Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech

Paula C. Stacey[1,2], and A. Quentin Summerfield[1]

[1] Department of Psychology, University of York, Heslington, York YO10 5DD, UK.

[2] MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK.

Suggested running title: Computer-based auditory training

Corresponding author:

Paula Stacey

Division of Psychology

Nottingham Trent University

Burton Street

Nottingham NG1 4BU

UK.

Telephone: (+44) 0115 858 5575

email: paula.stacey@ntu.ac.uk

Abstract

Five experiments were designed to evaluate the effectiveness of 'High-Variability' lexical training in improving the ability of normally-hearing subjects to perceive noise-vocoded speech that had been spectrally shifted to simulate tonotopic misalignment. Both word- and sentence-based auditory training improved the ability to identify words in sentences. Improvements following a single session (lasting 1-2 hours) of auditory training ranged between 7 and 12%pts and were significantly larger than improvements following a visual control task that was matched with the auditory training task in terms of the response demands. An additional three sessions of word- and sentence-based training led to further improvements, with the average overall improvement ranging from 13-18%pts. When a tonotopic misalignment of 3mm rather than 6mm was simulated, training with several talkers led to greater generalization to new talkers than training with a single talker. The results confirm that computer-based lexical training can help overcome the effects of spectral distortions in speech, and they suggest that training materials for use by adult cochlear-implant users should include several talkers.

I. Introduction

Cochlear implantation improves the speech perception abilities of post-lingually deafened adults with profound-to-total hearing loss (e.g. Summerfield & Marshall, 1995). However, outcomes following implantation are highly variable (Gantz *et al*., 1993). One way to improve the speech-perception skills of adult cochlear-implant users might be to administer auditory training. Initial investigations (Gagne *et al*., 1991; Dawson & Clark, 1997; Busby *et al*., 1991). provided only limited evidence in support of the effectiveness of such training. However, more recent studies have shown that computer-based auditory training can improve both timbre-recognition (Gfeller *et al*., 2002) and speech-perception skills (Fu *et al*., 2005[a]) of adults who use implants. Fu *et al*. (2005[a]) argued that previous studies had failed to find systematic benefits because insufficient training had been provided. Extensive training can now be delivered at low cost via personal computers. As a precursor to working with patients, the experiments reported here evaluated the effectiveness of two computer-based training approaches with normally-hearing subjects who listened to speech through a simulation of the information provided by a cochlear-implant system (Shannon *et al*., 1995).

A variable that has been associated with poor speech perception by users of cochlear implants is tonotopic misalignment between the frequency band transmitted by an electrode and the characteristic frequency of the location of that electrode (Skinner *et al*., 2002; Yukawa *et al*., 2004). One reason for providing auditory training would be to help overcome the difficulties in speech perception caused by tonotopic misalignment. The consequences of tonotopic misalignment for speech perception can be investigated with normally-hearing subjects who listen to speech through noise-band vocoders designed to simulate the information provided by a cochlear implant (Shannon *et al*., 1995; Baskent & Shannon, 2003; Dorman *et al*., 1997; Rosen *et al*., 1999). Accuracy of speech recognition declines when signals are spectrally shifted to simulate tonotopic misalignment (Baskent & Shannon, 2003; Dorman *et al*., 1997; Rosen *et al*., 1999). However, the decline can be ameliorated with auditory training (Rosen *et al*., 1999; Fu *et al*., 2005[b]; Faulkner *et al*., 2006).

Rosen *et al*. (1999) tested the ability of normally-hearing subjects to perceive spectrally-shifted speech which had been processed to simulate the consequences of tonotopic misalignment of 6.5mm. Fewer than 1% of words in sentences were identified correctly, compared with 64% correct performance with unshifted signals. However, after nine 20-minute sessions of auditory training using Connected Discourse Tracking (CDT, De Filippo & Scott, 1978), performance improved to 30% correct. In CDT, an experimenter reads a passage of text, and the subject attempts to repeat verbatim what was said, with corrective feedback from the reader. Although CDT is an effective training approach (see also Faulkner *et al*., 2006), it is labor intensive and expensive to administer clinically. Self-administered techniques could be a cost-effective alternative. Fu *et al*. (2005[b]) compared the effectiveness of two self-administered computer-based training protocols in improving the ability of normally-hearing listeners to discriminate spectrally-distorted vowel and consonant sounds. 'Word-based training' required subjects to identify the vowels in consonant–vowel–consonant monosyllabic words; 'sentence-based training' consisted of a computer-based CDT procedure. Both training approaches led to significant improvements in the ability to discriminate consonant sounds, but while word-based training led to significant improvements in the ability to discriminate vowel sounds, sentence-based training did not. Fu *et al*. (2005[b]) therefore concluded that word-based training might be more effective than sentence-based training in developing the speech-perception skills of cochlear-implant users.

There is a limitation with the study by Fu *et al*. (2005[b]) however, since they did not include a test of sentence perception. It is possible that sentence-based training leads to larger improvements in sentence perception than does word-based training. In addition, performance on a test of sentence perception might provide a more representative test of subjects' ability to communicate in everyday situations. Similar to Fu *et al*. (2005[b]), the present experiments examined the effectiveness of two computer-based lexical training approaches. The first training task required subjects to recognize isolated words, while the second training task required subjects to recognize words in sentences. The effectiveness of auditory training was assessed using tests of consonant and vowel discrimination. In addition, we assessed the extent to which each of the training tasks led to improvements on a test of sentence perception, thereby addressing the limitation of the study by Fu *et al*. (2005[b]). The experiments addressed two further issues, discussed below.

High-Variability training

Studies of perceptual learning for speech have shown that training with several talkers, usually termed 'High-Variability' training, is more effective than training with a single talker. High-Variability lexical training is effective in training Japanese Americans to distinguish /r/ and /l/ (Lively *et al*., 1993; Lively *et al*., 1994; Logan *et al*., 1991; Bradlow *et al*., 1997), for native English speakers to perceive Cantonese-accented speech (Bradlow & Bent, 2003), for American English speakers learning to classify American dialects (Clopper & Pisoni, 2004), and for North-American subjects seeking to learn Mandarin tones (Wang *et al*., 1999). These studies have demonstrated that training with several talkers is more generalizable to new talkers compared with training with a single-talker. Training with many talkers may help subjects to dissociate talker-specific information from lexically-specific information, since the two sorts of information are confounded when only one talker is used in training. The present experiments investigated whether High-Variability training is also advantageous in training subjects to perceive spectrally-distorted speech.

Training-related vs incidental learning

Although improvements on tests of speech perception following auditory training may be caused by the training task (training-related learning), the contribution of 'incidental learning' must also be considered. Incidental learning refers to improvements that occur independent of the auditory training task, through procedural learning of task demands (Robinson & Summerfield, 1996), or perceptual learning resulting from repeating exposure to test materials. A further type of incidental learning has also been documented. Amitay *et al*. (2006) reported larger improvements in frequency discrimination for control subjects who played a purely visual computer game between successive tests than for control subjects who did not engage in an intervening task. These results suggest that maintaining attention and arousal, without explicit training, may be sufficient to lead to improvements on perceptual tasks. In order to evaluate the extent to which a training task has contributed to improvements in performance, it is important to factor out improvements related to 'incidental learning'.

Rosen *et al*. (1999) did not include a control condition, so it was not possible to measure the extent to which auditory training, rather than incidental learning, led to the observed improvements in performance. Fu *et al*. (2005[b]) did include a control condition, in which one group of subjects undertook repeated tests of speech perception without undertaking training. However, this procedure does not control the effects of maintaining attention and arousal. In our experiments, we controlled incidental learning by comparing improvements following auditory training with improvements following a 'matched' visual control task that exposed subjects to the same vocabulary and imposed similar task demands as the auditory training task. We use the term 'matched' to refer to our goal of matching the visual control task with the auditory training task in terms of its content, nature, and task demands.

Aims and key hypotheses

Against this background, five experiments are reported in this paper. Experiments 1, 2, and 3 examined whether a single session (lasting 1 to 2 hours) of auditory training improved the perception of spectrally-shifted noise-vocoded speech more than did a matched visual control task. These experiments also tested the hypothesis that High-Variability training is more effective than training provided by a single-talker. A word-based training task was used in

Experiments 1 and 2, while a sentence-based training task was used in Experiment 3. Experiments 1 and 2 differed in the degree to which signals were spectrally shifted. In Experiment 1, speech was spectrally shifted to simulate a 6mm tonotopic misalignment, while Experiment 2 simulated a 3mm tonotopic misalignment. The aim was to measure the relationship between the degree of tonotopic misalignment and the amount of improvement. Experiments 4 and 5 included four High-Variability training sessions to establish whether performance continues to improve throughout 4 to 6 hours of word- (Experiment 4) and sentence-based (Experiment 5) training.

In summary, the experiments addressed three hypotheses:

(1) Word-based (Experiments 1 and 2) and sentence-based auditory training (Experiment 3) lead to larger improvements on tests of sentence, consonant, and vowel perception than does a matched visual control task.

(2) High-Variability training leads to larger improvements in speech perception than does Single-Talker training (Experiments 1, 2, and 3).

(3) Extending the amount of word-based (Experiment 4) and sentence-based (Experiment 5) auditory training leads to continued improvements in performance.

II. Methods common to all experiments

1. Design & procedure

The overall design and procedure were the same in each experiment. Subjects initially completed baseline tests of speech perception, followed by a variable number of sessions of auditory training, followed by further tests of speech perception. The two training procedures are described in the methods sections of Experiments 1 and 3. In all experiments, subjects completed a IEEE sentence test, then a test of vowel discrimination, followed by a test of consonant discrimination. These tests are described below. Training and testing took place in a double-walled sound-attenuated chamber. No feedback on accuracy was provided during testing.

Speech materials used for training and for tests of speech perception were recorded digitally (sample rate 44.1kHz, amplitude quantification 16 bits) in a carpeted double-walled sound attenuated chamber. The talkers had a range of British and Irish accents. Male talkers were aged between 27 and 58 years (mean 44 years); female talkers were aged between 27 and 40 years (mean age 30 years); and child talkers were aged 7 and 8 years. Stimuli were presented through an Audiomaster LS3/5A loudspeaker. Peak stimulus levels, measured with a Bruel and Kjaer Type 2260 sound level meter using a 1-s integration time and a Type 4189 half-inch microphone, ranged between 65 and 75 dB(A) across training and test materials.

2. IEEE sentence test

Four blocks of eighty sentences from the IEEE corpus (IEEE, 1969) were recorded by ten talkers (4 male, 4 female, 2 children). One block was used in each test session. All the sentences were different. There were five key words in each sentence. Subjects were asked to repeat all the words they heard, and the experimenter recorded which key words had been identified correctly.

3. Consonant test

Twenty /ɑː/-consonant-/ɑː/ nonsense syllables were included, incorporating the consonants /b ʧ d f g h ʤ k l m n p r s ʃ t θ v w z/. Presentation was computer controlled. Each consonant was displayed orthographically on a computer touch screen using its usual spelling (e.g. the sound /ʧ/ was written “CH”). Subjects reported the consonant in each stimulus by touching its orthographic transcription. There were 200 trials in each test. In Experiments 1, 2 and 4, materials were recorded by five talkers (2 male, 2 female, 1 child), and each talker recorded two tokens of each syllable. In Experiments 3 and 5, materials were recorded by 10 talkers (4 male, 4 female, 2 children), and each talker recorded a single token of each syllable.

4. Vowel test

Ten h-vowel-d words were included, containing 5 short vowels: /æ/ (had), /e/ (head), /ɪ/ (hid), /ɒ/ (hod), /ʊ/ (hood), and 5 long vowels: /ɑː/ (hard), /ɜː/ (heard), /iː/ (heed), /ɔː/ (hoard), /uː/ (who'd). Presentation was computer controlled. Each word was displayed orthographically on a computer touch screen. Subjects responded by touching the orthographic transcription of the appropriate word. There were 200 trials in each test. In Experiments 1, 2 and 4, materials were recorded by five talkers (2 male, 2 female, 1 child). Each talker recorded two tokens of each word which were presented twice. In Experiments 3 and 5, materials were recorded by 10 talkers (4 male, 4 female, 2 children). Each talker recorded two tokens of each word.

5. Speech processing

Speech processing was performed in real time with an 8-channel noise-excited vocoder (Shannon *et al.*, 1995) implemented on a SHARC digital processor (Analog Devices ADSP21065L). Following Rosen *et al.* (1999), speech signals were analyzed with 6th-order elliptical IIR filters with centre frequencies of 433, 642, 925, 1306, 1820, 2513, 3449, and 4712Hz. Filtered signals were half-wave rectified and low-pass filtered at 160Hz. The resulting waveform envelopes were multiplied by a white noise that had been low-pass filtered at 10kHz. The resulting signal in each channel was then filtered by a 6th-order elliptical IIR filter whose centre frequency had been shifted relative to the analysis filter in that channel in accordance with Greenwood's (1990) place-to-frequency function to simulate either a 3mm or a 6mm tonotopic shift. When a 6mm shift was simulated (in Experiments 1, 3, 4, 5), the centre frequencies of the reconstruction filters were 1206, 1685, 2332, 3205, 4382, 5971, 8115, and 11007Hz. When a 3mm shift was simulated (in Experiment 2), the centre frequencies were 741, 1057, 1485, 2061, 2839, 3889, 5305, and 7216Hz.

6. Subjects

All subjects had normal hearing (≤25dB HL at octave frequencies between 250 and 8000Hz, inclusive) measured according to British Society of Audiology (BSA) guidelines (BSA, 1981). All subjects were native speakers of British English, and were aged between 18 and 53 years (median age 20 years). Subjects were students or staff of the Universities of Nottingham and York. None of the subjects took part in more than one experiment.

III. Experiment 1

This experiment investigated the effectiveness of a word-based training task in improving the ability to identify words in sentences, vowel sounds, and consonant sounds. Improvements in speech perception following auditory training were compared with improvements following a matched visual control task. The effectiveness of High-Variability auditory training, in which training materials were recorded by ten different talkers, was compared with the effectiveness of Single-Talker auditory training.

A. Method

1. Subjects & speech processing

Sixteen subjects listened to speech through a noise-excited vocoder which simulated a 6mm tonotopic misalignment.

2. Training and control tasks

Training was provided by a 2-alternative forced-choice task. At the start of each trial, two words were presented orthographically on the left and right of a computer touch screen. The target word was then presented. Subjects responded by touching the word corresponding to the target. Visual feedback on accuracy was given, with a green check indicating that the subject had responded correctly, and a red cross indicating that an incorrect decision had been made. If subjects were incorrect, the trial was repeated until the correct response was given. During auditory training, the target was presented acoustically. During the control task, the target appeared orthographically in the centre of the screen degraded by visual noise.

To construct the training materials, 200 key words were selected from 40 IEEE sentences. Three foils were created for each key word, forming quasi-minimal pairs. Over the course of 1200 trials, each key word was presented as the target word itself with each of its three foils, and the three foils were presented as the target word with the key word as the alternative. For High-Variability auditory training, 10 talkers (4 male, 4 female, 2 children) recorded the 800 words, with each talker recording 80 words. A single male talker with a southern British accent recorded all 800 words for the Single-Talker condition. The control task was the same for both the High-Variability and Single-Talker conditions. The auditory training and visual control tasks took approximately one hour to complete.

3. Design & Procedure

Four groups of four participants participated in three sessions, Groups 1 and 2 received High-Variability training. Groups 3 and 4 received Single-Talker training. Groups 1 and 3 received auditory training between Test Sessions 1 and 2, while groups 2 and 4 received auditory training between Test Sessions 2 and 3 (Table 1). Sessions took place on consecutive days. During the first session, subjects completed baseline tests of speech perception (Test session 1). During the second session subjects completed the auditory training task or the visual control task, followed by further tests of speech perception (Test session 2). In the final session, subjects completed either the control task or the auditory training task, again followed by further tests of speech perception (Test session 3).

[TABLE 1]

4. Analyses

Analyses were based on changes in performance, measured in percentage points (%pts) (Footnote 1) between adjacent test sessions. We distinguished changes associated with auditory training from changes associated with visual control training. For Groups 1 and 3, the change following auditory training was the difference in score between Test Sessions 1 and 2, and the change following control training was the difference in score between Test Sessions 2 and 3. For Groups 2 and 4, the change following auditory training was the difference in score between Test Sessions 2 and 3, and the change following control training was the difference in score between Test Sessions 1 and 2. The first analysis tested whether either auditory or control training led to a significant improvement in performance. One-sample t-tests were performed on the changes in performance following auditory and, separately, control training. The second analysis tested whether auditory training was more effective than the control task in improving speech perception. This hypothesis was tested with paired-samples t-tests. The final analysis tested whether High-Variability auditory training led to larger improvements in performance than Single-Talker auditory training. This hypothesis was tested with independent samples t-tests on the change following auditory training according to whether subjects received High-Variability or Single-Talker training.

B. Results

1. IEEE Sentence test

Figure 1 (Panel A) shows the mean change, and the spread among subjects, in the ability to identify words in IEEE sentences following auditory and control training. The mean improvements and 95% confidence intervals are shown in Table 2. There was a significant improvement in sentence perception following auditory training (t_{15} = 8.09. $p<0.001$), but not following control training (t_{15} = 1.81). In addition, there was a significantly larger improvement following auditory training than following the control task (t_{15} = 3.13, $p<0.01$). There was no significant difference between the effectiveness of High-Variability and Single-Talker auditory training (t_{14} = 0.83). Figure 2 summarizes these effects by plotting the percentages of key words identified correctly in IEEE sentences according to test session, with data collapsed over variability.

[FIGURE 1]

[TABLE 2]

[FIGURE 2]

Five of the talkers who recorded the test of sentence perception also recorded the training materials ('old' talkers) and five did not ('new' talkers). There were no significant differences between the effectiveness of High-Variability and Single-Talker auditory training when talkers were 'old' (t_{14} = 1.10), or 'new' (t_{14} = 0.14).

2. Consonant test

There was a significant improvement in sentence perception following auditory training (t_{15} = 5.54. $p<0.001$), but not following control training (t_{15} = 1.43; Figure 1, Panel B; Table 2). However, the difference between the change in performance following the auditory training task compared with the control task just failed to reach significance at the <0.05 level (t_{15} = 2.06, p=0.057). There was no significant difference between the effectiveness of High-Variability and Single-Talker auditory training (t_{14} = 0.86; Table 2).

3. Vowel test

There was a significant improvement in vowel discrimination following auditory training (t_{15} = 4.62. p<0.001), but not following control training (t_{15} = 1.50; Figure 1, Panel C; Table 2). There was no significant difference between the change in performance following auditory training compared with the control task (t_{15} = 1.31). There was no significant difference between the effectiveness of High-Variability and Single-Talker auditory training (t_{14} = 1.10; Table 2).

C. Discussion

These results confirm previous demonstrations (Rosen *et al*.,1999; Fu *et al*., 2005[b]; Faulkner *et al*., 2006) that auditory training improves the ability to perceive spectrally-distorted speech. A computer-based, word training task, similar to that used by Fu *et al*. (2005[b]), was associated with significant improvements in the ability to identify words in sentences. Moreover, improvements in the ability to identify words in sentences were significantly larger following auditory training than following a visual control task which exposed subjects to the same vocabulary and imposed the same task demands as the auditory training task. By subtracting the improvement following control training (2.0%pts) from the improvement following auditory training (7.9%pts), an improvement of 5.9%pts in sentence perception can be attributed to perceptual learning resulting from auditory training.

Auditory training was also associated with significant improvements in consonant and vowel discrimination, whereas control training was not. Previously, Fu *et al*. (2005[b]) reported that word-based training led to significant improvements in consonant and vowel discrimination, while no significant improvements were found for a control group. However, the present experiment did not find that auditory training was associated with significantly *larger* improvements in consonant or vowel recognition than control training. Thus, the effect of auditory training on consonant and vowel discrimination was weaker than its effect on the ability to identify words in sentences. In contrast to these results, Fu *et al*. (2005[b]) reported a significantly larger improvement in consonant discrimination for a group who received word-based training than for a control group who undertook testing but not training. Part of the difference between the present result and Fu's may have arisen because we controlled for exposure to vocabulary and attention/arousal, along with exposure to test materials. Fu *et al*. (2005[b]) did not report whether the improvement for the word-based training group was significantly larger than the improvement for the control group on the test of vowel discrimination.

There was no evidence that High-Variability auditory training was more effective than Single-Talker auditory training, and High-Variability training did not lead to greater transfer to new talkers. An explanation for this difference may be found in the processing that generates noise-vocoded speech, which strips away many of the cues that distinguish talkers (Gonzalez & Oliver, 2005; Chinchilla-Rodriguez *et al*., 2004). Gonzalez and Oliver (2005) examined the ability to identify talker gender and speaker identity in noise-vocoded speech. With a similar noise-band vocoder as used in the present study (8-channel, 160Hz low-pass filtering of the envelopes within channels) subjects identified the gender of speakers with 89% accuracy, and could identify speakers with 78 to 84% accuracy. Thus, although discrimination of different speakers is possible with noise-vocoded speech, performance is far from perfect. In addition, we would expect poorer differentiation between talkers given the size of the tonotopic misalignment which we simulated. Chinchilla-Rodriguez *et al.* (2004) reported that 32 channels were required to discriminate voice gender when speech was spectrally shifted by an octave (to simulate a tonotopic misalignment between 4 and 5mm), compared with 16 channels for unshifted speech. It is possible therefore that differences between the effectiveness of training according to variability would be found if more cues that differentiate different talkers were retained by, for example, simulating a smaller degree of tonotopic misalignment.

IV. Experiment 2

Average levels of performance in Experiment 1 were low. Possibly, the effects of training would be larger if the amount of tonotopic misalignment was reduced so that subjects were operating on a steeper part of the psychometric function relating tonotopic misalignment to performance (Fu & Shannon, 1999). Experiment 2 examined the effectiveness of High-Variability and Single-Talker auditory training when a 3mm tonotopic misalignment was simulated.

A. Method

The methods were the same as in Experiment 1, with the following exceptions. 32 volunteers were tested, with 8 subjects in each group (Table 1). A tonotopic misalignment of 3mm was simulated. The test sentences used in Test Sessions 2 and 3 were counterbalanced across subjects.

B. Results

1. IEEE Sentence test

Figure 3 shows the percentages of key words in IEEE sentences reported correctly according to test session for Groups 1 and 3 combined, and Groups 2 and 4 combined. On average, 50.1% of key words were identified correctly at baseline. Both auditory (t_{31} = 9.70,

$p<0.001$) and control training ($t_{31} = 3.67$, $p<0.001$) were associated with significant improvements in the ability to identify words in sentences (Figure 1, Panel D; Table 2). Auditory training led to significantly larger improvements than control training ($t_{31} = 3.38$, $p<0.01$). High-Variability auditory training was associated with a significantly larger improvement in the percentage of key words correctly identified than Single-Talker auditory training ($t_{30} = 2.38$, $p<0.05$, Figure 4; Table 2).

[FIGURE 3]

[FIGURE 4]

An independent samples t-test on the improvement in accuracy of identifying words spoken by ‘old’ talkers revealed no significant difference between the effectiveness of High-Variability (mean improvement = 9.9%pts, 95% c.i. 5.7 to 14.1%pts) and Single-Talker (mean improvement = 5.9%pts, 95% c.i. 0.8 to 10.9%pts; $t_{30} = 1.31$, $p = 0.20$) auditory training. There was some evidence of a difference between the effectiveness of High-Variability (mean improvement = 13.1%pts, 95% c.i. 10.0 to 16.2%pts) compared with Single-Talker (mean improvement = 8.6%pts, 95% c.i. 5.0 to 12.3%pts) auditory training when talkers were ‘new’ ($t_{30} = 2.03$, $p = 0.051$).

2. Consonant test

There were significant improvements in consonant discrimination following both auditory ($t_{31} = 8.63$, $p<0.001$) and control training ($t_{31} = 2.52$; Figure 1, Panel E; Table 2). However, auditory training led to significantly larger improvements than control training ($t_{31} = 2.69$, $p<0.05$). The difference between the effectiveness of High-Variability and Single-Talker auditory training just failed to reach significance ($t_{30} = 1.89$, $p=0.069$; Table 2).

3. Vowel test

Vowel discrimination improved significantly following auditory (t_{31} = 6.60, $p<0.001$) and control training (t_{31} = 4.31, $p<0.001$; Figure 1, Panel F; Table 2). There was no significant difference between the effectiveness of auditory and control training (t_{31} = 1.63). The improvements following High-Variability and Single-Talker auditory training did not differ significantly (t_{30} = 0.36, p=0.72; Table 2).

C. Discussion

Experiments 1 and 2 show that training-related improvements in performance occur both when baseline performance is poor (≈10%, Experiment 1) and better (≈50%, Experiment 2). Perceptual learning in Experiment 2 accounted for a significant improvement of 5.8%pts in ability to identify words in sentences, compared with 5.9%pts in Experiment 1. In addition, Experiment 2 found that auditory training led to a significantly larger improvement in the ability to discriminate consonant sounds than control training. As in Experiment 1, auditory training was not significantly more effective than the control task in improving the ability to discriminate vowel sounds.

In contrast to Experiment 1, Experiment 2 found an advantage for High-Variability training over Single-Talker auditory training. The High-Variability auditory training task produced an improvement of 12%pts in the ability to perceive words in sentences, compared with an improvement of 7%pts following Single-Talker training. The advantage for High-Variability over Single-Talker auditory training was stronger when talkers were ‘new’ than when talkers were ‘old’. This result is compatible with earlier findings (Lively *et al*., 1993) that High-Variability auditory training can lead to greater transfer to novel talkers than Single-Talker training. It is possible that this result emerged in Experiment 2 but not in Experiment 1, because the degree of tonotopic misalignment simulated in Experiment 2 preserved more cues that distinguish talkers.

V. Experiment 3

Experiment 3 sought to establish whether larger improvements could be achieved with a different training task, which required subjects to discriminate words in sentences. We reasoned that a sentence training task might improve performance on a sentence test more than a word training task because there is evidence that auditory training generalizes best when training and test materials are similar (Greenspan *et al*., 1988; Hirata, 2004).

We retained the comparison between High-Variability and Single-Talker training because we expected there to be larger differences between talkers when they articulated entire sentences compared with single words. Accordingly, we hypothesized that a sentence training task including several talkers would be more effective than a task including only one talker. In Experiments 1 and 2, improvements in speech perception were assessed immediately following auditory training. In the present experiment, an additional testing session was included, approximately 2 weeks after the final training session, to establish whether learning was sustained over time. A spectral shift of 6mm rather than 3mm was simulated to avoid possible ceiling effects with male talkers, given that performance with the most intelligible male talker reached 91% key words correct in Experiment 2.

A. Method

1. Subjects & speech processing

16 volunteers took part. A 6mm tonotopic misalignment was simulated.

2. Training materials

Each trial of the auditory training task began with an acoustic presentation of the target sentence. Six orthographically presented words then appeared in random positions on the computer screen. Subjects were instructed to select the three words from this set which were present in the target sentence. Visual feedback on accuracy was given, with a green check indicating that the subject had selected a word which was in the sentence, and a red cross indicating that the subject had selected a word that was not present. If subjects selected a word which was not in the sentence, the sentence was presented again acoustically. Once all three target words had been selected, the target sentence was displayed orthographically at the top of the screen. Subjects were asked to study the sentence. Finally, the sentence was presented acoustically once more. Subjects were asked to listen carefully to the sentence, and attempt to pick out words in the sentence that they now knew were present. The aim was to maximize the amount of lexical feedback that subjects received. This protocol is analogous to the Distorted-Clear-Distorted (DCD) protocol which was found to maximize learning to perceive noise-vocoded speech by Davis *et al*. (2005). Our implementation includes an additional intervening task which allows performance to be monitored, and which maintains subjects' engagement. The control task was presented in the same format, except that the target sentences were presented orthographically, degraded by visual noise. The training and control tasks took between 1.5 and 2 hours to complete.

Three-hundred IEEE sentences that were not used as testing materials were selected as training sentences. Three words in each sentence were selected to be target words. We selected target words which were not highly semantically related, for example, in the sentence 'He wrote his last novel at this inn.', 'Wrote', 'Last' and 'Inn' were selected as target words, thus avoiding the semantically related words ‘Wrote’ and ‘Novel’ both being selected as targets. One foil was created for each target word, so as to form a quasi-minimal pair with the target (e.g. ‘Note’, ‘List’, and ‘It’ were selected as foils for the sentence above). For the High-Variability training condition, 10 talkers (4 male, 4 female, 2 children) recorded the 300 sentences, with each talker recording 30 sentences. A single male talker with a southern British accent recorded all 300 sentences for the single-talker condition.

4. Design, procedure, & analysis

There were four groups of subjects, with four subjects in each group. The design and procedure were the same as in Experiment 1 (Table 1), but with the addition of a fourth testing session (Test Session 4) which took place 9-18 days (median 13 days) after Test Session 3. In this final session subjects received no training. They just completed the tests of speech perception. The results were analyzed in the same way as in Experiment 1.

B. Results

1. IEEE Sentence test

Figure 5 shows the percentage of key words correctly identified in Test Sessions 1, 2, 3, and 4, according to whether subjects received auditory training or control training first. The figure suggests that improvements in performance followed auditory training, and that the level of performance was sustained for both groups over the 2-week interval between Test Sessions 3 and 4. There was a significant improvement following auditory training (t_{15} = 9.98, $p<0.001$), while the improvement following control training just failed to reach significance at the <0.05 level (t_{15} = 2.12, $p=0.051$; Figure 1, Panel G; Table 2). The improvement following auditory training was significantly larger than the improvement following control training (t_{15} = 3.64,

$p<0.01$). There was no significant difference between the effectiveness of High-Variability and Single-Talker auditory training (t_{14} = -1.87; Table 2).

[FIGURE 5]

Five of the ten talkers who recorded the test of sentence perception also recorded the training materials, giving five 'new' talkers and five 'old' talkers. There was no evidence that transfer of learning differed between the High-Variability and Single-Talker groups according to whether performance with new (t_{15} = -1.74) or old talkers (t_{15} = -1.31) was analyzed.

2. Consonant test

Auditory training was not associated with an improvement in the ability to discriminate consonant sounds (t_{15} = 1.29), but control training was (t_{15} = 3.99, $p<0.01$; Figure 1, Panel H; Table 2). Control training was not significantly more effective than auditory training however (t_{15} = 1.86). There was no significant difference between the effectiveness of High-Variability and Single-Talker auditory training (t_{14} = 0.85; Table 2).

3. Vowel test

Significant improvements in vowel discrimination followed both auditory (t_{15} = 2.39, $p<0.05$) and control training (t_{15} = 3.33, $p<0.01$; Figure 1, Panel I; Table 2). There was no significant difference between auditory and control training (t_{15} = -0.31), and no significant difference between High-Variability and Single-Talker auditory training (t_{14} = 0.54; Table 2).

C. Discussion

The results are partly consistent with those of Experiments 1 and 2. Sentence-based auditory training was significantly more effective than a visual control task in improving the ability to identify words in sentences. The average improvement of 7.5%pts which could be attributed to perceptual learning was comparable to the average improvements of 5.9%pts and

5.8%pts found in Experiments 1 and 2, respectively. Improvements in sentence perception were sustained over a two-week period during which no additional training was provided. Changes in mappings between acoustic-phonetic information and linguistic knowledge were therefore ‘relatively long-lasting’ – a requirement for perceptual learning to have occurred (Goldstone, 1998).

In contrast with Experiment 2, auditory training was no more effective than control training in improving the discrimination of consonants in nonsense syllables. This result may have occurred because the word-based training regime used in Experiment 2 provides more directed training in the discrimination of consonant sounds, being based on quasi-minimal pairs of words, the majority of which differed in terms of the consonant sound.

High-Variability auditory training was not more effective than Single-Talker training. As discussed in Experiment 1, this result may have arisen because the 6mm tonotopic misalignment was too severe to permit differentiation between talkers.

VI. Experiment 4

Experiments 1 and 2 demonstrated that word-based auditory training was more effective than a visual control task in improving the ability to identify words in sentences and consonants in nonsense syllables. Experiment 4 examined whether performance continued to improve if subjects repeated the word-based High-Variability auditory training task 4 times, for a total of approximately 4 hours of training.

A. Method

Four volunteers participated. They each completed four sessions of High-Variability word-based training, providing a total of 4,800 training trials. As in Experiment 1, during Test Session 1, subjects completed baseline tests of speech perception. Test Session 2 took place on the following day, and subjects completed the auditory training task followed by further tests of speech perception. On subjects' third and fourth visits they repeated the training task. In their fifth visit they completed the training task followed by further tests of speech perception (Test Session 3). The third to fifth visits were scheduled to take place at any time within a two-week period. The test sentences used in Test Sessions 2 and 3 were counterbalanced across subjects. Results were analyzed using repeated measures Analyses of Variance. A Greenhouse-Geisser correction was applied if the assumption of sphericity was violated (indicated by non-integer degrees of freedom). Planned comparisons were carried out using t-tests with a Bonferroni correction.

B. Results

1. IEEE Sentence test

Performance improved significantly over Test Sessions 1, 2, and 3 ($F_{1.0,3.1} = 60.32$, $p<0.01$; Figure 1, Panel J). In Test Session 1, subjects correctly reported 5.3% of words in sentences. There was a significant improvement of 7.1%pts following one training run between Test Sessions 1 and 2 (*planned* $t_3 = 5.34$, $p<0.05$) and of 7.4%pts following three further training runs between Test Sessions 2 and 3 (*planned* $t_3 = 14.60$, $p<0.001$; Figure 6, Panel A).

[FIGURE 6]

2. Consonant test

Performance on the consonant test also improved significantly over sessions ($F_{2,6} = 25.25$, $p<0.01$; Figure 1, Panel K). Initially, subjects identified 32.0% of consonant sounds correctly. There was an improvement in performance of 6.9%pts between Test Sessions 1 and 2, which failed to reach significance (*planned* $t_3 = 2.89$) and an improvement of 5.9%pts between Test Sessions 2 and 3 which just failed to reach significance (*planned* $t_3 = 3.98$, $p = 0.056$; Figure

6, Panel B). There was a significant improvement of 12.8%pts between Test Sessions 1 and 3 (*planned* $t_3 = 9.36$, $p<0.01$).

3. Vowel test

The ability to identify vowel sounds improved over sessions ($F_{2,6} = 14.08$, $p<0.01$; Figure 5, Panel C). There was a significant improvement of 12.6%pts between Test Sessions 1 and 3 (*planned* $t_3 = 6.50$, $p<0.05$, Figure 1, Panel L), but no significant improvements between adjacent test sessions.

C. Discussion

Experiment 4 established that performance continues to improve if additional auditory High-Variability training sessions are provided following an initial 1-hour session. Accuracy of identification of words in sentences improved from 5.3% to 19.8% correct after 4 training sessions. The improvement of 7.1%pts following the first training session is comparable to the improvement of 6.3%pts displayed by the equivalent group of participants in Experiment 1 who received a single session of High-Variability auditory training between Test Sessions 1 and 2. A similar pattern was found in the tests of consonant and vowel discrimination. Improvements after one training session were similar to those found in Experiment 1. The improvements roughly doubled following three more training sessions.

VII. Experiment 5

The aim of Experiment 5 was to establish whether additional sentence-based training sessions would produce further improvements after the initial session. As in Experiment 4, subjects repeated the High-Variability auditory training task 4 times, for a total of 6-7 hours of training.

A. Method

The methods were the same as those described in Experiment 4, except that the sentence-based training task was administered. Subjects completed a total of 1,200 training trials, with the same 300 sentences being presented during each of the 4 training sessions.

B. Results

1. IEEE Sentence test

A one-way ANOVA revealed that performance differed between Test Sessions 1, 2, and 3 ($F_{2,6}$ = 17.07, $p<0.001$; Figure 7, Panel A). Performance improved from 10.8% correct in Test Session 1, to 27.0% correct in Test Session 3. There was an overall significant (*planned* t_3 = 4.72, $p<0.05$) improvement of 16.3%pts between Test Sessions 1 and 3 (Figure 1, Panel M), but no significant improvement from Test Sessions 1 to 2, or 2 to 3.

[FIGURE 7]

2. Consonant test

Performance differed significantly between test sessions ($F_{2,6}$ = 13.16, $p<0.01$). Performance improved from 43.6% correct in Test Session 1, to 59.9% correct in Test Session 3. After adjusting for multiple comparisons, there was no significant improvement between adjacent test sessions, although there was a significant improvement of 16.3%pts between Test Sessions 1 and 3 (*planned* t_3 = 4.08, $p<0.05$; Figure 1, Panel N).

3. Vowel test

The percentage of vowel sounds identified correctly improved from 30.0% in Test Session 1 to 38.6% in Test Session 3. A one-way ANOVA on the performance in Test Sessions 1, 2, and 3 did not reveal a significant main effect ($F_{2,6}$ = 1.75). There were no significant improvements between adjacent test sessions, and no overall significant improvement between Test Sessions 1 and 3 (Figure 1, Panel O).

C. Discussion

Performance on the IEEE sentence test continued to improve with successive training sessions. The average improvement in the ability to identify words in sentences was 16.3%pts, with a range of improvements across subjects of 7.5 to 23.0%pts. The improvement of 23%pts is the largest found in any of the present series of experiments, and approach the gains found in other investigations that have used one-to-one training (e.g. Rosen *et al*., 1999; Faulkner *et al*., 2006). There was additionally a significant improvement of 16.3%pts in consonant discrimination by the end of training, but no significant improvement in vowel discrimination. As Experiment 3 found no significant difference between the effectiveness of sentence-based auditory training compared with the control task on the test of consonant discrimination, we cannot be confident that the improvement in consonant discrimination emerged because of auditory training, rather than because of repeated testing.

VIII. General discussion

The present study examined whether computer-based auditory training could lead to improvements in the ability of normally-hearing listeners to perceive spectrally-distorted speech that had been processed by a simulation of a cochlear-implant system. The five experiments have revealed that word- and sentence-based approaches to training lead to significant improvements in the ability to identify words in sentences, and these improvements are significantly larger than those that follow a visual control task. We have therefore extended Fu *et al*.'s (2005[b]) findings by showing that both word- and sentence-based training lead to significant improvements in the ability to identify words in sentences. Extending the amount of auditory training beyond the initial hour led to continued improvements in the ability to identify words in sentences, ranging between 8 and 23%pts. These improvements approach those reported by Rosen *et al*. (1999) who provided one-to-one training using CDT, although they are on average somewhat smaller. However, the evidence suggests that performance would have continued to improve if more training had been provided. (Footnote 2).

A further significant finding was that there was an advantage for High-Variability auditory training over Single-Talker auditory training in Experiment 2, which simulated a 3mm tonotopic misalignment. The results are compatible with the hypothesis (Lively *et al*., 1993) that High-Variability training generalizes better to new talkers than Single-Talker training. No significant differences between High-Variability and Single-Talker training emerged in Experiments 1 and 3, possibly because the simulation of a 6mm tonotopic misalignment, with 8 channels of information, was too extreme to permit subjects to differentiate between talkers (Chinchilla-Rodriguez *et al*., 2004).

Difficulty training vowel discrimination

An advantage for word-based training over the visual control task was found for consonant discrimination, but not for vowel discrimination. It seems therefore that vowel discrimination is more difficult to train than consonant discrimination. This might be because the speech processing preserved the predominantly temporal cues required for consonant discrimination to a greater extent than the predominantly spectral cues required for vowel discrimination. Vowel discrimination also depends on the interaction between formant frequencies and the fundamental frequency (f_0; Assman, Nearey, & Scott, 2002; Assman & Nearey, 2003). Assman and Nearey (2003) reported that vowel discrimination was adversely affected when f_0 remained constant and formants were shifted upward or downward in frequency. The decline in performance was counteracted however, when corresponding upward or downward shifts in f_0 were introduced. f_0 information is not strongly preserved in noise-vocoded speech, particularly in the speech of women and children, possibly contributing to the difficulties reported here.

Training-related versus incidental learning

In order to dissociate training-related learning from incidental learning, improvements following auditory training were compared with improvements following a ‘matched’ visual control task. The visual control task was matched with the auditory training task in terms of its content and task demands. In addition, the visual stimulus was impoverished somewhat to mirror the loss of spectral information in the auditory stimuli. However, the visual stimulus underwent no further distortion, and the consequences of spectral shift were not recreated in the visual control task. The visual control task might have acted as a more rigorous control condition if the consequences of both the loss of spectral detail and spectral shifting could have been recreated in the visual modality.

The suggestion that the auditory and visual control tasks could have been more fully matched is supported by differences in the level of performance between the auditory training task and the visual control task. Performance reached 98% correct in the visual control task in Experiment 1, compared with 90% in the auditory training task. Potentially, differences between the effectiveness of auditory compared with control training were simply due to differences in the difficulty of the training task. However, this suggestion is contradicted by the finding that there was no significant difference between the effectiveness of High-Variability and Single-Talker training in Experiment 1, despite the fact that performance averaged 85% correct in the High-Variability training task, compared with 94% correct in the Single-Talker training task.

Although we found that auditory training led to significantly larger improvements in speech perception than the visual control task, ‘incidental learning’ also contributed to improvements in performance. Figures 2, 3, and 5 show that the extent to which performance improved following the control task depended on the order in which the auditory training and visual control tasks were completed. Subjects who were exposed to the auditory training task first displayed a significant improvement between Test Sessions 1 and 2, but no further significant improvement between Test Sessions 2 and 3 following control training. In contrast, subjects who were exposed to the control task first displayed an improvement between Test Sessions 1 and 2, and then a further improvement between Test Sessions 2 and 3 following auditory training. This pattern of results is compatible with the idea that two processes contributed to the improvements in sentence perception. The first process is related to auditory training, while the second is related to incidental learning. Incidental learning happened early in the experiments, occurring between Test Sessions 1 and 2, but not between Test Sessions 2 and 3. Accordingly, both training-related and incidental learning caused the improvements between Test Sessions 1 and 2 for the group who received auditory training first. For the group who received control training first, it is likely that the improvement between Test Sessions 1 and 2 can be attributed to incidental learning, and that the improvement between Test Sessions 2 and 3 can be attributed to auditory training.

The present study sought to dissociate training-related learning from incidental learning in order to measure the effectiveness of our training tasks. If the contribution of incidental learning was large, and the contribution of training-related learning was very small, there would be little point in asking subjects to complete an extensive amount of auditory training. Arguably however, the more interesting distinction is between learning which generalizes across speech tests and translates into improvements in everyday life, versus learning which does not generalize. Given that this is a simulation study, and thereby can afford no advantages in terms of subjects' everyday lives, it was not possible to dissociate these two types of learning within the present experiment. In further work with cochlear-implant users, we are evaluating the effectiveness of auditory training using a questionnaire alongside tests of speech perception. Patients use the questionnaire to report the extent to which training has benefited them in everyday life. In this way, we shall determine whether training produces learning that generalizes beyond laboratory tasks.

Basis of improvements following training

There are three potential explanations for why performance improved following auditory training. First, it is possible that subjects learned to remap the novel auditory sensations onto their existing linguistic knowledge. Cochlear-implant users tend to exhibit quite marked improvements in speech perception during the first few months of implant use, and may continue to improve up to two years post implantation (Tyler & Summerfield, 1996; Tyler *et al.*, 1997). Evidence suggests that this improvement is driven primarily by improvements in the ability to map the novel sensations provided by a cochlear implant onto existing linguistic knowledge (Svirsky *et al.*, 2001, 2004). It is possible that a similar process underlies the improvements reported here. The word and sentence training tasks led to similar overall levels of improvement in the ability to identify words in sentences. It therefore seems that subjects abstracted general information about the mapping between acoustic properties and phonetic and/or lexical information. If performance had improved more following sentence training than following word training, this might have suggested that the cognitive skill required for sentence perception had improved, rather than a general improvement in the relationship between acoustic input and existing representations. Second, it is possible that subjects learned to hear differences between similar sounds. Goldstone (1998) describes differentiation as one of the major mechanisms of perceptual learning, explaining ‘stimuli that were once psychologically fused together become separated. Once separated, discriminations can be made between stimuli that were originally indistinguishable’ (p 596). This mechanism is thought to underlie the improvement in the ability of Japanese listeners to discriminate /r/ and /l/ (e.g. Lively *et al*., 1993). Third, it is possible that a component of the improvement following training arises from a general improvement in listening skills and in the ability to attend to auditory information. However, if a general improvement in listening skills was solely responsible for improvements following training, we would have expected improvements in all tests of speech perception.

Implications for cochlear-implant users

A motivation for this study was to establish whether self-administered computer-based training techniques would improve the ability of normally-hearing subjects to perceive spectrally-distorted speech, before evaluating whether the training procedures lead to improvements in speech perception amongst cochlear-implant users. The reasoning was that in order for a training regime to be valuable for cochlear-implant users, it should improve the ability of normally-hearing subjects to perceive speech which is reduced in spectral detail and is frequency shifted. The results indicate that our training procedures are successful in improving speech perception in normally-hearing subjects, and they also suggest that training materials should be recorded by several talkers. However, these results do not necessarily mean that the training procedures will improve speech perception amongst implantees. While cochlear-implant users listen to the distorted input during everyday communication and have the opportunity to learn the relationship between motor speech activity and the resulting auditory sensations, the normally-hearing listeners in our experiments only ever heard the distorted signal within the laboratory and they never hear their own speech through the simulation. Accordingly, the demonstration of learning with normally-hearing subjects can be considered a necessary, but insufficient, condition for auditory training to be successful with cochlear-implant users. However, despite this qualification, similar auditory training approaches have been associated with improved outcomes amongst adult users of cochlear implants (Fu *et al*., 2005[a]).

Conclusion

The present experiments show that self-administered computer-based auditory training regimes lead to improvements in the ability of normally-hearing listeners to perceive spectrally-distorted speech, even when exposure to the training vocabulary and test materials is controlled. The results suggest that the word- and sentence-based training tasks have the potential to improve the perception of speech by adult users of cochlear implants, and also suggest that training packages should contain materials spoken by multiple talkers. We are currently investigating whether these training materials lead to benefits for implantees.

IX. ACKNOWLEDGMENTS

Paula Stacey was supported by a studentship from the (UK) Medical Research Council. Experiment 1 was conducted at the MRC Institute of Hearing Research, Nottingham UK. Experiments 2 to 5 were conducted at the University of York where this paper was written. We thank John Chambers, Tim Folkard, and Silvia Cirstea for programming assistance, Dave Moore and Peter Bailey for support, and two reviewers for instructive criticisms which led to improvements in the paper.

X. Footnotes

Footnote 1

The size of changes in %-correct observed following training may depend on baseline performance levels, meaning that an improvement of 20%pts from 0% correct should not be equated to an improvement of 20%pts from 40% correct. Three transforms may reduce this problem; 1) differences in arcsine-transformed percentages, 2) percent reduction of error, or 3) calculating the ratio of the logarithms of error probabilities. We applied each of these transforms to our data, but found no difference in the pattern of results.

Footnote 2

The relationship between baseline performance and the overall improvement in performance was examined in Experiments 1, 2, and 3 for each outcome measure (sentences, consonants, and vowels). Four of the nine correlations were significant, with poorer performers at baseline showing larger improvements. However, there was much variability. Participants at the same baseline level showed improvements ranging from less than 10%pts to more than 20%pts, suggesting that baseline performance is only one of the variables that influence the size of the benefit following training.

XI. References

Amitay, S., Irwin, A., & Moore, D.R. (**2006**). "Discrimination learning induced by training with identical stimuli." Nat Neurosci, **9**, 1446-1448.

Assman, P. F., & Nearey, T. M. (**2003**). “Frequency shifts and vowel identification.” In M. J. Sole, D. Recasens, & J. Romero (Eds.), *Proc 15th Int Cong Phon Sci*, Barcelona, Spain, 1-4.

Assman, P. F., Nearey, T. M., & Scott, J. M. (**2002**). “Modelling the perception of frequency-shifted vowels.” *Proc 7th Int Con Spoken Lang Proc*, 425-428.

Baskent, D., & Shannon, R. V. (**2003**). “Speech recognition under conditions of frequency-place compression and expansion.” J Acoust Soc Am, **113**, 2064-2076.

Bradlow, A. R. and Bent, T. (**2003**). "Listener adaptation to foreign accented English." In M. J. Sole, D. Recasens, & J. Romero (Eds.), *Proc 15th Int Cong Phon Sci*, Barcelona, Spain, 2881-2884.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production." J Acoust Soc Am, **101**, 2299-2310.

BSA (**1981**). "Recommended procedures for pure-tone audiometry using a manually operated instrument." Br J Audiol, **15**, 213-216.

Busby, P. A., Roberts, S. A., Tong, Y. C., & Clark, G. M. (**1991**). "Results of speech perception and speech production training for three prelingually deaf patients using a multiple-electrode cochlear implant." Br J Audiol, **25**, 291-302.

Chinchilla-Rodriguez, S., Nogaki, G., & Fu, Q.-J. (**2004**). "Relative contribution of spectral and temporal cues and spectral profile to voice gender discrimination." J Acoust Soc Am, **116**, 2544-2545.

Clopper, C. G., & Pisoni, D. B. (**2004**). "Effects of talker variability on perceptual learning of dialects." Lang Speech, **47**, 207-239.

Davis, M. H., Hervais-Adelman, A., Taylor, K., McGettigan, C., & Jonsrude, I. S. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences." J Exp Psychol Gen, **134**, 222-241.

Dawson, P. W., & Clark, G. M. (**1997**). "Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant." Ear Hear, **18**, 488-501.

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

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

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

Fu, Q.-J., Galvin, J., Wang, X., & Nogaki, G. (**2005[a]**). “Moderate auditory training can improve speech performance of adult cochlear implant patients.” Acoust Res Lett Online, **6**, 106-111.

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

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

Gagne, J. P., Parnes, L. S., LaRocque, M., Hassan, R., & Vidas, S. (**1991**). “Effectiveness of an intensive speech perception training program for adult cochlear implant recipients.” Ann Otol Rhinol Laryngol, **100**, 700-707.

Gantz, B. J., Woodworth, G. G., Abbas, P. J., Knutson, J. F., & Tyler, R. S. (**1993**). “Multivariate predictors of audiological success with multi-channel cochlear implants.” Ann Otol Rhinol Laryngol, **102**, 909-916.

Gfeller, K., Witt, S., Ademek, M., Mehr, M., Rogers, J., Stordahl, J., & Ringgenberg, S. (**2002**). “Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients.” J Am Acad Audiol, **13**, 132-145.

Goldstone, R. L. (**1998**). “Perceptual learning.” Annu Rev Psychol, **49**, 585-612.

Gonzalez, J., & Oliver, J. C. (**2005**). “Gender and speaker identification as a function of the number of channels in spectrally reduced speech.” J Acoust Soc Am, **118**, 461-470.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (**1988**). “Perceptual learning of speech produced by rule.” J Exp Psychol Learn Mem Cogn, **14**, 412-433.

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

Hirata, J. (**2004**). “Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts.” J Acoust Soc Am, **116**, 2384-2394.

IEEE. (**1969**). *IEEE Recommended Practice for Speech Quality Measurements*. IEEE, New York.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (**1993**). “Training Japanese listeners to identify English /r/ and /l/. II. The role of phonetic environment and talker variablility in learning new perceptual categories.” J Acoust Soc Am, **94**, 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (**1994**). “Training Japanese listeners to identify English /r/ and /ll/. III. Long-term retention of new phonetic categories.” J Acoust Soc Am, **94**, 2076-2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (**1991**).” Training Japanese listeners to identify English /r/ and /l/: A first report.” J Acoust Soc Am, **89**, 874-886.

Robinson, K., & Summerfield, A. Q. (**1996**). “Adult auditory learning and training.” Ear Hear, **17**, 51S-65S.

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

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

Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., Neely, J. G., Kletzer, 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.” J Assoc Res Otolaryngol, **3**, 332-350.

Summerfield, A. Q., & Marshall, D. H. (**1995**). *Cochlear Implantation in the UK 1990-1994*: Main Report. HMSO Books, London.

Svirsky, M. A., Silveira, A., Suarez, H., Neuburger, H., Lai, T. T., & Simmons, P. M. (**2001**). “Auditory learning and adaptation after cochlear implantation: a preliminary study of discrimination and labeling of vowel sounds by cochlear implant users.” Acta Otolaryngol, **121**, 262-265.

Svirsky, M. A., Silveria, A., Neuburger, H., Teoh, S.-W., & Suarez, H. (**2004**). “Long-term auditory adaptation to a modified peripheral frequency map”. Acta Otolaryngol, **124**, 381-386.

Tyler, R. S., Parkinson, A. J., Woodworth, G. G., Lowder, M. W., & Gantz, B. J. (**1997**). “Performance over time of adult patients using the Ineraid or Nucleus cochlear implant.” J Acoust Soc Am, **102**, 508-522.

Tyler, R. S., & Summerfield, A. Q. (**1996**). “Cochlear implantation: relationships with research on auditory deprivation and acclimatization”. Ear Hear, **17**, 38S-50S.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (**1999**). “Training American listeners to perceive Mandarin tones.” J Acoust Soc Am, **106**, 3649-3658.

Yukawa, K., Cohen, L., Blamey, P., Pyman, B., Tungvachirakul, V., & O'Leary, S. (**2004**). “Effects of insertion depth of cochlear implant electrodes upon speech perception.” Audiol Neurootol, **9**, 163-172.

Table 1: Design of Experiments 1, 2, and 3.

Group	Day 1	Day 2		Day 3	
1	Test Session 1	High-Variability auditory training	Test Session 2	Visual control training	Test Session 3
2	Test Session 1	Visual control training	Test Session 2	High-Variability auditory training	Test Session 3
3	Test Session 1	Single-Talker auditory training	Test Session 2	Visual control training	Test Session 3
4	Test Session 1	Visual control training	Test Session 2	Single-Talker auditory training	Test Session 3

Table 2: Mean (and 95% confidence intervals) improvements (in %pts) in the auditory training task (auditory) and the visual control task (control), and in the High-Variability and Single-Talker versions of the auditory training task.

		Auditory	Control	High-Variability auditory	Single-Talker auditory
Experiment 1	Sentences	7.98 (5.88 to 10.09)	2.02 (-0.36 to 4.39)	7.16 (3.21 to 11.10)	8.81 (6.23 to 11.40)
	Consonants	6.16 (3.79 to 8.52)	1.75 (-0.86 to 4.36)	7.13 (3.59 to 10.66)	5.19 (1.23 to 9.14)
	Vowels	6.91 (3.72 to 10.09)	2.88 (-1.20 to 6.95)	7.06 (2.80 to 11.33)	6.75 (0.81 to 12.69)
Experiment 2	Sentences	9.38 (7.41 to 11.36)	3.63 (1.61 to 5.65)	11.53 (8.67 to 14.40)	7.23 (4.67 to 9.80)
	Consonants	6.02 (4.59 to 7.44)	2.27 (0.43 to 4.10)	7.28 (5.20 to 9.36)	4.75 (2.79 to 6.71)
	Vowels	7.05 (4.87 to 9.23)	4.17 (2.20 to 6.14)	6.66 (2.44 to 10.87)	7.44 (5.54 to 9.33)
Experiment 3	Sentences	10.44 (8.21 to 12.67)	2.89 (-0.02 to 5.80)	8.63 (4.97 to 12.28)	12.25 (9.50 to 15.00)
	Consonants	2.03 (-1.33 to 5.39)	7.75 (3.61 to 11.89)	3.38 (-2.88 to 9.63)	0.69 (-3.49 to 4.87)
	Vowels	3.25 (0.35 to 6.15)	3.97 (1.43 to 6.51)	4.00 (-0.78 to 8.78)	2.50 (-2.04 to 7.04)

Figure Captions

Figure 1: Results of Experiments 1 to 5. Changes in accuracy of identifying key words, consonants, and vowels by individual subjects. Improvements following auditory training are plotted as filled circles and improvements following the control task are plotted as open circles. The average improvement of each group is indicated by the dashed line. The numbers below the abscissa indicate the number of subjects in each group.

Figure 2: Results of Experiment 1: Percentage of key words correctly identified in IEEE sentences according to test session and training group. The mean value is represented by the dashed line in the box, the median by the solid line. The box spans the inter-quartile range. Outliers are plotted as dots beyond the 10th - 90th percentile whiskers. Groups 1 and 3 received auditory training between Test Sessions 1 and 2 and visual control training between Test Sessions 2 and 3. The order was reversed for Groups 2 and 4.

Figure 3: Results of Experiment 2: Percentage of key words correctly identified in IEEE sentences according to test session and training group.

Figure 4: Results of Experiment 2: Improvement in the percentage of key words correctly identified following the auditory training and control tasks according to whether subjects received High-Variability or Single-Talker auditory training. Error bars indicate 95% confidence intervals.

Figure 6: Results of Experiment 3: Percentage of key words in IEEE sentences correctly identified according to test session and training group.

Figure 5: Results of Experiment 4: Overall performance on the Sentence Test (Panel A), the Consonant Test (Panel B) and the Vowel Test (Panel C) in Test Sessions 1, 2, and 3.

Figure 7: Results of Experiment 5: Overall performance on the Sentence Test (Panel A), the Consonant Test (Panel B) and the Vowel Test (Panel C) in Test Sessions 1, 2, and 3.

Figure 1

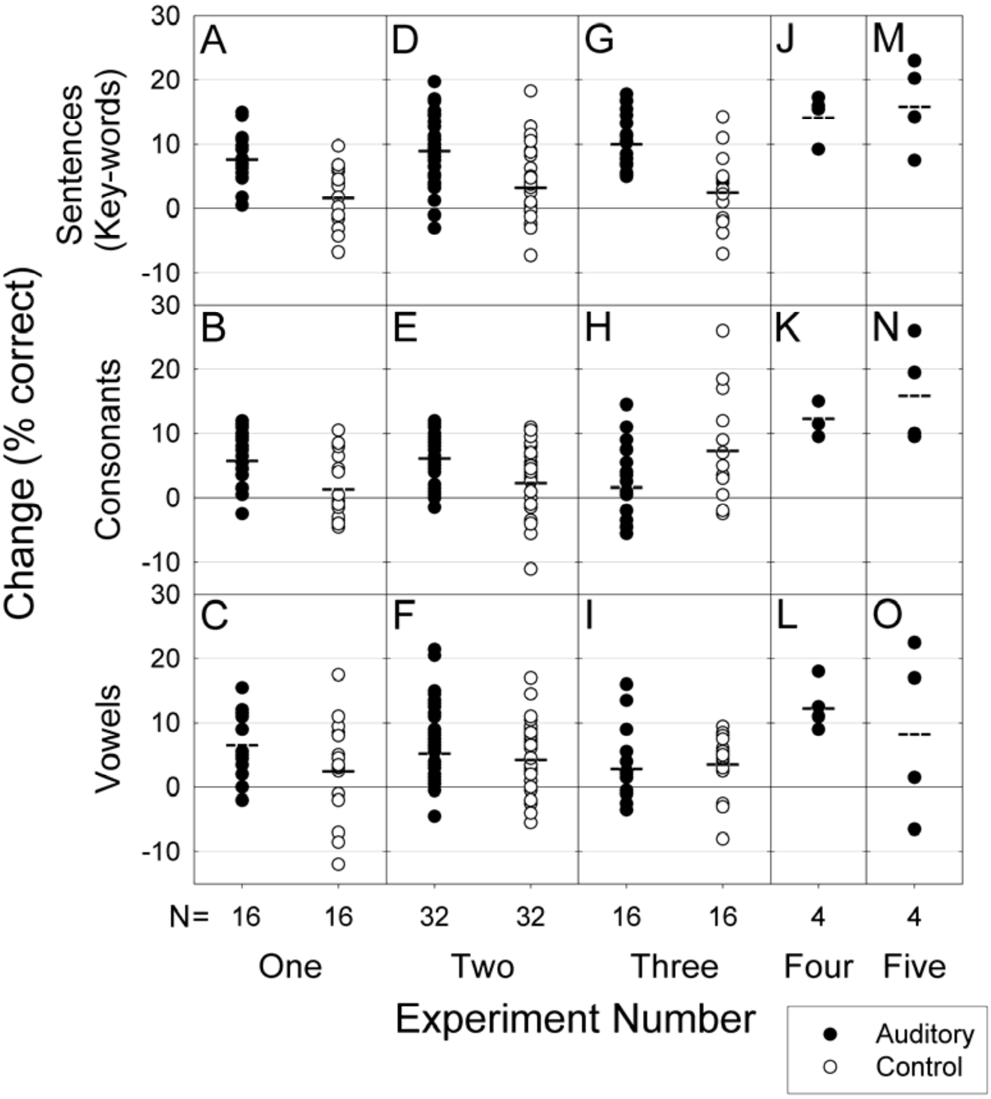

Change (% correct)
Sentences (Key-words)
Consonants
Vowels
30
20
10
0
-10
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
N=
16
16
32
32
16
16
4
4
One
Two
Three
Four
Five
Experiment Number
● Auditory
○ Control

Figure 2

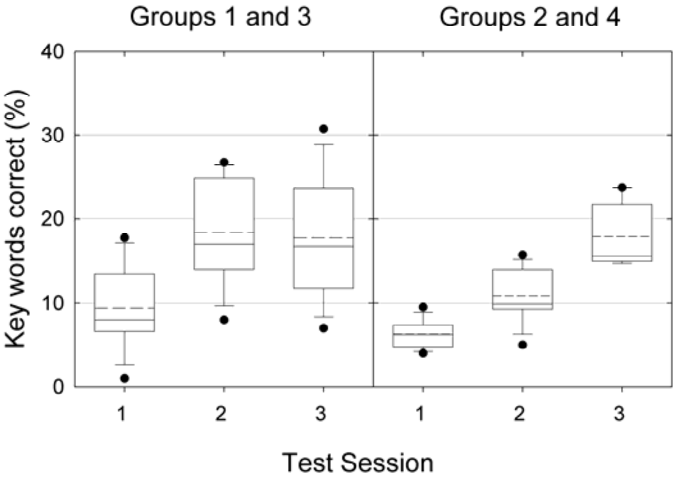

Groups 1 and 3
Groups 2 and 4
Key words correct (%)
40
30
20
10
0
1
2
3
1
2
3
Test Session

Figure 3

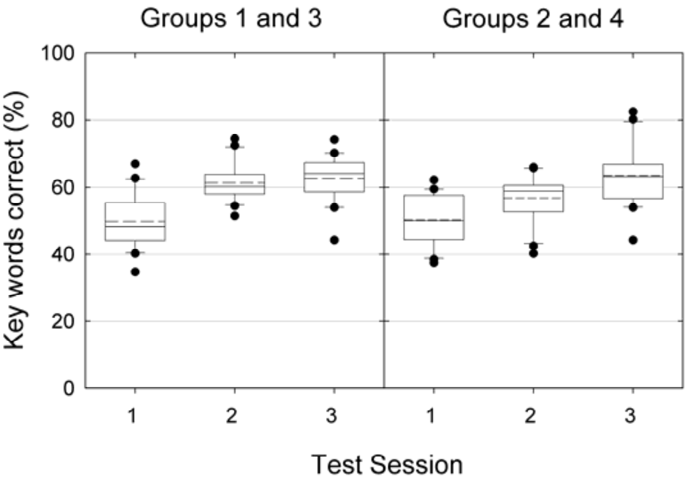

Groups 1 and 3
Groups 2 and 4
Key words correct (%)
100
80
60
40
20
0
1
2
3
1
2
3
Test Session

Figure 4

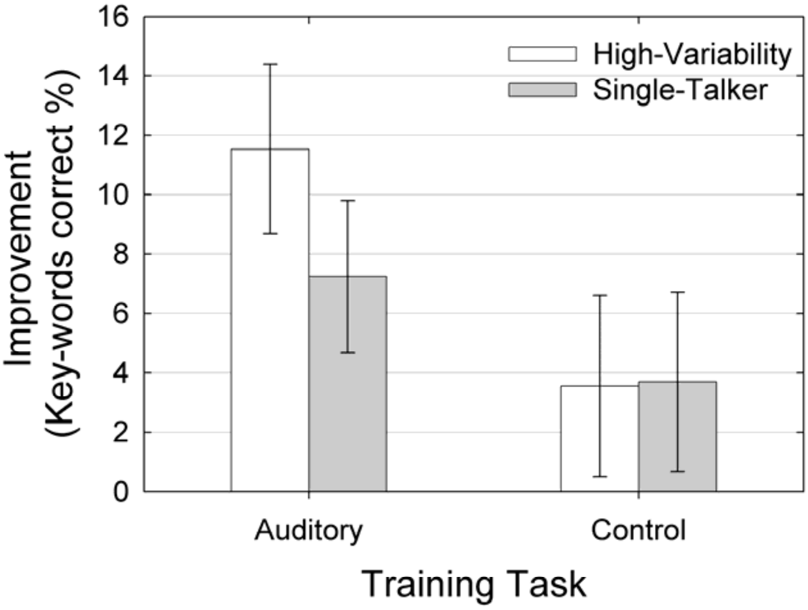

16
14
12
10
8
6
4
2
0
Improvement
(Key-words correct %)
High-Variability
Single-Talker
Auditory
Control
Training Task

Figure5

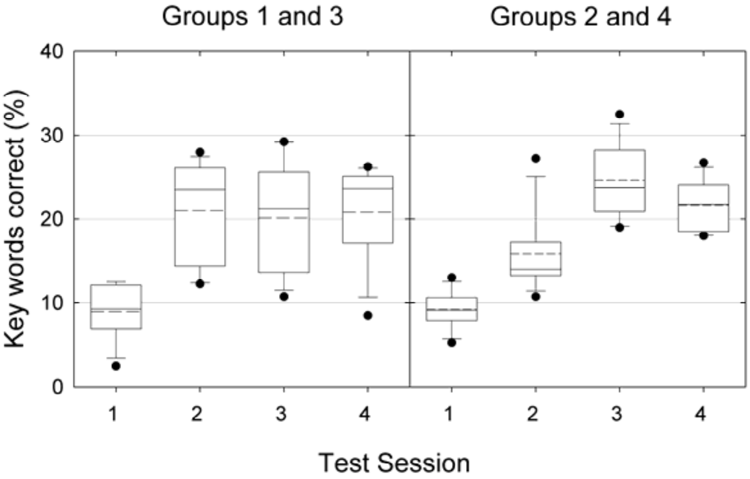

Groups 1 and 3
Groups 2 and 4
Key words correct (%)
40
30
20
10
0
1
2
3
4
1
2
3
4
Test Session

Figure 6

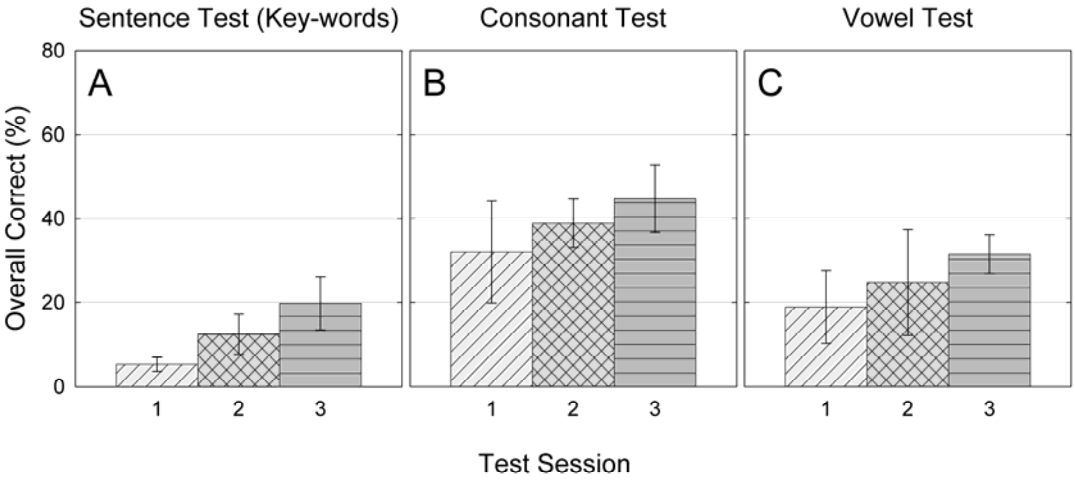
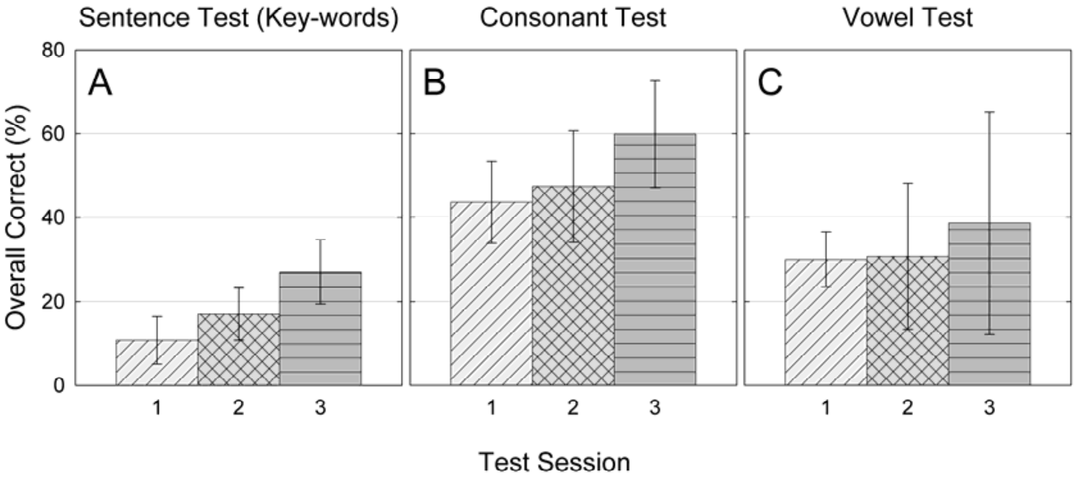

Sentence Test (Key-words)
Consonant Test
Vowel Test
A
B
C
Overall Correct (%)
0
20
40
60
80
1
2
3
Test Session

Figure 7


Sentence Test (Key-words)
Consonant Test
Vowel Test
A
B
C
Overall Correct (%)
0
20
40
60
80
1
2
3
Test Session