Speech, Hearing and Language: work in progress

Volume 13

The right information matters more than frequency-place alignment: simulations of cochlear implant processors with an electrode array insertion depth of 17 mm

Andrew FAULKNER, Stuart ROSEN and Clare NORMAN

Department of Phonetics and Linguistics
UNIVERSITY COLLEGE LONDON

The right information matters more than frequency-place alignment: simulations of cochlear implant processors with an electrode array insertion depth of 17 mm

Andrew FAULKNER, Stuart ROSEN and Clare NORMAN

Abstract

It has been claimed that speech recognition with a cochlear implant is dependent on the correct frequency alignment of analysis bands in the speech processor with characteristic frequencies (CFs) at electrode locations. However the cochlear position of the most apical electrode often has a CF of 1 kHz or more, and the use of filters aligned in frequency to relatively basal electrode arrays leads to significant loss of lower frequency speech information. This study simulates a cochlear implant array with 8 electrodes spaced 2-mm apart, inserted to a relatively shallow depth within the typical range, such that the most apical element is at a CF of 1851 Hz. Two noise-excited vocoder speech processors for this simulated electrode location have been compared, one with CF-matched filters, and one with filters matched to CFs at basilar membrane locations 6mm more basal than electrode locations.

An extended crossover training design examined pre- and post-training performance in the identification of vowels and words in sentences for both processors. The shifted processor led to higher post-training scores than the frequency-aligned processor for a male talker with both vowels and sentences. For a female talker, post-training vowel scores did not differ significantly between processors, while sentence scores were higher with the frequency-aligned processor. Training effects were significant only for the shifted processor. The effects of upward spectral shifting were significantly reduced with a few hours of experience. In the case of a shallow electrode insertion, it seems likely that speech processors should cover the most informative frequency range irrespective of electrode position and frequency misalignment.

1. Introduction

It has been claimed that speech recognition with a cochlear implant is adversely affected by a frequency mis-match of the analysis bands in the speech processor to the characteristic frequencies (CFs) at the implanted electrode locations (Dorman, Loizou, & Rainey, 1997; Shannon, Zeng, & Wygonski, 1998). Both of these studies employed simulations of cochlear implant speech processing in normally hearing listeners using vocoder-based processing in which the spectral envelope of speech was presented with an upward spectral shift. These simulations used a fixed speech processor and compared a relatively deep electrode insertion, for which CFs at electrode locations match the processor analysis filters (tonotopic mapping), to a shallower electrode insertion for which the CFs at electrode locations are higher in frequency than the processor analysis filters (upward shifted mapping). With upward shifts of 4 mm or more, speech scores were substantially reduced, for example sentence intelligibility was reduced from near 100% to 50% for a 4 mm shift (Dorman et al., 1997), and from 100% to virtually 0 with a 8mm shift (Shannon et al., 1998).

The notion that a tonotopic mapping is important for effective speech perception would, if substantiated, have important clinical implications. The consequences of a tonotopic mapping of the centre frequencies of speech processor analysis filters to CFs at electrode locations will inevitably depend on depth of array insertion. An in-

vivo CT study of electrode location in 19 patients implanted with the Nucleus 22-channel electrode showed that the position of the most apical electrode varied between 24 and 13.7 mm from the cochlear base, with a median of 20.3 mm (Ketten et al., 1998). All these electrode arrays were reported at surgery as fully inserted. The range of CFs at the most apical electrode in this patient group were estimated from Greenwood's (1990) formula as 400 to 2600 Hz, with a median of 1000 Hz (based on an average cochlear length of 33 mm that was derived from this same CT data). An acoustic simulation study of processors that are tonotopically mapped to an 8 element array with electrodes spaced 2mm apart and having the location of the most apical element at positions with CFs varying from 500 to 1851 Hz has shown significant deterioration of performance when lower frequency channels are lost (Faulkner, Rosen, & Stanton, 2000). When the most apical stimulation position simulated was 19 or 17 mm from the base of a 35 mm long cochlea (CFs of 1360 and 1850 Hz), identification of sentences, vowels, and consonants were all significantly poorer than for most apical locations 21, 23, and 25 mm from the base. This result is broadly consistent with predictions made from Articulation Index data (French & Steinberg, 1947; ANSI, 1998).

A further reason to doubt that processor filters should be matched to CFs at electrode locations is that the effect of such frequency mis-match has been shown in a simulation study to be markedly reduced with training. After less than three hours of experience of speech that is shifted upward to an extent corresponding to a 6.4 mm basalward basilar membrane shift, the performance of normally hearing listeners for such speech shows a substantial increase (Rosen, Faulkner, & Wilkinson, 1999). In that study, sentence intelligibility for upward shifted speech increased from virtually 0 to around 30% after experience. It seems very likely that cochlear implant users are also able to adapt to the clinical mapping of speech processor filters to their electrode locations given their extended experience. Harnberger et al, (2001) recently reported a study in experienced implant users that would be expected to reveal any lack of adaptation to upward spectral shift. In this study implant users who had had at least 12 months experience of implant use selected tokens from a set of synthesized vowel stimuli that best matched their expectation of a representative set of vowel qualities. An incomplete adaptation to spectral shifting would be expected to lead to choices of stimuli with lower 1^{st} and 2^{nd} formants than those of natural vowels. However, there was no evidence of such effects, suggesting that if these implant users were indeed subject to a basalward spectral shift, they had adapted to its effects.

Further evidence of adaptation in implant users comes from an acute study of modified analysis filter to electrode maps, in which a mapping that was familiar from extended use gave better speech performance than other mappings (Fu & Shannon, 1999a). This outcome, and comparable outcomes in similar acute studies of implant users (Fu & Shannon, 1999b; 1999c) suggest that the extended study of effects of frequency mapping in experienced implant users may be problematic because users are unlikely to tolerate an initial loss of benefit. However, it does appear that these effects can be investigated in simulation studies (Rosen, Faulkner, & Wilkinson, 1999).

The present study investigates upward spectral shifting as it might impact an individual cochlear implant user, that is, for a fixed electrode array insertion depth and alternative configurations of a speech processor. The aim is to compare after training the information loss that may arise from spectral shifting with the information loss

entailed by an unshifted mapping to a relatively shallow electrode insertion. In contrast, previous simulation studies have mostly investigated the effects of spectral shifting in conditions that simulate a fixed speech processor in conjunction with different electrode insertion depths (Dorman et al., 1997; Rosen et al., 1999; Shannon et al., 1998).

2. Experiment 1

The main study reported here used a crossover training design to compare simulations of tonotopically matched and upward shifted speech processors.

2.1 Method

2.1.1 Speech processing and equipment

Speech processing used eight-band noise-excited vocoders similar to those described by Shannon et al. (1995). Cross-over and centre frequencies for both the analysis and output filters were calculated using an equation (and its inverse) relating position on the basilar membrane to characteristic frequency, assuming a basilar membrane length of 35 mm (Greenwood, 1990):

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

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

The stages of processing in each band comprised an analysis filter, half-wave rectification, envelope smoothing with a 400 Hz low-pass filter, multiplication of a white noise by the envelope, and an output filter. Finally, the outputs of each band were summed together. Each channel of the processor received speech as input, without pre-emphasis.

The channel filter centre frequencies and –3 dB cut-off frequencies are shown in Table I. This series of centre frequencies represents cochlear locations each separated by a distance of 2 mm. Figure 1 represents the simulated electrode locations on cochlear position by CF coordinates. Two processing conditions were employed in training. Both simulated an electrode array having the most apical element located 16.9 mm from the cochlear base, through the use of output filters with centre frequencies between 1851 and 13783 Hz. The unshifted processor used analysis filters matching the output filters. This processing condition is termed *highpass* because of the loss of lower frequencies that it entails. The *shifted* processor used input filters with centre frequencies between 715 and 5923 Hz. For the shifted processor there is a mismatch between input and output filters equivalent to a 6 mm basalward shift along a 35 mm long cochlea. A third unshifted processor with input and output band cfs from 715 to 5923 Hz was also used in testing, but not for training, and is designated *normal*. The normal condition represents a tonotopically mapped speech processor for a simulated electrode with the most apical element located 22.9 mm from cochlear base.

Two implementations of the vocoder processing were employed. Training made use of real-time processing, while testing always employed off-line processing implemented in MATLAB.

Shift	High-pass	Centre frequency (Hz)	Cut-off (Hz)	Distance from base (mm)
Input band number				
			601	23.9
1		715	{	22.9
			845	21.9
2		995	{	20.9
			1167	19.9
3		1364	{	18.9
			1591	17.9
4	1	1851	{	16.9
			2150	15.9
5	2	2492	{	14.9
			2886	13.9
6	3	3338	{	12.9
			3857	11.9
7	4	4453	{	10.9
			5138	9.9
8	5	5923	{	8.9
			6826	7.9
	6	7861	{	6.9
			9050	5.9
	7	10416	{	4.9
			11983	3.9
	8	13783	{	2.9
			15850	1.9

Table I: *Centre and cut-off frequencies of input filters for shifted and high-pass processors. The output filters for both processors were identical to the input filters of the high-pass processor. The basilar membrane locations for a 35 mm long cochlea that match each centre and cut-off frequency are shown in the final column for reference.*

Off-line processing was executed at a 44.1 kHz sample rate. Prior to processing, all the recorded speech materials were band-limited to 11.05 kHz. Analysis filters in the off-line processing were Butterworth IIR designs with 3 orders per upper and lower side. The responses of adjacent filters crossed 3 dB down from the pass-band peak. Envelope smoothing used 2^{nd}-order low-pass Butterworth filters (400 Hz cut-off). A final low-pass filter was applied to the summed waveform from each of the eight bands at the upper cut-off frequency of the highest frequency channel (15.8 kHz) to limit the signal spectrum. This used a 6^{th}-order low-pass elliptical filter forwards and backwards to obtain the equivalent of a 12^{th}-order elliptical filter but with a zero phase shift.

Real-time processing ran at a 16 kHz sample rate on a DSP card (Loughborough Sound Images TMSC31), and was implemented using the Aladdin Interactive DSP Workbench (Hitech Development AB). To reduce the required computation, elliptical filter designs were used, with the same –3 dB crossover frequencies as those used for off-line processing. Analysis and output filters were 4^{th}-order band-pass designs, while the envelope smoothing filters were 3^{rd}-order low-pass. Because of the limited 8 kHz bandwidth, the uppermost three output bands could not be implemented in the real-time version of the shifted or highpass processors. Hence training only used the lower five bands of each processor. This limited the speech input bandwidth to be

from 601 to 2886 Hz (see table I) and would not be expected to have a major impact on performance in the connected discourse tracking task used in interactive training.

An equal-loudness correction was applied to each band of the shifted processor in both testing and training to make the loudness of the stimulation from each input band approximately the same as for the unshifted normal processor. An overall level correction was applied to the highpass processor to ensure that all processors led to similar SPLs for a given speech input.

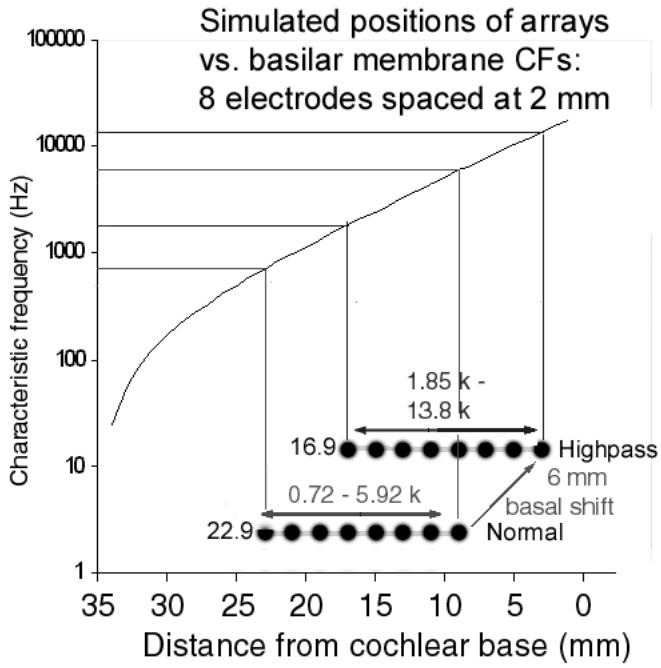

Figure 1: *Basilar membrane CFs against distance from cochlear base. The representations of the two electrode arrays represent the simulated positions of the electrodes for the highpass and normal processors. The CF range for each array is indicated. The speech processing filters for the shifted processor match the CFs of the electrode simulated in the normal processor, while the simulated electrode locations match the CFs of the electrode simulated in the highpass processor, i.e. there is a 6mm basalward shift.*

2.1.2 Stimuli

2.1.2.1 Vowel identification

17 b-vowel-d words from a male and female speaker of standard Southern British English were used, from digital anechoic recordings made at a 48 kHz sample rate. Presentation was computer-controlled. Each test run presented one token of each word from each of the two talkers, selected at random from a total set of six to ten tokens of each word from each talker. The vowel set contained ten monophthongs (in the words *bad, bard, bead, bed, bid, bird, bod, board, booed,* and *bud*) and seven diphthongs (in the words *bared, bayed, beard, bide, bode, boughed,* and *Boyd*). The spellings given here are those that appeared on the computer response buttons. During this test, subjects received visual feedback giving the identity of the stimulus after each response.

2.1.2.2 Sentence perception

Sentences produced by a further male and a further female talker with a Southern British accent were used. The female speech was from a 16 bit 48 kHz digital audio recording of the BKB sentences made simultaneously with an audio-visual recording (EPI Group, 1986; Foster *et al.*, 1993). The male speech was from an anechoic digital recording (16 bit, 44.1 kHz) of the IHR Adaptive Sentence Lists (MacLeod & Summerfield, 1990). Each test run used one list of sentences with 50 scored key words per list (45 scored words for the IHR sentences). No feedback was given in sentence testing.

2.1.3 Subjects

Eight adult native speakers of English took part. They were screened for normal hearing at 0.5, 1, 2 and 4 kHz, and were paid for their services.

2.1.4 Procedure

A cross-over training design was employed, with subjects trained and tested over two series of sessions with each of the shifted and highpass processing conditions Four subjects (group S-HP) were trained first with shifted processing followed by highpass processing. The order of the training conditions was reversed for the remaining four subjects (group HP-S). Table II displays the sequence of training and testing for group HP-S.

	Base-line 1	Highpass training				Base-line 2	Shifted training				Re-test
Session	1	2	3	4	5	6	7	8	9	10	11
CDT: min	-	H: 45	H: 35	H: 45	H: 35	-	S: 45	S: 35	S: 45	S: 35	-
Vowels	H: 2; S: 2	H: 2	H: 2	H: 2	H: 2	S: 2	S: 2	S: 2	S: 2	S: 2	H: 2
Sentences	H: 2m 2f; S: 2m 2f		H: 2m 2f		H: 2m 2f	S: 2m, 2f		S: 2m 2f		S: 2m 2f	H: 2m 2f

Table II: *Distribution of training and testing conditions over sessions for group HP-S. For group S-HP, sessions 2 to 5 and 11 used shifted processing, while sessions 6 to 10 used highpass processing. Processor conditions are indicated by H (highpass) and S (shifted). For CDT, the number of minutes of training in each session is shown. For vowel tests, the number of test lists per session is shown. For sentences, the number of lists in the session using the male (m) and female (f) talker is shown.*

The first session comprised familiarization and baseline testing. Subjects received first one test list of the vowel materials presented as unprocessed speech in order to familiarize them with the vowel task. This was followed by two vowel lists in each of the shifted and highpass conditions. Next, one list of sentences was presented, from the female talker, using the normal unshifted processor, again for the purpose of familiarization. Finally, two sentence lists from each of the male and female talkers were presented through both the shifted and highpass processors. The presentation order of the shifted and highpass processors was balanced across the 8 subjects within the vowel and sentence tests. For the condition trained first for each group, the vowel

and sentence scores from session 1 provided untrained performance baseline measures in that condition.

Sessions 2 to 5 comprised training and testing in the shifted condition (group S-HP) or the highpass condition (group HP-S). Vowel identification was tested at each session, while sentence tests were presented only in sessions 3 and 5. In session 6 subjects were retested on both vowel and sentence materials in the untrained condition. This established a baseline score for the second-trained condition measured one session prior to training in that condition. No training was included in session 6. Sessions 7 to 10 mirrored sessions 2 to 5, with the trained condition being reversed across groups S-HP and HP-S. The final 11th session contained no training, and comprised retests of vowel and sentence performance in the initially trained condition in order to assess the retention of any training effects over time.

Sentence (two lists per talker) and vowel testing (two lists) using the unshifted normal processor was also performed in sessions 6 and 11. These tests were included for two purposes. Firstly, to assess the effects of spectral shifting after training compared to a processor that delivered the same information to the tonotopically correct place. Secondly, to replicate a simulation of tonotopically-mapped processors for different electrode insertion depths (Faulkner, Rosen, & Stanton, 2000). The insertions simulated in that study were to basilar membrane CFs spanning 1851 to 13783 Hz (the highpass processor used here) compared to CFs spanning 715 to 5923 Hz (the normal processor used here) All testing and training took place in a sound-isolated room. The subject received diotic presentation of the processed speech stimuli over headphones (Sennheiser HD475 headphones for testing, AKG K240DF for training). Presentation levels were approximately 70 dBA.

2.1.5 Training procedure

Training was performed using connected discourse tracking (CDT: DeFilippo & Scott, 1978). Texts were chosen from the Heinemann Guided Readers series, elementary level. These texts, designed for learners of English as a second language, are controlled in syntactic complexity and vocabulary. The talker was author CN, a female speaker of standard southern British English. Talker and subject were in adjacent sound-isolated rooms, with a double-glazed communicating window that could be blinded. A constant masking noise at 45 dBA was present in the listener's room in order to mask any unprocessed speech transmitted through the intervening wall. The talker was able to hear the listener's responses over an intercom. The talker read from the text in phrases, and the listener repeated back what s/he had heard. If the listener's response was completely correct, the speaker moved on to the next phrase. Where any word or phrase was not correctly repeated after three presentations, the talker pressed a key to allow the listener to hear the word(s) as unprocessed speech. The first two 5-minute blocks of CDT training in each training session were auditory-visual. Subsequent 5-minute blocks (7 blocks in sessions 2, 4, 7 and 9; 5 blocks in sessions 3, 5, 8 and 10) used purely auditory presentation of processed speech.

2.2 Results

The main analyses of vowel and sentence scores were based on baseline scores collected immediately prior to training in each condition and scores after training at sessions 2 to 5 and 7 to 10. Hence, for a subject initially trained with the highpass

processor, the highpass baseline scores were from session 1, while the shifted baseline scores were those collected in session 6. The vowel and sentence data were analysed using repeated-measures ANOVA, with within-subjects factors of processing condition (shifted vs. highpass), talker, sessions of training, and the between-subject factor of training order. Hyunh-Feldt Epsilon corrections were applied to all F tests of factors with more than 1 degree of freedom.

2.2.1 Vowel identification

Vowel scores at baseline and over sessions of training are shown for the two talkers in figure 2. Figure 3 displays these scores separately for groups HP-S and S-HP. An ANOVA of the full data set showed main effects of talker and training, and a significant talker by processor interaction [F (1,6) =37.1, p =0.001, η^2 = 0.86, power = 1.00]. Hence, the primary analysis of this data was performed taking the male and female talkers separately.

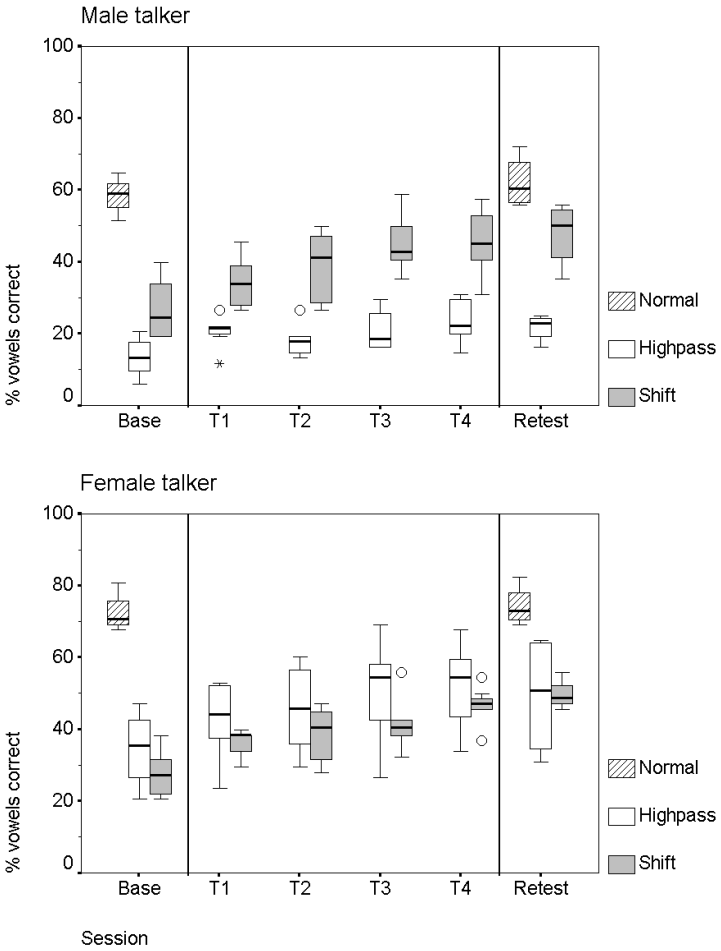

Figure 2: *Box and whisker plots showing vowel scores in highpass and shifted conditions at baseline, after 1, 2, 3 and 4 sessions of training (T1 to T4), and in the final retest session. Upper panel: male talker; lower panel: female talker. Performance with the normal processor in the baseline and final session is also shown.*

For the male talker, vowel identification was significantly more accurate in the shifted condition than in the highpass condition. [F (1,6) = 123, p <0.001, η^2 = 0.95, power = 1.00]. There was a significant effect of training [F (4,24) = 19.3, p < 0.001, η^2 = 0.76, power = 1.00]. Bonferroni-corrected paired comparisons (α = 0.05) showed that scores were higher in all post-training tests than at the first baseline, while scores after the fourth and final training period also exceeded those at the first two post-training

tests. There was a significant interaction between processor and training [F (4,24) = 5.20, p = 0.004, η^2 = 0.46, power = 0.93] that reflects the clear trend for a greater continuing increase in performance over training with the shifted processor. There was also an interaction between processor, training and training order [F (4,24) = 2.88, p = 0.044, η^2 = 0.33, power = 0.69]. This points to an effect of training for both groups in the shifted condition, but a training effect in the highpass condition only when this is the first condition trained. This appears to be largely due to baseline performance in the highpass condition, for group HP-S only, being lower than that after any degree of training (see figure 3)

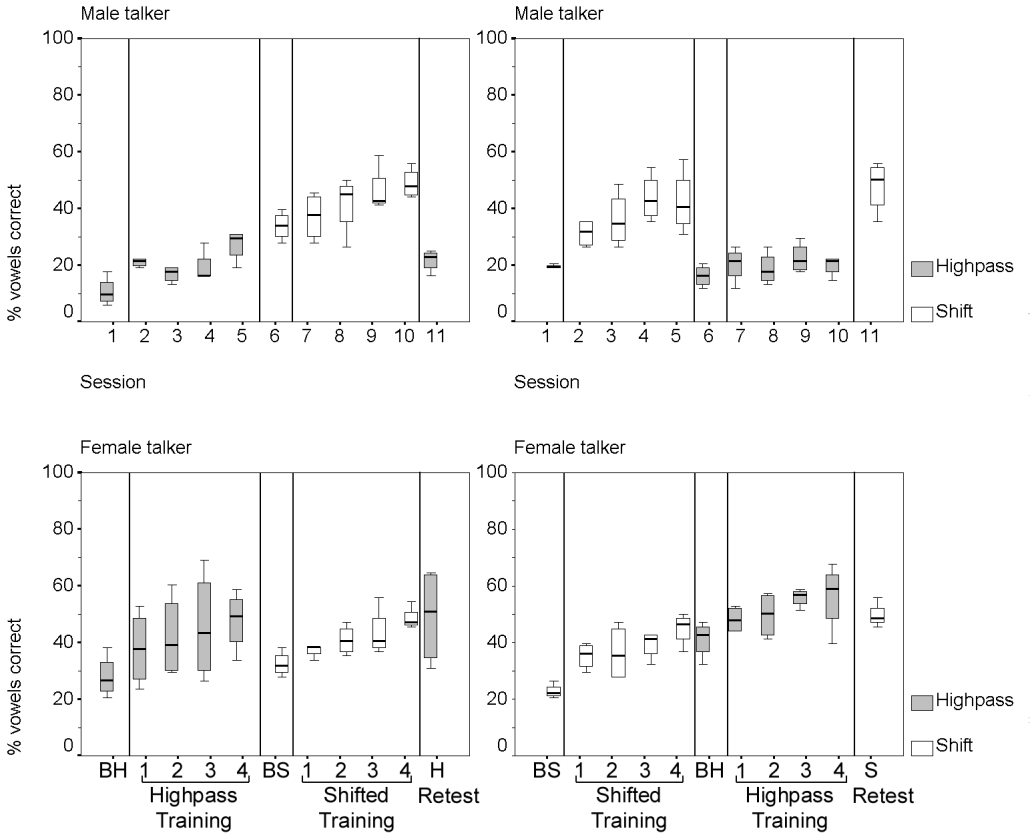

Figure 3: *Vowel scores in the shifted and highpass conditions over session by talker and training order. The left panels show scores from subjects trained first with the highpass condition; the right panels show scores from subjects trained first with the shifted condition. The ordinate is labelled with session number for the upper panels, while the lower panels indicate the training status at each session. Session 1 is the untrained baseline ("BS/BH"). Sessions 2 and 7 are after one session of training. Sessions 3 (8), 4(9) and 5(10) are after 2, 3 and 4 sessions of training respectively. Session 11 is a retest in the first trained condition.*

A one-way ANOVA was performed to compare post-training scores on male vowels with the highpass and shifted processors and also with the normal processor. Bonferroni-corrected paired comparisons confirmed that performance with the shifted processor exceeded that with the highpass processor, while performance with the normal processor was significantly exceeded that with each of the other two processors.

The pattern was different for the female talker. Here the effect of processor was not significant [F (1,6) = 4.01, p = 0.092, η^2 = 0.40, power = 0.39] although shifted scores tend to be lower than those in the highpass condition. The only significant effect was that of training [F (4,24) = 28.5, p < 0.001, η^2 = 0.83, power = 1.00]. As for the male talker, Bonferroni-corrected comparisons (α = 0.05) showed that scores at the first baseline session were significantly lower than at all post-training tests, and scores after the final fourth training session also significantly exceeded those after the first and second training sessions.

A one-way ANOVA was performed to compare female vowel scores at the final training session with the shifted and highpass processors and scores with the normal processor. Bonferroni-corrected paired comparisons confirmed that scores for the female talker did not differ between the shifted and highpass processors, while scores with the normal processor were higher than those from both the shifted and the highpass processors.

Talker and processor	Slope (SE) Units of % correct/sessions	R^2	df	F	p
Male, shifted	4.0 (1.3)	0.243	30	9.65	0.004**
Female, shifted	3.4 (0.92)	0.316	30	13.84	0.001**
Male, highpass	1.2 (0.80)	0.067	30	2.16	0.152
Female, highpass	3.2 (1.8)	0.092	30	3.05	0.091

Table III: *Linear regression of performance as a function of sessions of training*

2.2.2 Time course of training effects for vowels

A linear regression analysis of vowel scores against number of sessions of training (from 1 to 4, thus excluding the initial baseline data, and considering the data from the group as a whole rather from individual subjects) was performed for each talker in the shifted and highpass conditions (see table III and figure 4). In the shifted condition, there was a significant correlation of performance with amount of training, while correlations were not significant for either talker in the highpass condition. Logarithmic, logistic and exponential transformations of sessions of training yielded correlations that were virtually indistinguishable from those from linear regression. While noting that both highpass and shifted conditions after the first training session showed significant increases in performance from baseline levels in the ANOVA reported above (see section 2.2.1), regression analyses indicate that performance continues to increase with training only in the shifted condition. The highpass processor did lead to a non-significant trend of increasing performance with training for female vowels (see lower right panel of figure 4), but variability was high and trends were not consistent within subjects

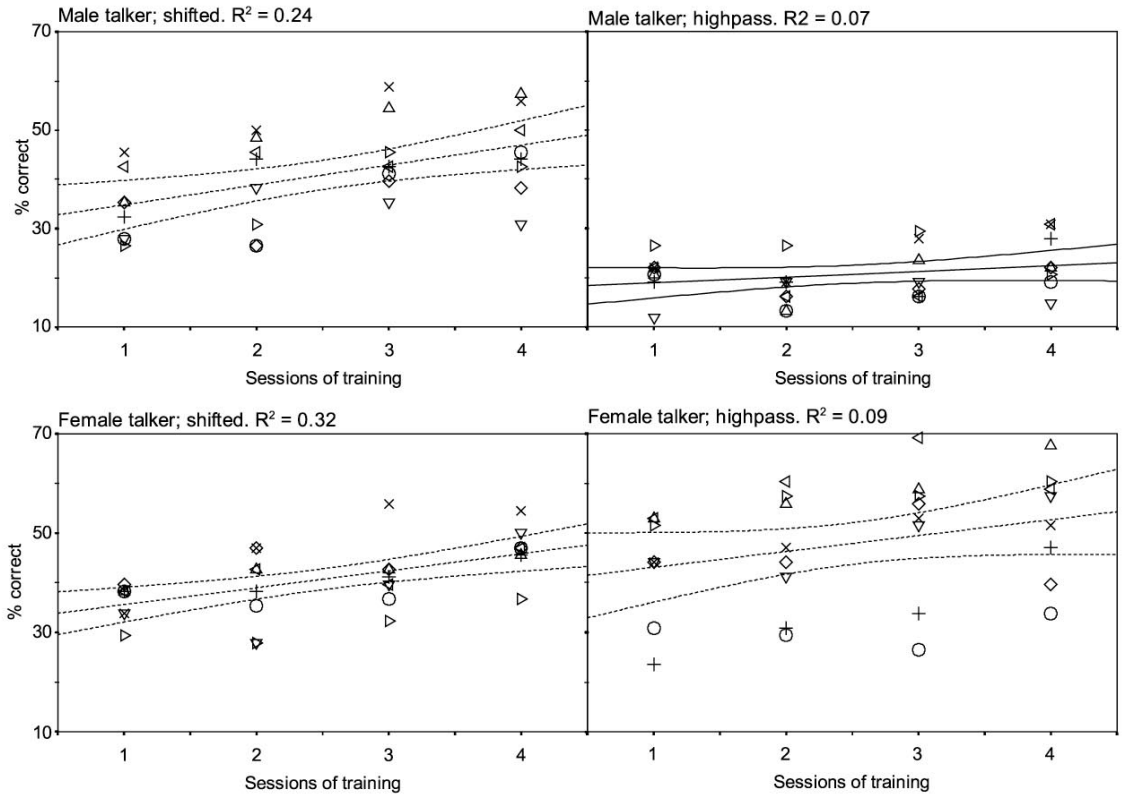

Figure 4: *Linear regression of vowel scores as a function of sessions of training. Symbols represent individual subjects. Lines represent a linear regression and 95% confidence limits.*

2.2.3 Sentence identification

Sentence performance at pre-training baseline, after 2 and 4 sessions of training, and in the final retest session is shown for each talker in figure 5. The same data is shown separately for groups HP-S and S-HP in figure 6. A similar analysis to that for the vowel data was performed, differing only in that sentence scores were not collected after the 2nd and 4th training sessions. Just as for vowels, the overall analysis showed significant effects of talker and of training, and a significant talker by processor interaction [$F(1,6) = 297$, $p < 0.001$, $\eta^2 = 0.98$, power = 1.00]. Hence, the sentence data were also analysed separately for each talker.

As for vowels, male sentence scores with the shifted processor were significantly higher than for the highpass processor [$F(1,6) = 166$, $p < 0.001$, $\eta^2 = 0.97$, power = 1.00]. There was a significant effect of training [$F(2,12) = 25.0$, $p < 0.001$, $\eta^2 = 0.81$, power = 1.00]. Bonferroni-corrected paired comparisons ($\alpha = 0.05$) showed that scores after 2 and 4 sessions of training were significantly higher than at the pre-training baseline, while scores after 4 sessions of training did not significantly exceed those after 2 sessions of training. In contrast to the male vowel data, here there was no significant interaction of training with processor. Even at baseline, performance with the shifted processor exceeded that with the highpass processor, and in later sessions, performance in the shifted condition approached ceiling levels.

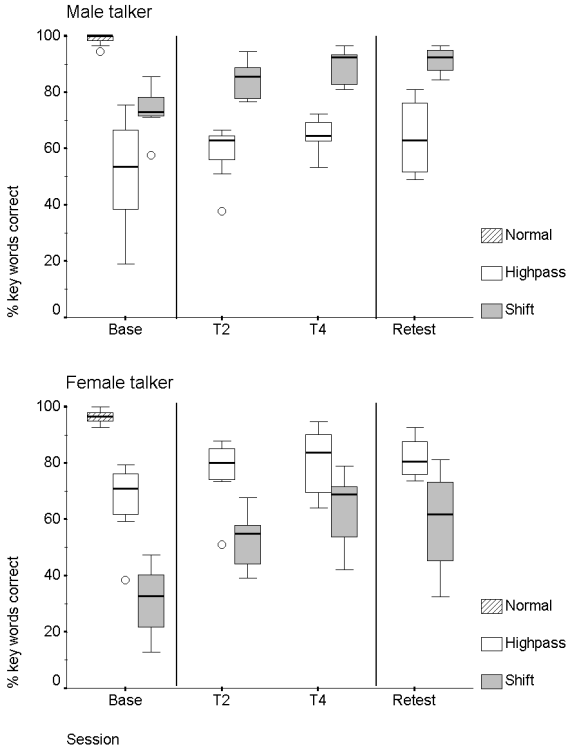

Figure 5*: Box and whisker plots showing sentence scores in the shifted and highpass conditions at baseline, after 2 and 4 sessions of training (T2, T4), and in the final retest session. Performance in the normal condition in the first and final session is also shown. Upper panel: male talker; lower panel: female talker.*

The order of training contributed to two interaction terms, these being training by training order [F (2,12) = 10.4, p = 0.002, η^2 = 0.63, power = 0.96] and processor by training by training order [F (2,12) = 17.4, p <0.001, η^2 = 0.74, power = 1.00]. These interactions relate, as in the male talker vowel data, to the presence of a training effect in the highpass condition only when this was the condition trained first (group HP-S: see figure 6).

A one-way ANOVA compared sentence scores for the male talker at the final training session with the shifted and highpass processor, and scores with the normal processor. Bonferroni-corrected paired comparisons confirmed that the shifted processor gave significantly higher trained performance than the highpass processor. Scores with the normal processor significantly exceed those with both of the other processors.

In contrast to the male talker, the highpass processor led to higher scores for the female talker than did the shifted processor [F (1,6) = 105, p < 0.001, η^2 = 0.97, power = 1.00]. Again there was a significant training effect [F (2,12) = 31.7, p < 0.001, η^2 = 0.84, power = 1.00]. As for male speech, Bonferroni-corrected comparisons showed that both post-training sessions gave higher scores than at baseline, while scores after 4 sessions of training were not higher than those after two sessions of training. For the female talker there was a significant processor by training interaction [F (2,12) = 5.82, p = 0.017, η^2 = 0.49, power = 0.77]. This interaction indicates a greater improvement in performance over training in the shifted condition than in the highpass condition.

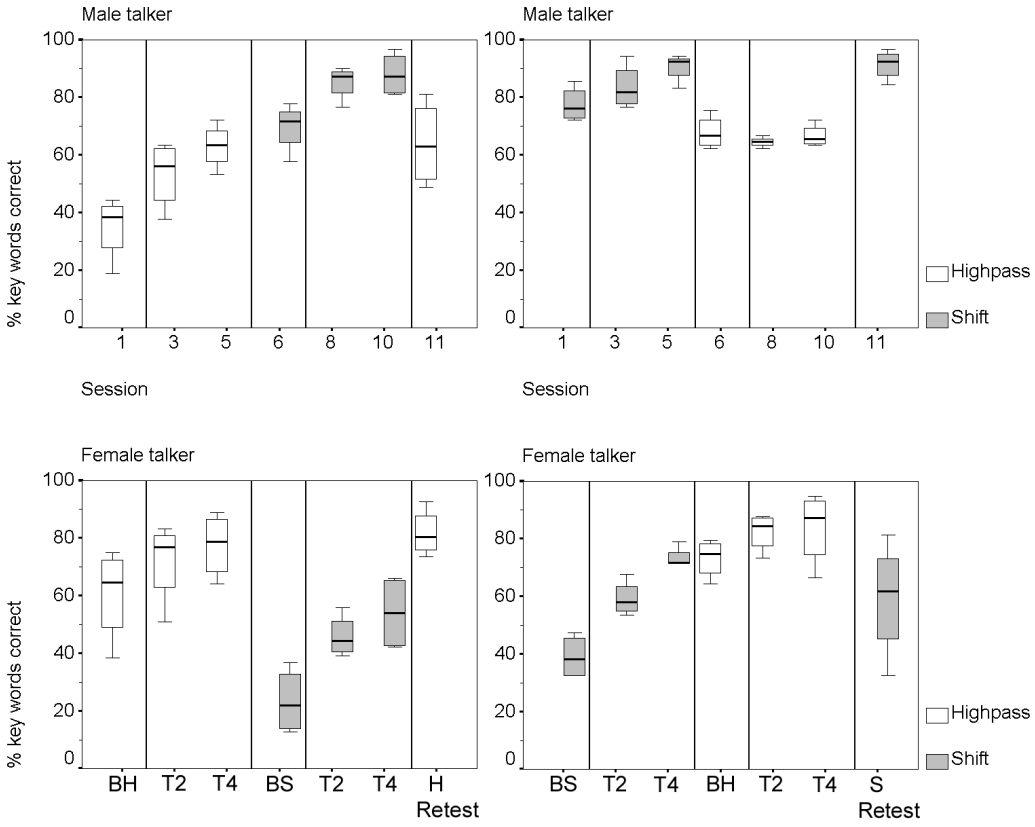

Figure 6: *Sentence scores over session by talker and training order. The left panels show scores from subjects trained first with the highpass condition; the right panels show scores from subjects trained first with the shifted condition. The ordinate is labelled with session number for the upper panels, while the lower panels indicate the training status at each session. Sessions 1 and 6 are the untrained baselines ("BS" for shifted baseline, "BH" for highpass baseline). Sessions 3 and 8 are after two sessions of training ("T2"). Sessions 5 and 10 are after 4 sessions of training ("T4"). Session 11 is a retest in the first trained condition.*

A one-way ANOVA compared sentence scores for the female talker at the final training session with the shifted and highpass processor and scores with the normal processor. Bonferroni-corrected paired comparisons showed that the highpass processor gave significantly higher trained performance than the shifted processor, although the interaction of processor and training in the previous ANOVA indicates that this difference is diminished compared to earlier in training. The scores with the normal processor significantly exceed those with both of the other processors.

CDT rates over training sessions for auditory-visual and auditory presentation modes are shown in figure 7. As would be expected in CDT, performance increased over sessions [$F(2.13, 12.8) = 58.5$, $p < 0.001$, $\eta^2 = 0.91$, power = 1.0]. This in part can be attributed to increasing experience of the talker and familiarity with the training text. There was also an expected main effect of presentation mode [$F(1,6) = 39.3$, $p = 0.001$, $\eta^2 = 0.87$, power = 1.0], with auditory-visual rates being generally higher, and often close to ceiling levels. A processor by mode interaction [$F(1,6) = 53.9$, $p < 0.001$, $\eta^2 = 0.90$, power = 1.0] indicates a greater increase from auditory to auditory-visual tracking rates with the shifted processor than for the highpass one.

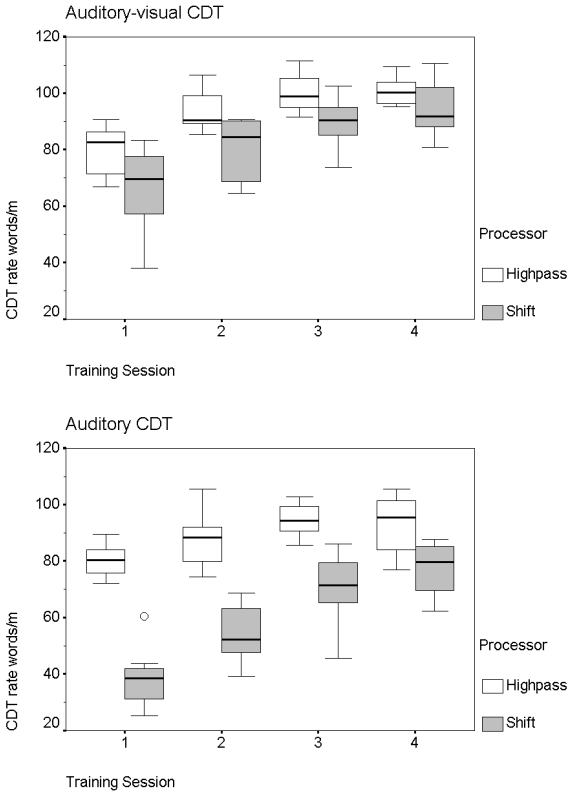

Figure 7: *CDT rates in training with auditory-visual (upper panel) and auditory (lower panel) presentation modes.*

2.2.4 Connected Discourse Tracking

Since auditory performance is of primary interest here, a second ANOVA was performed on auditory CDT rates. Here there was again a main effect of processor, with the highpass processor showing higher CDT rates overall with this female talker [$F(1,6) = 44.2$, $p = 0.001$, $\eta^2 = 0.88$, power = 1.0]. The effect of training was significant [$F(2.5, 14.8) = 67.6$, $p < 0.001$, $\eta^2 = 0.92$, power = 1.0}. There was also a processor by training interaction [$F(2.87, 17.2) = 8.776$, $p = 0.001$, $\eta^2 = 0.59$, power = 0.98], indicating a greater effect of training for the shifted processor. As is evident in the lower panel of figure 7, auditory tracking rates with the shifted processor in the last training session approached those with the highpass processor, although they were still significantly lower [$F(1,6) = 15.3$, 8, $p=0.008$, $\eta^2 = 0.72$]. There was no significant effect on CDT rates of the order in which the conditions were trained, or any significant interactions with this factor.

2.2.5 Retention of training over time

The extent to which subjects retained the effect of training in the shifted condition when this was the first trained condition can be assessed by comparing their performance after the 4th session of shifted training (session 5) with their performance in the shifted condition at session 11, after they have spent 4 sessions in training and testing with the highpass condition. These data are included in figure 3 for vowel identification and in figure 6 for the identification of words in sentences. This comparison was tested by repeated measures ANOVA, with factors of talker and test session (5 or 11). Neither for vowel nor for sentence materials was there a significant difference between scores at sessions 5 and 11 [vowels: $F(1,3) = 2.94$, $p = 0.19$,

power = 0.23; sentences: F (1,3) = 1.01, p = 0.39, power = 0.11]. Hence the effects of training accrued by session 5 appear to have been retained over the period of several days in which subjects had no exposure to the shifted condition. The power measures indicate a relatively low probability of missing a significant change in performance over time, despite the modest number of subjects (4) in this analysis.

2.2.6 Specificity of learning

If subjects are learning something that is not specific to the shifted condition during the its training, then we might expect that performance in the highpass condition may be further improved by training in the shifted condition. The specificity of learning in the shifted condition can be assessed when this was the second-trained condition by comparing performance at the end of the 4th session of highpass training (session 5) with performance in the highpass condition at session 11, after 4 sessions in training and testing with the shifted condition. These data too are displayed figure 3 (vowels) and figure 6 (sentences). A repeated measures ANOVA was again applied, with factors of talker and test session (5 or 11). Neither for vowel nor for sentence materials was there a significant difference between scores at sessions 5 and 11 [vowels: F (1,3) = 0.226, p = 0.67, power = 0.06; sentences: F (1,3) = 1.29, p = 0.34, power = 0.13]. Given the low power values seen here, a rather strong inference can be made that training in the shifted condition does not contribute to performance in the highpass condition; rather, the training in the shifted condition is specific to spectral shifting

3. Experiment 2

In order to further investigate the time-course of adaptation to an upward-shifted speech processor, a supplementary study was performed in which a single subject, who did not participate in experiment 1, was provided with extended training. Apart from the procedure and subjects, methods were identical to those used in experiment 1.

3.1 Procedure

The subject, an adult female, completed 11 sessions of training in the shifted condition. As in experiment 1, training in each session commenced with 10 minutes of auditory-visual CDT, followed by 45 minutes of auditory CDT in odd-numbered sessions and 35 minutes of auditory CDT in even numbered sessions. Vowel identification was tested after each training period (two lists of 68 words from the male and female talker). Sentence tests (two male talker and two female talker lists) were administered after training in each even numbered session. Testing was performed only in the shifted condition.

3.2 Results

Performance over training is shown in figure 7. Both vowel and sentences scores show continuing improvement for both male and female talkers, although the more sparsely sampled sentence data are less consistently increasing. These data were subjected to ANCOVA, with sessions of training as the (linear) covariate, and talker as a fixed factor. Both sentence scores [F (1,6) = 13.6, p = 0.01, η^2 = 0.70, power = 0.865] and vowel scores [F (1,18) = 63.8, p<0.001, η^2 = 0.78, power = 1.00] showed significant effects of training. Talker was also a significant factor for sentence scores

[F (1,6) = 19.8, p = 0.004, η^2 = 0.767, power = 0.96], with the male speaker giving substantially higher scores as in experiment 1. For vowels, talker had no significant effect [F (1,18) = 0.91, p = 0.354, η^2 = 0.048, power = 0.15]. In neither data set was there a significant interaction of talker with training.

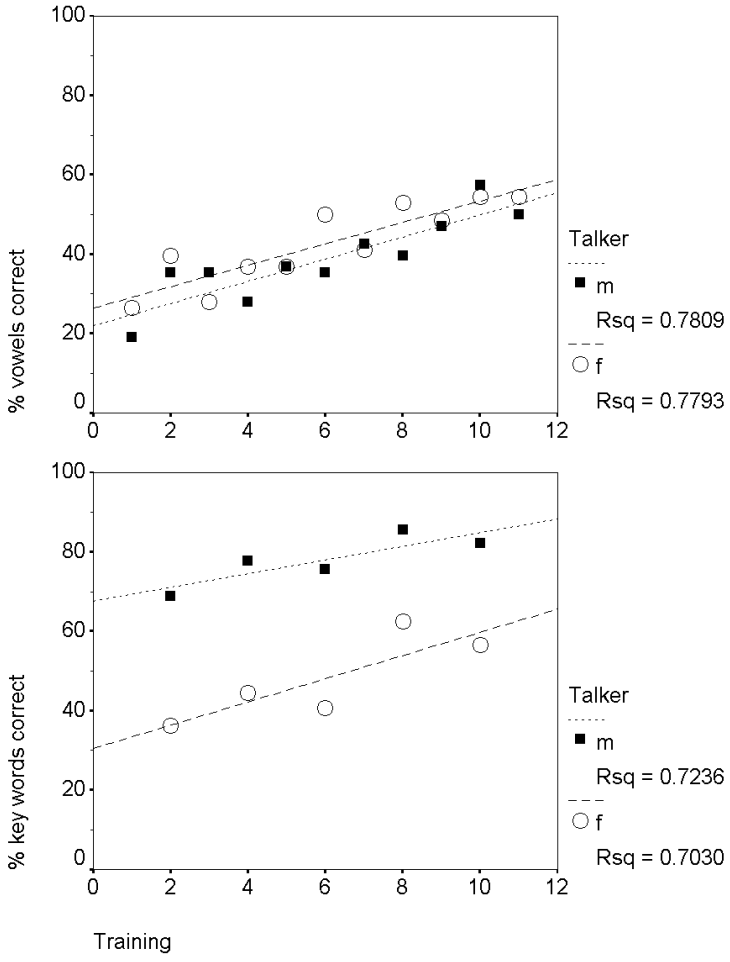

Figure 7: *Vowel identification performance (upper panel) and key words correct in sentences (lower panel) for the shifted processor as a function of number of training sessions in experiment 2. Solid symbols are for a male talker, unfilled symbols for a female talker. The lines show linear regression fits, for which the corresponding R^2 is shown in the legend.*

4. Discussion

Each of the measures made in this study confirm our earlier finding that normal listeners can learn to adapt to speech that is spectrally shifted upwards (Rosen, Faulkner, & Wilkinson, 1999). Furthermore, we find that for vocoder processed male speech, the identification of vowels and words in sentences is more accurate with an upward spectral shift corresponding to a 6 mm basalward basilar membrane shift than for a unshifted condition representing the same simulated electrode locations as the shifted condition, but with a tonotopically-matched processor that represents only speech information from a 1.6 kHz upwards. For female speech, we find no difference in vowel identification between this shifted processor and the highpass tonotopically-matched processor, while in CDT and the identification of words in sentences, an initial disadvantage with female speech for the shifted processor is significantly reduced with experience. Experiment 2 illustrates that more extended training than is given in experiment 1 appears to lead to a continuing improvement in performance for

both male and female speech, and we are likely, therefore, to be underestimating the ultimate degree of adaptation to upward-shifted speech spectra.

The 6 mm basalward basilar membrane shift simulated in the present study was similar to the 6.5 mm shift simulated in our earlier study (Rosen et al., 1999), and the degree of adaptation over a few hours of training was also similar in both cases. Unlike that earlier study, here training was also given in the unshifted highpass condition. Except that performance can increase between the first and second day of testing, there is no evidence here of adaptation to unshifted vocoder processed speech. A continuing adaptation over several hours of training was found only in the shifted condition, indicating that this adaptation appears to be specific to spectral shifting rather than a more general adaptation to noise-vocoded speech. As in this earlier study, performance with upward shifted speech does not reach the same levels as seen with the same information presented at the tonotopically-correct place (the normal processor in this study). However there is no reason to believe that further training would lead to an equivalence of performance with the same information shifted as with unshifted presentation.

A 6 mm basilar membrane shift is equivalent to a 1.37 octave shift for the lowest processor band used here (715 Hz centre frequency) and a 1.22 octave shift for the highest processor band (5923 Hz centre frequency). Such shifts are large by comparison with the normal range of formant frequencies, which scale approximately with vocal tract length. This is typically 17 cm in an adult male, 14 cm in an adult female, and 11 cm in a child 5 years of age, so that 5 year old children's formant frequencies are approximately 0.6 octaves higher than those of an adult male. We have found that the intelligibility of male speech is less affected by upward spectral shifting than is female speech. The same outcome (in the absence of training) has been reported previously for vowel identification (Fu & Shannon, 1999a). Fu & Shannon also reported that listeners tolerate larger downward spectral shifts of child and adult female speech than for male speech. The extent to which spectral shifts can be tolerated without training seems likely to depend upon the extent to which formant frequencies lie within the range of formants shown by human speech across talker sex and age.

4.1 Implications for cochlear implant processor fitting

For a cochlear implant patient with an electrode array whose most apical element is located 17 mm from the base of a 35mm long cochlea, the loss of lower frequency speech information that results from a tonotopically-matched speech processor with the lowest analysis filter band centred around 1850 Hz is significant, as we have shown previously (Faulkner, Rosen, & Stanton, 2000). The present study extends this conclusion to performance after several hours of training. In contrast, an upward shifted mapping to such an electrode position gives an implant user access to important speech information carried by frequencies below 1850 Hz. If implant users are able to adapt to such shifts, and evidence is accumulating to suggest that they are (Harnberger et al., 2001; McKay & Henshall, in process), then it would be preferable to deliver the most informative frequency range without regard to electrode position rather than to use a tonotopically matched mapping.

Acknowledgements
Clare Norman was supported by a Wellcome Trust Vacation Scholarship (ref. VS98/UCL/001/CH/TG/JG). We also gratefully acknowledge the support of the Royal National Institute for Deaf People (UK)

References

ANSI (1998). Draft ANSI standard for calculation of the Speech Intelligibility Index (SII) formerly the Articulation Index (AI) Accredited Standards Committee S3, Bioacoustics.

DeFilippo, C. L. & Scott, B. L. (1978). A method for training and evaluation of the reception of on-going speech. Journal of the Acoustical Society of America, 63, 1186-1192.

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

Faulkner, A., Rosen, S., & Stanton, D. (2000). Simulation of the effects of cochlear implant electrode insertion depth for tonotopically-mapped speech processors. Speech, Hearing and Language, work in progress, University College London, Department of Phonetics and Linguistics. (in process, Journal of the Acoustical Society of America)

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

Fu, Q. J. & Shannon, R. V. (1999c). Effects of electrode configuration and frequency allocation on vowel recognition with the nucleus-22 cochlear implant. Ear and Hearing, 20, 332-344.

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

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

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

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

Ketten, D. R., Vannier, M. W., Skinner, M. W., Gates, G. A., Wang, G., & Neely, J. G. (1998). In vivo measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays. Annals of Otology, Rhinology and Laryngology, 107, S175, 1-16.

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

McKay, C. M. & Henshall, K. R. Frequency-to-electrode allocation and speech perception with cochlear implants. Journal of the Acoustical Society of America (in process).

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

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

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