times were comparably longer to compensate for the anticipated increased difficulty. Yet despite the extended course of training in these experiments (5–10 h), subjects showed no evidence of a binaural advantage for the binaurally mismatched processing, even though there was evidence of adaptation to the monaural shifted speech, especially in Experiment 2. It thus seems unlikely that the lack of binaural advantage shown in these experiments is entirely attributable to a deficiency in the training procedure.

A novel aspect of the frequency-place mapping examined here, and a possible source of this resistance to learning, is the altered relative frequency order of the output bands in the binaurally mismatched processor. When combinations of shifted and unshifted bands were presented together, as in the 6DOS and 6OS processors, the output consisted of frequency information that had been shifted in both frequency and relative order. Ranking the output with reference to the relative frequency order of the input analysis bands yields, from low to high frequency, (2–1–4–3–6–5). Informative cross-frequency patterns, such as relative frequencies of the first and second formants, may as a result be destroyed. Evidence from the experiments here suggests that, despite the enriched frequency content, listeners do not learn to combine this altered output in the same way as for matched frequency-place maps. If the shifted channels contributed to intelligibility with 6DOS at all, they appeared to serve as a distraction, with the transmission of first formant information being particularly vulnerable.

The lack of binaural advantage for combinations of shifted and unshifted bands shown here may be indicative of a constraint on plasticity for speech perception to cases where relative frequency order has been preserved. This would be consistent with the finding that a frequency-warped but order-preserving frequency-to-place map is relatively easy to learn (Smith and Faulkner, 2006). There is, however, some evidence to the contrary: in his seminal study on adaptation to spectrally altered speech, Blesser (1972) trained listeners with speech that had been spectrally rotated around 1600 Hz so that low frequencies became high and vice versa. Even under such drastic conditions, some listeners did eventually learn to converse, although this did not appear until much later in the study, which consisted of 20 45 min sessions. Azadpour (2008) also found that listeners could adapt to aspects of spectrally rotated speech over a similar time course to that in Experiment 1.

With spectrally rotated speech, however, the relative ordering of frequency information is intact, even if it has been inverted. By contrast, the processing used here may selectively invert speech cues such as vowel formant patterns which may result in a novel percept, in addition to cues being presented to the wrong frequency region. In the present investigations, this appeared to hinder adaptation. Specifically, in Experiment 1, there was evidence of shift-specific adaptation to the 3OS condition for both groups for the vowel test, yet listeners still did not show a binaural advantage when listening in the 6DOS condition. This was despite the fact that in the absence of shifting, the three even (3EU) and three odd (3OU) processors resulted in significantly different vowel confusions, which would imply that there was room for improvement by combining the information if the listener had been able to do so.

B. Implications for bilateral stimulation with cochlear implants

Since performance did not reach asymptote in either of the present experiments, it remains to be determined whether continued learning beyond that demonstrated here would be shown, or whether this is too difficult a mapping to learn fully. In this respect, it is difficult to interpret the implications of this resistance for cochlear implant patients because the extent of their experience with any clinically fitted processor will be far greater than the 5–10 h examined in the laboratory here. For example, the group of patients investigated by Tyler and Summerfield (1997) reached asymptotic performance an average of 30–40 month post implantation.

Several studies have shown that some patients experience a synergistic improvement when using both implants over either implant on its own, even for speech in a quiet laboratory (Dorman and Dahlstrom, 2004; Litovsky *et al.*, 2006). The drastic spectral mismatch examined here may arguably be more representative of the mismatch with acoustic hearing and may be uncommon between two bilateral electrode arrays of the same design. Smaller mismatches than that studied here may be tolerated by CI users. Yet, not all bilateral CI users show an advantage over their best ear alone (Wackym *et al.*, 2007; Tyler *et al.*, 2007) and substantially mismatched frequency-place maps underpin this lack of advantage.[1] In a recent study of patients with bilateral cochlear implants, Mosnier *et al.* (2009) found that bilateral CI patients with asymmetrical monaural CI speech scores did not show a binaural advantage for speech in quiet, which suggests a difficulty with mismatched frequency-place maps.

Since the processor design considered here did not directly simulate typical bilateral implant systems currently in use, there may be limitations to generalizing the findings to CI patients. The speech processors were designed to minimize redundancy between the two ears in order to investigate global aspects of binaural place mismatch in the absence of local conflicts due to (1) the presentation of the same information to two different places or to (2) the presentation of different information to the same tonotopic location in each ear. However, typical bilateral cochlear implant systems do deliver roughly the same signal to each ear. Interleaving channels between the ears may have left insufficient cues to bind unshifted and shifted channels into a single auditory object. An increase in signal redundancy provided with "typical" bilateral processing may facilitate auditory grouping and thus the binaural combination of mismatched maps. However, the more spectrally rich signal provided by the better ear (i.e., the ear with the smallest degree of shift) with non-interleaved processing could also lead listeners to learn to attend to the better ear alone. Further research could explore whether amplitude comodulation of the shifted and unshifted channels would facilitate grouping of the mismatched signals as a single auditory object and thus allow for a binaural advantage (Carrell and Opie, 1992).

Current bilateral cochlear implants are fitted by mapping the full spectrum of speech information to each ear, but this is not a requirement for robust speech perception, as is clearly demonstrated by the speech intelligibility advantage with electro-acoustic stimulation (Ching, 2005). Alternative frequency-place mapping strategies could seek to align the frequency range delivered to each ear through assigning processor filter frequencies that were matched according to interaural location. For cases of asymmetrical insertion depths, this would likely entail that the highest and lowest channels were presented to only one ear. Such an alternative processing strategy would avoid drastic mismatches of frequency-place maps between the ears while maximizing signal redundancy and may ultimately lead to a stronger binaural advantage. To this end, robust methods for matching frequency-place maps between the ears for humans would need to be developed. New research into more objective techniques, such as the binaural interaction component of the evoked auditory brainstem response (EABR), as described by Smith and Delgutte (2007), could prove promising.

Many of the benefits of binaural hearing are realized for speech in noise, where differences in the sound signal at each ear can be used to obtain a better representation of what has been said. Indeed, of the two subjects tested by Dorman and Dahlstrom (2004), one showed binaural benefit over the better ear only in the noise condition. Better-ear effects may dominate in the absence of noise and thus limit any potential binaural advantage from the mismatched signals at each ear. If the signal here were presented in noise, listeners may have stronger motivation to rely on the shifted channels. When testing in noise, it may also be possible to use more spectral bands without achieving ceiling effects with the unshifted speech, and this increased spectral resolution may, in turn, facilitate learning.[2]

Current cochlear implant systems do not contribute the fine structure information thought to underpin many binaural hearing phenomena in normal hearing listeners. However, Long and colleagues showed at least one bilateral cochlear implantee to be sensitive to the interaural time difference (ITD) and interaural level difference (ILD) cues provided solely through envelope information. Crucially, this was limited to cases where interaural electrode pairs were matched in frequency (Long *et al.*, 2003). Normal hearing listeners also demonstrate best ITD (Nuetzel and Hafter, 1981) and ILD (Francart and Wouters, 2007) detection for similar interaural cochlear place regions. Though perhaps through a different underlying mechanism than has been considered here, evidence from psychophysics is thus consistent with the present research and suggests that maintaining similar frequency-place maps between the ears may be important for optimizing bilateral CI fittings.

V. CONCLUSION

This research examined whether listeners could learn to demonstrate a binaural advantage for speech information with a unilateral spectral shift. Despite undergoing significantly longer training periods than in previous studies with monaural spectral shifts, subjects in the present investigations never showed a binaural advantage for the binaurally mismatched frequency-place map. Post-training performance with the binaurally mismatched processor never exceeded that with the three unshifted channels, even after 10 h of training. This resistance to learning is suggestive of a constraint on speech perceptual plasticity to instances where relative frequency order is preserved. It may thus be important to keep frequency-place maps similar in the two ears when optimizing bilateral CIs for speech perception. The results are consistent with psychophysical studies in bilateral cochlear implant patients, which have also called for matching interaural electrode pairs in order to restore ITD and ILD sensitivities.

ACKNOWLEDGMENTS

Work supported by a grant from the European Commission (Grant No. FP6-004171 HEARCOM), Study Assistance from the Staff Development and Training Unit, Department of Human Resources, UCL awarded to C.M.S., and a Wellcome Trust Summer Vacation Scholarship to K.M. (Grant No. VS/06/UCL/A12).

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

[2] In a related study, Long *et al.* (2004) used 12 channel binaurally interleaved vocoders similar to those described here, with one ear processed in noise. Listeners in their experiment demonstrated a binaural advantage for consonant perception over either ear on its own. This could not be attributed to binaural redundancy since the interleaved channels ensured that each ear was getting different information. However, the processor in their study was tonotopically matched in both ears. It merits exploration whether this advantage remains if a spectral shift is introduced to one ear.

Azadpour, M. (**2008**). “Perceptual learning in speech processing: Comprehending spectrally-rotated speech,” Ph.D. dissertation, University of Trieste, Trieste, Italy.

Bench, J., Kowal, A., and Bamford, J. M. (**1979**). “The BKB (Bamford-Kowal-Bench) sentence lists for partially hearing children,” Br. J. Audiol. **13**, 108–112.

Blesser, B. (**1972**). “Speech perception under conditions of spectral transformation I. Phonetic characteristics,” J. Speech Hear. Res. **15**, 5–41.

Broadbent, D. E., and Ladefoged, P. (**1957**). “On the fusion of sounds reaching different sense organs,” J. Acoust. Soc. Am. **29**, 708–710.

Carrell, T., and Opie, J. M. (**1992**). “The effect of amplitude comodulation on auditory object formation in sentence perception,” Percept. Psychophys. **52**, 437–445.

Ching, T. Y. C. (**2005**). “The evidence calls for making binaural-bimodal fittings routine,” Hear. J. **58**, 32–41.

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

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

Cutting, J. E. (**1976**). “Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening,” Psychol. Rev. **83**, 114–140.

Davis, M. H., Johnsrude, I. S., and Hervais-Adelman, A. T. K. M. C. (**2005**). “Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences,” J. Exp. Psychol. Gen. **134**, 222–241.

De Filippo, C. L., and Scott, B. L. (**1978**). “A method for training and evaluating the reception of ongoing speech,” J. Acoust. Soc. Am. **63**, 1186–1192.

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

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997a**). “Simulating the effect of cochlear-implant electrode insertion depth on speech understanding,” J. Acoust. Soc. Am. **102**, 2993–2996.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997b**). “Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,” J. Acoust. Soc. Am. **102**, 2403–2411.

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

Faulkner, A., Rosen, S., and Stanton, D. (**2003**). “Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth,” J. Acoust. Soc. Am. **113**, 1073–1080.

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

Fishman, K. E., Shannon, R. V., and Slattery, W. H. (**1997**). “Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,” J. Speech Lang. Hear. Res. **40**, 1201–1215.

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

Francart, T., and Wouters, J. (**2007**). “Perception of across-frequency interaural level differences,” J. Acoust. Soc. Am. **122**, 2826–2831.

Fu, Q.-J., and Galvin, J. J. III (**2003**). “The effects of short-term training for spectrally mismatched noise-band speech,” J. Acoust. Soc. Am. **113**, 1065–1072.

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

Fu, Q.-J., Nogaki, G., and Galvin, J. J. III (**2005b**). “Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation,” J. Assoc. Res. Otolaryngol. **6**, 180–189.

Fu, Q.-J., and Shannon, R. V. (**1999**). “Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing,” J. Acoust. Soc. Am. **105**, 1889–1900.

Fu, Q.-J., Shannon, R. V., and Galvin, J. J. III (**2002**). “Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant,” J. Acoust. Soc. Am. **112**, 1664–1674.

Grant, K. W., Braida, L. D., and Renn, R. J. (**1994**). “Auditory supplements to speechreading: Combining amplitude envelope cues from different spectral regions of speech,” J. Acoust. Soc. Am. **95**, 1065–1073.

Greenwood, D. D. (**1990**). “A cochlear frequency-position function for several species — 29 years later,” J. Acoust. Soc. Am. **87**, 2592–2605.

Hamzavi, J., Pok, S. M., Gstoettner, W., and Baumgartner, W.-D. (**2004**). “Speech perception with a cochlear implant used in conjunction with a hearing aid in the opposite ear," Int. J. Audiol. **43**, 61–65.

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

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., and Neely, J. G. (**1998**). "In vivo measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays," Ann. Otol. Rhinol. Laryngol. **107**, 1–16.

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

Li, T., and Fu, Q.-J. (**2007**). "Perceptual adaptation to spectrally shifted vowels: Training with nonlexical labels," J. Assoc. Res. Otolaryngol. **8**, 32–41.

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

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

Loizou, P. C., Mani, A., and Tu, Z. (**2003**). "Dichotic speech recognition in noise using reduced spectral cues," J. Acoust. Soc. Am. **114**, 475–483.

Loizou, P. C., Dorman, M. F., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**, 2097–2103.

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

Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565–1574.

Macleod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation and recommendations for use," Br. J. Audiol. **24**, 29–43.

Mani, A., Loizou, P. C., Shoup, A., Roland, P., and Kruger, P. (**2004**). "Dichotic speech recognition by bilateral cochlear implant users," International Congress Series, **1273**, 466–469.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Mills, J. H., Dubno, J. R., and He, N. J. (**1996**). "Masking by ipsilateral and contralateral maskers," J. Acoust. Soc. Am. **100**, 3336–3344.

Mosnier, I., Sterkers, O., Bebear, J. P., Godey, B., Robier, A., Deguine, O., Fraysse, B., Bordure, P., Mondain, M., Bouccara, D., Bozorg-Grayeli, A., Borel, S., Ambert-Dahan, E., and Ferrary, E. (**2009**). "Speech performance and sound localization in a complex noisy environment in bilaterally implanted adult patients," Audiology and Neurotology **14**, 106–114.

Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112–1118.

Robinson, D., and Dadson, R. S. (**1956**). "A redetermination of the equal-loudness relations for pure tones," Br. J. Appl. Phys. **7**, 166–181.

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–3636.

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

Rubin, B. A., Uchanski, R. M., and Braida, L. D. (**1992**). "Integration of acoustic cues for consonants across frequency bands (A)," J. Acoust. Soc. Am. **91**, 2473–2474.

Shannon, R. V., Zeng, F.-G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2467–2476.

Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., Neely, J. G., Kletzker, G. R., Brunsden, B., and Blocker, B. (**2002**). "CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients," J. Assoc. Res. Otolaryngol. **3**, 332–350.

Smith, M. W., and Faulkner, A. (**2006**). "Perceptual adaptation by normally hearing listeners to a simulated "hole" in hearing," J. Acoust. Soc. Am. **120**, 4019–4030.

Smith, Z. M., and Delgutte, B. (**2007**). "Using evoked potentials to match interaural electrode pairs with bilateral cochlear implants," J. Assoc. Res. Otolaryngol. **8**, 134–151.

Souza, P., and Rosen, S. (**2009**). "Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech," J. Acoust. Soc. Am. **126**, 792–805.

Stacey, P., and Summerfield, Q. (**2007**). "Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech," J. Acoust. Soc. Am. **121**, 2923–2935.

Stakhovskaya, O., Sridhar, D., Bonham, B., and Leake, P. (**2007**). "Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants," J. Assoc. Res. Otolaryngol. **8**, 220–233.

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

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

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

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

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