Effects of Training on Speech Recognition Performance in Noise Using Lexically Hard Words

Matthew H. Burk
Larry E. Humes
Indiana University—Bloomington

Purpose: This study examined how repeated presentations of lexically difficult words within a background noise affect a listener’s ability to understand both trained (lexically difficult) and untrained (lexically easy) words in isolation and within sentences.

Method: In the 1st experiment, 9 young listeners with normal hearing completed a short-term auditory training protocol (5 hr) while 8 other young listeners with normal hearing completed a similar protocol lasting about 15 hr in the 2nd experiment. All training made use of multiple talkers and was in a closed-set condition. Feedback was provided on a trial-to-trial basis and consisted of either orthographic or orthographic and auditory feedback. Performance on both the trained and untrained words in isolation and within sentences was measured pre- and posttraining.

Results: Listeners’ performance improved significantly for the trained words in an open and closed-set condition, as well as the untrained words in the closed-set condition. Although there was no mean improvement in the number of keywords identified within sentences posttraining, 50% of the listeners who completed the long-term training showed improvement beyond the 95% critical difference.

Conclusions: With enough training on isolated words, individual listeners can generalize knowledge gained through isolated word training to the recognition of lexically similar words in running speech.

KEY WORDS: speech recognition, neighborhood activation model, training

The issue of auditory training has been in the forefront of auditory rehabilitation discussions over the past several years as both hearing aid and cochlear implant technologies evolve. Advances in hearing aids, designed to improve speech understanding (e.g., multiple microphones, noise reduction algorithms, improvements in signal clarity), still raise questions as to how best to benefit from such technologies in order to improve a user’s speech understanding. Is amplifying the speech signal with a hearing aid enough to optimize speech communication, regardless of the technologies in use, or is other training or counseling necessary during and after the fitting process?

The primary difficulties understanding speech for listeners with hearing impairment can be attributed to the reduced audibility of the signal, although reduced frequency selectivity, diminished temporal resolution, or cognitive declines associated with aging may also play an important role in speech perception abilities (Committee on Hearing and Bioacoustics and Biomechanics, 1988; Humes, 1996). Typically, for older adults, many years pass from when they first suspect a hearing problem to when they pursue...
help (Brooks, 1979; Smedley, 1990). Given the apparent plasticity of the auditory system, even in older adults (e.g., Willott, 1996), central changes might occur as a result of this long-standing peripheral hearing loss. If such changes do occur, it may not be realistic to assume that, upon insertion of a hearing aid, the brain can immediately make use of the restored information in the stimuli.

One way to potentially improve a hearing aid user’s speech understanding abilities is through auditory training. Currently, there are few data available to support the benefits of auditory training (Sweetow & Palmer, 2005). Sweetow and Palmer’s recent review regarding individualized auditory training examined six studies that had sufficient details and appropriate methodologies to meet the inclusion criteria for their review. Their general conclusion was that synthetic training (word or sentence level) may show some benefit in psychosocial situations, as well as improving speech recognition in noise. However, as they pointed out “there is a need for more evidence-based research to establish both the efficacy and efficiency of auditory training procedures” (Sweetow & Palmer, 2005, p. 502). The present study attempts to add to the information currently available by providing data on the magnitude of the overall training effect when variables such as the lexical properties of the words, the number of talkers, and the amount of training time are examined.

Previous work by Burk, Humes, Amos, and Strauser (2006) investigated the benefit of an auditory training protocol utilizing amplified speech and noise with both young listeners with normal hearing and older listeners with hearing impairment. A word-based training protocol consisting of 75 meaningful phonetically balanced monosyllabic CVC words (AB words; Boothroyd, 1995, 1999) was used to examine training-related improvements in word recognition and identification in noise, as well as generalization to the recognition of novel words, both in isolation and in sentences. They showed a closed-set training effect of about 11%–17% for approximately 3.5 hr of word-based training using orthographic feedback on the specific set of trained words. This training also generalized to novel talkers. However, sufficient generalization to novel words and sentences was lacking. Munro and Lutman (2005) also found similar improvements over time (approximately 10%–15%) for closed-set word identification performance in noise with trial-to-trial orthographic feedback. This study was not designed to evaluate a training regimen per se, so issues such as generalization of training to other words, talkers, sentences, and so forth were not examined. Based on the foregoing results, the present study further explored the use of words as the units of speech in the training protocol due to their ease of presentation and scoring.

Several issues potentially affecting a word-based training regimen were addressed in this study. The previous studies of Munro and Lutman (2005) and Burk et al. (2006), for example, demonstrated that young and older adults listening in noise can improve performance on a restricted set of words (75–80 words) with a relatively brief amount of training. If similar gains can be obtained for frequently occurring words typical in everyday speech, there might be some transfer of this training to conversational situations. Sets of words varying in word frequency have been developed based on the neighborhood activation model (NAM; Luce & Pisoni, 1998). This model maintains that the difficulty of a particular word in a word-recognition task is determined by its frequency of occurrence and neighborhood density. Neighborhood density refers to the number of similar-sounding items differing only by a phoneme from the target word. Based primarily on these two properties of word frequency and neighborhood density, the NAM can be used to separate words into categories ranging from easy to hard. Within this framework, a lexically easy word (e.g., loud) is one occurring frequently (high frequency of usage in the language) while also having few neighbors or rhyming words (low density), whereas a lexically hard word (e.g., wade) would be a rarely occurring word with many similar-sounding neighbors. The availability of both word lists (Takayanagi, Dirks, & Moshfeghi, 2002) and sentences (Bell & Wilson, 2001) based on the NAM enabled the use of hard words as training stimuli and were an ideal first step to investigate the influence of word frequency effects on training. For example, if training on hard words is shown to generalize to easy (high-frequency) words, typical running speech containing many high-frequency words might also improve.

Another issue needing further investigation was the number of talkers used within a formal training protocol. Although previous work using AB words in the Audiology Research Laboratory at Indiana University (Burk et al., 2006) did show good generalization from 1 talker to several novel talkers, a total of only 4 talkers were available for these materials, and only 1 was used for training to examine generalization to the other (3) talkers. Clopper and Pisoni (2004) have argued that more heterogeneous training stimuli can produce better generalization to novel stimuli (both other talkers and other materials). The idea of increasing talker variability has also been explored when training Japanese listeners to identify the English /t/ and /l/ with similar beneficial results (Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991). Although training using multiple talkers could conceivably increase training time, the potential improvements in generalization to stimuli outside the laboratory would be worth the additional time involved.

Although the ultimate goal of these experiments is to help in the design of an effective auditory training
protocol for new hearing aid users, the experiments described here involve only young listeners with normal hearing. Previous work on auditory training found that with adjustments to the speech-to-noise ratio (SNR), older listeners with hearing impairment performed similarly to young listeners with normal hearing (Burk et al., 2006). This, together with a long history of equating speech recognition performance between groups of varying hearing loss and age (e.g., Dubno & Schaefer, 1992, 1995; Humes, Dirks, Bell, & Kincaid, 1987), through adjustments in SNR, led to the use of normal-hearing adults listening in background noise in this study.

**Experiment 1**

Experiment 1 addressed the effects of training using lexically hard words on the identification of other words and on the recognition of sentences. As previously discussed, auditory training using words in isolation has been shown to be effective at improving recognition of trained words but did not generalize to novel words or sentences (Burk et al., 2006). In this case, training on hard words (a more difficult task) might have the added benefit of generalizing to easy words, as well as sentences containing either hard words (similar to those trained) or sentences containing easy words.

A second issue addressed in this experiment was the effect of training using multiple talkers \((n = 6)\) versus a single talker. The use of multiple talkers during training should show even greater generalization to unfamiliar talkers outside the laboratory due to the use of more heterogeneous training stimuli. The following describes a short-term auditory training protocol based on lexically hard words presented by multiple talkers.

**Method**

**Participants.** Nine young normal-hearing (YNH) listeners (3 men and 6 women) participated in the study. All listeners had air-conduction thresholds equal to or better than 20 dB HL (American National Standards Institute, 1996) at octave frequencies from 250 through 8000 Hz and ranged in age from 18 to 34 years \((M = 25.3\) years). Listeners were recruited through general postings and e-mails within the Department of Speech and Hearing Sciences at Indiana University Bloomington. Listeners were paid $10 per session as well as a $35 completion bonus upon collection of all data. During the training aspect of the study, a listener received a bonus of $2 for each training block in which his or her score was higher than his or her initial closed-set baseline performance, for a possible bonus of $12 per session. Participants were not informed of this performance incentive until after the baseline measures were completed.

**Stimuli.** Monosyllabic CVC words based on the NAM, as recorded by Takayanagi et al. (2002), were used as the training materials. These materials contained 75 lexically hard words (high neighborhood density, low frequency of usage) and 75 lexically easy words (low neighborhood density, high frequency of usage) spoken by 12 talkers (6 men and 6 women), with 11 of the 12 talkers ranging in age from 25 to 50 years, plus one 10-year-old girl (see Table 1). A speech-shaped noise was created to match the long-term average spectrum \((±2\) dB from 50 to 10000 Hz) of the NAM words (see Figure 1). The long-term average spectrum was established by concatenating all of the stimulus files \((150\) words \(×\) 12 talkers) into one large stimulus file and computing an averaged fast Fourier transform using a 50-ms window. The NAM words for each talker and 12 random samples of the speech-shaped noise were stored as individual WAV files.

**Table 1.** Fundamental frequency measured for the vowels /i/, /e/, /a/ \((n = 12)\) of each talker.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Talker</th>
<th>Gender</th>
<th>Range</th>
<th>Min</th>
<th>Max</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiar talkers</td>
<td>1</td>
<td>F</td>
<td>18.8</td>
<td>213.4</td>
<td>232.1</td>
<td>221.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>F</td>
<td>22.0</td>
<td>179.6</td>
<td>201.6</td>
<td>191.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>F</td>
<td>13.9</td>
<td>168.1</td>
<td>182.0</td>
<td>173.7</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>M</td>
<td>12.2</td>
<td>113.1</td>
<td>125.3</td>
<td>117.9</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>M</td>
<td>11.2</td>
<td>99.9</td>
<td>111.1</td>
<td>105.8</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>M</td>
<td>7.8</td>
<td>88.2</td>
<td>96.0</td>
<td>91.5</td>
</tr>
<tr>
<td>Unfamiliar talkers</td>
<td>7 (10-year-old)</td>
<td>F</td>
<td>17.6</td>
<td>195.1</td>
<td>212.7</td>
<td>202.8</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>F</td>
<td>15.2</td>
<td>176.6</td>
<td>191.8</td>
<td>183.3</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>M</td>
<td>15.1</td>
<td>97.7</td>
<td>112.8</td>
<td>105.0</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>M</td>
<td>6.7</td>
<td>89.3</td>
<td>96.0</td>
<td>93.3</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>M</td>
<td>5.8</td>
<td>71.5</td>
<td>77.3</td>
<td>74.1</td>
</tr>
<tr>
<td>52-item novel talker</td>
<td>12</td>
<td>F</td>
<td>6.5</td>
<td>138.5</td>
<td>144.9</td>
<td>141.9</td>
</tr>
</tbody>
</table>

Note: “Familiar” or “unfamiliar” talkers indicates whether the specific talkers were used within the training sessions or for the generalization tasks.
(16-bit, 48.8 kHz sampling frequency) on the computer hard drive for future presentation. Sentences based on the NAM were extracted from the Department of Veterans Affairs Sentence Test (VAST; Department of Veterans Affairs, 1998) CD and stored as individual WAV files. A detailed description of the sentence materials can be found in Bell and Wilson (2001). A total of 40 lexically easy sentences (each containing 3 lexically easy keywords) and 40 lexically hard sentences (each containing 3 lexically hard keywords) were used, all spoken by a single female talker. Sentences were not specifically controlled in that some of the hard sentences contained keywords on which the listeners were specifically trained (approximately 9%). The same can be said for the easy words in that some of the keywords within the sentences were contained within the easy word list used for the generalization tasks (approximately 7%). The speech-spectrum noise for the sentences was created by cutting out (at zero crossings of the noise waveform) the steady-state portion of the first 16 noise segments from the VAST CD and concatenating them to create a single 50-s noise segment (see Figure 1). This noise segment was then divided into 12 smaller segments of exactly 4 s each, which were randomly selected as the competition for each sentence presentation. Although the noise from the VAST CD was not reshaped, the edits described earlier were necessary in order to obtain the random 4-s samples of noise as needed for the laboratory presentation system.

All testing took place using insert earphones (ER-3A, Etymotic Research) in a double-walled sound-treated room (Industrial Acoustics Company) that met or exceeded American National Standards Institute guidelines for permissible ambient noise (American National Standards Institute, 1999). The listeners were seated in front of a touch screen monitor to receive orthographic feedback during training and to select appropriate responses. The WAV files were presented digitally (16-bit D/A, 48.8 kHz sampling rate) via Matlab (The MathWorks, Version 7) running Tucker Davis Technologies (TDT) System-III equipment. All stimulus presentation parameters (randomization, SNR, presentation earphone, etc.) were controlled by a custom Matlab program. The target stimulus, either a word or a sentence, was always centered within the competing noise. This yielded a total presentation time of either 2 s or 4 s per trial for the words or sentences, respectively. All stimuli (words, sentences, and noise) were mixed and presented to the right ear only. The left ear was occluded during testing with the left insert earphone. Overall levels were adjusted using circuits controlling a TDT RP2.1 real-time processor and a headphone buffer (TDT HB7) to produce an overall level of 90 dB sound pressure level for the speech and noise stimuli combined, as measured in an HA-2 2-cm³ coupler. All stimuli were presented within the speech-shaped noise at an overall SNR of −5 dB for words and −7 dB for sentences, based on the average root mean square of the concatenated stimuli. This SNR was chosen based on pilot testing with a goal of avoiding both ceiling and floor effects for all of the varied baseline conditions (i.e., open and closed set, easy and hard stimuli).

Procedure. Listeners took part in seven 75- to 90-min sessions over approximately a 2-week period. No more than one session per day was allowed, with a maximum time between sessions of 3 days. The specific conditions administered in each session are summarized in Table 2. As can be seen in Table 2, there were both open-set and closed-set response conditions. For the open-set conditions, each listener wrote his or her response on an answer sheet, whereas for the closed-set conditions, all
75 words were presented alphabetically in list form on the touch screen from which the listener selected the appropriate response by touching the word on the screen.

Baseline testing for Session 1 consisted of a list of 52 words (26 hard and 26 easy words) presented by a male talker (Talker 4), one who was among the talkers used for training, and the same list presented by a female talker (Talker 12), one who was not included in the subsequent training. These 52 words, also supplied by Takayanagi et al. (2002), were originally supplied as practice items for each of the 12 recorded talkers. The 2 talkers’ average fundamental frequency (for /i/, /e/, /a/) differed by 24 Hz, while their average amplitude across all words differed by about 0.1 dB. These 52 words were not presented again until the completion of training. These two lists were followed by the hard and easy sentence sets spoken by a single female talker (listed as female in Table 2), who was not included in the group of 12 talkers shown in Table 1, presented all sentences.

Each day of training consisted of six blocks of the same 75 hard words for a total of 24 training blocks (6 blocks by four training sessions). This enabled the listener to hear every word spoken by every talker (Talkers 1–6) within each training session. Presentation order for the words was randomized, while each talker presented a different 12–13 words for each of the 6 training blocks (e.g., Talker 1 presented Words 1–12 during Block 1, Words 13–25 during Block 2). Posttraining scores for the 75 hard words (trained) and the 75 easy words (untrained) presented by the familiar talkers (1–6) were measured in both an open- and closed-set condition during Session 6. Although the trial-by-trial word randomization was different, the same 12–13 words were spoken by the same talkers as in Session 1. The last session consisted of generalization tasks. The listeners heard the set of 75 easy and 75 hard words presented by Talkers 7–11 (talker generalization) and the set of 52 hard/easy words heard only once prior to training presented by Talker 4 and Talker 12 (word and talker generalization). Finally, listeners were presented with a new set of hard and easy sentences spoken by the same female talker as the Session 1 sentences to assess generalization of training to novel sentences.
Results and Discussion

Figure 2 presents the individual percentage correct scores across the 24 training blocks (upper panel) and group (bottom) percentage correct scores from pretest through the posttest sessions for the hard word set. Individual pre-to posttraining percentage correct scores are also presented in Table 3. Prior to statistical analysis, all percentage correct scores were transformed into rationalized arcsine units (RAUs; Studebaker, 1985) to stabilize the error variance. With regard to the individual data (top panel), Cronbach’s alpha was used to evaluate the consistency of performance across trial blocks. Visual inspection of the individual data suggests that those listeners who performed best at pretest tended to perform best throughout. Likewise, those who performed worst initially tended to remain...
near the bottom of the group throughout. This was confirmed by a Cronbach’s alpha value of .98.

With regard to the group data (bottom panel), the main effect of training block on word-recognition performance was significant, $F(24, 192) = 11.43$, $p < .001$, as indicated by a general linear model repeated-measures analysis. Post hoc paired-sample t tests, using a Bonferroni adjustment for multiple comparisons, were then computed as needed, with all differences designated as significant when $p < .05$. Performance on 9 of the 24 training blocks was significantly better as a group than the listeners’ initial baseline closed-set performance (left filled circle), as noted by the asterisks in the bottom panel of Figure 2. As can be seen, by Training Block 18, or after approximately 3.5 hr of training, listeners were consistently performing better than their baseline score.

Figure 3 shows mean pre- and posttraining word identification (closed-set) and word-recognition (open-set)
performance for the hard (trained) and easy (untrained) words. Mean performance posttraining was always greater than the mean performance prior to training. In particular, performance for the hard words improved by 50.9% in the open-set format and 19.7% in the closed-set format versus easy word improvements of 5.8% in the open-set format and 13% in the closed-set format. Differences in performance from pre- to posttraining were examined via paired-sample t tests. Performance on the trained words improved significantly for both the open- and closed-set response formats (p < .001) while performance on the untrained words improved significantly for the closed-set condition (p < .001) but not the open-set condition (p = .098). Upon completion of the training, listeners were recognizing a greater percentage of the (trained) hard words compared with the (untrained) easy words (82.5% vs. 57.8%). In this case, the hypothesis that training on lexically hard words would generalize to easy words proved false in an open-response format but true in the closed-response format.

Figure 4 shows mean posttraining word-recognition scores for the 75 trained (hard) and untrained (easy) words presented by familiar (n = 6) versus unfamiliar talkers (n = 5) in an open-set response format. For the trained words, listeners performed significantly better than the initial pretraining score when they heard the list presented by either the familiar talkers, t(8) = 15.46, p < .001, or the unfamiliar talkers, t(8) = -10.71, p < .001. Listeners did perform significantly better overall on the trained words presented by the familiar talkers versus those same words presented by unfamiliar talkers, t(8) = 3.29, p = .011, but there was a large improvement in word recognition regardless of the talkers. When listening to the easy (untrained) words, there was a significant increase in performance from pretraining to posttraining for the words presented by the unfamiliar talkers, t(8) = 2.70, p = .027, but no improvement for the words presented by the familiar talkers, t(8) = -1.87, p = .098, nor was there a difference in posttraining performance for the list presented by the familiar talkers relative to the unfamiliar talkers, t(8) = -1.83, p = .104. This generalization to talkers, but not words, is consistent with a lexical approach to learning (rather than specific acoustical learning).

Figure 5 shows mean open-set performance on the 52-item list of hard/easy words presented by a familiar talker and an unfamiliar talker broken down into the 26 hard and 26 easy words. These words were not included in the training materials, so not only was one talker unfamiliar, but the words had not been presented since the baseline testing in Session 1. Listeners significantly improved on novel word recognition, both hard, t(8) = -2.78, p = .024, and easy, t(8) = -3.96, p = .004, when presented by a familiar talker. The two conditions
on the right of Figure 5 were not significant and constituted the use of both an unfamiliar talker presenting hard words, $t(8) = -1.30, p = .229$, and an unfamiliar talker presenting easy words, $t(8) = -0.114, p = .912$, pre- to posttraining.

From the results shown in Figures 4 and 5, it is apparent that listeners showed the most improvement when listening to the list of trained words presented by the familiar talkers and only slightly poorer performance when those same trained words were presented by unfamiliar talkers (see Figure 4). Listeners also improved significantly when presented with novel words spoken by a familiar talker (easy–familiar), but overall performance was reduced for the unfamiliar talkers relative to the familiar talkers (see Figure 5). Listeners showed no significant improvement when novel hard or easy words were presented by an unfamiliar talker.

The last aspect of this experiment investigated the ability of listeners to generalize from word-based auditory training to sentences. Mean percentage-correct scores for the keywords identified within hard and easy Sentence Set A (pretraining) were compared with comparable scores for Sentence Set B (posttraining). As shown in Figure 6 there were no significant improvements in keyword identification for running speech posttraining for either the hard, $t(8) = -0.19, p = .854$, or easy, $t(8) = -1.13, p = .291$, sentences. Individual difference scores in RAU are shown in Figure 7 for each of the 9 participants, with the dotted line representing the 95% critical difference (CD) of 11.7 RAU for a score based on 120 items (Studebaker, 1985). For the hard sentences, no listeners improved beyond the 95% CD; whereas on the easy sentences, 1 listener significantly improved while 1 listener did significantly worse upon completion of the training protocol.

The purpose of Experiment 1 was to examine the ability to generalize improvements in the identification of lexically hard words to both easy (untrained) words and running speech. Based on previous work in the second author’s lab (Burk et al., 2006), training with isolated words is effective at improving performance specifically for the trained words. In this case, the use of lexically hard words as the training stimuli was thought not only to provide an opportunity to improve on the specific words used for training, but also to generalize to lexically easy words. Unfortunately, generalization to easy words was mixed, and there were no improvements in sentence recognition as a result of the isolated word training.

One issue that deserves further investigation revolves around the overall length of training time. Although these data suggest training with words is not an effective approach to improving the recognition of sentences, there was a significant positive correlation ($r = .78$) between the improvement in hard-word identification during training and the corresponding improvement in
hard-word sentence recognition. This might suggest more training time is necessary in order to significantly improve sentence recognition. This issue was investigated in Experiment 2.

**Experiment 2**

**Method**

*Participants.* A new group of 8 YNH listeners (3 women and 5 men) ranging in age from 19 to 25 years old ($M = 21.6$ years) took part in Experiment 2. Listeners were again paid $10 per session, and a $35 bonus was paid upon completion of every seven sessions. The listeners also received a $5 performance bonus when their mean score for the current session’s training (across the six blocks) was higher than their mean score from the previous session’s training. Due to the potentially higher pay associated with the longer training protocol (i.e., $10/session plus $35 every seven sessions), the performance bonus was decreased to $5 per session. Because the bonus was now based on the listener’s average score across all six blocks, it would still be in the listener’s best interest to perform at his or her best on all six blocks, in much the same way as in Experiment 1, in which the bonus was based individually on each block. Again, they were not informed of the performance incentive until after the baseline measures had been completed.

*Procedure.* The stimuli and presentation methods were identical to those used in Experiment 1 with the following exceptions. First, for some of the listeners, the type of feedback used during the training portion of the study differed in this experiment from that used in Experiment 1. Five of the listeners received the same orthographic feedback as in Experiment 1, while the 3 remaining listeners received both orthographic and auditory feedback. There were originally 5 members of the auditory feedback group as well, but 2 withdrew after the first few sessions. Upon missing a response, listeners from the auditory feedback group were presented with the correct response orthographically and they could choose to listen to either the response they chose or the correct target stimulus. The feedback was presented at the same SNR used throughout the training ($-5 \text{ dB SNR}$). The listeners could choose to hear either the target stimulus or the stimulus corresponding to their response for up to a total of five stimulus presentations, or they could bypass the auditory feedback entirely on a trial-by-trial basis.

The second difference between Experiments 1 and 2 was in the length of training. The procedures of Sessions 1 (pretraining), 6, and 7 (posttraining) from Experiment 1 remained the same, as did the number of training blocks per session during the training (1 session contained six blocks). However, the amount of intervening training varied across the 8 listeners from a minimum of 8 training sessions (approximately 12 hr) for all listeners in the
auditory feedback group to a maximum of 20 training sessions (approximately 25 hr) for 1 participant in the orthographic feedback group. The 5 listeners in the orthographic feedback group all completed a minimum of 12 training sessions (approximately 15 hr), which was at least 3 times the total training time used in Experiment 1. Because the training extended for upwards of 2 months, most participants finished 2–4 training sessions per week, with occasional breaks of up to a week due to vacation or illness.

One last difference between Experiments 1 and 2 involved the sentences used to examine the benefit of word training on running speech. Like Experiment 1, Sentence Set B was used as the posttest sentence stimuli. However, within this experiment, the original sentences (Sentence Set A) were also repeated posttraining. Because of the longer time frames between Session 1 and the posttraining sessions (mean duration of 6.6 weeks), the likelihood that the participants would remember the initial sentences was reduced considerably, and this allowed for a more direct comparison of posttraining benefit.

Results and Discussion

Figure 8 shows word-recognition performance as a function of training block for the 8 listeners in the long-term training protocol. Initially, two groups of listeners (orthographic feedback and auditory/orthographic feedback) were to be examined in order to address the effect feedback type might have on training performance. However, informal observation suggested that the listeners chose not to use the auditory feedback after the first session of training. Due to the overall lack of auditory feedback used, as well as the limited control over the amount of auditory feedback supplied, its effects on training could not be properly examined. However, examining differences in performance between the two feedback groups showed only the main effect of trial as significant, $F(47, 282) = 7.35, p < .001$, with neither the group, $F(1, 6) = 1.41, p = .280$, nor the Group × Trial, $F(47, 282) = 0.78, p = .849$, effects reaching significance, as indicated by a general linear model, mixed-model analysis with a repeated-measures variable of trial and a between-subjects variable of group. Therefore, the data were combined for all further analyses of the long-term training protocol with all significant differences designated at the $p < .05$ level.

Table 4 shows individual word-recognition performance pre- to posttraining for both the hard and easy words, while mean improvements are shown in Figure 9. As in Experiment 1, there was a significant improvement in word recognition of 55% for the trained words in the open response set ($p < .001$) versus 7.7% the untrained words ($p = .057$). Significant improvements were also shown for both the trained ($p < .001$) and the untrained words ($p = .006$) in the closed-set response condition (24.2% and 6.8%, respectively). These results are essentially the same as the short-term training group in Experiment 1. The increased training time had little effect on increasing generalization to easy (untrained) words. The pre- and posttraining scores for the hard and easy words in both the open- and the closed-set conditions were also compared across Experiments 1 and 2 (Figure 3 and Figure 9, respectively) using an independent samples $t$ test. As expected, there were no significant differences between the pretest scores for the short-term training compared with the long-term training. A comparison of the posttest scores also showed no increases in overall word-recognition abilities as a consequence of increasing the overall training time.

Figure 10 examines generalization to talkers and shows results similar to those shown in Figure 4 from the previous experiment. That is, listeners performed significantly better than the initial pretraining score when they heard the trained list presented by either the familiar talkers $t(7) = -14.74, p < .001$, or the unfamiliar talkers, $t(7) = -17.33, p < .001$. Listener performance was
poorer when listening to the trained list presented by unfamiliar talkers. However, the overall improvement on the trained list, even when presented by unfamiliar talkers, was still 43.9% relative to the baseline score. As in the short-term training, listeners again improved significantly when listening to the easy words presented by the unfamiliar talkers, \( t(7) = -2.62, p = .034 \), but there was no improvement for the list presented by the familiar talkers, \( t(7) = -2.28, p = .057 \), although the difference in scores for the familiar versus unfamiliar talkers was only 0.5%. Again, the trends regarding generalization to talkers are very similar to those observed over the shorter term training and maintain the idea of lexical learning being the dominant factor within the auditory-training protocol used.

Figure 11 depicts mean open-set performance on the novel 26 hard and 26 easy words presented by either the familiar or the unfamiliar talker, relative to the short-term training, where listeners only improved on the novel words presented by the familiar talker (Figure 4),

**Table 4.** Word recognition in percentage correct for lexically hard (trained) and lexically easy (untrained) words in an open-set and closed-set response format after a long-term training protocol.

<table>
<thead>
<tr>
<th>Word recognition</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trained words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preopen</td>
<td>41.3</td>
<td>24.0</td>
<td>29.3</td>
<td>32.0</td>
<td>38.7</td>
<td>28.0</td>
<td>33.3</td>
<td>36.0</td>
</tr>
<tr>
<td>Postopen</td>
<td>97.3</td>
<td>82.7</td>
<td>94.7</td>
<td>94.7</td>
<td>89.3</td>
<td>90.7</td>
<td>77.3</td>
<td>80.0</td>
</tr>
<tr>
<td>Preclosed</td>
<td>61.3</td>
<td>70.7</td>
<td>57.3</td>
<td>70.7</td>
<td>74.7</td>
<td>78.7</td>
<td>60.0</td>
<td>65.3</td>
</tr>
<tr>
<td>Postclosed</td>
<td>92.0</td>
<td>86.7</td>
<td>93.3</td>
<td>96.0</td>
<td>93.3</td>
<td>97.3</td>
<td>84.0</td>
<td>89.3</td>
</tr>
<tr>
<td><strong>Untrained words</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preopen</td>
<td>45.3</td>
<td>44.0</td>
<td>44.0</td>
<td>52.0</td>
<td>45.3</td>
<td>42.7</td>
<td>48.0</td>
<td>54.7</td>
</tr>
<tr>
<td>Postopen</td>
<td>58.7</td>
<td>48.0</td>
<td>46.7</td>
<td>46.7</td>
<td>58.7</td>
<td>65.3</td>
<td>61.3</td>
<td>52.0</td>
</tr>
<tr>
<td>Preclosed</td>
<td>78.7</td>
<td>80.0</td>
<td>70.7</td>
<td>76.0</td>
<td>76.0</td>
<td>85.3</td>
<td>76.0</td>
<td>76.0</td>
</tr>
<tr>
<td>Postclosed</td>
<td>88.0</td>
<td>80.0</td>
<td>80.0</td>
<td>90.7</td>
<td>80.0</td>
<td>92.0</td>
<td>77.3</td>
<td>85.3</td>
</tr>
</tbody>
</table>
listeners in the long-term group improved significantly on both the novel easy list presented by the familiar talker, $t(7) = -2.77, p = .028$, and the novel hard list presented by the unfamiliar talker, $t(7) = -3.11, p = .017$. However, the pre- to posttraining scores did not improve significantly for the novel hard words presented by the familiar talker, $t(7) = -1.82, p = .112$, as in Experiment 1, or for the novel easy words presented by an unfamiliar talker, $t(7) = -1.00, p = .350$, the trends in overall scores were very similar between the short- and long-term training.

Up to this point, the extension of training time made no appreciable difference in the results relative to generalization to untrained words or unfamiliar talkers. However, the primary reason for extending the training was to potentially improve generalization to running speech. When looking at mean improvements in the number of keywords identified (see Figure 6), the only significant improvement was an increase of 10.9% when Sentence Set A (hard) was repeated at the end of the training sessions, $t(7) = -3.11, p = .017$. Keyword identification improved by 6.9% pre- to posttraining for the Set A hard sentences to the Set B hard sentences ($p = .140$) and by 5.5% for the Set A easy sentences to the same Set A easy sentences ($p = .149$), while performance dropped slightly by −0.8% when comparing the Set A easy sentences to the Set B easy sentences ($p = .136$).

However, the individual difference scores suggest that several listeners did improve significantly in their recognition of keywords in sentences. Figure 12 shows the individual difference scores in RAUs for the keywords identified correctly when listening to either the Set A sentences posttraining or the Set B sentences posttraining, as in Experiment 1. As shown in the top left panel, 6 of the 8 listeners improved beyond the 95% CD when retested on the Set A sentences containing the hard words. When the sentence testing was directly comparable to Experiment 1, as in the upper right panel, 4 of the 8 listeners improved significantly with the new set of hard sentences, whereas no such improvements

Figure 9. Word-recognition performance pretraining versus posttraining after long-term training on lexically hard words compared with lexically easy words in both an open- and closed-set response format. Error bars represent one standard deviation. *Significant differences pre- to posttraining significant at the .05 level.

Figure 10. Open-set word-recognition performance pretraining versus posttraining after long-term training on lexically hard (trained) words and lexically easy (untrained) words presented by familiar and unfamiliar talkers. Significant differences identified compare posttraining performance with pretraining performance. Error bars represent one standard deviation. *Significant at the .05 level.

Figure 11. Open-set word-recognition performance on a set of 52 novel words divided into the 26 hard and 26 easy stimuli after long-term training, presented by a familiar talker used during the training and an unfamiliar talker excluded from the training. Error bars represent one standard deviation. Asterisks identify pre- to posttraining differences significant at the .05 level.

Figure 12. Individual difference scores in RAUs for the keywords identified correctly when listening to either the Set A sentences posttraining or the Set B sentences posttraining, as in Experiment 1.
were noted for the shorter term training in Experiment 1 (see Figure 7). As might be expected due to the lack of generalization to easy words, there were fewer significant improvements pre- to posttraining for the easy sentences with overall trends similar to those of Experiment 1.

**General Discussion**

The purpose of these experiments was to examine ways to improve a listener’s speech understanding in noise through the use of a word-based auditory training protocol. Two broad issues were examined: the ability to generalize training based on hard words to both novel easy and hard words and the ability to generalize word-based training to running speech.

The hypothesis that training using lexically difficult words would generalize to easy novel words was not supported. Both the short- and long-term groups showed significant improvements in closed-set recognition on the easy (untrained) words but no improvements when tested in an open-set condition. Once again, this would limit the utility of using only words, as the training seems very word specific, if a primary goal is generalization to other novel words. Ultimately, improvements in the recognition of running speech would benefit a hearing aid user to a greater degree outside the laboratory. In this case, the extension of training for longer periods did improve generalization to sentences containing lexically hard keywords. The length of the training protocol allowed the original sentences (Sentence Set A) to be repeated posttraining with less worry of the sentences being memorized. When looking at this comparison, 75% of the listeners improved their hard keyword recognition significantly (beyond the 95% CD). When comparing hard keyword recognition pre- to posttraining for Sentence Set A to the novel Sentence Set B, 50% of the listeners in the long-term group improved beyond the 95% CD, versus 0% for the short-term group. As previously mentioned, about 7%-9% of the keywords in each sentence set were among the trained words, which is about the size of the relative improvement from training. Development of additional sentence sets made up of the same words used in training should be explored further when examining generalization of the training effects to running speech.

![Figure 12.](image-url)

*Figure 12. Individually difference scores in RAU pretraining to post-training for lexically hard sentence Set A to the lexically hard sentence Set A (top left), hard sentence Set A to hard sentence Set B (top right), easy sentence Set A to easy sentence Set A (bottom left), and easy sentence Set A to easy sentence Set B (bottom right) after long-term training. The 95% critical difference is indicated by the dotted line. * Indicates a significant (p < .05) mean improvement in keyword identification pre- to posttraining.*
The issue of talker variability was also examined in both experiments. Previous work with AB words (Burk et al., 2006) showed good generalization from 1 talker used for training to 4 novel talkers. The extension to training with 6 talkers within this study presented similar results. That is, listeners still maintained their large improvements for the trained words regardless of whether the talker was familiar or unfamiliar. Although the overall magnitude of the improvements in word recognition for the novel easy and hard words in Experiment 2 was smaller than that of trained words, the familiarity of the talker seemed to have minimal impact on the results.

Isolated word training had shown no generalization to running speech when using either AB words (Burk et al., 2006) or lexically hard words for training of only a few hours (Experiment 1). By increasing the training time from about 5 hr to a minimum of 12–15 hr, listeners improved hard keyword identification within lexically difficult sentences by 50%–75%. The results of this investigation lend support to the use of words as a means to improve sentence recognition.

Several issues regarding the use of words in an auditory training protocol are in need of continued investigation. The first involves the auditory feedback used during the long-term training. As stated previously, those listeners in the long-term training who could receive auditory feedback generally chose to skip the option on a trial-by-trial basis. Although an older listener with hearing impairment would have more personal motivation to improve their speech understanding, and therefore might use the feedback to a greater extent, the option to skip the feedback entirely has been eliminated from ongoing work. This will allow for a more direct comparison of the various types of available feedback. Another issue requiring further study involves the number of words used for training. The current experiments trained listeners on only 75 words. Therefore, it may be beneficial to train the listeners on hundreds or thousands of words. Previous and ongoing work suggest that listeners can effectively be trained on a list of words and maintain this improvement while subsequently being trained on new words. It may be that training a listener on a new set of words each week would create a large enough set of trained words from which the listener could draw for generalization to sentences. Although the sample sizes of the current study were limited, they were sufficient for initial exploratory studies of various training parameters. However, larger samples will be needed to establish the efficacy of the final training protocol once developed.

The current data from YNH listeners suggest that sufficient training on isolated words can have an impact on the listener’s ability to understand running speech in noise. The impact of auditory training on a listener’s speech understanding outside the clinic will have to be evaluated in a more formal manner while also finding better ways to objectively measure generalization to everyday speech. Although previous work has shown comparable performance for YNH and older hearing-impaired listeners within a short-term auditory training protocol (Burk et al., 2006), the rate of training, ability to make use of context in sentences, personal motivation, and other cognitive issues may all play a role in the overall training effect for older listeners. Research is ongoing to both replicate Experiment 2 with older listeners with hearing impairment and investigate the effect of increased set size on generalization to running speech.

Acknowledgments

A portion of this article was presented at the 2005 Acoustical Society of America Meeting in Vancouver, British Columbia, Canada. This work was supported, in part, by National Institutes of Health (NIH) Grant R01 AG08293, awarded to the second author, and by NIH Postdoctoral Training Grant T32 DC00012, awarded to the first author. The authors would like to thank Lauren Strauser for her assistance in data collection and to the participants, especially those in the long-term training protocol used in Experiment 2, for their dedication.

References


Received January 17, 2006
Accepted June 2, 2006
DOI: 10.1044/1092-4388(2007/003)

Contact author: Matthew H. Burk, Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, IN 47405-7002.

E-mail: maburk@indiana.edu.