Published in final edited form as:

J Speech Lang Hear Res. 2008 June ; 51(3): 759–771. doi:10.1044/1092-4388(2008/054).

Effects of Long-Term Training on Aided Speech-Recognition Performance in Noise in Older Adults

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

Abstract

Purpose—This study examined how repeated presentations of words in noise affected understanding of both trained and untrained words in noise (in isolation and in sentences).

Method—Eight older listeners with hearing impairment completed a word-based auditory training protocol lasting approximately 12 weeks. Training materials were presented in a closed-set condition with both orthographic and auditory feedback on a trial-to-trial basis. Performance on both trained and untrained lexically easy and hard words, as well as generalization to sentences, was measured. Listeners then returned for an additional 14 weeks to monitor retention of the trained materials.

Results—Training listeners on 1 set of words improved both their open- and closed-set recognition of the trained materials but did not improve performance on another set of untrained words. When training switched to the other set, performance for the new set of words improved significantly, whereas significant improvements on the previously trained words were maintained. Training generalized to unfamiliar talkers but did not generalize to untrained words or untrained keywords within running speech. Listeners were able to maintain improved performance over an extended period.

Conclusion—Older listeners were able to improve their word-recognition performance in noise on a set of 150 words with training.

Keywords

speech recognition; training; hearing impaired

Many older individuals seeking help from an audiologist have a sloping sensorineural hearing loss. This type of hearing loss will often become apparent to the patient due to decreased speech understanding in noisy environments. The problem of speech understanding, particularly in noise, is twofold. There is the obvious issue of reduced audibility of the signal in the high frequencies, where the listener's hearing loss is typically most severe, as well as the interfering effects of the competing stimulus. Due to the particularly deleterious effects of background noise for older hearing-impaired (OHI) listeners relative to young normal-hearing (YNH) listeners (Frisina & Frisina, 1997; Gordon-Salant & Fitzgibbons, 1995), it is often desirable to improve the signal-to-noise ratio (SNR) for older adults. Hearing aids have been an effective means for improving the audibility of the speech signal but have been less successful at improving the SNR sufficiently. Technologies such as directional microphones and noise-reduction algorithms, designed to improve the acoustical SNR, continue to have promise. Another approach that

might also provide benefit, however, is to train the listener to make better use of the existing SNR.

Research in auditory training has a long history, with several key studies conducted in the 1970s and 1980s (Rubinstein & Boothroyd, 1987; Sweetow & Palmer, 2005; Walden, Erdman, Montgomery, Schwartz, & Prosek, 1981; Walden, Prosek, Montgomery, Scherr, & Jones, 1977). As will be discussed below, results of past studies dealing with both auditory and visual (i.e., lip-reading) training were encouraging (Walden et al., 1977, 1981). However, technical limitations at the time may have restricted their utility. Although some of these early results were promising, showing improved sentence recognition due to auditory or visual consonant training, the rehabilitation programs required a patient to return to the clinic for several hours over an extended time period. Such programs are difficult to justify from a cost-benefit analysis. Patients may not be willing to return numerous times to a clinic for results that are perceived as small for the time expended.

Still, patients may be willing to take part in self-paced training that could be completed at home. The prospects for a home-based training system have given rise to a renewed interest in auditory training, due mainly to the ease with which a training protocol could be administered using personal computers or multi-media players. Even with these advances in available technologies, few objective data exist regarding the benefits of an auditory training program, whether based on consonants, words, or running speech (Sweetow & Palmer, 2005).

Walden et al. (1981) focused on consonant training as a means to improve a listener's ability to comprehend sentences. Their rationale for consonant training was that improved consonant recognition within a sentence combined with the increased context of meaningful sentences helps to decrease the overall options available from which the listener can choose (Walden et al., 1981). They trained three groups of adult males on auditory or visual consonant recognition for approximately 7 hr. All of the participants were enrolled in a standard 2-week inpatient group rehabilitation program, with two groups receiving either extra auditory or extra visual consonant recognition training. Although all three groups improved significantly pre- to postrehabilitation, the groups receiving the extra auditory or visual training improved significantly over the control group. When examining the training effect on sentence recognition, they found a moderate correlation between improvements in consonant recognition and improvements in sentence recognition. Others have examined auditory training based on high-context materials, such as sentences or phrases (Blamey & Alcantara, 1994; Montgomery, Walden, Schwartz, & Prosek, 1984; Rubinstein & Boothroyd, 1987). Rubinstein and Boothroyd (1987), for example, trained one group of listeners on both consonant recognition and sentence perception, whereas another group of listeners received training only on sentence perception. Although the authors did find a significant but small improvement posttraining in speech-recognition performance (about 5%, averaged across both groups), they did not find any significant difference in performance between the two training paradigms (sentences alone or consonants plus sentences).

In recent years, the Audiology Research Laboratory (ARL) at Indiana University has been investigating various word-based training programs. Variables investigated to date include the number and type of words used during training, the length of the training program, and the number of talkers used for training (Burk & Humes, 2007; Burk et al., 2006). Although the overall training time and stimuli have varied across experiments, there have been several consistent outcomes in these studies. First, both YNH and OHI listeners are able to effectively improve their word recognition performance on trained sets of words ranging in size from 75 to 150 items. With approximately 3.5 hr of training, YNH listeners were able to

significantly improve their open- and closed-set word recognition in noise (by 52.5% and 16.7%, respectively) for a set of 75 phonetically balanced monosyllabic CVC AB words (Boothroyd, 1995, 1999) presented by 1 female talker. Once listeners were trained on the set of 75 words, their improved performance was maintained when listening to the same words presented by 3 unfamiliar talkers (Burk et al., 2006). Although large improvements were shown on the trained set of words regardless of the talker, there was much less (although significant) improvement for a novel set of 75 AB words (11.1% and 9.2% for open- and closed-set responses, respectively) and no improvement for running speech. Replicating the measurements with OHI listeners while listening to amplified speech simulating a well-fit hearing aid showed improvements in performance almost identical to that of the YNH listeners. These data point out the ability to improve word recognition in noise for OHI listeners on a limited set of words, yet changes were obviously necessary if any generalization to running speech outside the laboratory was to occur. Based on this first study with AB words, there were several questions needing further investigation: (a) Would a different set of words provide greater generalization to untrained words and/or running speech? (b) would training using several talkers, rather than 1, provide greater generalization to untrained materials? and (c) would increasing the amount of training have an effect on generalization to novel words and sentences?

Several of the above issues were investigated by Burk and Humes (2007) by examining the benefits of a word-based auditory training protocol based on lexically hard words within a group of YNH listeners. *Lexically hard words* were defined, in accordance with the neighborhood activation model (NAM; Luce & Pisoni, 1998), as words that are relatively rare in frequency of occurrence and have many similar sounding or “neighboring” words. It was conjectured that if listeners could improve their recognition of lexically hard words in noise through training, they might also improve their recognition of untrained lexically easy words. Because everyday communication consists of many “easy” words, overall generalization to both novel “easy” words and running speech might improve. The results were once again similar to those found when examining training improvements with AB words—that is, listeners improved significantly on the trained set of 75 lexically hard words (50.9% and 19.7% for open- and closed-set items, respectively), while improving to a much smaller degree on the untrained lexically easy words (approximately 6%–13%). Increasing the training time from approximately 5 hr to approximately 15 hr and training with multiple talkers ($n = 6$) had little effect on the overall improvements in word recognition but did seem to increase the listeners’ abilities to generalize the word-based training to running speech. None of the listeners in the short-term training group improved their hard keyword identification within a novel set of sentences, compared with 50% of those in the long-term training group. Although the same sentences were not repeated at the end of the short-term training, 75% of the listeners improved their hard keyword identification beyond the 95% critical difference when listening to the same sentences presented prior to training. Because the increased training time did seem to affect overall generalization to running speech, the next step was to replicate the long-term training protocol using lexically hard words with OHI listeners. For OHI listeners, it was unknown whether the extended protocol discussed in Burk and Humes (2007) would provide similar or potentially better results. The following experiment examines the ability of OHI listeners to improve their word recognition in noise for lexically hard words and whether such improvements, if they occur, generalize to novel words and sentences. Of further interest was the effect that training on a new set of words has relative to retention of previously trained materials—that is, does learning a second set of 50–75 words impact the retention of the training benefits for the initial set of 50–75 words?

As will be discussed below, training the OHI listeners on 75 lexically hard words did not generalize to a set of 75 lexically easy words. As a result, the training was extended to

include several sessions with lexically easy words. The impact of learning 75 additional lexically easy words on the retention of the initially trained set of 75 (lexically hard) words was investigated. Small decreases in performance for the original set of 75 words were observed following training with the new set of 75 words. Therefore, additional measurements were completed to determine whether this was simply due to forgetting the materials over time or due to some form of interference of the new set of words with those learned originally.

Method

The following stimuli and procedures were identical in most regards to those described in Burk and Humes (2007). A condensed version summarizing the stimuli and procedures is supplied here, with any changes to previous designs described in greater detail.

Participants

Eight older listeners with hearing impairment (4 men and 4 women) ranging in age from 58 to 78 years ($M = 69.5$) took part in this study. Listeners had sensorineural hearing loss ranging from mild to moderately severe at octave frequencies from 250 to 6000 Hz. Individual audiograms for the right (test) ear of each listener, as well as the mean hearing threshold across all listeners, are shown in Figure 1.

Listeners were recruited from either a list of previous study participants, excluding prior training studies, or clinic records of patients who indicated their willingness to participate in hearing studies from within the Department of Speech and Hearing Sciences at Indiana University Bloomington. Listeners were paid \$10 per session with a \$35 bonus paid upon completion of every seven sessions. During sessions dedicated solely to training (i.e., not baseline or posttraining measures), listeners also had the opportunity to receive an additional \$5 performance incentive per session. The performance incentive was not revealed to the participant until pretraining baseline performance had been completed and was earned when a listener's mean score for a current session's training (across six blocks of 75 trials) was higher (by any amount) than their mean score from the previous session's training. Depending on performance incentives, listeners were paid approximately \$15–\$20 per hour. An inability to improve and therefore receive some number of performance incentives had no bearing on a listener's continuation in the study or his/her normal session payments.

Stimuli

Both word and sentence materials based on the neighborhood activation model (Luce & Pisoni, 1998) were used for training and generalization tasks. The training tasks used 75 lexically hard and 75 lexically easy monosyllabic CVC words as recorded by Takayanagi, Dirks, and Moshfegh (2002). A lexically hard word such as *dill* is a word that has a low frequency of occurrence and has many close neighbors (words differing by one phoneme; e.g., *bill*, *dip*, *doll*). A lexically easy word is one that is used frequently but has few neighbors (e.g., *food*). The original recordings consisted of 12 talkers (6 men and 6 women), which allowed for the presentation by two groups: one group of talkers used for training and a separate group of talkers used for generalization purposes. Other lexically hard and easy words from the same recordings, spoken by both a familiar (trained) and unfamiliar (untrained) talker but distinct from the 150 words cited above, were used to examine generalization to novel words.

The sentence materials, recorded by Bell and Wilson (2001), were extracted from the Veterans Affairs Sentence Test (VAST) CD (U.S. Department of Veterans Affairs, 1998). A total of 160 sentences spoken by 1 female talker was used, with one set of 80 sentences

containing three lexically hard keywords and one set of 80 sentences containing three lexically easy keywords per sentence. The background noises used during presentation of both the word and sentence stimuli consisted of 20 randomly selected steady-state speech-shaped noises filtered to match the RMS long-term average spectrum of the respective speech stimuli. Again, a more detailed description of the stimuli, talkers, and noise creation can be found in Burk and Humes (2007).

After the individual word, sentence, and noise files were created, the stimuli were further shaped according to each listener's hearing loss to ensure audibility of the speech signals through 4000 Hz. Custom filters based on a quasi-DSL approach (Seewald, Hudson, Gagne, & Zelisko, 1992) were designed so that the RMS spectrum of the speech stimuli was at least 20 dB SL relative to the listener's hearing thresholds at 1/3-octave bands from 200 to 4000 Hz. If the 1/3-octave band levels of the matched speech-shaped noise compared with the listener's hearing thresholds showed that the stimuli exceeded the hearing threshold by 20 dB or more at any 1/3-octave band, no amplification was applied. If the signal level was insufficient to maintain a 20-dB SL signal at any 1/3-octave band, appropriate gain was then applied (see Figure 2). Once the individual filters were created for each listener, the stimuli were filtered using Adobe Audition (Adobe Systems Incorporated, Version 1.0) and stored for subsequent playback during testing and training. The same spectral shaping was applied to both the speech and the noise.

Equipment

All listening sessions took place in a sound-treated booth meeting American National Standards Institute (ANSI, 1999) specifications for ears-covered threshold testing. Participants were run one or two at a time, seated at computer stations separated by a divider. The stimuli were presented via a personal computer running MATLAB (Version 7.0; The MathWorks) interfaced to a Tucker Davis Technologies (TDT) System III real-time processor (RP2.1). A sampling rate of 48.8 kHz and 16-bit resolution were used. All presentation parameters including SNR, overall levels, and randomization were controlled through custom MATLAB programs. The signals were amplified to the maximum undistorted output of the system before being attenuated by a TDT headphone buffer (HB7) en route to ER-3A insert earphones (E.A.R. Corporation). Both earphones were inserted during testing, with all stimuli presented to the right ear only. As stated previously, the spectra of the unshaped speech and noise stimuli (overall level approximately 68 dB SPL) were adjusted offline using a quasi-DSL approach.

Procedures

Listeners completed a total of 20–24 sessions at an average of 3 sessions per week (range of 2–4). The sessions were broken down into a preliminary session to arrive at a suitable SNR, baseline measures, training sessions with lexically hard words, a midpoint evaluation, and training sessions with lexically easy words, followed finally by a session of posttraining measures (see Table 1). Upon completion of the entire training protocol, listeners were asked to return weekly for up to 14 weeks to monitor retention of the training materials.

Prior to starting the full training protocol, listeners completed a practice session of 52 lexically easy and hard words presented by 2 different talkers in an open-set condition at an SNR of 0 dB and listened to 40 sentences containing keywords of medium lexical difficulty (high usage–high density and low usage–low density) at an SNR of –2 dB. The SNR was then adjusted for listeners into two broad categories based on their scores. If open-set word recognition performance fell below 25%, SNR was increased by 3 dB to an overall SNR of +3 dB for words and +1 dB for sentences, as was the case for 5 of the 8 listeners. For those listeners ($n = 3$) who performed in the 30%–50% range on the preliminary open-set word

items, the SNR was not changed. A summary of SNRs used throughout the study is provided in Table 2.

Once the initial SNR was set, Session 1 included baseline measures for both the 75 easy and 75 hard words in an open-set and closed-set condition presented by the 6 talkers used throughout the training, as well as 40 lexically easy and 40 lexically hard sentences in an open-set condition presented by 1 female talker not used during the training. Open-set baseline measures were always obtained prior to closed-set measures. Two lists of 52 more words (26 easy/26 hard) were also presented by 1 familiar talker (to be used within the training) and 1 unfamiliar talker excluded from the training. These two lists of 52 items were the same words presented during the preliminary session, although potentially at a more advantageous SNR.

Upon completion of the baseline measures, listeners completed 9–11 sessions of training, with each session containing six blocks of the 75 hard words in a closed-set condition. During closed-set training, listeners selected from the set of 75 words on the computer screen and were given both orthographic and auditory feedback (at the same SNR) upon an incorrect answer. After training, participants were evaluated on their open-set word-recognition and closed-set word-identification performance for the trained words presented by the familiar talkers as well as their ability to generalize to 5 novel talkers presenting the same trained (hard) and untrained (easy) words. The number of training sessions to complete prior to the midpoint evaluation was determined based on previous work with YNH listeners (Burk & Humes, 2007) and overall performance of the individual participants (i.e., whether or not they reached ceiling or appeared to be at an asymptotic point on their learning curve). After the midpoint evaluation, participants were switched from training on lexically hard words to training on the 75 lexically easy words. The number of training sessions for the easy words (9–11) was equal to the number of training sessions completed previously by that listener for the hard words. Due to the decreased difficulty of the easy words, SNR was again adjusted in order to eliminate ceiling effects. Performance was adjusted by a 5%/dB correction (determined during pilot testing) to attempt to reduce closed-set easy word identification scores to approximately 65%. Approximating this criterion resulted in two groups, those needing a –3 dB correction and those needing a –6 dB SNR correction. Again, Table 2 summarizes all SNRs used for each portion of the study.

Following this second training phase, open- and closed-set performance was measured for both the 75 easy and the 75 hard words presented by familiar talkers as well as open-set performance on the hard and easy words presented by unfamiliar talkers. This was followed in the last session by the presentation of 160 lexically easy and hard sentences, 80 of which were identical to those presented prior to training (repeat set) and 80 that were new to the listeners (new set). This allowed a direct pre- to post comparison for 80 of the sentences (40 containing lexically hard keywords and 40 containing lexically easy keywords) while also assessing performance for a set of new, but similar, sentences. Last was a repeat presentation of the 52 novel words presented by 1 familiar and 1 unfamiliar talker.

Upon completion of the training protocol, listeners were asked to return once a week for up to 14 weeks to follow any declines in posttraining word recognition performance. For the first 7 weeks, the listener's word recognition performance was measured by presenting the 75 easy words in an open-set condition. Because the easy words were the most recently trained stimuli, primary interest was in any decrease in performance due to time, absent competition from newly trained words. If a listener could not return at their weekly appointment time, he or she simply skipped the week and returned the following week. Other than the first week after training, when only 2 participants could return, all other weeks had a minimum of 5 listeners returning, with all 8 returning at the 7-week

posttraining point. After measuring easy-word open-set performance over the 7 weeks, listeners were presented with the hard words in both an open- and closed-set condition. Any listener not performing within the 95% critical difference of their hard-word posttraining scores from the midpoint training evaluation was retrained by completing either one or two sessions of closed-set training. Once trained to restore performance to within the 95% critical difference of their original posttraining score at the midpoint evaluation, open-set performance for the hard words was measured once a week for 7 weeks. Six of the 8 listeners completed this portion of the study, with only 1 participant missing his appointment at Week 4.

Results and Discussion

All statistical analyses were completed after converting the percent correct scores to rationalized arcsine units (RAUs; Studebaker, 1985). This was necessary in order to stabilize the error variance in scores, particularly as listeners performed in the 90%–100% range posttraining for certain conditions. A Bonferroni adjustment for multiple comparisons was used to establish significance for all t tests using an overall alpha value of .05. For ease of comparison across recent articles regarding auditory training from the ARL, the figures are intentionally similar to those published in past articles. However, all the data reported here are new data not published previously.

Word Recognition Improvements

Figure 3 provides the individual data for the training portion of this study. The left panel contains the data for the training with the 75 lexically hard words, and the right panel shows the data for the subsequent training with the 75 lexically easy words. Based on preliminary pilot testing with other words, SNRs were adjusted individually for hard words and again for easy words in an effort to bring the initial closed-set identification scores at the start of training to about 70%–75% correct. As noted previously, the specific SNRs required for each participant and training phase are provided in Table 2. In general, except for the top performer for hard words (left panel), the individual adjustments in SNR, designed to allow sufficient room for improvement during training while not making the training frustratingly difficult, were successful. The data in Figure 3 also demonstrate that although there are differences in the extent and time course of learning among listeners, every participant shows improved performance with training. All listeners improved their open-set word recognition performance beyond the 95% critical difference (14.8 RAUs for the set of 75 items) on both the hard and easy words posttraining, and 6 of the 8 listeners improved upon their baseline closed-set performance for both sets of words.

Pre- to posttraining improvements—Figure 4 summarizes the mean open-set word-recognition and closed-set word-identification performance for the hard and easy words before and after training. Comparison of the pretraining baseline scores (black bars) to those obtained after hard-word training (midway evaluation; gray bars) enables examination of the impact of hard-word training. When participants were trained on hard words, their performance improved significantly from the pretraining baseline of 33.8% to the midpoint evaluation score of 81.2% for the open-set condition, $t(7) = -8.65$ ($p < .001$) and from 68.3% to 84.7% in the closed-set condition, $t(7) = -10.18$ ($p < .001$). Training on hard words did not improve easy word open-set performance, $t(7) = -1.40$ ($p > .004$), or easy-word closed-set performance, $t(7) = -3.47$ ($p > .004$).

The striped bars in Figure 4 show mean performance at the end of the subsequent training using lexically easy words. Once the training was switched to the easy words after the midpoint evaluation, listeners improved to a much greater degree on those words relative to

the increases in performance when training consisted of only hard words. Open-set word recognition for the easy words improved significantly from 48.8% at the initial baseline (black bar) to 89.2% after easy word training (striped bar), $t(7) = -12.73$ ($p < .001$), whereas the closed-set easy word recognition showed a significant improvement of 17.2% from baseline to posttraining, $t(7) = -5.43$ ($p < .001$). The additional training on easy words improved performance on those specific words but decreased performance slightly (13.4%) and significantly, $t(7) = 6.32$ ($p < .001$), for the open-set recognition of hard words; however, it did not significantly impact the closed-set identification of hard words (see Figure 4). For both cases, performance for hard words at the end of hard-word and easy-word training sessions (striped bars) remained significantly greater than pretraining baseline performance (black bars), $t(7) = -6.11$, hard open; $t(7) = -4.84$, hard closed ($p < .004$).

After training on all 150 words, significant improvements were shown for both the easy and hard words in the open- and closed-set conditions (black bars compared with striped bars in Figure 4)—that is, listeners were able to maintain most of the benefit obtained from the initial training on the 75 hard words while increasing the overall trained set size from 75 to 150 words. Burk et al. (2006) found similar results with YNH listeners who were first trained on a set of 75 monosyllabic words followed by training on a new set of 75 monosyllabic words. Training trends for the second set of words showed equally large improvements in word recognition performance, whereas mean word recognition on the originally trained words dropped by only 2.7% over the fairly short 1.5-week protocol. This is important, as one of the potentially limiting factors of word-based training appears to be the listeners' focus on the lexical properties of the words themselves (i.e., memorizing the words) rather than a global improvement in their abilities to understand speech of any type in noise. If listeners are essentially memorizing the words, regardless of the talker, it would be important to train them on larger set sizes than the 75 used previously within the ARL (Burk & Humes, 2007; Burk et al., 2006). The data in Figure 4 indicate that it is possible for older hearing-impaired adults to improve their open-set recognition of words in noise for a set size of at least 150 words. The limit to the number of words a listener would be able to learn remains to be seen.

Generalization to novel talker—As has been previously shown (Burk & Humes, 2007; Burk et al., 2006), listening to isolated words spoken by specific talkers generalized well to the same words spoken by unfamiliar talkers. Figure 5 shows mean posttraining performance for the open-set recognition of trained words spoken by both the familiar (gray bars) and unfamiliar (white bars) talkers. Listeners performed significantly better than their baseline performance, improving by 40.2% for the hard words, $t(7) = -13.8$ ($p < .001$) and 35% for the easy words, $t(7) = -13.42$ ($p < .001$) even when presented by unfamiliar talkers. Generalization of trained words to novel talkers after isolated word training has been a consistent result throughout the word-based training studies within the ARL. This consistent observation supports the notion that the underlying mechanism is in the lexical representation of the words and not merely memorizing the talkers' pronunciation of those words.

Generalization to novel words—An important issue for word training is the ability to generalize to new, untrained words, whether spoken by familiar or unfamiliar talkers. The issue of word generalization was evaluated via a set of 52 easy/hard novel words presented pre- and posttraining. These words were presented by 1 talker, either a familiar talker used within the training or an unfamiliar talker not included within the training. Figure 6 displays the mean performance for familiar and unfamiliar talkers. No significant improvement was noted pre- to posttraining for the novel words when presented by either the familiar talker, $t(7) = 0.52$ ($p > .05$), or unfamiliar talker, $t(7) = -2.17$ ($p > .05$). In general, the results in Figure 6 are similar to those obtained using other words and listeners in prior studies. For

example, Burk et al. (2006) trained OHI listeners on one set of 75 words, and recognition for a separate set of 75 words showed a significant, albeit small, 6.9% improvement. Burk and Humes (2007) also found small but significant improvements in easy word recognition after training on hard words. Because of the small but significant improvement in word recognition of untrained words in previous studies (Burk & Humes, 2007; Burk et al., 2006), there was some optimism that larger improvements in novel word recognition might be demonstrated in the current study after training on larger sets of both easy and hard words. However, as shown in Figure 6, this was not the case. For the most part, substantial gains in open-set word recognition performance in noise are confined to those words that have been trained.

Sentence Keyword Improvements

Figure 7 shows the mean open-set keyword recognition scores obtained in noise for the OHI listeners. Mean keyword recognition performance in noise improved from pretraining baseline (black bars) to posttraining (gray and striped bars) by only 4%–8%. Posttraining performance was assessed using sentences identical to those used in the pretraining baseline (gray bars) and a new set of sentences not heard previously by the listeners (striped bars).

Although mean data showed no significant improvements in keyword sentence recognition ($p > .0125$; $0.05/4$), individual data did show small but significant improvements. Figure 8 shows improvement in keyword identification in RAUs pretraining to posttraining for the individual listeners. Five of the 8 listeners improved beyond the 90% critical difference when listening to the same hard sentences presented a second time at the end of the training protocol. When comparing pretraining performance on the Set A hard sentences to the Set B hard sentences posttraining, 4 of the 8 older hearing-impaired listeners again improved beyond the 90% critical difference. Only 2 of the 8 listeners (S4 and S8) did not improve to some degree pre- to posttraining on sentence keyword identification. It is interesting to note that Listener 4 in Figure 8 was the best-performing listener on the hard-word training (black line in Figure 3). Her poorer performance when listening to sentences post-training may possibly be explained by her exceptional use of context due to her background as a poet (both writing and memorizing for recitals), combined with her unwillingness to repeat sentences during posttraining measures that she thought did not “make sense,” even when the correct words were heard. (Given the constraints of the NAM sentences, many of the sentences could be judged as having relatively low interword predictability.) Subject 8, on the other hand, had the most severe hearing loss within the group (thin solid line in Figure 1) and had difficulty both understanding instructions and using the computer.

Although there were significant improvements on the trained words in isolation, the ability to generalize knowledge gathered from word-based training to running speech would be a better test regarding the benefits of an auditory training protocol. Based on the results shown with isolated words, it might be expected that posttraining performance would be dependent on the number of trained keywords versus the overall number of novel keywords—that is, generalization to running speech may occur for a trained word presented by an unfamiliar talker but might not occur for novel (untrained) words in the same sentence. In this case, only about 6%–9% of the sentence keywords were words in which the listeners received training.

Because the mean sentence improvement was in line with the overall number of trained keywords, the proportion correct pretraining to posttraining for the approximately 9% trained hard keywords was examined. Figure 9 provides scatterplots of these individual data when only the trained keywords in the sentences were examined. Each panel represents a separate sentence set (A or B). Although the proportions of trained words pre- to posttraining show a general trend toward improved performance, with the exception of a

couple of participants, the mean hard-word improvement was 7.9% for Set A (range = -12.5% to 21.7%) and 5.8% for Set B (range = -20% to 17.5%). Easy word performance was not examined because so many baseline scores were near ceiling for these keywords. The small number of trained keywords makes interpretation difficult, but these data point to the potential for improved keyword performance within sentences. Again, this observation is generally consistent with the notion that the learning is lexical in nature, confined to the specific words within a sentence that have been trained. Given the format of the sentences and the small percentage of trained keywords comprising the sentences, improved recognition of the trained keywords did not assist in the recognition of untrained keywords.

Retention

“Aided” word-recognition performance in noise and, to some degree, individual sentence keyword identification improved for some listeners during the laboratory training protocol. It is important to examine retention of the trained materials once the active training protocol was complete. There are two primary issues that could cause a decline in performance from the listener’s maximum posttraining scores: time and competition from newly introduced materials.

When listeners completed their hard-word training, an average time of 7 weeks passed from the midpoint evaluation (after training on hard words) to the post-training evaluation administered upon completion of the easy word training. Figure 10 shows the mean open-set word recognition performance across time for both the hard (filled circles) and easy (unfilled circles) words. Regarding the hard words, open-set recognition improved from about 30% to about 80% following hard-word training. Subsequent easy-word training resulted in a drop in hard-word performance of about 15%. Because this drop in performance followed the easy-word training, it was unknown whether the drop was due to time, interference from the new easy words, or a combination of the two. This issue could best be addressed by following the listeners’ retention over the same 7-week time period when no training was being conducted. Because the most recently completed training made use of easy words, following the posttraining measures, listeners’ performance for the recently trained easy words was measured weekly for 7 weeks. Easy word-recognition performance dropped by an average of 4.4% over the 7-week time frame. Next, listener’s hard-word open-set recognition scores were measured and, if significantly below the corresponding score at the midway evaluation, additional hard-word training (one to three sessions) was provided. Weekly measures of open-set hard word-recognition in noise were obtained, and the mean data for 6 of the 8 (2 withdrew from the study) listeners is on the far right in Figure 10. Although the mean scores from week to week varied by approximately 4%, there was essentially no change in open-set hard-word recognition over the 7 weeks. The stability of both the easy and hard word scores over the respective 7-week retention periods, compared with the much larger drop of 13.4% in open-set hard-word recognition after training on easy words, points to interference and not time as the dominant factor in reducing overall performance for the hard words after easy-word training.

As discussed previously, Burk et al. (2006) found no significant drop in performance on a set of 75 monosyllabic words after training on a new set of 75 words. Because the training protocol was only 1.5 weeks long, time was not an issue, although interference may have been. Although there was no drop in performance, overall word recognition performance was lower due to the more limited training protocol, absent of any feedback. The reduced training time limited maximum performance, in turn possibly limiting the drop in performance upon the inclusion of a new set of words. In summary, there appears to be an interference effect that, when examined across studies, is relatively small, amounting to a loss of 2.7%–13.4% following training with a second set of 75 words. The drop appears to

be a little bigger when more time is spent in learning both sets of words, but it does not appear that the amount of time lapsed is responsible for the loss.

Although introducing new sets of words does seem to affect performance to some degree on previously trained words, listeners were able to quickly match their previous posttraining performance with only a small amount of retraining. As stated previously, when the listeners returned on Week 8 of the retention study to measure hard-word retention over a 7-week time frame, they were first retrained to within the 95% critical difference of their maximum hard word open-set performance at the midpoint evaluation. This retraining took only four blocks (approximately 1 hr) for 50% of the listeners, with a maximum of 12 blocks of the 75 words for 2 of the listeners. The ability to quickly return to posttraining levels of performance was also shown in Burk et al. (2006). They measured performance following training on monosyllabic words after a 6-month break. These listeners were not only performing significantly better than their baseline performance after 6 months but maintained over half of the improvement measured immediately following training. Even after 6 months, they were again able to be retrained to within the 95% critical difference in about 1 hr, or four blocks of 75 words.

Based on this information, it would be beneficial for auditory word-based training protocols to periodically review previously trained words. These review sessions or blocks would help in eliminating any decreases in performance due to either time or interference, as any originally trained words would never be completely expunged from memory.

Summary

This experiment was designed as an extension of previous work in the ARL using words as a basis for an auditory training protocol. The primary goals of this training protocol were to increase generalization to both words and sentences over previous training protocols (Burk & Humes, 2007; Burk et al., 2006) and to examine long-term retention of the trained materials. First, when examining improvements in isolated word recognition, listeners showed a large open-set improvement on the trained set of 75 hard words ($\approx 40\%$) while still maintaining a significant improvement over baseline when trained on a new set of 75 easy words. Listeners maintained these improvements for the 150 total words in an open-set condition for upwards of 3.5 months. As has been shown previously, both training using 1 talker (Burk et al., 2006) and training using the multiple talkers in this experiment and in Burk and Humes (2007) generalized well to the same trained words presented by unfamiliar talkers. Although generalization to unfamiliar talkers has been a consistent finding, generalization to new words has been minimal through the varied training protocols, whether in isolation or within sentences.

Listeners showed improved performance for trained keywords when embedded within a sentence as well as the ability to effectively recognize what were now 150 words in noise. If this training protocol was extended further to incorporate more individual words, particularly commonly used words, there is potential for improvements in running speech. Listeners have shown that with a brief refresher, they are able to maintain their performance on previously trained words while learning a new set of words. If the number of trained words could be extended beyond 150, to perhaps hundreds of words, sentence recognition could improve to a greater degree, particularly if the trained words consisted of the most frequently used words in everyday speech.

It is important to note that the current focus has been on collecting data in order to isolate factors that may be of importance when creating an auditory training protocol for older listeners with hearing impairment. Current work within the ARL has pointed to the

importance of longer training regimens (i.e., weeks versus hours), and therefore a tradeoff has been made in choosing to focus on smaller groups of listeners taking part in longer training protocols versus larger groups taking part in short training regimens. However, as increasing knowledge of what is or is not effective emerges, replication with larger groups of listeners will be necessary prior to generalization to other populations. Once the materials and protocol have been established through laboratory research, it will be possible to create a version of the training paradigm that can be used conveniently at home by older adults. This will enable older adults to devote even more time to training than in these laboratory-based studies and would also increase the frequency of such training beyond that possible with repeated visits to the laboratory (for example, daily training for 60–90 min).

Acknowledgments

This work was supported by National Institutes of Health Grant R01 AG08293, awarded to the second author. The authors would like to thank Lauren Strauser and Charles Barlow for their assistance in data collection.

References

- American National Standards Institute. Specifications for audiometers (ANSI S3.6-1996). New York: Author; 1996.
- American National Standards Institute. Maximum permissible ambient noise levels for audiometric test rooms, (S3.1-1999). New York: Author; 1999.
- Bell T, Wilson R. Sentence recognition materials based on frequency of word use and lexical confusability. *Journal of the American Academy of Audiology*. 2001; 12:514–522. [PubMed: 11791938]
- Blamey P, Alcantara J. Research in auditory training. *Journal of the Academy of Rehabilitative Audiology*. 1994; (Suppl 27):161–191.
- Boothroyd, A. Speech perception tests and hearing-impaired children. In: Plant, G.; Spens, KE., editors. *Profound deafness and speech communication*. London: Whurr Publishers; 1995. p. 345–371.
- Boothroyd, A. *Computer-Assisted Speech Perception Assessment (CASPA v2.2)*. San Diego, CA: Author; 1999.
- Burk M, Humes L. Effects of training on speech-recognition performance in noise using lexically hard words. *Journal of Speech, Language, and Hearing Research*. 2007; 50:25–40.
- Burk M, Humes L, Amos N, Strauser L. Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners. *Ear and Hearing*. 2006; 27:263–278. [PubMed: 16672795]
- Frisina D, Frisina R. Speech recognition in noise and presbycusis: Relations to possible neural mechanisms. *Hearing Research*. 1997; 106(1–2):95–104. [PubMed: 9112109]
- Gordon-Salant S, Fitzgibbons P. Recognition of multiply degraded speech by young and elderly listeners. *Journal of Speech and Hearing Research*. 1995; 38:1150–1156. [PubMed: 8558883]
- Luce P, Pisoni D. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*. 1998; 19:1–36. [PubMed: 9504270]
- Montgomery A, Walden B, Schwartz D, Prosek R. Training auditory-visual speech reception in adults with moderate sensorineural hearing loss. *Ear and Hearing*. 1984; 5:30–36. [PubMed: 6706024]
- Rubinstein A, Boothroyd A. Effect of two approaches to auditory training on speech recognition by hearing-impaired adults. *Journal of Speech and Hearing Research*. 1987; 30:153–160. [PubMed: 3599947]
- Seewald R, Hudson S, Gagne J, Zelisko D. Comparison of two methods for estimating the sensation level of amplified speech. *Ear and Hearing*. 1992; 13:142–149. [PubMed: 1397753]
- Studebaker G. A rationalized arcsine transform. *Journal of Speech and Hearing Research*. 1985; 28:455–462. [PubMed: 4046587]

- Sweetow R, Palmer CV. Efficacy of individual auditory training in adults: A systematic review of the evidence. *Journal of the American Academy of Audiology*. 2005; 16:494–504. [PubMed: 16295236]
- Takayanagi S, Dirks D, Moshfegh A. Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. *Journal of Speech, Language, and Hearing Research*. 2002; 45:585–597.
- U.S. Department of Veterans Affairs. Veterans Administration Sentence Test (VAST). Mountain Home, TN: VA Medical Center; 1998.
- Walden B, Erdman S, Montgomery A, Schwartz D, Prosek R. Some effects of training on speech recognition by hearing-impaired adults. *Journal of Speech and Hearing Research*. 1981; 24:207–216. [PubMed: 7265936]
- Walden B, Prosek R, Montgomery A, Scherr C, Jones C. Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research*. 1977; 20:130–145. [PubMed: 846196]

Figure 1.
Pure tone thresholds (in dB HL) for the right ear of the 8 older listeners with hearing loss.
Mean threshold shown by solid thick bar.

Figure 2. Example of the amplification provided to maintain an audible signal of at least 20 dB SL from 200 Hz to 4000 Hz for the mean hearing loss of the 8 older listeners. LTA = long-term average.

Figure 3.

Closed-set word-identification performance during long-term training (in percent correct) for hard and easy words as a function of training block. The data have been smoothed by taking a running average of each listener's performance based on his/her preceding seven trials. For example, the first data points for each participant at Training Block 8 are the average of Blocks 1–7. SNR = signal-to-noise ratio.

Figure 4.

Open-set word-recognition and closed-set word-identification performance pretraining versus either the mid- or posttraining evaluation of lexically hard words and lexically easy words. Error bars represent 1 *SD*. *Differences between conditions marked by horizontal bars significant at the .0042 (.05/12) level.

Figure 5. Open-set word-recognition performance pretraining versus posttraining after long-term training on lexically hard words and lexically easy words presented by familiar and unfamiliar talkers. Significant differences identified compare posttraining performance to pretraining performance. Error bars represent 1 *SD*. *Significant at the .0125 (.05/4) level.

Figure 6.

Pretraining to posttraining open-set word-recognition performance on a novel set of 52 lexically easy/hard words (26 easy, 26 hard), spoken by a single familiar talker used throughout the training and a single unfamiliar talker excluded from the training. Error bars represent 1 *SD*. Alpha level .025 (.05/2).

Figure 7. Percentage of lexically easy and lexically hard keywords identified correctly within Sentence Set A and Sentence Set B pre-training versus posttraining. Error bars represent 1 *SD*. Alpha level .0125 (.05/4).

Figure 8.

Individual difference scores in RAU pretraining to posttraining 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 horizontal dotted line, and the 90% critical difference is indicated by the horizontal solid line.

Figure 9.

Top: The proportion of correctly recognized easy and hard trained words within Sentence Set A, pretraining to posttraining. Bottom: The proportion of correctly recognized easy and hard trained words within Sentence Set B, pretraining to posttraining.

Figure 10.

Open-set word recognition performance for the trained hard and easy words as a function of time in weeks posttraining. Performance is tracked from the pretraining baseline, through training on the respective word lists, to 7 weeks after completion of the training protocol.

Table 1

Presentation conditions for each testing session.

Session	Talker #	Familiar talker	List	# Items	Response	Feedback
Baseline pretraining (1.5 hr)	Female	No	Hard sentences	40	Open set	No
	Female	No	Easy sentences	40	Open set	No
Hard-word training (12–15 hr)	4	Yes	Easy/hard words	52	Open set	No
	12	No	Easy/hard words	52	Open set	No
	1–6	Yes	Hard words	75	Open set	No
	1–6	Yes	Easy words	75	Open set	No
	1–6	Yes	Hard words	75	Closed set	No
	1–6	Yes	Easy words	75	Closed set	No
	1–6	Yes	Hard words	6 × 75	Closed set	Yes
	1–6	Yes	Hard words	75	Open set	No
	1–6	Yes	Easy words	75	Open set	No
	1–6	Yes	Hard words	75	Open set	No
Midpoint evaluation (1.5 hr)	7–11	No	Hard words	75	Open set	No
	7–11	No	Easy words	75	Open set	No
	1–6	Yes	Hard words	75	Closed set	No
	1–6	Yes	Easy words	75	Closed set	No
	1–6	Yes	Easy words	6 × 75	Closed set	Yes
	1–6	Yes	Hard words	75	Open set	No
	1–6	Yes	Easy words	75	Open set	No
	1–6	Yes	Hard words	75	Open set	No
	1–6	Yes	Easy words	75	Open set	No
	1–6	Yes	Hard words	75	Open set	No
Easy-word training (12–15 hr)	1–6	Yes	Easy words	75	Closed set	No
	1–6	Yes	Easy words	6 × 75	Closed set	Yes
	1–6	Yes	Hard words	75	Open set	No
	1–6	Yes	Easy words	75	Open set	No
	1–6	Yes	Hard words	75	Open set	No
	1–6	Yes	Easy words	75	Open set	No
	1–6	Yes	Hard words	75	Closed set	No
	1–6	Yes	Easy words	75	Closed set	No
	1–6	Yes	Hard words	75	Closed set	No
	1–6	Yes	Easy words	75	Closed set	No
Posttraining evaluation (1.5 hr)	1–6	Yes	Hard words	75	Open set	No
	1–6	Yes	Easy words	75	Open set	No
	1–6	Yes	Hard words	75	Open set	No
	1–6	Yes	Easy words	75	Open set	No
	1–6	Yes	Hard words	75	Closed set	No
	1–6	Yes	Easy words	75	Closed set	No
	1–6	Yes	Hard words	75	Closed set	No
	1–6	Yes	Easy words	75	Closed set	No
Generalization novel words (1.5 hr)	Female	Repeated	Hard sentences	40	Open set	No
	Female	Repeated	Easy sentences	40	Open set	No
	Female	No	Hard sentences	40	Open set	No
	Female	No	Easy sentences	40	Open set	No
	4	Yes	Easy/hard words	52	Open set	No
	12	No	Easy/hard words	52	Open set	No
	12	No	Easy/hard words	52	Open set	No
	12	No	Easy/hard words	52	Open set	No

Note. “Familiar talker” indicates whether the specific talker(s) were used within the training sessions. A single female talker, who was not included in the group of 12 talkers, presented all sentences.

Table 2
Signal-to-noise ratio used for each individual participant through all aspects of the training protocol.

Participant	Preliminary		Baseline and hard-word training		Easy-word training		Posttraining	
	Words	Sent.	Words	Sent.	Words	Sent.	Words	Sent.
1	0	-2	0	-2	-6	0	0	-2
2	0	-2	0	-2	-3	0	0	-2
3	0	-2	0	-2	-3	0	0	-2
4	0	-2	+3	+1	-3	0	0	-2
5	0	-2	+3	+1	0	0	0	-2
6	0	-2	+3	+1	0	0	0	-2
7	0	-2	+3	+1	0	0	0	-2
8	0	-2	+3	+1	0	0	0	-2

Note. Sent. = sentences. There is no “Sent.” column under “Easy-word training,” because that condition was not run.