The right information may matter more than frequency-place alignment: simulations of frequency-aligned and upward shifting cochlear implant processors for a shallow electrode array insertion

Andrew FAULKNER, Stuart ROSEN AND Clare NORMAN

Department of Phonetics and Linguistics, University College London.

Wolfson House, 4 Stephenson Way, London NW1 2HE, UK.

Manuscript for Ear and Hearing #MM030: Revision 2 – dated 1 November 2005.

Corresponding author: Andrew Faulkner.

tel +44(0)2076797408, fax +44(0)2073830752, email andyf@phon.ucl.ac.uk

Running head: Simulations of frequency-aligned and shifted CI processors

Abstract

Objective

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 use of filters aligned in frequency to a relatively basal electrode array position leads to significant loss of lower frequency speech information. This study uses an acoustic simulation to compare two approaches to the matching of speech processor filters to a electrode array having 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 are compared, one with CF-matched filters, and one with filters matched to CFs at basilar membrane locations 6mm more apical than electrode locations.

Design

An extended crossover training design examined pre- and post-training performance in the identification of vowels and words in sentences for both processors. Subjects received about three hours of training with each processor in turn.

Results

Training improved performance with both processors, but training effects were greater for the shifted processor. For a male talker, the shifted processor led to higher post-training scores than the frequency-aligned processor 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.

Conclusions

Even for a shallow electrode insertion, we conclude that a speech processor should represent information from important frequency regions below 1 kHz and that the possible cost of frequency misalignment can be significantly reduced with listening experience.

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. The simulations represented a fixed speech processor used with either a relatively deep electrode insertion, for which CFs at electrode locations matched the processor analysis filters (tonotopic mapping), or with a shallower electrode insertion for which the CFs at electrode locations were 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 frequency shift approximating a 4 mm mismatch (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. This study was designed to compare tonotopic and upward shifted mappings to an electrode array that is not inserted to an ideal depth.

The consequences of a tonotopic mapping of the centre frequencies of speech processor analysis filters to CFs at electrode locations will depend on the 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 of 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 derived from this same CT data). A second study of a further 13 Nucleus-22 patients using similar methods showed a range of apical electrode depths from 11.9 to 25.9 mm, with a similar median insertion depth of 19.1 mm (Skinner et al., 2002). An acoustic simulation study of processors that are tonotopically mapped to an 8-element 2 mm-spaced electrode array with 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, 2003). When the most apical simulated electrode position was 19 or 17 mm from the base of a 35 mm long cochlea (CFs of 1360 or 1850 Hz), identification of sentences, vowels, and consonants were all significantly poorer than for most apical locations at 21, 23, and 25 mm from the base. This result is broadly consistent with predictions of the effects of the loss of lower frequency information according to the Articulation Index (ANSI, 1997; French & Steinberg, 1947; Fletcher & Steinberg, 1930). Such a loss of low frequency information may make the matching of filters to electrode position CFs undesirable.

A further reason to doubt that processor filters should be matched to CFs at electrode locations is that the effect of such a frequency mis-match has been shown in a simulation study to be markedly reduced with training. After less than three hours of experience with speech that is shifted upward to an extent corresponding to a 6.5 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 speech processed through an upward-shifted four-band noise-excited vocoder increased from virtually 0 to around 30% after experience. It seems 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. Harnsberger et al. (2001) recently reported a study in experienced implant users that should have been sensitive to any lack of adaptation to upward spectral shift. In this study implant users, all with 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 1st and 2nd formants than those of natural vowels. However, there was no evidence of such effects, even for those users with shallower electrode placements, suggesting that these implant users had adapted to any effects of a basalward spectral shift. A similar question was asked in a recent study of the vowel confusions made by 19 Finnish implant users at between 6 and 24 months after implantation (Valimaa, Maatta, Lopponen, & Sorri, 2002a). Once again, vowel judgments were consistent with the hypothesis that users had largely adapted to any basalward shift.

More specific evidence of adaptation to shifts in place of stimulation in implant users comes from a recent study of three subjects who were given 3 months experience with a speech processor whose analysis filter centre frequencies were all lowered by between 0.68 and 1 octave (Fu, Shannon, & Galvin, 2002). Both consonant and vowel identification were markedly reduced initially. After two weeks, all three subjects showed increases in performance that approximately halved the initial decrement. However, there was little further improvement in performance after three-months, when scores remained significantly below the levels observed with the original speech processor filters. Hence, implant users can indeed learn to make better use of spectrally shifted speech cues, but on the evidence of this study, this adaptability may be limited.

Perhaps the most direct evidence of perceptual re-learning of cochlear place to frequency mapping in implant users comes from a longitudinal study of perceptual vowel spaces (Svirsky, Silveira, Neuburger, Teoh, & Suarez, 2004). Using a task similar to that employed by Harnsberger et al. (2001), this study observed the development of the perceptual vowel spaces of four post-lingually deafened adult implant users over a two year period. While vowel spaces were very different from those of normal hearing listeners at switch-on, all four subjects showed near-normal vowel spaces after periods of time that varied from one day to 24 months.

Previous simulation studies of the effects of spectral shifting have mostly investigated the conditions that simulate a fixed speech processor in conjunction with different electrode insertion depths (Dorman et al., 1997; Rosen et al., 1999). In contrast, 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. Hence, the aim is to compare, after a period of training, the cost to speech perception of spectral shifting to the cost of lower frequency information loss entailed by an unshifted mapping to a relatively shallow electrode insertion.

2. Experiment 1

2.1 Method

2.1.1 Speech processing and equipment

Speech processing used eight-band noise-excited vocoders similar to those described by Shannon, Zeng, Kamath, Wygonski, & Ekelid (1995). The choice of 8 channels for the simulation was motivated by indications that most adult implant users have the use of only 6-8 effective spatial channels, even if their implant has as many as 24 electrodes (Fishman, Shannon, & Slattery, 1997; Friesen, Shannon, Baskent, & Wang, 2001). The spatial extent of the cochlear stimulation from each channel in the normal listeners should, therefore, be similar to that expected in a cochlear implant user.

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: Note 1):

$$frequency = 165.4\left(10^{0.06x} - 1\right), \text{ and the inverse, } x = \frac{1}{0.06}\log\left(\frac{frequency}{165.4} + 1\right) \qquad \text{[Equation 1]}$$

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.

Insert Table 1 about here

Figure 1 about here

The channel filter centre frequencies and –3 dB cut-off frequencies are shown in Table 1. This series of centre frequencies represents cochlear locations each separated by 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 *high-matched* 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 center frequencies from 715 to 5923 Hz was also used in testing, but not for training, and is designated *low-matched*. The low-matched condition represented a tonotopically mapped speech processor for a simulated electrode with the most apical element located 22.9 mm from cochlear base. In this condition, the input information is the same as that for the shifted processor: its inclusion allows a pure measure of the effect of shifting.

Two implementations of the vocoder processing were employed. Training made use of live speech and hence required real-time processing, while testing always employed off-line processing implemented in MATLAB which ensured that stimuli could be identically repeated and also allowed the use of a higher audio bandwidth.

Off-line processing was executed at a 44.1 kHz sample rate. Prior to processing, all the recorded speech materials were band-limited to 11.025 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 2nd-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 was a 6th-order low-pass elliptical filter applied forwards and backwards to obtain the equivalent of a 12th-order elliptical filter but with a zero phase shift.

Real-time processing ran on a DSP card (Loughborough Sound Images TMSC31), and was implemented using the Aladdin Interactive DSP Workbench (Hitech Development AB). Because the computational power of the DSP was limited, the sampling rate was restricted to 16 kHz, and elliptical rather than Butterworth filter designs were used. Analysis and output filters were 4th-order band-pass designs, with the same –3dB cut-offs as those used in off-line processing, while the envelope smoothing filters were 3rd-order low-pass. Because of the limited 8 kHz bandwidth of the real-time processing, the uppermost three output bands could not be implemented in the real-time version of the shifted or high-matched processors. Hence training only used the lower five bands of each processor. This limited the speech input bandwidth to between 601 and 2886 Hz (see Table 1). Although there is significant speech information above 2.9 kHz, the loss of this would not be expected to have a major impact on performance in the connected discourse tracking task used in interactive training [Note 1]. Nor would it be expected to affect any learning of vowel identity cues in upward-shifted speech, because this bandwidth limit did not eliminate information in the vowel F1 and F2 frequency regions.

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 low-matched processor and hence to preserve relative loudness across the spectra of unshifted and shifted speech. The equal-loudness correction was half of the difference (in dB) between the minimal audible field threshold (MAF) at the centre frequency of the analysis filter and the MAF at the centre frequency of the shifted output filter [Note 2]. An overall level correction was applied to the high-matched processor to ensure that all processors led to similar SPLs for a given speech input.

2.1.2 Training method

Training was performed using connected discourse tracking (CDT: DeFilippo & Scott, 1978). This technique engages the listener in a communicative task, and been found effective as a training method for spectrally shifted speech (Rosen et al., 1999). The task produces a measure of communication rate, expressed in words/minute, indicating the number of words correctly repeated by the listener. The female talker was author CN, a speaker of standard southern British English. CN's speech was not used for any of the testing. 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 repeated it a final time after pressing a key that allowed the listener to hear unprocessed speech. After 5 minutes indicated by an electronic timer, the talker marked the last word correctly repeated so that the words repeated in each 5-minute block could be counted.

CDT 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. 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 to mask any unprocessed speech transmitted through the intervening wall. The talker was able to hear the listener's responses over an intercom.

2.1.3 Test Materials

2.1.3.1 Vowel identification

Seventeen 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 and subsequently resampled to 44.1 kHz. 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. There were, therefore, 34 tokens in each test list presentation. The vowel set contained ten monophthongs, /æ/ (bad); /ɑː/ (bard); /i:/ (bead); /ɛ/ (bed); /ɪ/ (bid); /ɜː/ (bird); /ɒ/ (bod); /ɔː/ (board); /u:/ (booed); /ʌ/ (bud)) and seven diphthongs, /eə/ (bared); /eɪ/ (bayed); /ɪə/ (beard); /aɪ/ (bide); /əʊ/ (bode); /aʊ/ (boughed); /ɔɪ/ (Boyd). Subjects made their response by selecting one of 17 response buttons with the mouse. The spellings given here are those that appeared on the buttons. During the vowel identification test, subjects received visual feedback giving the identity of the stimulus after each response. This feedback was expected to reinforce the effects of the training provided by connected discourse tracking.

2.1.3.2 Sentence perception

Sentences were produced by additional male and female speakers of standard Southern British English. 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, Reference Note 1; 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). These two sentence lists are very similarly constructed and of similar difficulty. Each test run used one list of sentences (16 sentences with 50 scored key words per list for the BKB sentences and 15 sentences with 45 scored words for the IHR sentences). No feedback was given in sentence testing.

2.1.4 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.

2.1.5 Procedure

A cross-over training design was employed, with subjects trained and tested over two series of sessions with each of the shifted and high-matched processing conditions. Subjects completed one or two sessions on any one day. The complete set of 11 sessions for a subject took place over a period of two to three weeks. Four subjects (group S-HM) were trained first with shifted processing followed by high-matched processing. The order of the training conditions was reversed for the remaining four subjects (group HM-S). Table 2 displays the sequence of training and testing for group HM-S.

Insert Table 2 about here

The first session comprised familiarization and baseline testing. In this session the presentation order of tests using shifted and the high-matched processors was balanced across the eight subjects within the blocks of vowel and sentence tests. For the first training condition 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-HM) or the high-matched condition (group HM-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-HM and HM-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 low-matched 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 et al., 2003). The insertions simulated in that study were to basilar membrane CFs spanning 1851 to 13783 Hz (as the high-matched processor used here) compared to CFs spanning 715 to 5923 Hz (as the low-matched processor used here).

The first two 5-minute blocks of CDT training in each training session were auditory-visual, ensuring that the listener was able to understand most of the training text. 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, forcing the listener to extract as much as possible from the auditory input.

All testing and training took place in a sound-isolated room. The subjects received diotic presentation of the processed speech stimuli over headphones (Sennheiser HD475 for testing, AKG K240DF for training). Presentation levels were approximately 70 dBA.

2.2 Results

2.2.1 Standard features of analyses

The CDT, vowel and sentence data were analysed using repeated-measures ANOVA, with within-subject factors of processing condition (shifted vs. high-matched), talker, session number, 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. The main analyses of vowel and sentence scores used pre-training baseline scores collected immediately prior to training in each condition and trained scores as measured after training at sessions 2 to 5 and 7 to 10. Hence, for a subject initially trained with the high-matched processor, the high-matched baseline scores were from session 1, while the shifted baseline scores were those collected in session 6.

Insert Figure 2 about here

2.2.2 Connected Discourse Tracking

CDT rates over training sessions for auditory-visual and auditory presentation modes are shown in Figure 2. As would be expected in CDT, performance increased over sessions [$F(2.13, 12.8) = 58.5$, $p < 0.001$, $\eta^2 = 0.91$]. 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 [$F(1,6) = 39.3$, $p = 0.001$, $\eta^2 = 0.87$], with auditory-visual rates being generally higher, and often close to ceiling levels. A processor by mode interaction [$F(1,6) = 53.9$, $p < 0.001$, $\eta^2 = 0.90$] indicates a greater benefit from visual cues with the shifted processor than with the high-matched one.

Since auditory performance is of primary interest here, a second ANOVA was performed on auditory CDT rates. There was again a main effect of processor, with the high-matched processor showing higher CDT rates overall with this female talker [$F(1,6) = 44.2$, $p = 0.001$, $\eta^2 = 0.88$]. The effect of training was significant [F (2.5, 14.8) = 67.6, $p < 0.001$, $\eta^2 = 0.92$]. There was also a processor by training interaction [F (2.87, 17.2) = 8.776, $p = 0.001$, $\eta^2 = 0.59$], consistent with a greater effect of training for the shifted processor. As is evident in the lower panel of Figure 2, auditory tracking rates with the shifted processor in the last training session approached those with the high-matched 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, nor any significant interaction with this factor.

Insert Figure 3 and Figure 4 about here

2.2.3 Vowel identification

Vowel scores at baseline and over sessions of training are shown for the two talkers in Figure 3. Figure 4 displays these scores separately for groups HM-S and S-HM. An ANOVA showed main effects of talker and training, and a significant talker by processor interaction [F (1,6) =37.1, $p =0.001$, $\eta^2 = 0.86$]. Because of this interaction, the primary analysis of this data was performed taking the male and female talkers separately.

For the male talker, vowel identification (see upper panel of Figure 3) was significantly more accurate in the shifted condition than in the high-matched condition [F (1,6) = 123, $p <0.001$, $\eta^2 = 0.95$]. There was a significant main effect of training [F (4,24) = 19.3, $p < 0.001$, $\eta^2 = 0.76$]. Bonferroni-corrected paired comparisons 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 also a significant interaction between processor and training [F (4,24) = 5.20, $p = 0.004$, $\eta^2 = 0.46$] that reflects the clear trend for a greater continuing increase in performance over training with the shifted processor. An interaction between processor, training and training order [F (4,24) = 2.88, $p = 0.044$, $\eta^2 = 0.33$] was also found (see upper panels of Figure 4). This interaction is attributed to a lack of a training effect in the high-matched condition when this is the second condition trained, while there is evidence of training in this condition when it is trained first.

A one-way ANOVA was performed to compare post-training scores on male vowels with the high-matched and shifted processors and also with the low-matched processor (data shown in Figure 3). Bonferroni-corrected paired comparisons confirmed that performance with the shifted processor exceeded that with the high-matched processor, while performance with the low-matched processor significantly exceeded that with each of the other two processors.

The pattern was different for the female talker (lower panels of Figures 3 and 4). Here the effect of processor was not significant [$F(1,6) = 4.01$, $p = 0.092$, $\eta^2 = 0.40$], although shifted scores tended to be lower than those in the high-matched condition. The only significant effect was that of training [$F(4,24) = 28.5$, $p < 0.001$, $\eta^2 = 0.83$]. As for the male talker, Bonferroni-corrected comparisons 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. Unlike the male talker vowel results, there was no significant interaction of training with processor, nor were there any significant interactions involving the training order factor.

A one-way ANOVA was performed to compare female talker vowel scores at the final training session with the shifted and high-matched processors and scores with the low-matched processor (see lower panel of Figure 3). Bonferroni-corrected paired comparisons confirmed that scores for the female talker did not differ between the shifted and high-matched processors, while scores with the low-matched processor were higher than those from both the shifted and the high-matched processors.

Insert Figure 5 about here

2.2.4 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 than as individuals) was performed for each talker in the shifted and high-matched conditions (see Figure 5). In the shifted condition (left panels) there was a significant correlation of performance with amount of training for both the male and the female talkers, while correlations were not significant for either talker in the high-matched condition (right panels). Logarithmic, logistic and exponential transformations of sessions of training each yielded correlations that were virtually indistinguishable from those from linear regression. While noting that both high-matched and shifted conditions after the first training session showed significant increases in performance from baseline levels in the ANOVA reported above (see section 2.2.3), regression analyses indicate that performance continues to increase with training only in the shifted condition. The data plotted in the lower right-panel of Figure 5 seem to suggest a trend of increasing performance with training for the female talker vowels, but this pattern is shown consistently by only 2 of the 8 subjects (+ and △ symbols).

Insert Figure 6 and Figure 7 about here

2.2.5 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 6. The same data is shown separately for groups HM-S and S-HM in Figure 7. 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 of performance with the shifted and high-matched processors showed significant effects of talker and of training, and a significant talker by processor interaction

[F (1,6) = 297, $p < 0.001$, $\eta^2 = 0.98$]. Hence, the sentence data were, like the vowel data, analysed separately for each talker.

As was observed for vowels, the male-talker sentence scores, shown in the upper panel of figure 6, were significantly higher with the shifted processor than with the high-matched processor [F (1,6) = 166, $p < 0.001$, $\eta^2 = 0.97$]. There was a significant effect of training [F (2,12) = 25.0, $p < 0.001$, $\eta^2 = 0.81$]. Bonferroni-corrected paired comparisons 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 talker vowel data, here there was no significant interaction of training with processor. Even at baseline, performance with the shifted processor exceeded that with the high-matched processor, and in later sessions, performance in the shifted condition approached ceiling levels. The between-subject order of training effect contributed to two interaction terms, these being training by training order [F (2,12) = 10.4, $p = 0.002$, $\eta^2 = 0.63$] and processor by training by training order [F (2,12) = 17.4, $p < 0.001$, $\eta^2 = 0.74$]. These interactions relate, as in the male talker vowel data, to the presence of a training effect in the high-matched condition only when this was the condition trained first (group HM-S: see upper panel of Figure 7).

A one-way ANOVA compared male talker sentence scores between the three processors after training. Scores for the shifted and high-matched processors were from the final session at which that processor was trained, while scores with the untrained low-matched processor were from session 6 (data shown in upper panel of Figure 6). Bonferroni-corrected paired comparisons confirmed that the shifted processor gave significantly higher trained performance than the high-matched processor. Scores with the low-matched processor significantly exceeded those with both of the other processors.

In contrast to the male talker, sentence scores for the female talker, as shown in the lower panel of Figure 6, were higher with the high-matched processor than with the shifted processor [F (1,6) = 105, $p < 0.001$, $\eta^2 = 0.97$]. Again there was a significant training effect [F (2,12) = 31.7, $p < 0.001$, $\eta^2 = 0.84$]. As for male speech, Bonferroni-corrected comparisons showed that both post-training sessions gave higher scores than at baseline, while scores after four 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, $p = 0.017$, $\eta^2 = 0.49$]. This interaction indicates a greater improvement in performance over training in the shifted condition than in the high-matched condition. There were no significant interactions with the training order factor for the female talker sentence data.

As with the male talker, sentence scores for the female talker at the final training session with the shifted and high-matched processor were compared to each other and to scores with the low-matched processor (data shown in lower panel of Figure 6). Bonferroni-corrected paired comparisons showed that the high-matched 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 low-matched processor significantly exceeded those with both of the other processors.

2.2.6 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 high-matched condition. These data are included in Figure 4 for vowel identification and in Figure 7 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; sentences: F (1,3) = 1.01, p = 0.39]. 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 and moreover were exposed to training in the high-matched condition.

2.3 Interim Discussion: Evidence for adaptation specific to upward-spectral shifting.

The design adopted here allows an assessment of two distinct aspects of adaptation to the speech-processing used in simulations of cochlear implant speech processors. Even in the absence of upward-spectral shifting, there is some degree of adaptation to noise-excited vocoder processing evident both here and in a previous study of unshifted processors (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Faulkner et al., 2003). The results of the present study show evidence of an adaptation that is specific to upward spectral shifting. Interactions in ANOVA indicate greater improvements with training with the shifted processor compared to the high-matched processor for male talker vowels, female talker sentences, and in auditory CDT with a female talker. Improvements in vowel identification for both the male and the female talker subsequent to the very first hour of exposure to these processors correlated significantly with the degree of training for the shifted processor, but not for the high-matched processor. It is only for the male talker sentences, where performance in the shifted condition was close to ceiling levels, that we find no evidence for greater adaptation in the shifted condition compared to the high-matched processor.

3. Experiment 2

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 more 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.

Insert Figure 8 about here

3.2 Results

Performance over training is shown in Figure 8. Both vowel and sentence 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, $\eta^2 = 0.70$] and vowel scores [F (1,18) = 63.8, p<0.001, $\eta^2 = 0.78$] showed significant effects of training. Talker was also a significant factor for sentence scores [F (1,6) = 19.8, p = 0.004, $\eta^2 = 0.77$], 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, $\eta^2 = 0.05$]. In neither data set was there a significant interaction of talker with training.

The single subject tested here shows comparable performance after four sessions of training to the final trained scores of the subjects taking part in experiment 1 with the same processor, For example, vowel accuracy has reached about 40% correct both for this subject and for those in experiment 1.

4. Discussion

This study confirms earlier findings that normal listeners can learn to adapt to speech that is spectrally shifted upwards (Fu & Galvin, 2003; Rosen et al., 1999). For vocoder-processed male speech, after about three hours of training in each condition, 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 an unshifted condition representing the same simulated electrode locations. For female speech, we find no difference in post-training vowel identification between this shifted processor and the high-matched 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 can 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.

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 high-matched condition. With unshifted processing, performance does increase with training for those subjects who were trained first in this condition, but there is no evidence of adaptation to unshifted vocoder-processed speech for those subjects receiving training in this condition subsequent to training in the shifted condition. For the shifted condition, however, continuing adaptation over several hours of training was found even after subjects had previously been trained in the unshifted condition. This indicates that some of the improvement in performance seen here is specific to an adaptation to spectral shifting that continues to occur well after a shorter period of adaptation to the effects of noise-excited vocoding with a limited number of spectral bands. As in our earlier study (Rosen et al., 1999), performance with upward shifted speech does not reach the same levels as seen with the same information presented at the tonotopically-correct place. It remains unclear whether further training would lead to equivalent performance with the same information shifted or unshifted.

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, 1999). Fu and 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 key spectral features related to formant frequencies lie within the range of formants shown by human speech across talker, sex and age.

It might be argued that the spectral shifts applied here are examples of a more general variation of spectral envelope associated with changes in vocal tract size. However, a 6 mm basilar membrane shift is very large compared to the between-talker variation 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. 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), and vowel formants are therefore shifted upwards to degree that considerably exceeds the formant shifts from adult male to 5 year old child speech.

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 et al., 2003). As we have shown in the present study, the consequences of this lower frequency information loss cannot be overcome by 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, 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. While it is not clear that implant users can fully adapt to shifted patterns of stimulation, a number of studies indicate that at least partial adaptation can occur (Dorman & Ketten, 2003; Fu et al., 2002; Harnsberger et al., 2001; McKay & Henshall, 2002; Skinner et al., 2002; Valimaa et al., 2002a; Valimaa, Maatta, Lopponen, & Sorri, 2002b). It may well be that implant users with relatively shallow electrode insertions can adapt sufficiently well that the effects of shifting become less damaging than the loss of low frequency information that would be inevitable with a processor matched to the positions of their electrodes. It has even been suggested that implant users can adapt to different frequency-to-place maps between two ears (Dorman & Dahlstrom, 2004), a possibility that merits further research.

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). We thank section editor Melanie Matthies for a number of helpful suggestions and two anonymous reviewers for their constructive comments.

References

ANSI (1997). Methods for the Calculation of the Articulation Index (Rep. No. ANSI S3.5-1997 (R 2002)). New York: American National Standards Institute.

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

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.

Dorman, M. F., & Ketten, D. (2003). Adaptation by a cochlear-implant patient to upward shifts in the frequency representation of speech. Ear and Hearing, 24, 457-460.

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

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

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

Fletcher, H., & Steinberg, J. C. (1930). Articulation testing methods. Journal of the Acoustical Society of America, 1, 1-48.

Foster, J. R., Summerfield, A. Q., Marshall, D. H., Palmer, L., Ball, V., & Rosen, S. (1993). Lip-reading the BKB sentence lists; corrections for list and practice effects. British Journal of Audiology, 27, 233-246.

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

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

Frijns, J. H. M., Briaire, J. J., & Grote, J. J. (2001). The importance of human cochlear anatomy for the results of modiolus-hugging multichannel cochlear implants. Otology and Neurotology, 22, 340-349.

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

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

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

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.

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

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. (2002). Frequency-to-electrode allocation and speech perception with cochlear implants. Journal of the Acoustical Society of America, 111, 1036-1044.

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

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

Sherbecoe, R. L. & Studebaker, G. A. (2002). Audibility-index functions for the connected speech test. Ear and Hearing, 23, 385-398.

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

Svirsky, M. A., Silveira, A., Neuburger, H., Teoh, S.-W., & Suarez, H. (2004). Long-term auditory adaptation to a modified peripheral frequency map. Acta Oto-Laryngologica, 124, 381-386.

Valimaa, T. T., Maatta, T. K., Lopponen, H. J., & Sorri, M. J. (2002a). Phoneme recognition and confusions with multichannel cochlear implants: Vowels. Journal of Speech, Language, and Hearing Research, 45, 1039-1054.

Valimaa, T. T., Maatta, T. K., Lopponen, H. J., & Sorri, M. J. (2002b). Phoneme recognition and confusions with multichannel cochlear implants: Consonants. Journal of Speech, Language, and Hearing Research, 45, 1055-1069.

Notes

1. The Greenwood equation may not be completely accurate as a model of the nerve fibres stimulated by a given cochlear implant electrode, since the stimulated neural tissues may not be those that are normally excited by motion at a particular basilar membrane location in response to acoustic input (Frijns, Briaire, & Grote, 2001), However, any error in simulated electrode position is approximately constant, at least in the outer two turns of the cochlea.

2. Articulation Index importance functions for connected speech materials show that the importance of information below 2.9 kHz is in excess of 80% of the total (e.g., Sherbecoe & Studebaker, 2002).

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

Reference Note

1. Video Recording. The BKB (Bamford-Kowal-Bench) standard sentence lists. 1986. London. EPI Group, Department of Phonetics and Linguistics, University College London.

Figure legends

Figure 1: Basilar membrane CFs against distance from cochlear base. The simulated positions of the two electrode arrays and the CF range for each array are shown for the high-matched and low-matched processors. The speech processing filters for the shifted processor match the CFs of the electrode simulated in the low-matched processor, while the simulated electrode locations match the CFs of the electrode simulated in the high-matched processor, i.e., there is a 6mm basalward shift.

Figure 2: CDT rates in training with auditory-visual (upper panel) and auditory (lower panel) presentation modes. In this and subsequent box and whisker plots, the boxes show the inter-quartile range (IQR), the bars show the median, and the whiskers show the range excluding outliers: outlying points are those more than 1.5 IQR units outside of the IQR, and are shown by symbols (○ or ✳).

Figure 3: Box and whisker plots showing vowel scores in high-matched 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 low-matched processor from session 1 (shown here as baseline) and session 6 (shown here as retest) is also included.

Figure 4: Vowel scores in the shifted and high-matched conditions over session for each talker and training order. The left panels show scores from subjects trained first with the high-matched condition; the right panels show scores from subjects trained first with the shifted condition. The abscissa 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.

Figure 5: Linear regression of vowel scores as a function of sessions of post-baseline training. Upper panels: male talker, lower panels: female talker. Left panels: shifted processor, right panels: high-matched processor. Symbols represent individual subjects. Lines represent a linear regression and 95% confidence limits.

Figure 6: Box and whisker plots showing sentence scores in the shifted and high-matched conditions at baseline, after 2 and 4 sessions of training (T2, T4), and in the final retest session. Performance in the low-matched condition as measured in session 6 is also shown alongside the retest data. Upper panel: male talker; lower panel: female talker.

Figure 7: Sentence scores over session by talker and training order. The left panels show scores from subjects trained first with the high-matched condition; the right panels show scores from subjects trained first with the shifted condition. The abscissa 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 high-matched 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

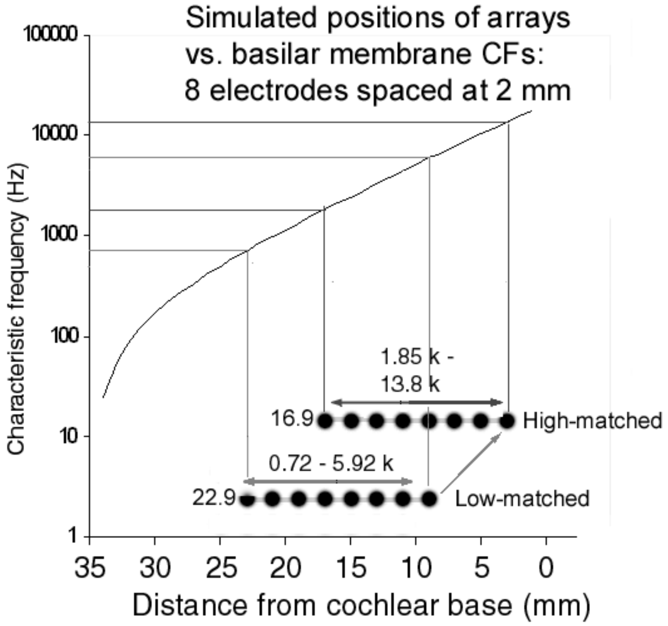
Figure 8: Vowel identification performance (upper panel) and key words correct in sentences (lower panel) for the shifted processor as a function of number of training sessions in experiment 2. Solid symbols are for a male talker, unfilled symbols for a female talker. The lines show linear regression fits, for which the corresponding R^2 is shown in the legend.

Input band		Centre frequency (Hz)	Cut-off (Hz)	Distance from base (mm)
Shifted	High-matched			
			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

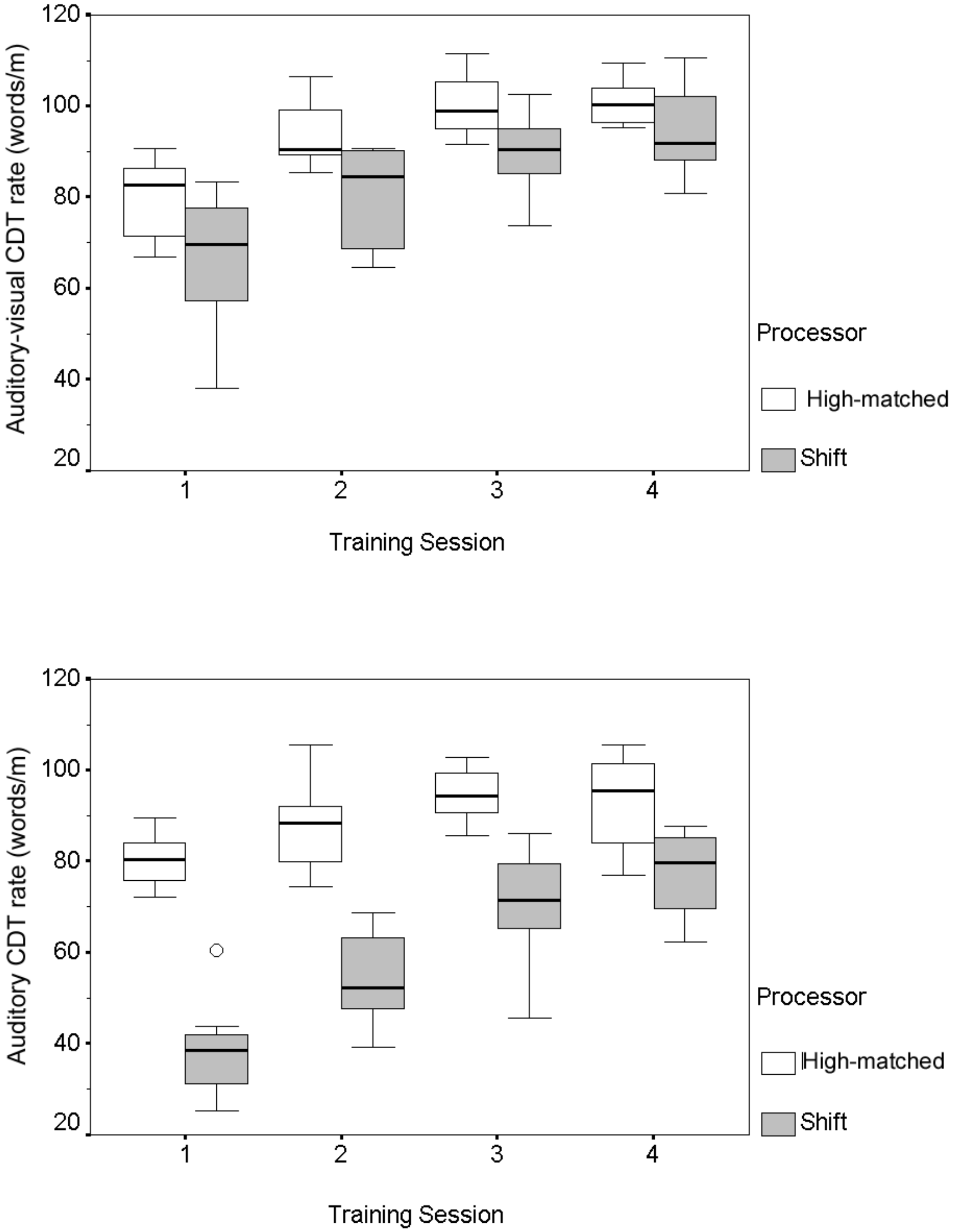
Table 1: Centre and cut-off frequencies of input filters for shifted and high-matched processors. The output filters for both processors were identical to the input filters of the high-matched processor. The basilar membrane locations for a 35 mm long cochlea that match each centre and cut-off frequency are shown in the right hand column.

Session	Training	Testing
1: Baseline 1	None	Familiarisation - 1 list vowels: unprocessed - 1 sentence list: low matched (LM) 1st baseline measures - 2 lists vowels: high-matched: (HM) - 2 lists each of female and male talker sentences: HM - 2 lists vowels: shifted (S) - 2 lists each of female and male talker sentences: S
2: T1: HM	CDT 45m: HM	- 2 lists vowels :HM
3: T2: HM	CDT 35m: HM	- 2 lists vowels :HM - 2 lists each of female and male talker sentences: HM
4. T3: HM	CDT 45m: HM	- 2 lists vowels :HM
5. T4: HM	CDT 35m: HM	- 2 lists vowels :HM - 2 lists each of female and male talker sentences: HM
6. Baseline 2	None	Low-matched control tests - 2 lists vowels: LM - 2 lists each of female and male talker sentences: LM Baseline for 2nd trained condition - 2 lists vowels: S - 2 lists each of female and male talker sentences: S
7. T1: S	CDT 45m: S	- 2 lists vowels: S
8. T2: S	CDT 35m: S	- 2 lists vowels: S - 2 lists each of female and male talker sentences: S
9. T3: S	CDT 45m: S	- 2 lists vowels: S
10. T4: S	CDT 35m: S	- 2 lists vowels: S - 2 lists each of female and male talker sentences: S
11. Retest	None	Retest of 1st trained condition - 2 lists vowels: HM - 2 lists each of female and male talker sentences: HM

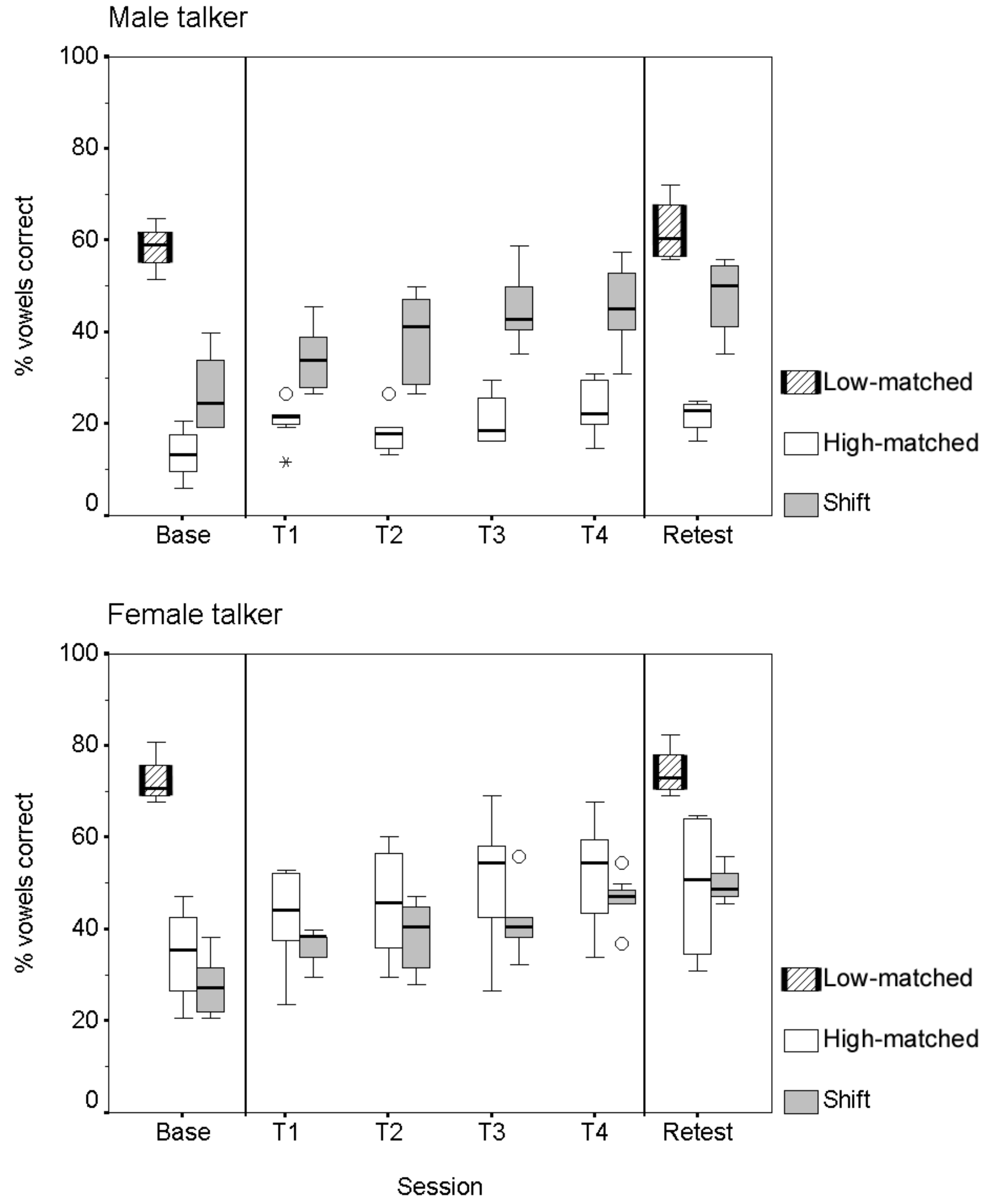
Table 2: Summary of training and testing over sessions for group HM-S. For group S-HM, sessions 2 to 5 and 11 used shifted in place of high-matched processing, while sessions 6 to 10 used high-matched in place of shifted processing. For CDT, the number of minutes of training in each session is also shown.



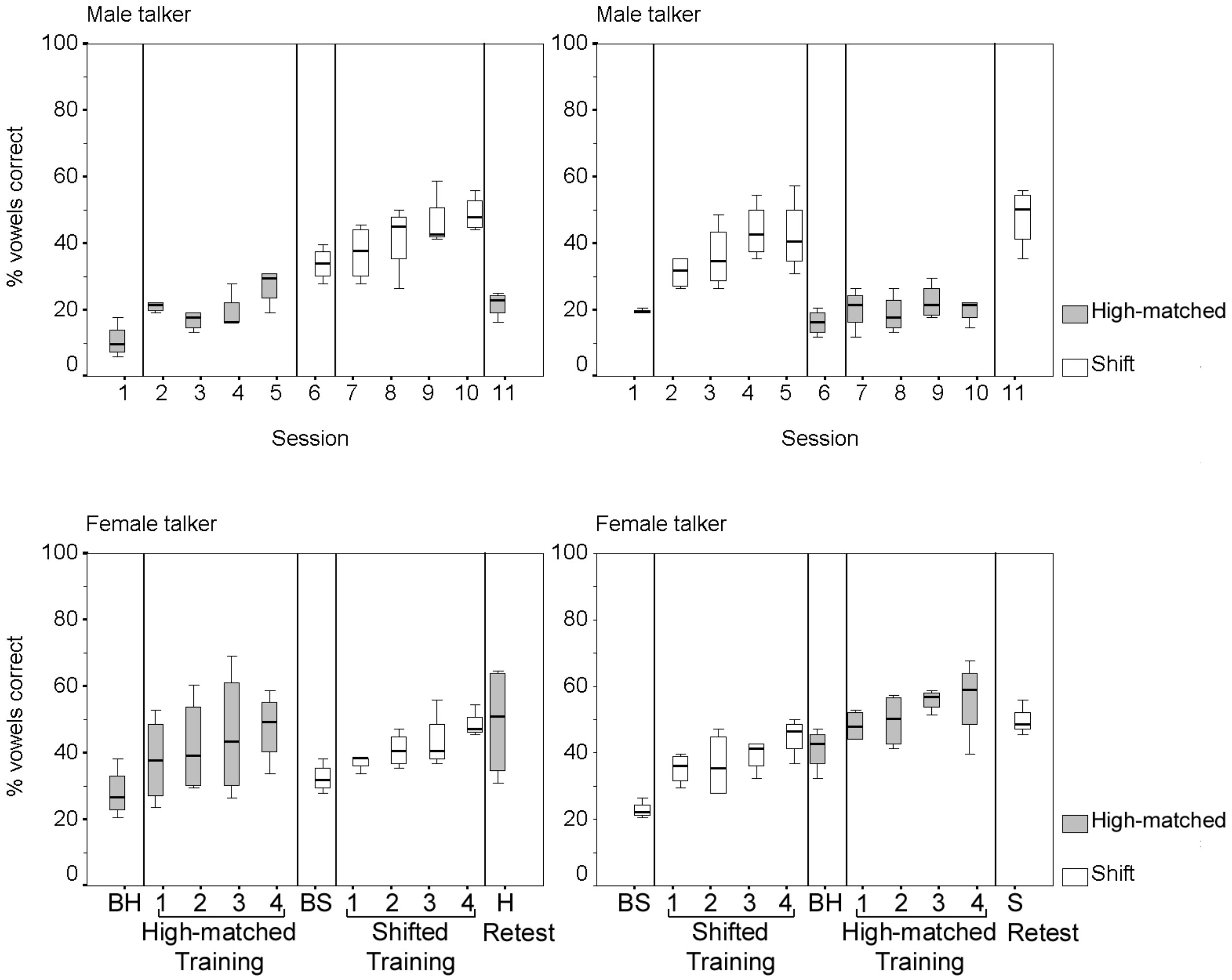
Faulkner Rosen and Norman Figure 1



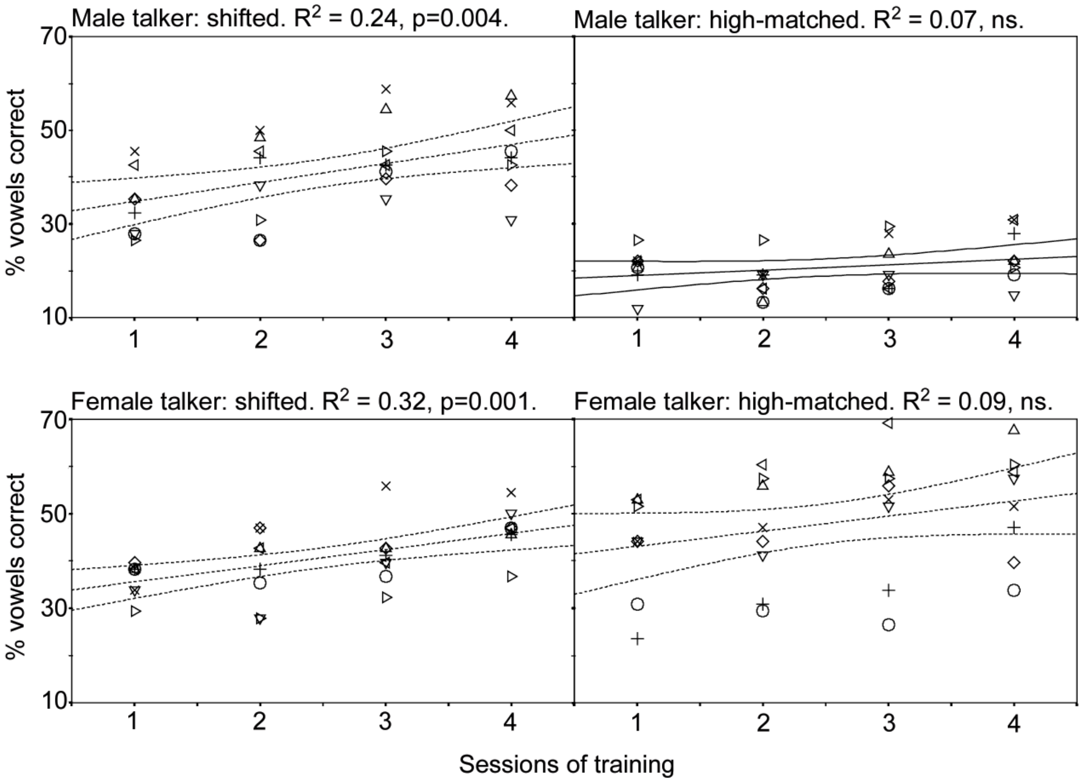
Faulkner Rosen and Norman, Figure 2:



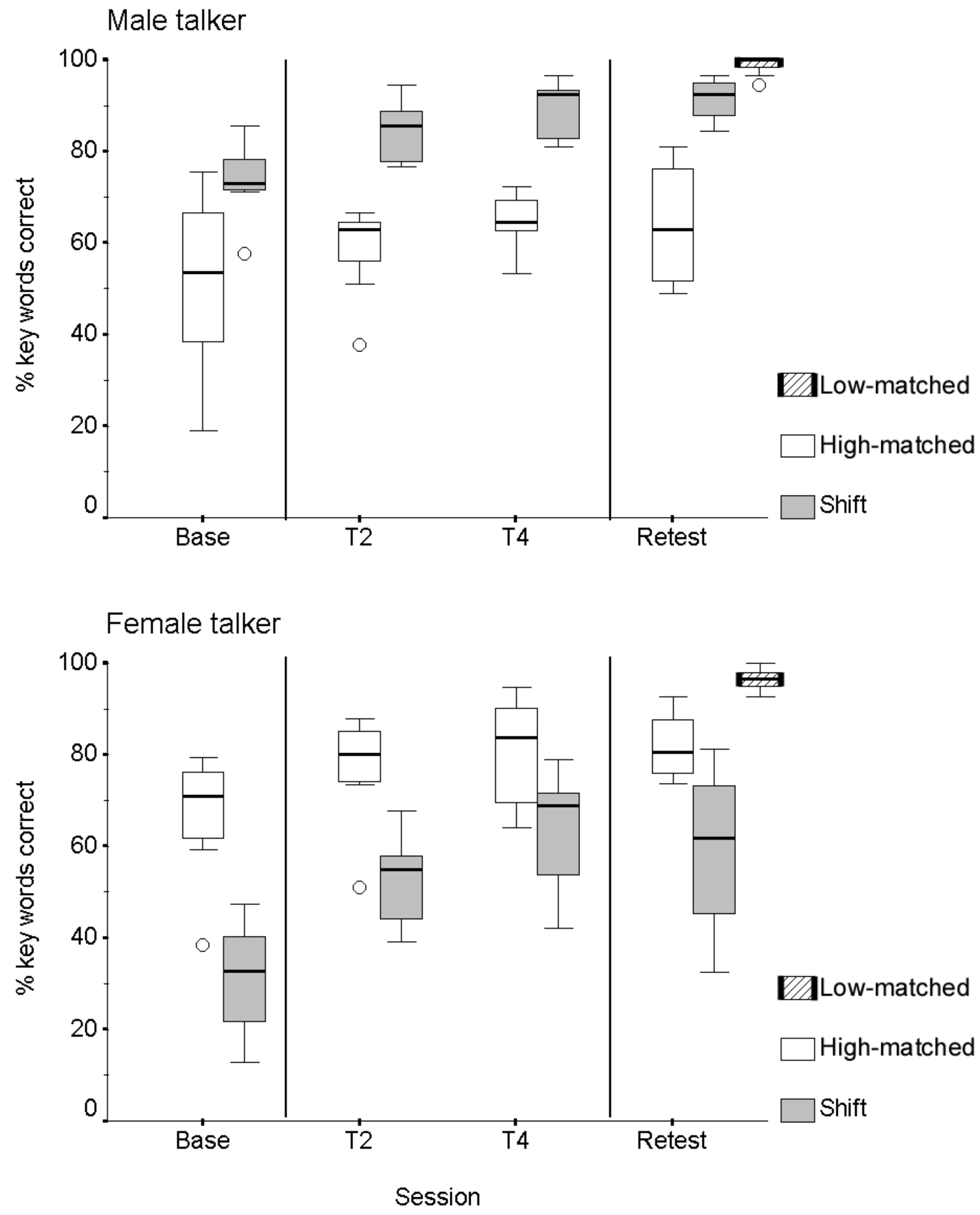
Faulkner Rosen and Norman, Figure 3



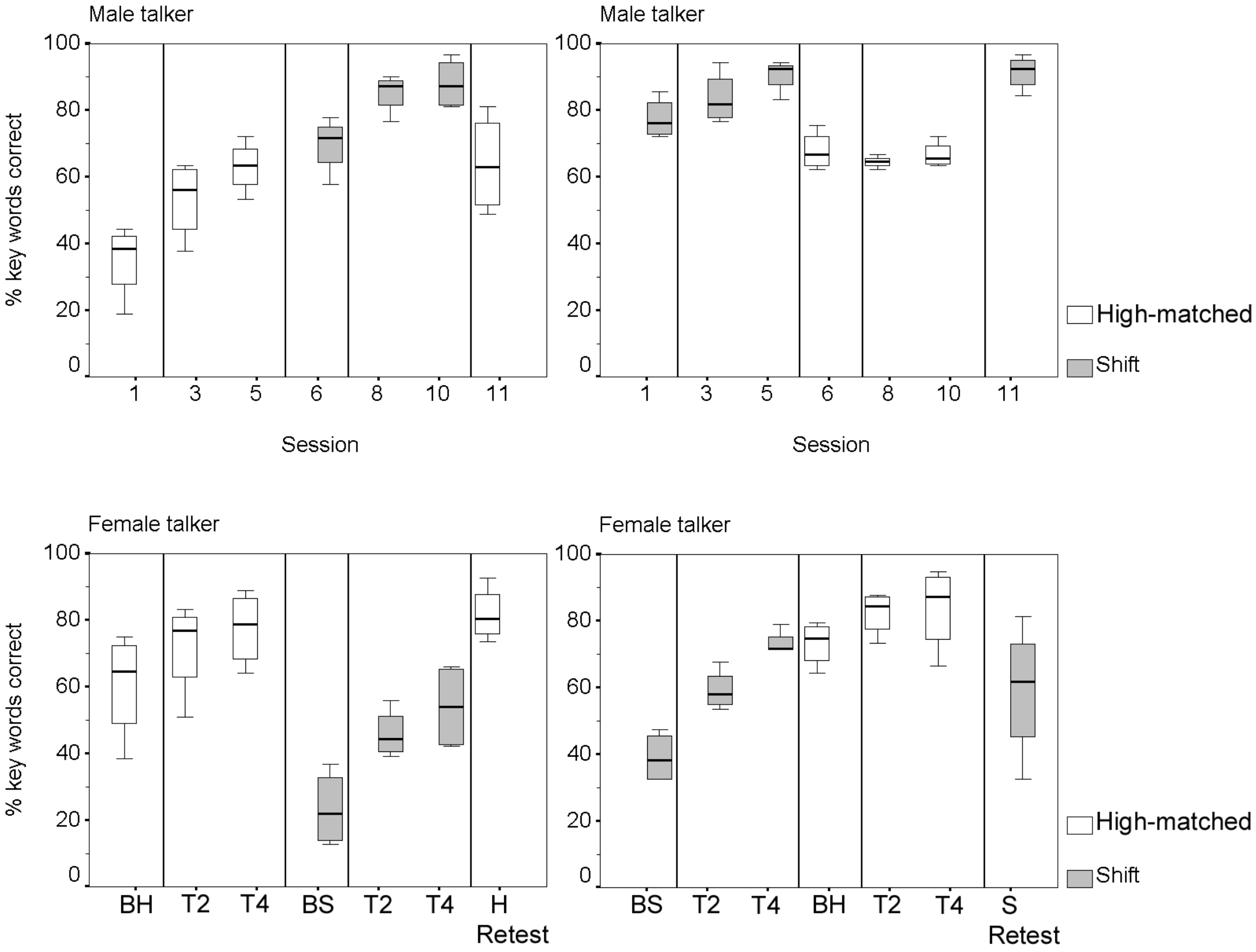
Faulkner Rosen and Norman, Figure 4:



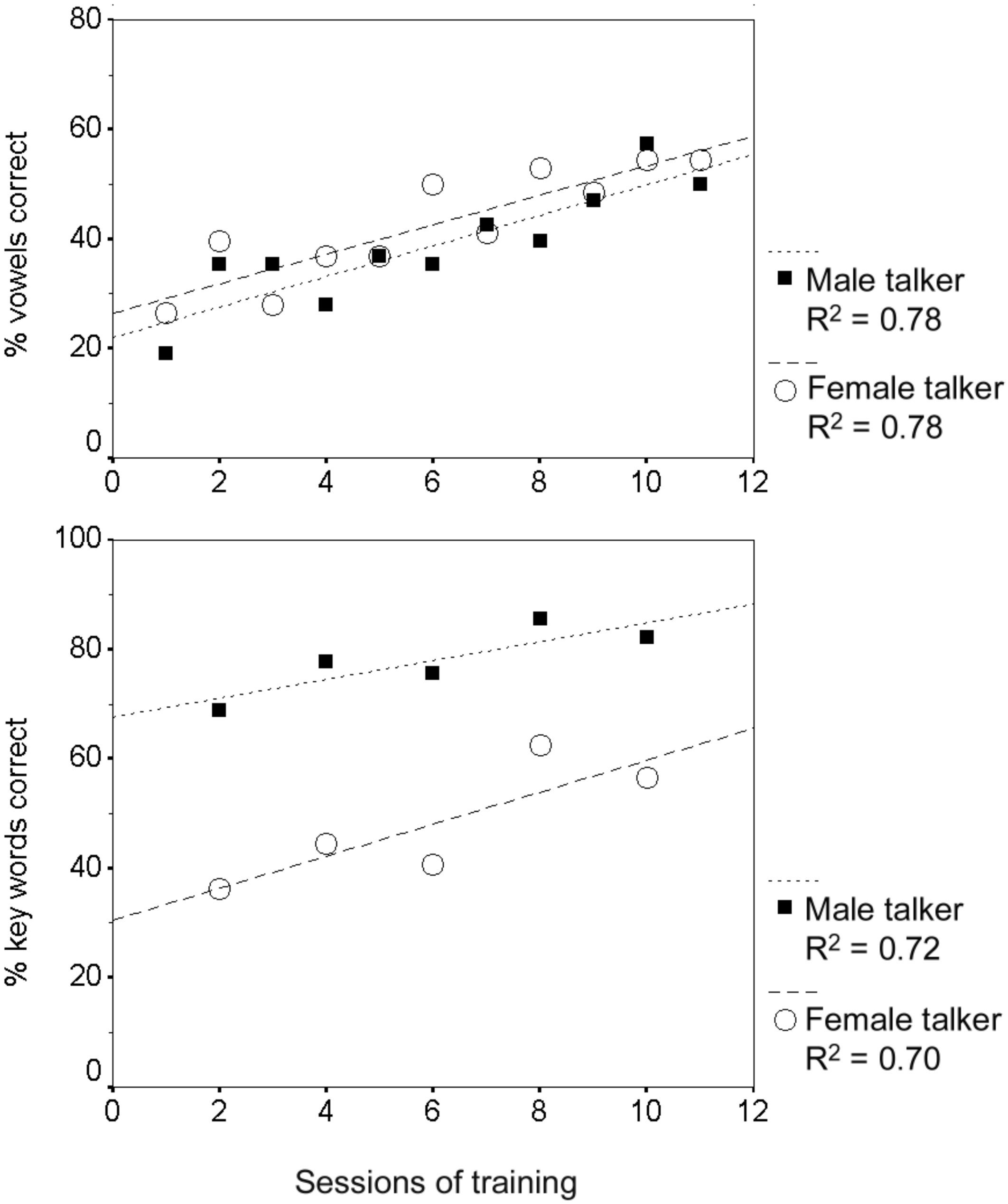
Faulkner Rosen and Norman, Figure 5



Faulkner Rosen and Norman, Figure 6:



Faulkner Rosen and Norman, Figure 7:



Faulkner Rosen and Norman, Figure 8