Running head: STRATEGIES FOR AUDITORY TRAINING

Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally-distorted speech

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

[1] Department of Psychology, University of York, Heslington, York YO10 5DD, UK and MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK.

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

Corresponding author:

Paula Stacey

Division of Psychology

Nottingham Trent University

Burton Street

Nottingham NG1 4BU

e-mail: paula.stacey@ntu.ac.uk

Abstract

Purpose: To compare the effectiveness of three self-administered strategies for auditory training that might improve speech perception by adult users of cochlear implants. The strategies are based, respectively, on discriminating isolated words, words in sentences, and phonemes in nonsense syllables.

Method: Participants were 18 normally-hearing adults who listened to speech processed by a noise-excited vocoder to simulate the information provided by a cochlear implant. They were assigned randomly to word-, sentence-, or phoneme-based training and underwent nine 20-minute training sessions on separate days over a 2-3-week period. The effectiveness of training was assessed as the improvement in accuracy of discriminating vowels and consonants, and identifying words in sentences, relative to participants' best performance in repeated tests prior to training.

Results: Word- and sentence-based training led to significant improvements in the ability to identify words in sentences that were significantly larger than the improvements produced by phoneme-based training. There were no significant differences between the effectiveness of word- and sentence-based training. No significant improvements in consonant or vowel discrimination were found for the sentence- or phoneme-based training groups, but some improvements were found for the word-based training group.

Conclusions: The word- and sentence-based training strategies were more effective than the phoneme-based strategy at improving the perception of spectrally-distorted speech.

Keywords: Auditory training, cochlear implants, perceptual learning

Comparison of word-, sentence-, and phoneme-based training strategies in improving the perception of spectrally-distorted speech

Auditory training can improve the accuracy of speech perception by several groups with communication difficulties, including language-learning impaired children (Tallal, Miller, Bedi, Byma, Wang, Nagarajan, Schreiner, Jenkins, & Merzenich, 1996), adult second-language learners (Lively, Logan, & Pisoni, 1993), and hearing-impaired adults (Walden, Erdman, Montgomery, Schwartz, & Prosek, 1981). In addition, evidence from Fu, Galvin, Wang, and Nogaki (2004, 2005[a]) suggests that computer-based auditory training can improve the accuracy of speech perception by adult users of cochlear implants. There is uncertainty, however, about how training materials for use by adult cochlear-implant users should be structured to achieve maximum effectiveness. The goal of the present study was to compare the effectiveness of three computer-based auditory training regimes that might be used to improve the speech-perception skills of adult users of cochlear implants.

Overall approach

Experiments on auditory training for users of cochlear implants could be conducted directly with patients, or by first evaluating the effectiveness of training procedures with normally-hearing participants. Like some other researchers (Faulkner, Rosen, & Norman, 2006; Fu, Nogaki, & Galvin, 2005[b]; Rosen, Faulkner, & Wilkinson, 1999), we chose to minimise the involvement of patients in evaluating ineffective strategies and we instead trained participants with normal hearing to discriminate speech which had been processed by a noise-excited vocoder (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). A noise-excited vocoder can be used to mimic the speech processing that occurs in a cochlear-implant system, and it allows

the spectral and temporal information that is transmitted to the listener to be manipulated. In addition, signals can be spectrally shifted to simulate the consequences of tonotopic misalignment between the frequency band transmitted by an electrode and the characteristic frequency of the location of that electrode. Tonotopic misalignment is one of the causes of poor speech perception by users of cochlear implants (Skinner, Ketten, holden, Harding, Smith, Gates, Neely, Kletzer, Brunsdn, & Blocker, 2002). The term 'spectrally-distorted speech' will be used to describe the resulting signal.

Strategies for auditory training

Rosen *et al*. (1999) evaluated the effectiveness of Connected Discourse Tracking (CDT, De Filippo & Scott, 1978) for improving the perception of spectrally-distorted speech. In CDT, an experimenter reads a passage of text, and the participant attempts to repeat verbatim what was said, with corrective feedback from the experimenter. The strategy improved the perception of spectrally distorted vowels, consonants, and sentences (see also Faulkner *et al*., 2006). However, CDT is labour intensive and therefore expensive to administer clinically. Self-administered training could be more cost-effective. Fu *et al*. (2005[b]) compared two such regimes. 'Word-based training' required participants to identify the vowels in consonant–vowel–consonant monosyllabic words; 'sentence-based training' was a computer-based CDT procedure. Fu *et al.* (2005[b]) compared the effectiveness of the two regimes in improving the ability of normally-hearing listeners to discriminate spectrally-distorted vowels and consonants. Word-based training led to significant improvements in the ability to discriminate both types of material, whilst sentence-based training led to significant improvements only in the ability to discriminate consonants. Fu *et al*. (2005[b]) concluded that word-based training might be more effective than sentence-

based training in improving the speech-perception skills of cochlear-implant users. However, Fu *et al.* (2005[b]) did not include a test of sentence perception as an outcome measure. This omission may limit the scope of their conclusions, given that training with sentences leads to larger improvements in sentence perception than training with isolated words (Greenspan, Nusbaum, & Pisoni, 1988; Hirata, 2004), and that the perception of words in sentences is more representative of everyday listening situations than the perception of isolated words. For this reason, we included tests of the ability to identify words in sentences as outcome measures in the present experiment.

The evidence reviewed above suggests that perceptual learning of speech can be specific to trained stimuli and training contexts. However, generalisation to untrained stimuli is important if auditory training is to be a useful therapeutic intervention for impaired populations. Some authors, including Moore, Rosenberg, and Coleman (2005), have argued that training to discriminate phonemes in nonsense syllables might be the optimal strategy for achieving generalisation to natural speech, because phonemes can be thought of as 'the building blocks of language'. Moore *et al.* (2005) evaluated the effectiveness of a phonetic training game, *Phonomena*[1], at improving phonological awareness in typically-developing children. Phonological awareness was measured with the Phonological Assessment Battery (Frederickson, Frith, & Reason, 1997), which includes tests of the ability to manipulate sounds in words, identify rhymes, and read non-words. *Phonomena* consists of eleven sets of synthetic speech sounds, each of which ranges across a series of intervening sounds from one consonant-vowel nonsense syllable or isolated vowel to another (for example, from "bee" to "dee", or "i" to "e"). Participants are trained to discriminate

[1] Mindweavers Ltd

members of each sound set. The difficulty of the training task varies adaptively to provide training at 'the edge of competence' (Moore *et al*., 2005). Moore *et al*. (2005) demonstrated that twelve 30-minute sessions of training using *Phonomena* were associated with significant, and large, improvements on the Phonological Assessment Battery by typically-developing children aged 8 to 10 years. Given the success of *Phonomena* in this context, we evaluated its effectiveness in improving the perception of spectrally-distorted speech.

Role of lexical information in perceptual learning of speech

Although Moore *et al*. (2005) demonstrated that phonetic training can lead to improvements in the ability to manipulate speech sounds, recent evidence shows that lexical information plays an important role in the perceptual learning of speech (Davis, Hervais-Adelman, Taylor, McGettigan, & Jonsrude, 2005; McQueen & Mitterer, 2005; Norris McQueen & Cutler, 2003). These results suggest that there may be advantages for word- and sentence-based training strategies over phoneme-based strategies. For example, Davis *et al*. (2005) showed that learning to identify words in noise-vocoded sentences is enhanced if participants receive lexical feedback. Learning was compared between two groups of participants. The first group heard a noise-vocoded sentence, followed by the sentence unprocessed, and finally the noise-vocoded sentence again (Distorted-Clear-Distorted, D_1CD_2). The second group heard a noise-vocoded sentence twice, followed by the same sentence unprocessed (Distorted-Distorted-Clear, D_1D_2C). The accuracy with which D_1 was identified improved in both groups from approximately 45% of words correct in the first block of 10 sentences, to approximately 64% of words correct in the second block of 10 sentences. However, learning was significantly greater in the D_1CD_2 group than the D_1D_2C group. Performance was equivalent between the groups for the very first

sentence that was presented (mean words correct: D_1CD_2 = 4.5%, D_1D_2C = 5.1%), but during the second block of 10 sentences the D_1CD_2 group reported 69% of words correctly, while the D_1D_2C group reported 58% of words correctly. Thus, repeated presentation of the original stimulus facilitated learning, provided that unambiguous feedback had already been provided. The same pattern was found with written feedback, but was not found when participants received training with non-words. These results are compatible with the idea that a top-down, lexically driven mechanism is involved in the perceptual learning of noise-vocoded speech. For this reason, we included lexical training regimes in the present experiment.

Controlling for incidental learning

Although improvements on tests of speech perception following auditory training may be caused by the training task itself ('training-related' learning), 'incidental learning' may also contribute. Incidental learning refers to improvements in performance on outcome tests that occur for reasons other than learning produced by the training task. Improvements might arise through procedural learning of the methods and requirements of tests (Robinson & Summerfield, 1996), or perceptual learning resulting from repeated exposure to test materials. It is difficult to be sure that incidental learning has been fully controlled. Some studies designed to improve the ability of normally-hearing participants to understand spectrally-distorted speech (e.g. Fu *et al*., 2005[a]) have sought to account for incidental learning by including control groups whose members undertook repeated tests of speech perception without undertaking training. However, evidence from Amitay, Irwin, and Moore (2006) suggests that it may not be sufficient to exercise control for incidental learning in this way. Amitay *et al*. (2006) found larger improvements in frequency discrimination by participants who played a purely visual computer game between successive tests than

by participants who did not engage in an intervening task. Possibly, maintaining attention and arousal, without explicit training, may be sufficient to induce improvements on auditory perceptual tasks. Thus, control groups may need to undertake forms of training as well as testing. That approach was adopted by Stacey and Summerfield (2007) to examine the short-term effects of auditory training, but is inefficient when an experimental design requires extensive amounts of training and testing.

As an alternative, the present study attempted to control for the effects of incidental learning by repeatedly administering tests of speech perception at baseline until an asymptote in performance was reached. A subsequent improvement in performance following auditory training was interpreted as evidence that training was effective. The design assumes that incidental learning is a short-term phenomenon that can occur fully in a single session in which participants are exposed to a large amount of spectrally-shifted speech. Although this procedure is likely to control for a large component of incidental learning, we cannot be certain that the effects of incidental learning have been completely eliminated. Therefore, hereafter we state that we have 'partially controlled' for incidental learning.

Aims of study

The present study had two main aims. Firstly, we investigated the effectiveness of word, sentence, and phonetic training strategies at improving the perception of spectrally-distorted speech. Tests of sentence, consonant, and vowel perception were included to investigate the generalisation of training to a range of tests of speech perception. Given the role of lexical information in the perceptual learning of speech, we hypothesised that larger improvements on tests of speech perception would be produced by word- and sentence-based strategies than by the

phonetic strategy. In addition, we expected sentence-based training to lead to larger improvements on tests of sentence perception than word-based training. Secondly, we investigated the extent to which incidental learning contributes to improvements in performance on tests of speech perception. We anticipated material improvements related to incidental learning, but we also expected significant training-related learning once incidental learning had been partially controlled.

Method

Participants

Participants were 18 students and staff from the University of York with normal hearing (≤25dB HL at octave frequencies between 250 and 8000Hz, inclusive, in both ears) measured according to British Society of Audiology (BSA) guidelines (BSA, 1981). All participants were native speakers of British English, and were aged between 18 and 28 years (median 19 years).

Speech recordings and presentation of stimuli

Speech recordings used as training materials and in tests of speech perception were recorded digitally (sample rate 44.1kHz, amplitude quantization 16 bits) in a carpeted double-walled sound attenuated chamber. Stimuli were presented to participants through an Audiomaster LS3/5A loudspeaker in a single-walled sound attenuated chamber. Sound levels were measured at the place occupied by the participant's head, with the participant absent. The mean peak level was 70dB(A), ranging between 65 and 75 dB(A) within training and test materials.

Training tasks

Word training task. 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

acoustically. Participants responded by touching the word corresponding to the target. Visual feedback on accuracy was given by a green tick or a red cross. If participants responded correctly, the next trial began. If their response was incorrect, the trial was repeated until the correct response was given. Repeating trials if participants responded incorrectly allowed them to hear the word again while knowing what the correct answer should be. This should facilitate the remapping between auditory sensations and linguistic knowledge.

To construct the training materials, 200 key words were selected from 40 IEEE sentences not used in the test of sentence perception. Three foils were created for each key word, forming quasi-minimal pairs. Most of the words were monosyllabic (e.g. hot, ship, sell), but some words were longer (e.g. shimmered, friendly). Materials were recorded by a single male talker with a southern British accent. There were 1200 training trials.

Sentence training task. Three-hundred IEEE sentences were used for training. These sentences were different from those used in the IEEE sentence test. Each trial of the sentence training task began with an acoustic presentation of the target sentence. Six words then appeared orthographically on the computer screen. Participants were instructed to select the three words from this set which were present in the target sentence. Visual feedback on accuracy was given by a green tick next to a selected word which was in the sentence, or a red cross next to a selected word that was not present. If participants selected a word which was not present, the sentence was presented again acoustically. Participants continued to select words until all three target words had been selected. Once all three target words had been selected, the target sentence was displayed orthographically at the top of the screen and participants were asked to study the sentence. Finally, the sentence was presented

acoustically once more. Participants were asked to listen carefully and to attempt to pick out words in the sentence that they now knew were present. The aim was to maximise the amount of lexical feedback that participants received. This training task is analogous to the Distorted-Clear-Distorted (D_1CD_2) condition which was found to facilitate learning by Davis *et al*. (2005), supplemented by an intervening task which allows performance to be monitored and which maintains participants' engagement. Materials were recorded by a single male talker with a southern British accent; this was the same talker who recorded the materials for the word training task.

Phonetic training task.

Description of task

The phonetic training task was based on *Phonomena* (Moore *et al*., 2005). *Phonomena* consists of 11 sets of sounds, each of which ranges across a series of intervening sounds from one syllable to another. The sets range either from one vowel to another (e.g. "i" to "e") or from one consonant-vowel syllable to another (e.g. "va" to "wa", or "sa" to "sha"; Table 1). The sets were designed to exemplify a wide range of the phonemic contrasts found in British English. At either extreme of a sound set is a synthesised example, derived from a naturally-spoken utterance. To create the sets, Moore *et al.* (2005) warped the extreme examples acoustically into one another in equal steps to create continua each consisting of 96 stimuli. Based on pilot testing, the sound set 'd_g' was excluded from the present study, because listeners were unable to discriminate stimuli in this sound set reliably.

[TABLE 1]

The training task consisted of an XAB two-alternative forced-choice procedure, in which participants heard a target sound (X) and were asked to decide which of two following sounds (A or B) was the same as the target. Three boxes,

labelled “Target”, “A”, and “B”, were displayed on the computer screen. Each box was illuminated by changing its background colour while the corresponding sound was presented. At the top of the display, participants were reminded that their task was to decide “Which is the same as the target sound, A or B?”. Participants responded by pressing keys labelled “A” and “B” on a computer keyboard. Visual feedback on accuracy was given by a green tick or a red cross.

Adaptive procedure

The separation of the pairs of stimuli from the middle of the sound set varied according to participants’ performance, thus allowing the Just Noticeable Difference (JND) to be estimated. At the beginning of each block of trials, stimuli were selected from towards the end points of sound sets (stimuli 8 and 89). These stimuli should be easy to discriminate. The procedure then reduced the separation of stimuli following correct responses to make discrimination more difficult, or increased the separation of stimuli following incorrect responses to make discrimination easier. The switch between the separation of stimuli being reduced to being increased (or vice versa) was labelled a *reversal*. The adaptive procedure was run in three phases. First, the separation between stimuli was reduced by 20 steps following a correct response and was increased by 20 steps following an incorrect response. This rule was used until two reversals had occurred. Second, the separation was reduced by 10 steps following two correct responses and was increased by 10 steps following an incorrect response. This rule was used until a further three reversals had occurred. Third, this ‘two-down, one-up’ rule was used with a step size of 2 steps until participants had completed 60 trials in total. The ‘two-down, one-up’ procedure yields a JND corresponding to 71% correct performance (Levitt, 1971). The JND was calculated as the average of the

separations at the reversals during the third phase of the task. Figure 1 reproduces the separations traversed during an example run of the task.

[FIGURE 1]

Testing materials

BKB sentence test. Eight blocks of 32 sentences from the BKB corpus (Bench & Bamford, 1979) were recorded by two adult talkers of British English (1 male, 1 female). Blocks contained 16 sentences recorded by each talker. One block was used during each test. Sentences were not repeated. There were three key (content) words in each sentence. Participants were asked to repeat all the words they heard, and the experimenter recorded which key words had been identified correctly, using the 'tight' scoring procedure (Bamford & Wilson, 1979). An example of a BKB sentence, with the key words underlined, is: "The clown had a funny face".

IEEE sentence test. Four blocks of eighty sentences from the IEEE corpus (IEEE, 1969) were recorded by ten talkers with a range of British and Irish accents (4 male, 4 female, 2 female children). Blocks contained 8 sentences recorded by each of the 10 talkers. One block was used during each test. Sentences were not repeated. There were five key words in each sentence. Participants were asked to repeat all the words they heard, and the experimenter recorded which key words had been identified correctly, using the tight scoring procedure. An example of an IEEE sentence, with the key words underlined, is: "The wharf could be seen from the opposite shore".

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").

Participants reported the consonant in each stimulus by touching its orthographic transcription. Materials were recorded by 10 talkers (4 male, 4 female, 2 female children) who each recorded single token of each syllable.

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. Participants responded by touching the orthographic transcription of the appropriate word. There were 200 trials in each test. Materials were recorded by 10 talkers (4 male, 4 female, 2 female children). Each talker recorded two tokens of each word.

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). Speech signals were analysed 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 a 6mm tonotopic shift on the basilar membrane on the cochlea. The centre frequencies of these reconstruction filters were 1206, 1685, 2332, 3205, 4382, 5971, 8115, and 11007Hz.

Design & procedure

There were 3 training conditions (word training, sentence training, phonetic training), with 6 participants in each training group. Apart from the training regime, the design and procedure were the same for each group. Participants participated over the course of 10 sessions, which took place on 10 different (not necessarily consecutive) days. Figure 2 illustrates the sequence of training and test sessions.

[FIGURE 2]

Baseline session (Session 1). Pure-tone audiometry was conducted to measure air-conduction thresholds in each ear. Then participants undertook the BKB sentence test, the vowel test, and the consonant test, in that order. At least three runs of each test were administered, until an asymptote in performance was reached. A participant was declared to have reached asymptotic performance if performance was stable, within a 3% margin of error, on adjacent runs. If performance on the third run was more than 3% better than performance in the first or second run, a test was repeated, either until performance was stable, within a 3% margin of error, or until a test had been administered five times. Finally, participants completed the IEEE sentence test once.

Training sessions (Sessions 2, 3, 5, 6, 8, & 9). Twenty minutes of auditory training were administered in each training session. Participants in the phonetic-training group were exposed to the sound sets in the order shown in Table 1. On average, participants completed 5 or 6 sound sets in each session. To ensure that participants in the word- and sentence-training groups were not exposed to the same materials repeatedly, trials that had been completed were excluded from subsequent training sessions.

Training and testing sessions (Sessions 4, 7, & 10). Participants began by completing 20 minutes of auditory training. They then completed the tests of speech perception. The IEEE sentence test was administered first, followed by the vowel test, the consonant test, and finally the BKB sentence test.

Results

Baseline performance

Table 2 shows the average baseline performance for each of the training groups on each of the tests of speech perception. One-way between groups ANOVAs revealed no significant differences in baseline performance according to training group, for any of the speech tests.

[TABLE 2]

Data from training tasks

Table 3 shows the average amount of training completed by participants in each training group, in each training session. Performance on all the training tasks improved significantly over time (Figure 3).

[TABLE 3]

[FIGURE 3]

Improvements following auditory training

In order to assess whether performance had improved following auditory training, a conservative measure of baseline performance was adopted, which consisted of each participant's *best* performance during any of the baseline tests (referred to as the 'highest baseline'). For example, if a participant scored 5% the first time a test was completed, 15% the second time the test was completed, and 10% the third time a test was completed, their 'highest baseline' was recorded as 15%.

Figure 4 shows the improvements on the IEEE, BKB, consonant, and vowel tests following one, two, and three hours of training for each of the training groups. One-sample t-tests with a Bonferroni correction for nine comparisons were used to test whether improvements were significant. Adjustments were made for nine comparisons because we wished to test whether improvements following one, two, and three hours of training were significant for three training groups. Table 4 shows the results of these analyses. Bonferroni corrected p-values are reported. There were no significant improvements for the phonetic training group on any of the tests. There were significant improvements on the IEEE and BKB sentence tests for the groups who received word and sentence training. Other significant improvements were not sustained after three hours of training. There were significant improvements in consonant discrimination for the word training group following one and two hours of training; improvements following three hours of training narrowly missed significance for the word and sentence groups. A significant improvement in vowel discrimination was found only for the word training group following two hours of training.

[FIGURE 4]

[TABLE 4]

Comparison between training conditions

This section compares the effectiveness of the different training strategies. The dependent variable was the amount of improvement, relative to the highest baseline, following auditory training. Results were analysed with 3 (training time: one, two, three hours) x 3 (training strategy: word, sentence, phoneme) mixed Analyses of Variance. When there was a significant interaction between training time and training strategy, planned comparisons were carried out using one way Analyses of Variance to establish whether performance improved over time for each of the training groups.

A Bonferroni correction for three comparisons was applied (indicated by the notation *adjusted* F). When there was a significant main effect of training strategy, planned comparisons were carried out on the differences between the training groups following three hours of training. A Bonferroni correction for three comparisons was applied (indicated by the notation *adjusted* t). Bonferroni corrected p-values are reported.

IEEE sentence test. There were significant main effects of training time ($F_{2,30}$ = 14.13, p<0.001) and training strategy ($F_{2,15}$ = 11.22, p<0.01), along with a significant interaction between training time and training strategy ($F_{4,30}$ = 3.66, p<0.05). Significant improvements over time were found for the word (*adjusted* $F_{2,10}$ = 10.29, p<0.05; linear trend = $F_{1,5}$ = 16.03, p<0.05) and sentence training groups (*adjusted* $F_{2,10}$ = 11.32, p<0.01; linear trend = $F_{1,5}$ = 19.09, p<0.01), but not for the phonetic training group (*adjusted* $F_{2,10}$ = 0.14, linear trend = $F_{1,5}$ = 0.13; Figure 4). Following three hours of training, both the word- and sentence-training groups displayed significantly larger improvements than the phonetic-training group (word training vs phonetic training: *adjusted* t_{10} = 3.73, p<0.05; sentence training vs phonetic training: *adjusted* t_{10} = 7.28, p<0.001; Table 5). There was no significant difference between the improvements displayed by the word and sentence training groups following three hours of training.

BKB sentence test. There were significant main effects of training time ($F_{2,30}$ = 24.57, p<0.001) and training strategy ($F_{2,15}$ = 7.56, p<0.01), but no significant interaction between the variables ($F_{4,30}$ = 0.57). There were significant linear ($F_{1,5}$ = 24.92, p<0.001) and quadratic ($F_{1,5}$ = 23.05, p<0.001) components to the effect of training time. On average, participants identified 2.67% (95% c.i. -0.24 to 5.58) of key words correctly following one hour of training. This increased to 10.38% (95%

c.i. 8.16 to 12.61) following two hours of training, and then levelled off at 10.72% (95% c.i. 7.0 to 14.45) key words correct following three hours of training. The group who received word-based training displayed a significantly larger improvement following three hours of training than the group who received phonetic training (*adjusted* $t_{10} = 4.50$, $p<0.01$; Table 5). There were no other significant differences between the training groups following three hours of training.

Consonant test. There was a significant effect of training time ($F_{2,30} = 6.63$, $p<0.01$), but no significant effect of training strategy ($F_{2,15} = 2.97$) and no significant interaction between training time and training strategy ($F_{4,30} = 0.45$). There was a significant linear component to the effect of training time ($F_{1,5} = 13.29$, $p<0.01$). Following one hour of training, 5.58% (95% c.i. 3.25 to 7.92) of consonant sounds were discriminated correctly. This rose to 7.00% (95% c.i. 4.99 to 9.01) following two hours of training, and to 9.56% (95% c.i. 6.61 to 12.50) following three hours of training. Following three hours of training, all of the training groups improved between 7 and 12% (Table 5).

Vowel test. On the vowel test, there were no significant effects of training time ($F_{2,30} = 2.93$), training strategy ($F_{2,15} = 1.48$) and no significant interaction ($F_{4,30} = 0.60$).

[TABLE 5]

Effect of not controlling for incidental learning

Figure 5 shows the amount of improvement between the first time tests were completed in the baseline session (Base 1) and the final testing session for the BKB, consonant, and vowel tests after three hours of training. These improvements are labelled ‘uncontrolled’ because the effects of repeated testing were not controlled. The amount of improvement between the ‘highest baseline’ and the final testing

session is also shown. These improvements are labelled ‘controlled’ because the effects of repeated testing were partially controlled. Significant improvements are highlighted. Bonferroni corrections were applied within each test, separately for ‘uncontrolled’ and ‘controlled’ comparisons. (Therefore, adjustments were made for 3 comparisons.) Figure 5 shows that if no control is exercised over the effects of repeated testing, all the groups would be judged to display significant improvements on the BKB and consonant tests. When control is exercised, however, only the word and sentence training groups display improvements that reach significance on the BKB and consonant tests.

[FIGURE 5]

Discussion

In this study, word- and sentence-based training strategies led to significantly larger improvements on tests of sentence perception than did a phoneme-based strategy. Contrary to the expectation that larger improvements on sentence tests would follow sentence training than word training (Hirata, 2003; Greenspan *et al.*, 1988), both types of training improved the accuracy of identifying words in sentences. There were no significant differences between the training strategies in improving consonant or vowel discrimination. However, there were significant improvements on the consonant and vowel tests for the word training group but not for the sentence or phonetic training groups. The word training group displayed significant improvements on the consonant test following one and two hours of training, and on the vowel test following two hours of training. These findings lend tentative support to Fu *et al.*’s (2005[b]) findings that word training was more effective at improving phonemic discrimination than sentence training. In addition, the present study demonstrated quite large improvements in performance from simply repeating the outcome tests. If

we had not partially controlled for these effects of incidental learning, each of the training strategies would have been judged to produce significant improvements in the BKB sentence test and the consonant-discrimination test.

Why was phonetic training relatively unsuccessful?

There are three non-exclusive explanations for why phonetic training produced improvements in performance in the study reported by Moore *et al*. (2005) but not in the present study. The first explanation is associated with the nature of the outcome measures. Moore *et al*. (2005) investigated the effectiveness of phonetic training using the Phonological Assessment Battery (Frederickson *et al*. 1997). Good performance on the tests in that battery requires participants to be able to identify and manipulate phonemes. In contrast, good performance on the sentence and vowel tests used in the present experiment requires participants to use lexical knowledge. Phonetic training led to only minimal improvements on these tests. The consonant test however, taps phonetic rather than lexical knowledge, and the phonetic training task was associated with larger improvements on this test. This pattern of results is compatible with the idea that the phonetic training task produces learning that transfers to the perception of phonemes in nonsense syllables more than to the perception of real words.

The second explanation is associated with the nature of the training stimuli. In the study by Moore *et al*. (2005), children listened to un-distorted speech. It is possible that the phonetic training task was effective in this context because the mapping between the input and phonetic representations was straightforward, allowing participants to label stimuli as one phoneme or another. However, the noise-vocoded versions of the stimuli that were used in the present study may have been more difficult to map onto existing phonetic representations. Although performance improved over time on the phonetic training task itself, no improvement in speech

perception would be expected if the training stimuli were not mapped onto phonetic representations. Moreover, while the word- and sentence-based training tasks required participants to label stimuli (as one word or another), the XAB forced-choice task in the phonetic training task did not require phonemes to be labelled. It is possible that larger improvements following phonetic training with distorted speech would have been found if participants had been required to label stimuli and thereby remap representations for sounds onto existing phonetic representations. To test this explanation, phonetic training could be administered in a similar 2-alternative forced-choice task as was used for word-based training (e.g. participants hear 'le' and are required to classify the stimulus as 'le' or 're').

The third explanation is associated with the differences between the participants. Moore *et al.*'s (2005) study was carried out with pre-adolescent children, whose attentional skills are likely to be more variable than those of the adult participants in the present study. The children in Moore *et al.*'s study may have displayed improved phonological awareness because the training task improved their attentional skills, rather than their perceptual skills. Indeed, Moore *et al*. found no significant improvement on the phonetic training task over time, which suggests that no perceptual learning occurred. Although the study included a control group, those children did not receive an alternative to training. The trained group might have improved more than the control group because maintaining attention and arousal alone can be sufficient to lead to improvements on perceptual tasks (Amitay *et al*., 2006).

Importance of lexical information

The present results are consistent with the hypothesis that lexical information is important in the perceptual learning of distorted speech (Davis *et al*., 2005).

However, there are two alternative explanations for why the word- and sentence-based strategies led to larger improvements in speech perception than the phoneme-based strategy. The first is that the word- and sentence-training tasks required participants to label stimuli, whereas the phonetic training task did not. Labelling stimuli might be particularly important with distorted speech since the mapping between stimuli and existing linguistic representations might not be straightforward. The second alternative explanation is that word and sentence training tasks exposed participants to a larger amount of distorted speech, and a wider range of phonemes, than did the phonetic training task. However, if this was the only reason for differences between the effectiveness of word and sentence training compared with phonetic training, we would have expected larger differences between the word and sentence training tasks than we found. There were on average 8 words in each training sentence, which meant that participants heard approximately 2,400 words in the sentence training task, compared with the 800 words in the word training task.

Basis of improvements following training

Word- and sentence-based training in the present study may have helped participants to map the novel input provided by a noise-excited vocoder onto existing linguistic representations. Cochlear-implant users also display quite marked improvements in performance over time after their devices are switched on (Tyler, Parkinson, Woodworth, Lowder, & Gantz, 1997; Tyler & Summerfield, 1996), with larger improvements during the first year of implant use, and smaller improvements thereafter. Svirsky, Silveira, Suarez, Neuburger, Lai, and Simmons (2001) asked whether longitudinal improvements in vowel identification are driven by improved discrimination of stimulation delivered to different electrodes, or by improvements in labelling speech sounds. They measured electrode discrimination and vowel

identification in seven postlingually deafened cochlear-implant users immediately after their devices were switched on, and then 6.5 to 32 months thereafter. They found that improvements in electrode discrimination were insufficient to explain the improvements in vowel recognition, and concluded that the main factor driving improvements in speech perception following implantation is an improvement in labelling speech sounds.

The finding that word and sentence training tasks led to similar overall levels of improvement in the ability to identify words in sentences suggests that participants 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 skills required for sentence perception had improved, rather than a general improvement in the relationship between acoustic input and existing representations. In addition, the relative failure of phonetic training supports the hypothesis that improvements in speech perception arose due to changes in the mapping between sounds and lexical representations. The phonetic training task did not require sounds to be labelled, and did not exploit lexical information, both of which may play a role in altering the relationship between sensations and linguistic knowledge.

Importance of controlling for effects of repeated testing

The improvements without control for repeated testing were approximately twice as large as those found when partial control was exercised. If no control had been exercised over the effects of repeated testing, each of the training strategies would have been judged to lead to significant improvements on the BKB sentence test and the consonant test. Whilst we found partially controlled improvements of 13% and 16% on the BKB sentence test following word- and sentence-based training, these

improvements would have been 27% and 28% if no control had been exercised over the effects of repeated testing. Such improvements are comparable to the uncontrolled improvement of 30% on the BKB test following three hours of one-to-one training via Connected Discourse Tracking reported by Rosen *et al*. (1999).

Conclusion

The word- and sentence-based training strategies were more effective than the phoneme-based strategy at improving the perception of spectrally-distorted speech. These results have implications for the design of training procedures for use by adult cochlear-implant users. It is possible that the word- and sentence-training strategies will be more effective than phonetic training strategies that, like the one examined in this paper, do not require participants to assign phonetic labels to sounds

References

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

Bamford, J., & Wilson, I. (1979). Methodological considerations and practical aspects of the BKB sentence lists. In: J. Bench, & J. Bamford, (Eds) *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children*. London: Academic Press, 147-187.

Bench, J., & Bamford, J. (1979). *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children*. London: Academic Press.

BSA. (1981). Recommended procedures for pure-tone audiometry using a manually operated instrument. *British Journal of Audiology, 15*, 213-216.

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. *Journal of Experimental Psychology: General, 134*, 222-241.

De Filippo, C. L., & Scott, B. L. (1978). A method for training and evaluating the reception of ongoing speech. *Journal of the Acoustical Society of America, 63*, 1186-1192.

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 & Hearing, 27*, 139-152.

Frederickson, N., Frith, U., & Reason, R. (1997). *The Phonological Assessment Battery, Standardised Edition.* Windsor: NFER–Nelson.

Fu, Q.-J., Galvin, J., Wang, X., & Nogaki, G. (2004). Effects of auditory training on adult cochlear implant patients: a preliminary report. *Cochlear Implants International, 5*(Supplement 1), 84-90.

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

Fu, Q.-J., Nogaki, G., & Galvin, J. (2005[b]). Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation. *Journal of the Association for Research in Otolaryngology, 6*, 180-189.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 421-433.

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.

Hirata, J. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *Journal of the Acoustical Society of America, 116*, 2384-2394.

IEEE. (1969). *IEEE Recommended Practice for Speech Quality Measurements.* New York: Institute for Electrical and Electronic Engineers.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*, 467-477.

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 variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242-1255.

McQueen, J. M., & Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. *Proceedings of ISCA Workshop on Plasticity in Speech Perception*, 233-236.

Moore, D. R., Rosenberg, J. F., & Coleman, J. S. (2005). Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children. *Brain and Language, 94,* 72-85.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47,* 204-238.

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

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *Journal of the Acoustical Society of America, 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. *Journal of the Association for Research in Ololaryngology, 3, 332-350.*

Stacey, P.C., & Summerfield, A.Q. (2007). Effectiveness of computer-based auditory training in improving the perception of noise-vocoded speech. *Journal of the Acoustical Society of America, 121,* 2923-2935.

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 Oto-laryngologica, 121*, 262-265.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science, 271*, 81-84.

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. *Journal of the Acoustical Society of America, 102*, 508-522.

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

Walden, B. E., Erdman, S. A., Montgomery, A. A., Schwartz, D. M., & Prosek, R. A. (1981). Some effects of training on speech recognition by hearing-impaired adults. *Journal of Speech and Hearing Research, 24*, 207-216.

Table 1.

Phonemic contrasts in *Phonomena*. Adapted from Moore *et al*. (2005).

Sound set name	**Phonetic transcription**	**Phonemic contrast**	**Informal description of syllables**	**Training order**
a_uh	/æ/_/ʌ/	± back	'a'_'uh'	1
b_d	/biː/_/diː/	± labial	'bee'_ 'dee'	2
l_r	/liː/_/ɹiː/	± lateral	'lee'_'ree'	3
e_a	/e/_/æ/	± low	'eh'_'a'	4
m_n	/mɑː/_/nɑː/	± labial	'mar'_'nar'	5
s_sh	/sɑː/_/ʃɑː/	± anterior	'sar'_'shar'	6
er_or	/ɜː/_/ɔː/	± round	'err'_'or'	7
s_th	/sɑː/_/θɑː/	± distributed	'sar'_'thar'	8
v_w	/vɑː/_/wɑː/	± sonorant	'var'_'wah'	9
i_e	/ɪ/_/e/	± high	'ih'_'eh'	10
d_g	/dɑː/_/gɑː/	± coronal	'dar'_'gar'	Not used

Table 2.

Average (and standard deviation) baseline performance (% correct) for each training group on each speech test. For the BKB, consonant, and vowel tests, the highest baseline score is shown.

		Training group		
		Word	Sentence	Phonetic
Speech test	IEEE sentences	11.79	12.88	11.75
		(5.47)	(7.65)	(3.14)
	BKB sentences	42.17	38.67	41.33
		(10.23)	(13.82)	(2.42)
	Consonants	47.83	45.42	48.33
		(9.06)	(10.93)	(6.20)
	Vowels	31.92	27.75	29.00
		(8.69)	(8.12)	(7.31)

Table 3.

Average (and standard deviation) amount of training completed by participants in each training group during each 20 minute training session. Data for the word and sentence training groups are the average number of trials completed. Data for the phonetic training group are the average number of sound sets completed. The overall average is the average total number of trials or sound sets completed.

		Training group		
		Word	Sentence	Phonetic
Training session	1	379.0 (24.3)	62.7 (5.2)	4.83 (0.75)
	2	418.2 (35.7)	64.7 (5.6)	4.83 (0.41)
	3	390.8 (23.6)	68.7 (4.6)	5.33 (0.52)
	4	415.5 (40.6)	68.7 (7.1)	5.00 (0.63)
	5	407.3 (18.8)	72.3 (11.8)	5.67 (0.52)
	6	403.0 (30.2)	77.7 (7.3)	5.00 (0.00)
	7	417.5 (32.0)	75.5 (5.2)	5.67 (0.52)
	8	441.8 (26.8)	71.8 (5.5)	5.67 (0.52)
	9	421.8 (51.2)	76.0 (9.3)	5.17 (0.41)
Overall average		3695.0 (159.4)	638.0 (39.5)	31.44 (2.13)

Table 4.

Values of *adjusted* t with five degrees of freedom calculated on the improvements following one, two, and three hours of auditory training relative to the highest baseline score for training group on each test of speech perception. A Bonferroni correction for nine comparisons was applied to each test. Bonferroni corrected p-values are reported (* = $p<0.05$, ** = $p<0.01$, ***, $p<0.001$).

(continued on next page)

(from previous page)

Speech test	**Overall time spent training**	**Training strategy**		
		Word	Sentence	Phonetic
IEEE	One hour	3.29	6.16 *	1.21
	Two hours	4.73 *	6.75 **	1.64
	Three hours	5.44 *	14.82 ***	1.50
BKB	One hour	1.58	4.00	-1.87
	Two hours	6.97 **	6.11 *	4.44 (p=0.06)
	Three hours	7.19 **	3.40	2.32
Consonant	One hour	5.05 *	2.93	1.61
	Two hours	5.81 *	3.88	2.99
	Three hours	4.28 (p=0.07)	4.38 (p=0.06)	3.35
Vowel	One hour	1.73	3.58	0.19
	Two hours	5.13 *	2.65	2.38
	Three hours	2.30	2.48	1.13

Table 5.

Average (and standard deviation) improvement (%) following three hours of training on each of the tests of speech perception, for each of the training groups.

		Training group		
		Word	Sentence	Phonetic
Speech test	IEEE sentences	12.46	15.63	2.25
		(5.61)	(2.58)	(3.68)
	BKB sentences	13.00	16.17	3 .00
		(4.43)	(11.63)	(3.16)
	Consonants	12.33	9.75	6.58
		(7.06)	(5.45)	(4.81)
	Vowels	5.83	7.75	1.83
		(6.20)	(7.65)	(3.98)

Figure Captions

Figure 1: Example of a run of the phonetic training task.

Figure 2: The sequence of training and testing sessions undertaken by participants with the approximate duration of each session.

Figure 3: Improvements over time on the training tasks. For the word training task, the average percentage of trials discriminated correctly is plotted. For the sentence training task, the average *number* of errors made (i.e. the average number of non-target words selected) is plotted. For the phonetic training task, the average JND isn plotted.

Figure 4: Improvements on the IEEE sentence test, the BKB sentence test, the consonant test, and the vowel test following one, two, and three hours of word-based training, sentence-based training, and phonetic training. Error bars denote Bonferroni corrected (99.44%) confidence intervals. Open circles plot improvements for individual participants.

Figure 5. Improvement in performance on the BKB sentence test (Panel A), the consonant test (Panel B), and the vowel test (Panel C). The light grey bars (uncontrolled) show the overall level of improvement between the first time tests were completed in the baseline session and the final testing session. The dark grey bars (controlled) show the level of improvement between the 'highest baseline' and the final testing session. Error bars denote 95% confidence intervals. Significant improvements (tested with one-sample t-tests with a Bonferroni correction for three comparisons) are highlighted (* = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$).

Figure 1

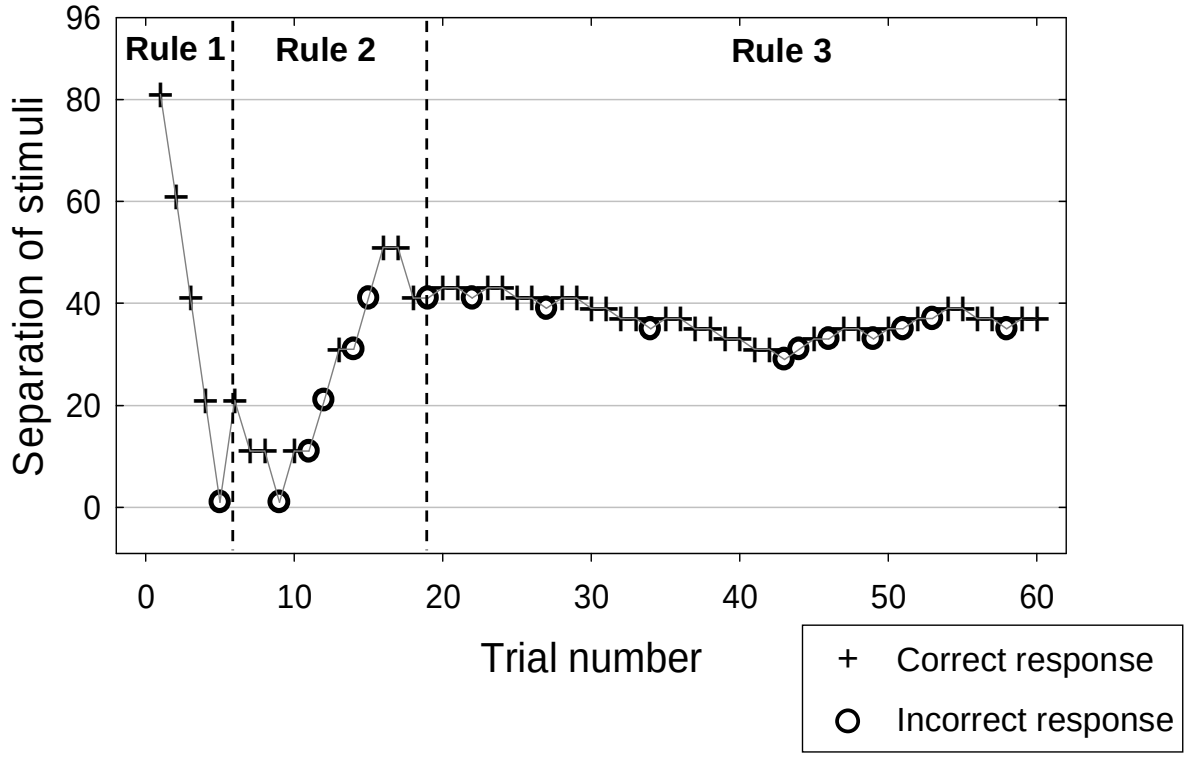

Figure 2

Figure 3

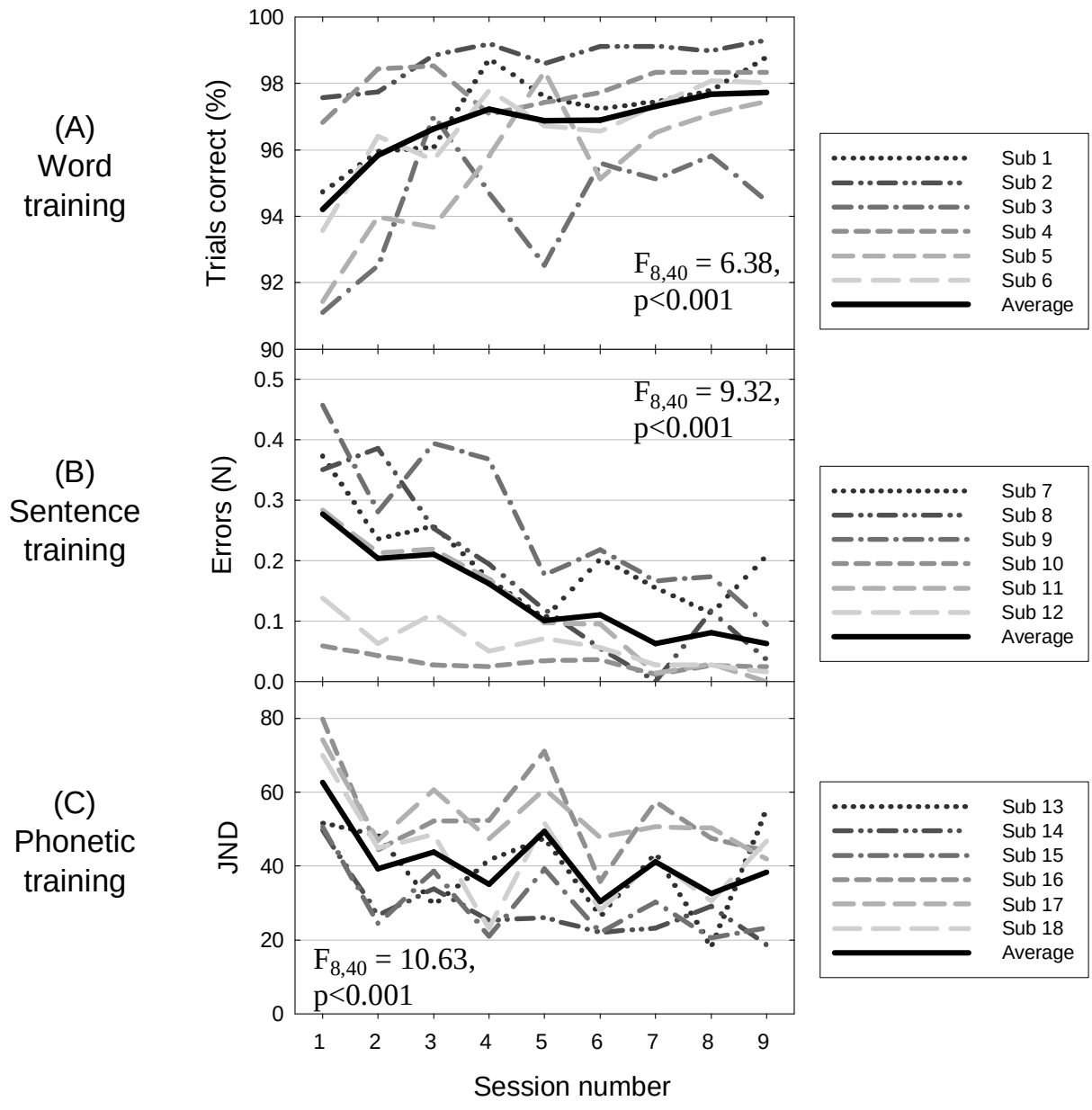

Figure 4

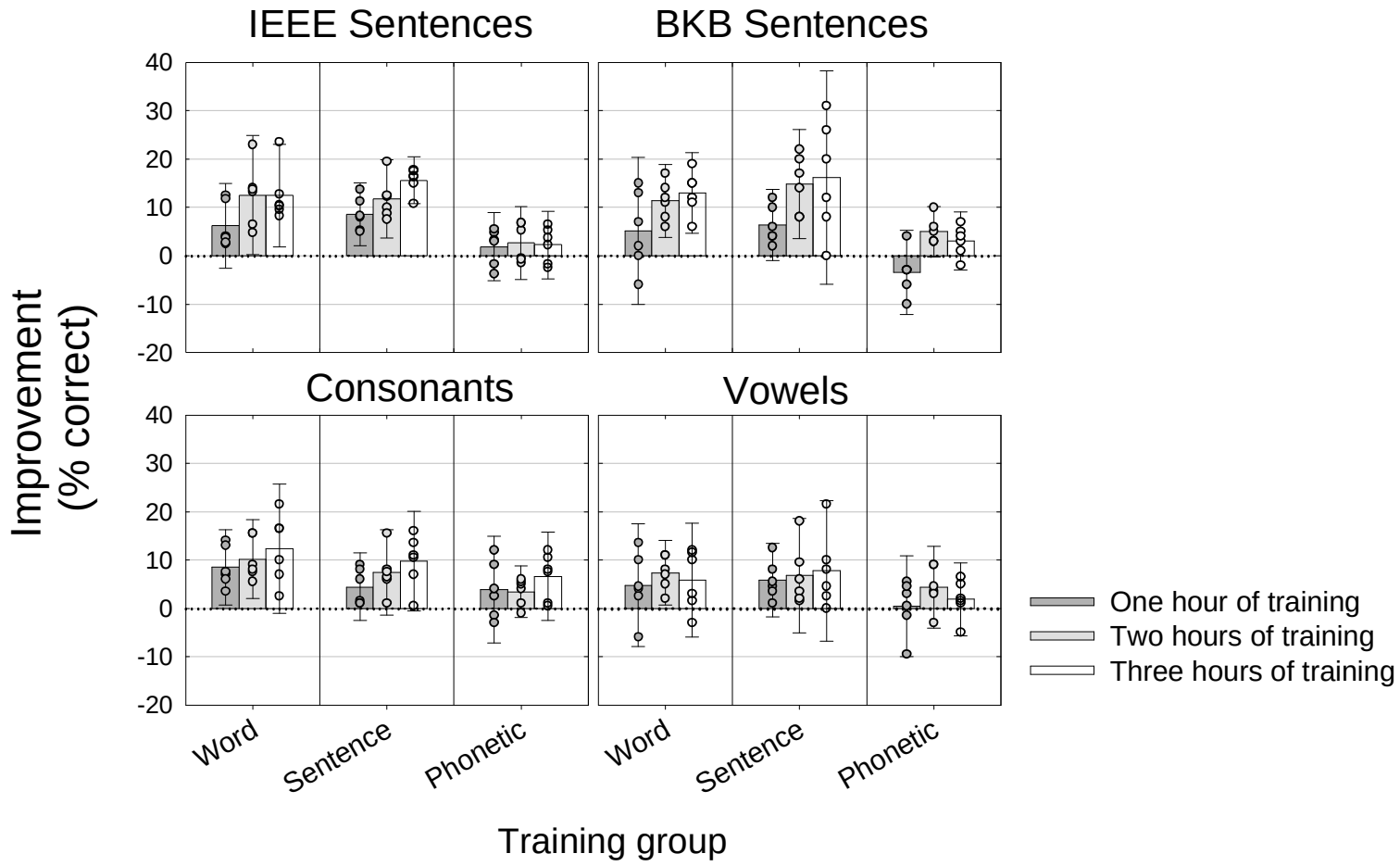

Figure 5