

Perceptual adaptation by normally hearing listeners to a simulated “hole” in hearing

Matthew W. Smith^{a)} and Andrew Faulkner

Department of Phonetics and Linguistics, UCL, Wolfson House, 4 Stephenson Way, London NW1 2HE, United Kingdom

(Received 28 September 2005; revised 7 September 2006; accepted 7 September 2006)

Simulations of cochlear implants have demonstrated that the deleterious effects of a frequency misalignment between analysis bands and characteristic frequencies at basally shifted simulated electrode locations are significantly reduced with training. However, a distortion of frequency-to-place mapping may also arise due to a region of dysfunctional neurons that creates a “hole” in the tonotopic representation. This study simulated a 10 mm hole in the mid-frequency region. Noise-band processors were created with six output bands (three apical and three basal to the hole). The spectral information that would have been represented in the hole was either dropped or reassigned to bands on either side. Such reassignment preserves information but warps the place code, which may in itself impair performance. Normally hearing subjects received three hours of training in two reassignment conditions. Speech recognition improved considerably with training. Scores were much lower in a baseline (untrained) condition where information from the hole region was dropped. A second group of subjects trained in this dropped condition did show some improvement; however, scores after training were significantly lower than in the reassignment conditions. These results are consistent with the view that speech processors should present the most informative frequency range irrespective of frequency misalignment. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2359235]

PACS number(s): 43.71.Ky, 43.71.Es, 43.66.Ts [KWG]

Pages: 4019–4030

I. INTRODUCTION

Regions of the cochlea where there are no functioning inner hair cells and/or neurons have been referred to as “dead regions” (Moore, 2004; Moore and Glasberg, 1997) and “holes in hearing” (Shannon, Galvin, and Baskent, 2002). The transduction of basilar membrane vibrations into auditory neural impulses is not possible within dead regions. Even so, if a sound contains frequencies corresponding to the characteristic frequencies (CFs) of the dead region, it is still possible for these frequencies to be detected. If the acoustic intensity of this part of the signal is sufficiently great, the vibration pattern of the basilar membrane will spread to locations neighboring the dead region, where the hair cells are still functioning. Consequently, the frequency information is received but activates neurons in the “wrong” tonotopic location. It is questionable whether “off-frequency” listening of this nature is beneficial to speech perception. For example, a number of studies have examined the benefits of amplification for individuals with high frequency hearing loss and suspected associated dead regions. In these studies the gain of the subjects’ hearing aids was set to make the high frequencies audible. Amplification, however, often did not improve speech intelligibility and sometimes even resulted in poorer performance (Ching, Dillon, and Byrne, 1998; Hogan and Turner, 1998; Turner and Cummings, 1999). In the studies of Vickers, Moore, and Baer (2001) and Baer, Moore, and Kluk (2002), most listeners benefited from the amplification of frequencies somewhat above (1.7 times) the estimated

edge-frequency of the dead region. However, when frequencies higher than this were amplified, performance hardly changed, and even worsened for some individuals. In most cases it was assumed that those individuals who did not benefit from high frequency amplification lacked functional hair cells and/or neurons in the basal part of the cochlea, and that signal detection was being mediated by hair cells at the boundary of, rather than within, the dead region.

Dead regions, or “holes in hearing,” may also have a significant impact on speech perception by cochlear implant listeners, for whom the effective performance of their devices relies on there being functioning neurons at electrode locations. Elevated electrical thresholds may be evident for implant electrodes situated within a dead region, and the normal clinical solution in this sort of situation would be to increase the stimulation levels to the relevant electrodes (Shannon *et al.*, 2002). However, this may cause the electrical activation to spread away from the hole region to neighboring areas of surviving neurons, once again resulting in warping of the tonotopic representation of spectral information. Bearing in mind the previously described results from hearing aid users, Shannon *et al.* (2002) hypothesized that spectral warping in cochlear implants may result in poorer levels of speech intelligibility than if the spectral information from the hole was simply lost. The authors assessed the impact of spectral holes of varying size (1.5–6 mm) and location with both implant patients and normally hearing listeners, the latter using noise-vocoder simulations of CI signal processing. Holes were created by turning off certain electrodes, or by eliminating the relevant output noise bands. Also examined was whether the frequency information from the hole regions

^{a)}Electronic mail: matthew@phon.ucl.ac.uk

could be usefully preserved by remapping the relevant analysis bands to electrodes/output noise bands apical to the hole, basal to the hole, or both sides of the hole; however, these remapping conditions also entailed a warping of the place code of the spectral envelope. The results showed that speech recognition scores were significantly reduced when the holes were at least 4.5 mm wide. Holes in the apical region were the most damaging, while speech information that is highly dependent on spectral cues—vowel identity and consonantal place of articulation—were affected more than information that is predominantly temporally cued—voicing and manner of consonant articulation. Crucially, conditions that attempted to preserve frequency information by remapping it around the hole—and thus warping the place code—appeared to produce results no better than those achieved when the information from the hole region was simply dropped. The authors concluded that these results were not encouraging for the prosthetic restoration of lost information due to a hole in the receptor array, suggesting that the pattern of spectral information becomes unusable when not presented to the correct tonotopic location.

Similar conclusions have been drawn from studies into absolute frequency-place shifting in implants, arising from shallow electrode array insertions. Here, the tonotopic mismatch between the analysis bands in the speech processor and the CFs at implanted electrodes is equivalent to a basalward basilar membrane shift. A number of simulation studies have shown that listeners could tolerate shifts of 3 mm, but that larger shifts produced large decrements in speech recognition performance (Dorman, Loizou, and Rainey, 1997; Shannon, Zeng, and Wygonski, 1998; Fu and Shannon, 1999). However, it is important to note that the subjects in these experiments were given no training and, therefore, had little time to adapt to the effects of tonotopic mismatch. In simulation studies where normally hearing subjects have been given a few hours of training with spectrally shifted speech, the detrimental effects of tonotopic mismatch have been significantly reduced (Rosen, Faulkner, and Wilkinson, 1999; Faulkner, Rosen, and Norman, 2001). Therefore, one might expect that implant users, who are listening with their devices for several hours every day, would also adapt to the frequency mapping provided by their speech processor. This suggestion is supported by the study reported by Harnsberger, Svirsky, Kaiser, Pisoni, Wright, and Meyer (2001). Experienced implant users were asked to provide goodness ratings for synthetic vowel stimuli varying in formant frequencies F1 and F2, in order to construct individual perceptual vowel spaces. If the subjects were failing to adapt to the spectrally shifted information presented by their devices, they would be expected to choose stimuli with lower F1 and F2 formant frequencies than those of natural vowels. However, there was no evidence of any systematic shift in the perceptual vowel spaces.

Shannon *et al.*'s (2002) acute study into “holes in hearing” did not assess the effects of learning. However, listeners might be able to adapt to the distorted representation of information in spectrally warped speech in the same way that they can adapt to spectrally shifted speech. Furthermore, if cochlear implant speech processors were to use analysis fil-

ters frequency-aligned to CFs at electrode locations, those patients with large dead regions would be likely to suffer a significant loss of speech information. Kasturi, Loizou, Dorman, and Spahr (2002) examined the effect of the location and size of spectral holes in a cochlear implant simulation, using speech processed through six frequency bands and synthesized as a sum of sine waves. When a 6.25 mm hole was created in the mid-frequency region, between 790 and 2100 Hz, consonant intelligibility fell from more than 90% correct (when all bands were present) to less than 70% correct. Vowel recognition fell to less than 50% accurate in this particular condition. It seems likely then that dropping the information from a hole region is a less than ideal option.

In the present study, an implant simulation with normally hearing listeners was conducted, similar to that of Shannon *et al.* (2002); however, in this study subjects were given training and time to adapt to each condition. Because of the time-consuming nature of the training, a single hole of fixed size (10 mm) and location (mid-frequency region) was examined; acutely, this produced a pronounced detrimental effect on speech intelligibility. The main hypothesis is that, in the context of a hole in hearing, the detrimental effects of a tonotopic mismatch can be significantly reduced with experience. Consequently, it is also hypothesized that, with training, differences might become apparent between conditions which attempt to preserve frequency information by remapping it around a hole, and conditions in which the information from the hole region is lost. In the first experiment subjects were trained in two “preservation” conditions to see if the precise way in which information from the hole region is remapped (warped) around the hole has a significant affect on speech perception. One of these conditions was similar to that examined by Shannon *et al.*, with the warping effect concentrated at the edges of the hole. In the second, however, the warping effect was spread over the entire frequency range. Performance in these conditions was compared to that in a baseline condition in which information from the hole region was simply dropped. In light of the results from this experiment, a second experiment was conducted in which a different group of subjects was trained in the “information dropped” condition.

II. METHOD

A. Test materials

For each processing condition, three tests were used to evaluate performance: Recognition of keywords in sentences, medial vowel identification, and medial consonant identification. No lip-reading cues were provided. Test materials were digitally recorded, at a sampling rate of 48 kHz, by a male and female speaker using standard Southern British English pronunciation. One female speaker recorded the sentence stimuli and another, the vowel and consonant stimuli. The same male speaker was used for all three tests.

The IEEE sentences (IEEE, 1969) were used for the word recognition task. These comprise 72 blocks of ten sentences, with five scored keywords per sentence. The vowel stimuli comprised ten monophthongs (/æ a ɔ i ε ɪ ʒ ɒ u ʌ/) and six diphthongs (/eʊ eɪ ɪə aɪ əʊ oɪ/) and were presented in

a /b/-vowel-/d/ context: *bad bard bawd bead bed bid bird bod bood bud bared bayed beard bide bode boyd*. These words were presented in the carrier phrase “say /bVd/ again.” Test blocks consisted of 48 stimuli from one speaker with three tokens of each word selected at random from a pool containing five tokens of each word.

Each of the nineteen consonant stimuli (/b tʃ d f g k l m n p r s ʃ t θ v w y z/) were recorded in three vCv contexts: /aCa/, /iCi/, and /uCu/; and were presented in the carrier phrase “say /vCv/ again.” Both speakers recorded five tokens of each consonant in each vowel context. Test blocks consisted of 57 stimuli from one speaker with one token of each consonant in each vowel context, selected at random from the full recorded set.

B. Signal processing

Four speech processors were designed, each utilizing noise-band vocoding similar to that described by Shannon, Zeng, Kamath, Wygonski, and Ekelid (1995). The speech signals were band-pass filtered into a number of contiguous frequency bands, from which amplitude envelopes were extracted via half-wave rectification and low-pass filtering. Each envelope was used to modulate a white noise which was subsequently band-pass filtered. All bands were then summed and presented to the listener. Thus, within each band, temporal and amplitude cues were preserved while spectral detail was removed. The center frequencies and -3 dB cut-off frequencies of the analysis and output filters (Table I) were calculated using Greenwood’s (1990) equation relating basilar membrane location to characteristic frequency, assuming a basilar membrane length of 35 mm. Signal processing was implemented in two distinct ways: Off-line processing of the recorded test materials, and on-line processing of live speech during training.

Off-line processing was implemented in MATLAB. Analysis and output band-pass filters were sixth-order Butterworth IIR designs, while the low-pass filter used in envelope extraction was a third-order Butterworth filter with a cut-off frequency of 400 Hz. Finally, to limit the signal spectrum, the summed waveform was low-pass filtered at the upper cut-off of the highest frequency band using a sixth-order elliptic filter. Real-time processing was implemented using the Aladdin Interactive DSP Workbench (Hitech AB) and a DSP card (Loughborough Sound Images TMS31) running at a sampling rate of 11025 Hz. For each processor, the characteristics of the filters were the same as in the off-line processing but in order to reduce the computational load, sixth-order elliptic rather than Butterworth filters were used.

Three of the four processor designs simulated a “hole in hearing” extending from a point on the basilar membrane 25.8 mm from the base (CF 424 Hz) to a point 15.8 mm from the base (CF 2182 Hz). The first of these three processors divided the speech signals into 12 contiguous frequency bands spanning 130–4518 Hz (Table I) and had tonotopically matched analysis and output filters. To create the hole, the middle six output bands were eliminated [Fig. 1(b)]. This processor is subsequently referred to as the “Dropped” condition since spectral information from the hole region was not available to the listener. It should be emphasized that this

TABLE I. Filter center and cut-off frequencies, representing: (a) Analysis and output bands for the Matched processor and analysis bands for the S-warp processor; (b) analysis bands for the Dropped and A-warp processors; (c) output bands for the Dropped, A-warp, and S-warp processors.

(a)		(b)		(c)	
Center (Hz)	Cut-off (Hz)	Center (Hz)	Cut-off (Hz)	Center (Hz)	Cut-off (Hz)
	130		130		130
(1) 207		(1) 166	207	(1) 166	207
	303	(2) 252	303	(2) 252	303
(2) 424		(3) 360	424	(3) 360	424
	577	(4) 496	577		
(3) 769		(5) 667	769		
	1011	(6) 883	1011		
(4) 1316		(7) 1154	1316		
	1699	(8) 1496	1699		
(5) 2182		(9) 1926	2182		2182
	2789	(10) 2468	2789	(4) 2468	2789
(6) 3555		(11) 3150	3555	(5) 3150	3555
	4518	(12) 4008	4518	(6) 4008	4518

processor used the same output bands as the two reassignment processors, all having three output bands apical to the hole and three basal to the hole (Table I). However, the three designs differed in the mapping of analysis filters to output filters. In attempting to preserve the spectral information encoded in the frequencies of the hole region, the designs of processors two and three warped the place code of the spectral envelope. Processor two—referred to as “S-warp” (for *Spread warping*)—had six contiguous bands spanning 150–4513 Hz (Table I) and mapped these to the three output bands either side of the hole, entailing a downward spectral shift of the lower three input bands and an upward spectral shift of the three higher input bands [Fig. 1(d)]. Processor three—referred to as “A-warp” (for *Adjacent warping*)—

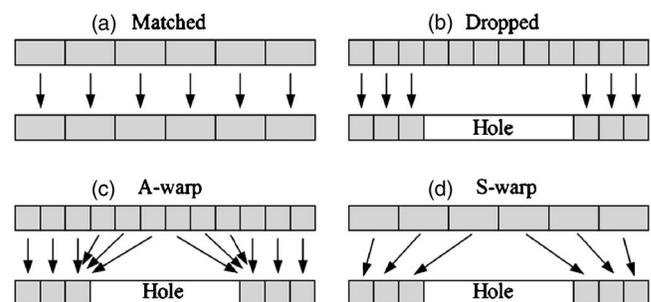


FIG. 1. Mapping of analysis to output bands for the four processors.

analyzed signals into 12 bands; unlike the Dropped processor, however, the middle six bands were not discarded. For both the upper and lower halves of the hole region, envelopes from the three contributing bands were summed and added to the envelope of the band adjacent to the corresponding side of the hole. These two summed envelopes were divided by the number of contributing bands—so that the amplitude was of a similar level to that in neighboring bands (following the method used by Shannon *et al.*, 2002)—and then used to modulate the relevant hole-adjacent output noise-band. All other analysis bands were mapped to tonotopically matched output bands [Fig. 1(c)].

The final processor—referred to as the “Matched” condition—provided an indication of performance level when no hole was present. This was a six-channel, tonotopically matched processor with analysis filters identical to those used in the S-warp condition [Table I and Fig. 1(a)]. As a result of this design the input information to the Matched and S-warp processors was kept the same and enabled a direct test of the effects of warping as implemented in S-warp (an *a priori* assumption was that performance with the S-warp processor would be at least as good as the performance with the A-warp and Dropped processors).

C. Procedure

1. Experiment 1

Experiment 1 investigated the effects of training on the performance with the A-warp and S-warp processors. Eight normally hearing native speakers of English participated in the experiment. Their ages ranged from 18 to 43 and each was paid for taking part. During testing and training, processed speech was presented at a comfortable listening level, via Sennheiser HD25-1 headphones. Presentation was always purely auditory. No feedback was provided in testing and, for the sentence material, subjects never encountered the same sentence twice.

Baseline testing was undertaken in all four processing conditions, the order of presentation being balanced across subjects. First, subjects were familiarized with the sentence task by listening to two lists of BKB sentences (Bench and Bamford, 1979) presented through the Matched processor. The subject attempted to repeat each sentence. A loose scoring method was employed; answers with the same morphological root as that of the target keyword were treated as correct. For familiarization with the vowel task subjects heard an abbreviated test block (16 stimuli) using unprocessed speech. The 16 /b/-vowel-/d/ words were displayed in a grid on a computer screen and subjects attempted to identify each word heard by mouse-clicking on the correct box. A similar familiarization procedure was carried out for the consonant task but with the screen grid displaying orthographic forms of the nineteen consonant stimuli. In the experimental conditions, subjects completed all three tasks before moving on to the next; this meant listening to four lists of IEEE sentences (two each with the male and female speaker), two blocks of vowels (one from each speaker), and two blocks of consonants (one from each speaker) through each processor.

Training commenced once baseline testing was com-

pleted. As all subjects trained with both processors, a crossover design was used to control for order effects—four subjects starting with the A-warp processor and four with the S-warp processor. Training consisted of Connected Discourse Tracking (De Filippo and Scott, 1978), an interactive procedure using connected speech in real-time communication. Speaker (author M.S.) and subject sat in separate sound-isolated rooms connected by an audio link, the voice of the speaker being processed in real time. Stories from the Heinemann Guided Readers series were read to the subject who was required to repeat what the speaker had read before he could move on to the next sentence, phrase, or word. If there was an error in the response the speaker re-read the section until the subject could repeat it correctly. A pragmatic approach to correcting errors was taken: The speaker could use intonation and stress to focus the subject’s attention on the location of the error, and/or read the next phrase or sentence to provide additional contextual cues. However, three incorrect attempts on the part of the subject were followed by a fourth reading but this time with the audio link switched to unprocessed speech. This way, any potential impasse could be avoided. Subjects were familiarized with the CDT procedure using the Matched processor.

For both the A-warp and S-warp processors, training consisted of five 35 min sessions of CDT, broken down into seven 5 min blocks. After each training session, subjects were tested in the same condition using four lists of sentences (two male and two female), two blocks of vowels (one from each speaker), and two blocks of consonants (one from each speaker). After completing training and testing in the first-trained condition, subjects were tested in the other condition to be trained; this provided a fresh baseline measure for that condition. At the end of the experiment, subjects had five minutes of CDT with the Matched processor; this provided some indication of the optimum tracking rate attainable with a six-channel noise-excited vocoder. Finally, re-testing was performed with the Matched and Dropped processors, and with the processor used in earlier training sessions. The order of presentation of the conditions was balanced across subjects. Table II presents the testing and training schedule for a subject trained first with the A-warp processor. Subjects differed in the elapsed time required by the training regime. One week was the maximum time required to complete five sessions for either condition.

2. Experiment 2

Experiment 2 investigated the effects of training on performance with the Dropped processor. A different group of eight normally hearing subjects took part. Again, all were native speakers of English and each was paid for taking part. Their ages ranged from 19 to 43.

Except that subjects were trained in only one condition, and that consonant testing was not conducted (only small training effects being found for these materials in Experiment 1), the procedure was similar to that outlined above. Baseline testing with all four processors was followed by five sessions of training and testing with the Dropped processor. Finally, subjects were re-tested with the Matched, A-warp, and S-warp processors.

TABLE II. Schedule for a subject trained first with the A-warp processor and then with the S-warp processor (m=Matched, d=Dropped).

Session	CDT (35 min)	Number of male and female talker		
		Sentence lists	Vowel blocks	Consonant blocks
Baseline 1	1	m: 2m,2f	m: 1m,1f	m: 1m,1f
	2	d: 2m,2f	d: 1m,1f	d: 1m,1f
		a: 2m,2f	a: 1m,1f	a: 1m,1f
		s: 2m,2f	s: 1m,1f	s: 1m,1f
A-warp training	3	✓	a: 2m,2f	a: 1m,1f
	4	✓	a: 2m,2f	a: 1m,1f
	5	✓	a: 2m,2f	a: 1m,1f
	6	✓	a: 2m,2f	a: 1m,1f
	7	✓	a: 2m,2f	a: 1m,1f
Baseline 2	8		s: 2m,2f	s: 1m,1f
	9	✓	s: 2m,2f	s: 1m,1f
S-warp training	10	✓	s: 2m,2f	s: 1m,1f
	11	✓	s: 2m,2f	s: 1m,1f
	12	✓	s: 2m,2f	s: 1m,1f
	13	✓	s: 2m,2f	s: 1m,1f
Re-test	14		m: 2m,2f	m: 1m,1f
			d: 2m,2f	d: 1m,1f
			a: 2m,2f	a: 1m,1f

III. RESULTS

For both experiments data were analyzed by repeated-measures ANOVA ($\alpha=0.05$) using within-subject factors of processor, sessions of training, and speaker gender, and, in Experiment 1, the between-subjects factor of training order. Huynh-Feldt corrections were applied to all F tests, and *post hoc* tests were carried out using Bonferroni-corrected pairwise comparisons. For the linear regression analyses, scores were adjusted in order to remove between-subject variability, while preserving differences across training, by subtracting the difference between the grand mean and the subject's mean score over the five training sessions.

A. Experiment 1

1. Sentences

The results are summarized in Fig. 2. Significant effects of processor were apparent both at baseline [$F(2.61, 18.2) = 121, p < 0.001$] and at the end of the experiment [$F(3, 18) = 72.8, p < 0.001$]—where A-warp and S-warp scores from the final training session were compared with retest scores for the Matched and Dropped processors. At both points in time *a priori* contrasts indicated that scores with the Dropped processor were significantly lower than scores in the other three conditions. In addition, *post hoc* testing showed that scores in the Matched condition were significantly higher than scores in the other conditions at baseline; however, performance in the Matched and S-warp processors was statistically equivalent at the end of the experiment, though this may have been due to a ceiling effect with the Matched processor.

A comparison of sentence scores between the A-warp and S-warp processors, based on scores from the relevant

baseline session and all five training sessions, revealed main effects of processor [$F(1, 6) = 85.5, p < 0.001$], speaker [$F(1, 6) = 115, p < 0.001$], and training [$F(5, 30) = 31.0, p < 0.001$]. Performance for S-warp was significantly better than for A-warp, the mean and median scores of the former being higher at every stage, while higher scores were achieved with the female speaker in both conditions. From Fig. 2, it is evident that scores for S-warp and A-warp were much better after the first training session than at baseline. Nevertheless, there were indications that some performance improvements took place between Sessions 1 and 5 of training. *A priori* contrasts showed that scores after Session 5 were greater than scores not only at baseline, but also at Sessions 1 and 2, while linear regression—excluding baseline scores—indicated a significant correlation between performance and training session for S-warp, though not for A-warp (Table III). Even though no training was given with the Dropped processor, scores were significantly better at retest than at baseline [$F(1, 7) = 9.70, p < 0.05$].

The absence of a main effect of training order indicated that there was no difference in the overall performance of subjects who trained first with A-warp and those who trained first with S-warp (Fig. 3). Subjects generally performed better in the condition trained second, suggesting a generalized learning effect for noise-vocoded speech being carried over from the condition trained first. Nevertheless, this advantage was small when the second condition was A-warp and somewhat larger when it was S-warp (5% and 12% differences, respectively). In addition, a condition by training by training order interaction [$F(4.69, 28.16) = 4.44, p < 0.01$] suggested that S-warp scores improved irrespective of training order, while A-warp scores improved only when A-warp was trained first.

2. Vowels

The results are summarized in Fig. 2. Baseline scores represent a significant effect of processor [$F(3, 21) = 58.0, p < 0.001$], as do scores at the end of the experiment [$F(3, 18) = 61.4, p < 0.001$]. Again, *a priori* contrasts indicated significantly poorer performance in the Dropped condition than in the other conditions. At the baseline, *post hoc* contrasts showed that scores with the Matched processor were significantly better than those with the other processors but, as with the sentences, Matched and S-warp processor scores were statistically equivalent at the end of the experiment.

Comparison of A-warp and S-warp scores from the baseline to the end of training, demonstrated significant effects of processor [$F(1, 6) = 49.0, p < 0.001$], speaker [$F(1, 6) = 17.8, p < 0.01$], and training [$F(5, 30) = 14.4, p < 0.001$]. S-warp performance was significantly better than A-warp performance, mean and median scores for S-warp being higher at every stage of training, and scores with the male speaker were significantly higher than with the female speaker. However, there was a significant talker by condition interaction [$F(1, 6) = 58.5, p < 0.001$], a separate analysis for the two speakers indicating that although S-warp scores were significantly better than A-warp scores with the male speaker [$F(1, 6) = 66.8, p < 0.001$], the difference was not significant

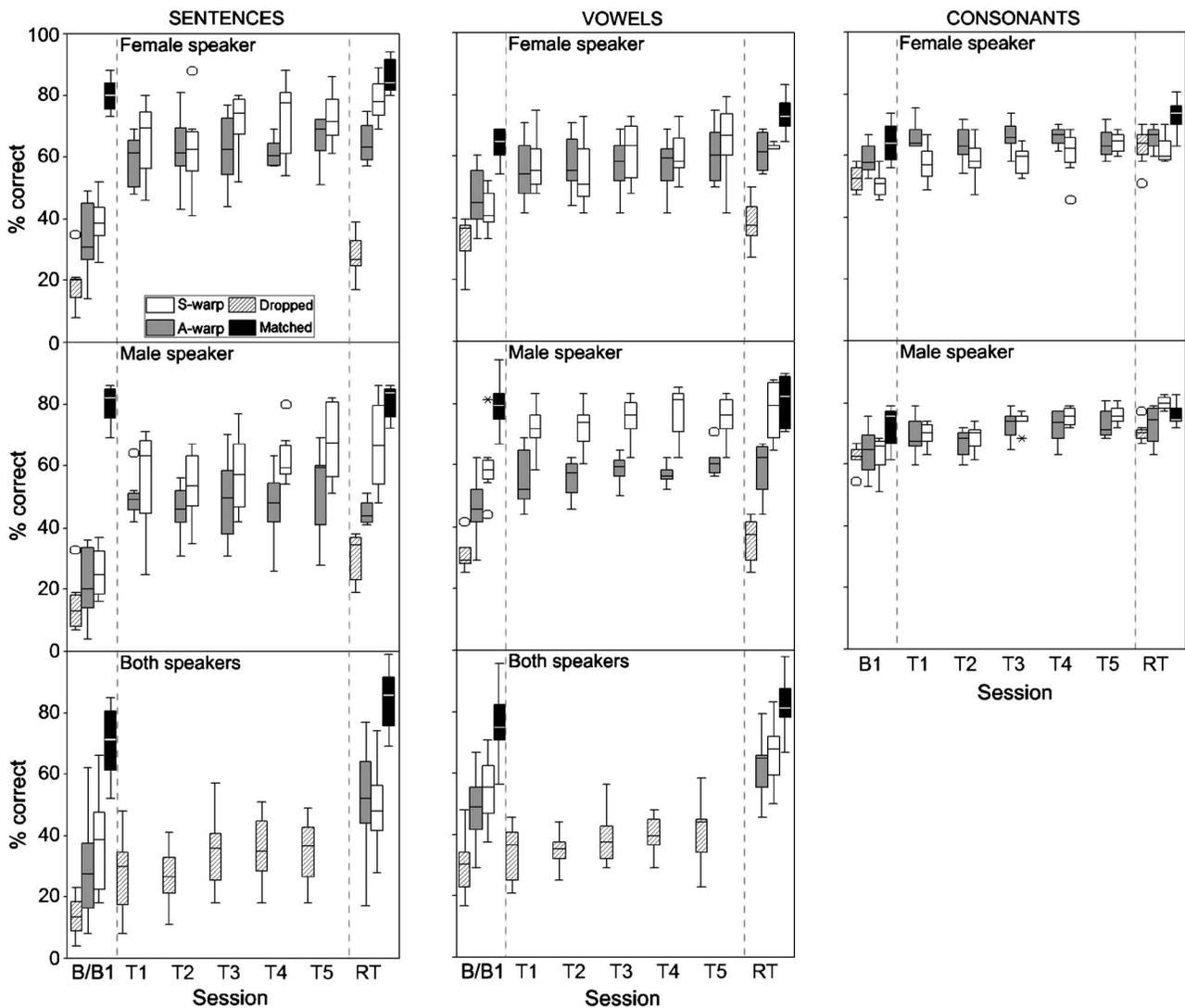


FIG. 2. Performance as a function of session: Baseline B1 (experiment 1) or B (experiment 2), training sessions T1 to T5, retest RT. The upper six panels show data from Experiment 1, and the lower two panels data from Experiment 2 (where consonant testing was not carried out). No effect of speaker was observed in Experiment 2, so data from both speakers is presented in each panel. Baseline and retest data for the “untrained” processors are also shown.

for the female speaker [$F(1,6)=4.31, p=0.083$]. With regard to the training effect, there were indications that, as with the sentences, some improvement in performance occurred after the first training session. *A priori* contrasts showed that scores from the final session were significantly higher than scores from all prior sessions except the fourth, while linear regression across the training sessions demonstrated a significant correlation between training session and scores for both processors (Table III). Comparing the baseline and retest scores of the untrained processors, there was a significant effect of session for the Matched processor [$F(1,7)=5.71, p<0.05$] but not for the Dropped processor.

There was no main effect of training order. However, a significant condition by training order effect was observed [$F(1,6)=37.8, p<0.005$]; subjects who trained second with A-warp achieved almost identical scores for both processors (Fig. 3), while those who trained second with S-warp performed much better in this condition than in the A-warp condition (19% difference in the means).

3. Consonants

Performance was influenced in only minor ways by the different types of spectral manipulation in the four conditions (Fig. 2). Consonant recognition is especially influenced by temporal cues, and these are reasonably well-preserved in noise-band representations of speech, even when only a small number of contiguous bands are available (Shannon *et al.*, 1995). Although main effects of the processor were seen both at the baseline [$F(2.34, 16.4)=13.8, p<0.001$] and at the end of training [$F(3, 18)=6.52, p<0.005$], the only significant *a priori* contrasts were between the Dropped and Matched conditions. At the end of the training there was no significant difference between Matched, A-warp, and S-warp processors according to *post hoc* testing.

In an ANOVA comparing S-warp and A-warp scores across baseline and training sessions, the effect of the processor narrowly missed significance [$F(1,6)=5.69, p=0.054$]. However, male speaker scores were significantly higher than female speaker scores [$F(1,6)=286, p<0.001$] and a significant talker by condition interaction was also ob-

TABLE III. Performance as a function of training sessions T1 to T5. Data from linear regression analyses.

	Processor/ Speaker	R	(R^2)	F	p	Regression line (% increase in score/session)
Sentences	S-warp (f)	0.435	0.189	8.86	$p < 0.01$	2.35
	S-warp (m)	0.487	0.237	11.8	$p < 0.005$	3.28
	A-warp (f)	0.285	0.081	3.36	n.s.	1.31
	A-warp (m)	0.148	0.022	0.85	n.s.	0.68
	Dropped (f)	0.432	0.187	8.71	$p < 0.01$	2.31
	Dropped (m)	0.475	0.225	11.1	$p < 0.005$	2.15
Vowels	S-warp (f)	0.560	0.313	17.3	$p < 0.001$	2.16
	S-warp (m)	0.406	0.165	7.49	$p < 0.01$	1.23
	A-warp (f)	0.368	0.135	5.94	$p < 0.05$	1.01
	A-warp (m)	0.383	0.147	6.55	$p < 0.05$	1.20
	Dropped (f)	0.511	0.261	13.4	$p < 0.005$	1.74
	Dropped (m)	0.571	0.326	18.4	$p < 0.001$	1.78
Consonants	S-warp (f)	0.483	0.233	11.6	$p < 0.005$	1.58
	S-warp (m)	0.643	0.414	26.9	$p < 0.001$	1.90
	A-warp (f)	0.082	0.007	0.26	n.s.	-0.20
	A-warp (m)	0.535	0.286	15.2	$p < 0.001$	1.34

served [$F(1,6)=83.3$, $p < 0.001$], indicating that performance with the A-warp processor was significantly better than performance with the S-warp processor for the female speaker but not for the male speaker. A significant effect of training [$F(5,30)=17.3$, $p < 0.001$] and a just significant condition by session interaction [$F(5,30)=2.63$, $p=0.044$] were observed. Consequently, a separate analysis was carried out for the two conditions. S-warp scores at the final training session were significantly higher than at all sessions apart from the fourth, while *post hoc* tests indicated that baseline scores were significantly lower than scores at Sessions 3 and 4. For the A-warp processor, scores at the final training session were significantly higher than those at the baseline and at the second training session. Scores from Session 4 were also higher than at Session 2 according to *post hoc* contrasts. Linear regression analysis (Table III) showed a significant correlation between scores and training session for the S-warp processor. However, for the A-warp processor the correlation was only significant with the male speaker. Again there was evidence of generalized learning for the untrained conditions, with retest scores significantly better than baseline scores for both the Matched [$F(1,7)=11.0$, $p < 0.05$] and the Dropped processor [$F(1,7)=75.4$, $p < 0.001$].

A significant interaction of condition with training order [$F(1,6)=23.5$, $p < 0.005$] showed that scores were higher in the condition trained second. Even so, differences between the two processors were small regardless of training order (Fig. 3).

B. Experiment 2

As the baseline session was essentially identical to that in Experiment 1, it is not surprising that, for both sentence and vowel data, the same pattern of processor effects was seen prior to training (Fig. 2).

1. Sentences

For the Dropped processor there was a main effect of training [$F(5,35)=22.8$, $p < 0.001$] but not of the speaker.

Again, there were indications that some performance improvements took place between Sessions 1 and 5 of training. *A priori* contrasts showed that Session 5 scores were higher than baseline scores and scores from Sessions 1 and 2. A linear regression analysis of scores from the training sessions—excluding baseline scores—indicated a significant correlation between training session and scores (Table III). Nevertheless, at the end of the experiment (when retest scores for the Matched, S-warp, and A-warp processors were compared with the session five score for the Dropped processor) performance in the Dropped condition was still significantly poorer than performance in the other three conditions. This is at least partly due to the fact that performance also improved in the untrained conditions. There were significant effects of session (baseline versus retest) for the Matched and A-warp processors [$F(1,7)=17.8$, $p < 0.005$ and $F(1,7)=27.2$, $p < 0.005$, respectively], while the effect of the session narrowly missed significance for S-warp [$F(1,7)=5.38$, $p=0.053$].

2. Vowels

Here, there was a main effect of training [$F(4.93,34.5)=14.1$, $p < 0.001$] but not of the speaker. The pattern of contrasts was similar to that seen with sentences. Performance at Session 5 was significantly better than at baseline and at training Sessions 1 and 2, while *post hoc* comparisons showed that baseline scores were significantly lower than scores after the third and fourth training sessions. The results of linear regression analysis (Table III) showed that there was still a significant correlation between the performance and training session, even when baseline scores were excluded. However, as with sentences, a significant effect of processor at the end of the experiment [$F(1.68,11.7)=65.0$, $p < 0.001$] indicated that scores in the Dropped condition were still significantly lower than in the other conditions, despite subjects only having had training in the Dropped condition. Once again, S-warp and A-warp scores rose sig-

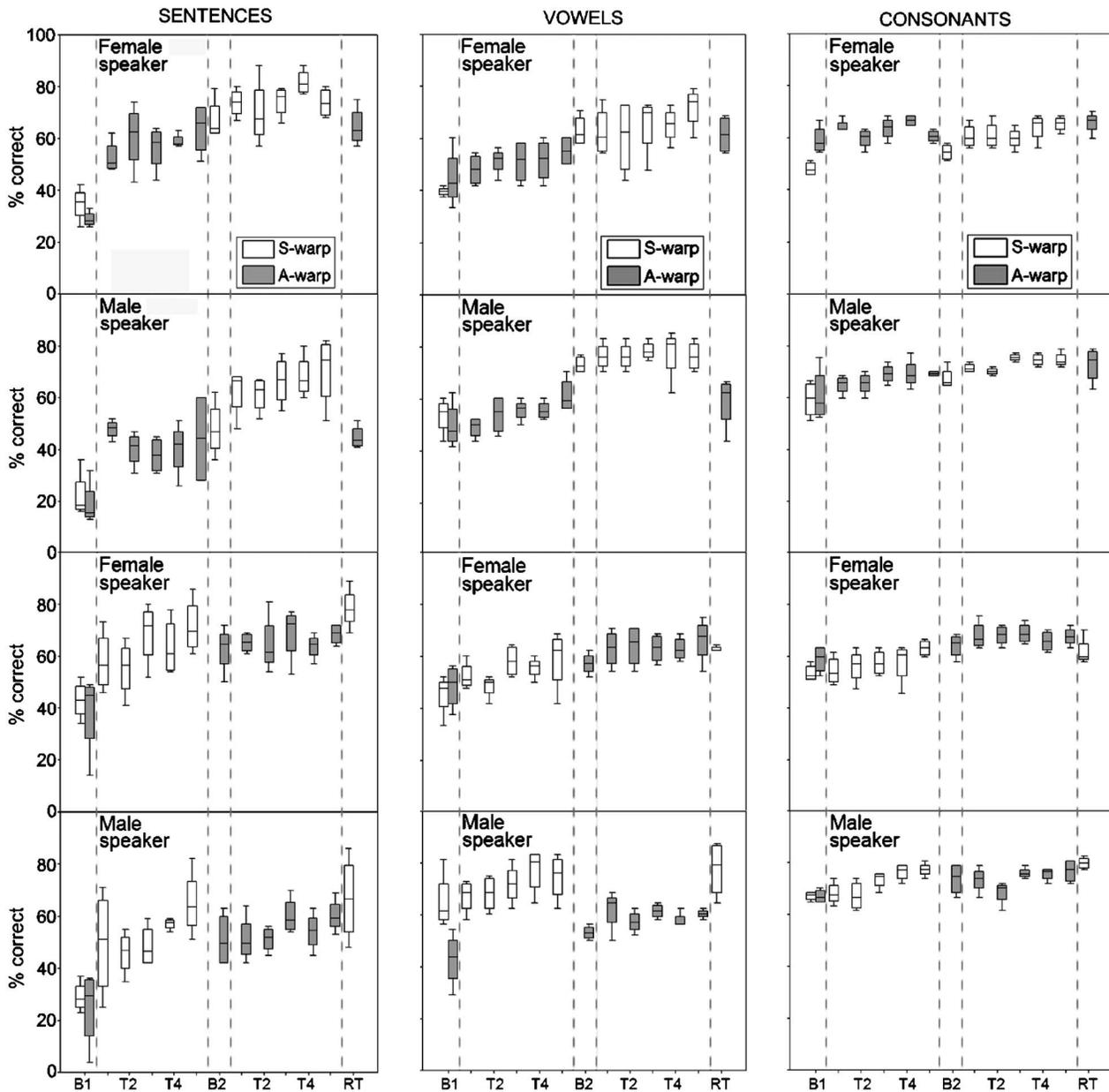


FIG. 3. Experiment 1: The upper six panels represent scores for subjects who trained first with the A-warp processor, and the lower six panels scores for subjects who trained first with the S-warp processor (B2=baseline for the second condition trained). Male and female scores are presented separately.

nificantly between the baseline and retest [$F(1,7)=19.6$, $p < 0.005$ and $F(1,7)=41.9$, $p < 0.001$, respectively].

C. Between-subjects comparisons: Experiment 1 vs Experiment 2

Subjects in both experiments received the same amount of processor-specific training; however, subjects in Experiment 1 received twice the overall amount of training as subjects in Experiment 2. Despite this, the complete data set from Experiment 1 was incorporated into the following analyses in order to increase their power. This is taken into consideration in the discussion of the results.

Between-subjects comparisons across experiments—incorporating data from baseline and all five training sessions—revealed a main effect of processor for sentences [$F(2,21)=30.6$, $p < 0.001$] and vowels [$F(2,21)=43.2$, $p < 0.001$]. *Post hoc* tests revealed that S-warp and A-warp

scores were significantly higher than scores with the Dropped processor for both types of test material.

D. Connected discourse tracking

It should be noted that “corrections”—sections of the training text read with the audio link switched to unprocessed speech—were counted in the final “words per minute” scores. The general pattern of results seen in the analysis of the sentence, vowel and consonant data is also reflected in the CDT rates (Fig. 4). Between-subjects comparisons across experiments showed that overall CDT rates for the Dropped processor were significantly lower than rates for the A-warp and S-warp processors [$F(1,14)=18.0$, $p < 0.005$ and $F(1,14)=23.9$, $p < 0.001$, respectively], while in Experiment 1, a significant effect of the processing condition indicated that S-warp tracking rates were a little higher than those of

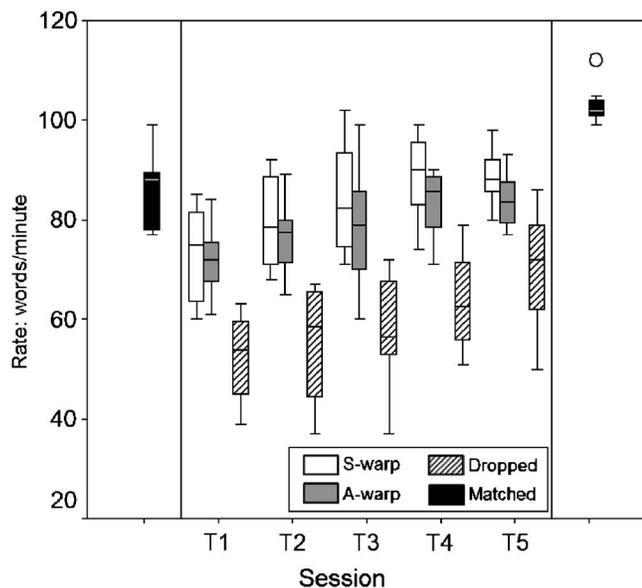


FIG. 4. CDT rates for subjects in experiment 1 (A-warp and S-warp processors) and experiment 2 (Dropped processor). Overall, subjects in experiment 1 had twice as much training as those in experiment 2. Performance with the Matched processor, before and after training, is also shown.

A-warp [$F(1, 6)=7.30, p < 0.05$]. The rates achieved with the Matched processor at the end of Experiment 1 were the highest of all. When compared with CDT rates from the final training session, a main effect of the condition was seen [$F(2, 10)=70.5, p < 0.001$], *post hoc* tests showing that CDT rates with the Matched processor (103 words per minute) were significantly higher than S-warp rates (90 wpm) and A-warp rates (85 wpm). In addition, a main effect of the training on CDT rates was seen in Experiments 1 and 2 [$F(3.74, 22.4)=77.7, p < 0.001$ and $F(4, 28)=21.7, p < 0.001$, respectively] though some of this improvement may be attributable to increasing experience with the speaker (author M.S.) and the particular training texts used.

IV. DISCUSSION

This simulation of cochlear implant signal processing confirmed that a hole in the spectral representation of speech can have a significant effect on intelligibility. When the information from the hole region was dropped altogether performance was poor, even after three hours of training, with only 34% accuracy achieved in sentence keyword identification, and 41% in vowel identification. The frequency region defined by the hole (424–2182 Hz) included much of the F1 and F2 formant frequency information that is crucial to vowel identification (Peterson and Barney, 1952; Nearey, 1989). However, vowel recognition remains well above chance in noise-vocoder processing, even when there are no spectral cues, because nonspectral cues such as duration and overall amplitude are preserved (Shannon *et al.*, 1995); this supports the conclusion that the Dropped processor transmitted little of the spectral detail required for vowel identification. Consonant scores were less severely affected by the hole, as demonstrated by the fact that scores in the A-warp, S-warp, and Dropped conditions were similar to scores in the Matched, “no hole,” condition. This finding is similar to

TABLE IV. Mean scores at the end of training (Session T5). Matched processor scores and consonant scores for the Dropped processor (in *italics*) are taken from the Experiment 1 Retest session.

	A-warp	S-warp	Dropped	Matched
Sentences	59%	70%	34%	83%
Vowels	61%	70%	41%	77%
Consonants	69%	70%	67%	75%

those of Lippmann (1996) and Breeuwer and Plomp (1984, 1985, 1986) for nonvocalized speech signals, which found that high levels of consonant recognition could be achieved with widely separated bands of high and low frequency information. These results and the findings of the present study show how the preserved temporal and durational cues in speech (including noise-vocalized speech) can provide reasonably good consonant intelligibility. In the present study it can also be inferred that the between-condition differences seen in sentence testing were due primarily to different levels of vowel intelligibility (see Table IV).

Subjects in Experiment 1 received twice the overall amount of training received by subjects in Experiment 2. Nevertheless, it seems unlikely that five more training sessions with the Dropped processor would have produced levels of intelligibility equivalent to those observed with the S-warp and A-warp processors: The differences between the mean scores, post-training, were great—S-warp 70%, A-warp 59%, Dropped 34% for sentences; S-warp 70%, A-warp 61%, Dropped 41% for vowels—while the gradients of the regression lines for the Dropped processor were around 2% per session (Table III). Therefore, the critical finding of this study was that speech intelligibility was much higher in the “preservation” conditions than in the condition where information from the hole region was simply dropped, a conclusion supported by the CDT data for the three processors. Indeed, for the sentence and vowel data, scores in the Dropped condition were significantly lower than S-warp and A-warp scores even before training, probably because the size and location of the hole had an especially deleterious effect when listening with the Dropped processor.

Although the S-warp and A-warp processors mapped frequencies to the wrong tonotopic location in the cochlea subjects clearly learned to adapt to the distorted patterns of spectral information; for instance, with the S-warp processor accuracy of keyword identification improved by nearly 40%. Despite it not being possible to conclude that a longer period of training would have led to complete adaptation, it is interesting to note that at the end of Experiment 1, test scores in the S-warp condition were statistically equivalent to scores in the Matched condition; this, however, may be partly due to ceiling effects in the sentence data, and the degree of variability observed among subjects in the S-warp condition. In addition, post-training scores with the Matched processor might have been higher if subjects had been trained in this condition. Scores for the Dropped processor in Experiment 1 were significantly better at retest than at baseline even without training. Thus it was felt necessary to in-

TABLE V. Average frequencies of the first and second formants of the male and female speaker vowels.

		æ	ɑ	ɔ	i	ɛ	ɪ	ʊ	ʊ	u	ʌ
		bad	bard	board	bead	bed	bid	bird	bod	booed	bud
(M)	F1	683	639	473	294	619	416	500	647	317	613
	F2	1608	1080	810	2337	1871	2080	1601	987	1654	1452
(F)	F1	1140	808	496	303	803	573	678	557	333	763
	F2	1899	1166	912	2708	2196	2340	1811	1125	1956	1532

investigate whether a greater degree of improvement might be seen if training was provided in this condition; however, the scores achieved after five sessions of training were not dissimilar to those recorded at retest in Experiment 1. Therefore, the improvement in performance with the Dropped processor probably reflected a general learning effect for noise-vocoded speech (performance improvements—with relatively little exposure—with tonotopically matched noise-vocoded speech have been reported elsewhere, e.g., Faulkner, Rosen, and Stanton, 2003; Davis *et al.*, 2005). However, the superior scores achieved with the S-warp and A-warp processors suggest that the pattern of spectral information and the primary speech cues were still usable to some considerable degree in these conditions despite being presented to the wrong tonotopic location.

By the end of training S-warp scores were approximately 10% better than A-warp scores on the sentence and vowel tests, though the conditions were not significantly different at the baseline. In both conditions the frequencies above and below the center of the frequency range (1011 Hz) were compressed, and warped in opposite directions. For S-warp, the warping effect was spread over the entire frequency range; however, the compression factor was nonuniform with frequencies near the center of the hole being shifted by just over an octave while those in the most apical and basal bands were shifted by under one-third of an octave. In contrast, warping in the A-warp condition was localized to the hole-adjacent bands, with tonotopically matched analysis and output bands apical and basal to these. In order to distinguish spectral patterns when speech signals are reduced to a small number of bands, as they are in cochlear implants and implant simulations, listeners must be able to detect differences between the envelope-modulated levels of neighboring output bands. That the ratios between formant frequencies serve as important cues to speech sounds has been demonstrated in many studies; for instance, Peterson and Barney, 1952; Ladefoged and Broadbent, 1957. In the S-warp condition equal acoustic frequency ranges were mapped to output bands representing constant distances along the cochlea. With the A-warp condition much larger frequency ranges were mapped to the hole-adjacent output bands than to other output bands. It appears that this mapping design, and the consequent reduction in the resolution of hole region frequencies, was less effective at preserving (within the apical and basal portions of the output signal) the information carried by the spectral shape of the original signal.

For the preservation conditions performance was better with the male speaker in the vowel and consonant tests but

better with the female speaker for words in sentences. The female speaker sentences were delivered at a slightly slower and more deliberate pace than the male speaker sentences. It is possible that the female speaker advantage would have been nullified, or even reversed, if the speaking rate of the two speakers had been the same. It is not entirely clear why scores in the vowel and consonant tests were better with the male speaker. The ratios between the frequencies of formant peaks will have been particularly disrupted if the first formant was warped to the apical side of the hole and the second formant was warped to the basal side. However, an examination of the formant frequencies in Table V (with reference to the processor filter cut-off frequencies in Table I) indicates that the likelihood of “formant splitting” is no greater for the female than the male speaker. Perhaps a more straightforward explanation for the speaker differences is that the CDT was conducted with the same male speaker who recorded the test materials; subjects, therefore, had more experience of listening to this speaker.

Drawing comparisons with the acute investigation into the effect of spectral holes of Shannon *et al.* (2002) is not straightforward because of methodological differences. In that study a larger frequency range was divided into 20 contiguous bands, and the largest hole examined was 6 mm. Their study also examined the effect of holes varying in size and location, and both cochlear implant patients and normally hearing listeners took part. Notwithstanding these differences, it is still interesting to compare the main findings: Shannon *et al.* found no significant differences between conditions that reassigned information from the hole region—including a condition similar to A-warp—and a condition in which the information from the hole region was simply dropped. Their conclusion was that the pattern of spectral information becomes unusable if it is not presented to the correct tonotopic location. However, the results of the present study indicate otherwise, at least for the size and location of hole investigated here. Subjects were clearly able to make use of warped spectral information and performed better in the preservation conditions than in the Dropped condition, even before training. Furthermore, the detrimental effect of distorting the spectral representation of speech was significantly reduced with training, an outcome consistent with results from other implant simulation studies (e.g., Rosen *et al.*, 1999; Faulkner *et al.*, 2001).

A. Implications for Cochlear Implants

Evidence is accumulating to suggest that implant users are able to adapt when their speech processors present infor-

mation to the wrong tonotopic location (Harnsberger *et al.*, 2001; McKay and Henshall, 2002; Fu, Shannon, and Galvin, 2002). If the speech perceptual difficulties resulting from frequency-to-place mismatch can be readily overcome with experience, it would appear preferable that the most informative frequency range is delivered by the speech processor. Evidence from another study simulating holes in hearing suggests that the location and pattern of holes affect mostly vowel recognition (Kasturi *et al.*, 2002). These authors conclude that neuronal “dead” regions ought to account for some of the variability in vowel recognition performance among cochlear implant listeners. They also go on to suggest that speech processor frequency spacing should be customized for each implant user, with frequency-importance functions—a measure of the proportion of speech information in each channel—having smaller weights on less functional electrodes that may be located within dead regions.

However, it is still not clear whether hearing-impaired individuals gain any benefit from acoustic information presented to a dead region. Several studies with hearing aid users have questioned the benefits of providing high-frequency amplification for individuals with steeply sloping losses (e.g., Murray and Byrne, 1986; Ching *et al.*, 1998; Hogan and Turner, 1998; Turner and Cummings, 1999). In these studies some subjects demonstrated no improvement or even reduced speech intelligibility with high-frequency amplification. Vickers *et al.* (2001) found that amplifying frequencies up to one octave above the estimated edge frequency of a dead region provided some benefit to listeners but amplification of frequencies well within the dead region did not benefit listeners and actually led to poorer performance in some cases. For individuals whose high frequency hearing loss may be associated with a neuronal dead region, Turner and colleagues have suggested that such amplification produces a spread of excitation in the cochlea near the amplified region and a consequent distortion of the spectral pattern. Something similar may happen in cochlear implants where there are electrodes located in a dead region. It may be possible to present an audible signal to the patient by increasing the level of electrical stimulation to the electrode(s); however, the result might be a spread of excitation with functional auditory neurons neighboring the dead region being activated. Again, this distortion of the spectral pattern of information might lead to poorer speech recognition performance. Nevertheless, the results of the present study suggest that, because cochlear implant listeners may well be able to adapt to spectrally warped speech (in a controlled reassignment of frequency to cochlear location), effective speech processors could be designed which route spectral information away from electrodes in the hole region.

The findings of the present study are also relevant to the design of speech processing strategies that combine cochlear implant use with a hearing aid in the nonimplanted ear. For instance, it is conceivable that an implant user could have a relatively shallow electrode array insertion in one ear and some residual low frequency hearing, stimuable with a hearing aid, in the opposite ear. If there were no overlap between the stimuable regions of the two cochleae—ignoring the possibility of a precise abutment of these regions—

effectively, there would be a “hole” in hearing *across the two ears*. Although benefits, in terms of improved speech recognition, sound localization and sound quality, have been demonstrated for bilateral-bimodal listening (e.g., Ching *et al.*, 2001 and 2004; Hamzavi *et al.* 2004; Dettman *et al.* 2004; Tyler *et al.* 2002; Armstrong *et al.* 1997), there is considerable variability in performance, and many patients revert to using the implant on its own because they receive no benefit, or even experience interference, when using the acoustic device at the same time. Studies, to date, have given little consideration to the likely incongruities in the frequency-to-place code mapping of the two devices and their effect on speech recognition performance. Recently, “soft surgery” techniques have been developed with the aim of preserving residual low frequency hearing by using a relatively shallow insertion electrode array in the same ear (combined electric and acoustic stimulation, EAS). The results of an EAS simulation study (Dorman *et al.*, 2005) demonstrated that speech intelligibility scores were highest in conditions that minimized the frequency gap between low frequency acoustic hearing and simulated electrode locations. The authors also report the findings of a supplementary experiment where the frequency components from the gap (or “hole”) region were shifted up instead of being removed. For several simulated electrode insertion depths intelligibility scores were poorer than when “holes” were left in the spectrum; however, subjects received little practice with the up-shifted conditions.

Simulation experiments like the ones described in the present study could also be used to examine questions related to signal processing in bilateral cochlear implants. Differences in insertion depth of bilateral electrode arrays are likely to lead to frequency-place maps that are in conflict between the two ears. Recent studies (e.g., Dorman and Dahlstrom, 2004) suggest that implant users can make use of mismatched information from the two auditory peripheries, such that speech intelligibility scores with both implants are better than when either is used on its own. However, little analytic work appears to have been carried out regarding the effects of parametric manipulations of the two implant speech processor maps.

V. CONCLUSION

In the context of speech recognition with cochlear implants, the present findings suggest that rerouting spectral information around a hole may be better than simply dropping it, even though the pattern of information is warped. Post-training performance levels in two preservation conditions were significantly better than in a condition where information from the hole region was simply dropped. The deleterious effect caused by warping the speech spectrum was significantly reduced with just three hours of training—a period of time that is inconsequential compared to the amount of experience cochlear implant patients have with their devices. Although subjects also adapted somewhat to the Dropped condition, the effect was small by comparison; this is not surprising since much of the formant frequency detail was lost. Of the two preservation conditions, performance was better when the warping was spread over the

entire frequency range; when warping was localized to hole-adjacent bands, sentence and vowel recognition scores were around 10% lower despite the fact that, beyond the hole region, analysis bands were mapped to the correct tonotopic location. This suggests that there may be performance benefits for warping schemes that map equal acoustic frequency ranges to electrodes representing constant distances within the cochlea; however, further experiments, with holes in different locations, would be required to confirm this.

Although the present results cannot predict complete adaptation to spectral warping with further training, they do imply that acute studies probably underestimate the degree to which listeners can adapt to this type of distortion. The findings are also consistent with results from other simulation studies that have suggested that speech processors should present the most informative frequency range irrespective of frequency misalignment.

ACKNOWLEDGMENTS

The authors are grateful to the Royal National Institute for Deaf People for supporting this work through the awarding of a PhD studentship to Matthew Smith.

Armstrong, M., Pegg, P., James, C., and Blamey, P. (1997). "Speech perception in noise with implant and hearing aid," *Am. J. Otol.* **18**, S140–S141.

Baer, T., Moore, B. C. J., and Kluk, K. (2002). "Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies," *J. Acoust. Soc. Am.* **112**, 1133–1144.

Bench, R. J., and Bamford, J. M. (Eds.) (1979). *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children* (Academic, London).

Breeuwer, M., and Plomp, R. (1984). "Speechreading supplemented with frequency-selective sound-pressure information," *J. Acoust. Soc. Am.* **76**, 686–691.

Breeuwer, M., and Plomp, R. (1985). "Speechreading supplemented with formant-frequency information from voiced speech," *J. Acoust. Soc. Am.* **77**, 314–317.

Breeuwer, M., and Plomp, R. (1986). "Speechreading supplemented with auditorily presented speech parameters," *J. Acoust. Soc. Am.* **79**, 481–499.

Ching, T. Y. C., Dillon, H., and Byrne, D. (1998). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," *J. Acoust. Soc. Am.* **103**, 1128–1140.

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., Psarros, C., Hill, M., Dillon, H., and Incerti, P. (2001). "Should children who use cochlear implants wear hearing aids in the opposite ear?" *Ear Hear.* **22**, 365–380.

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

Dettman, S., D'Costa, W. A., Dowell, R. C., Winton, E. J., Hill, K. L., and Williams, S. S. (2004). "Cochlear implants for children with significant residual hearing," *Arch. Otolaryngol. Head Neck Surg.* **130**, 612–618.

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. (1997). "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," *J. Acoust. Soc. Am.* **102**, 2993–2996.

Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J., and Schmidt, J. S. (2005). "Acoustic simulations of combined electric and acoustic hearing," *Ear Hear.* **26**, 1–10.

Faulkner, A., Rosen, S., and Norman, C. (2001). "The right information matters more than frequency-place alignment: Simulations of cochlear implant processors with an electrode array insertion depth of 17 mm," *Speech, Hearing and Language: Work in Progress* **13**, 52–71.

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.

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. (2002). "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," *J. Acoust. Soc. Am.* **112**, 1664–1674.

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. F., 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.

Harnsberger, J. D., Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., and Meyer, T. A. (2001). "Perceptual 'vowel spaces' of cochlear implant users: Implications for the study of auditory adaptation to spectral shift," *J. Acoust. Soc. Am.* **109**, 2135–2145.

Hogan, C. A., and Turner, C. W. (1998). "High-frequency audibility: Benefits for hearing-impaired listeners," *J. Acoust. Soc. Am.* **104**, 432–441.

IEEE (1969). "Recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **17**, 225–246.

Kasturi, K., Loizou, P. C., Dorman, M., and Spahr, T. (2002). "The intelligibility of speech with 'holes' in the spectrum," *J. Acoust. Soc. Am.* **112**, 1102–1111.

Ladefoged, P., and Broadbent, D. E. (1957). "Information conveyed by vowels," *J. Acoust. Soc. Am.* **29**, 98–104.

Lippmann, R. P. (1996). "Accurate consonant perception without mid-frequency speech energy," *IEEE Trans. Speech Audio Process.* **4**, 66–69.

McKay, C. M., and Henshall, K. R. (2002). "Frequency-to-electrode allocation and speech perception with cochlear implants," *J. Acoust. Soc. Am.* **111**, 1036–1044.

Moore, B. C. J. (2004). "Dead regions in the cochlea: Conceptual foundations, diagnosis, and clinical applications," *Ear Hear.* **25**, 98–116.

Moore, B. C. J., and Glasberg, B. R. (1997). "A model of loudness perception applied to cochlear hearing loss," *Aud. Neurosci.* **3**, 289–311.

Murray, N., and Byrne, D. (1986). "Performance of hearing-impaired and normal hearing listeners with various high frequency cut-offs in hearing aids," *Aust. J. Audiol.* **8**, 21–28.

Nearey, T. M. (1989). "Static, dynamic, and relational properties in vowel perception," *J. Acoust. Soc. Am.* **85**, 2088–2113.

Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.

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.

Shannon, R. V., Galvin, J. J., and Baskent, D. (2002). "Holes in hearing," *J. Assoc. Res. Otolaryngol.* **3**, 185–199.

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

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.

Turner, C. W., and Cummings, K. J. (1999). "Speech audibility for listeners with high-frequency hearing loss," *Am. J. Audiol.* **8**, 47–56.

Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (2002). "Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization," *Ear Hear.* **23**, 98–105.

Vickers, D. A., Moore, B. C. J., and Baer, T. (2001). "Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies," *J. Acoust. Soc. Am.* **110**, 1164–1175.