



Published in final edited form as:

Ear Hear. 2009 October ; 30(5): 613–627. doi:10.1097/AUD.0b013e3181b00d90.

Development and Efficacy of a Frequent-Word Auditory Training Protocol for Older Adults with Impaired Hearing

Larry E. Humes, Ph.D., Matthew H. Burk, Ph.D., Lauren E. Strauser, M.A., and Dana L. Kinney, M.A.

Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN 47405-7002

Abstract

Objectives—The objective of this study was to evaluate the efficacy of a word-based auditory-training procedure for use with older adults who have impaired hearing. The emphasis during training and assessment is placed on words with a high frequency of occurrence in American English.

Design—A repeated-measures group design was used with each of the two groups of participants in this study to evaluate the effects of the word-based training regimen. One group was comprised of 20 young adults with normal hearing and the other consisted of 16 older adults with impaired hearing. The group of young adults was not included for the purpose of between-group comparisons. Rather, it was included to demonstrate the efficacy of the training regimen should efficacy fail to be demonstrated in the group of older adults and also to estimate the magnitude of the benefits that could be achieved in younger listeners.

Result—Significant improvements were observed in the group means for each of five measures of post-training assessment. Pre-training and post-training performance assessments were all based on the open-set recognition of speech in a fluctuating speech-like background noise. Assessment measures ranged from recognition of trained words and phrases produced by talkers heard during training to the recognition of untrained sentences produced by a talker not encountered during training. In addition to these group data, analysis of individual data via 95% critical differences for each assessment measure revealed that 75–80% of the older adults demonstrated significant improvements on most or all of the post-training measures.

Conclusions—The word-based auditory-training program examined here, one based on words having a high frequency of occurrence in American English, has been demonstrated to be efficacious in older adults with impaired hearing. Training on frequent words and frequent phrases generalized to sentences constructed from frequently occurring words whether spoken by talkers heard during training or by a novel talker.

In the 2000 U.S. Census, 35 million Americans over 65 years of age were counted, representing 12.4% of the U.S. population (Hetzl & Smith, 2001). Approximately 30% of those over age 65 in the U.S., or more than 9 million older Americans, have a significant hearing loss that is sufficient to make them hearing-aid candidates (Schoenborn & Marano, 1988). Yet, only about 20% of those older Americans who could benefit from hearing aids actually seek them out and, of those who do seek them out, only about 40–60% are satisfied with them and use their hearing aids regularly (Kochkin, 1993a, 1993b, 1993c, 2000, 2005). These figures, moreover, are typical of those in other countries, such as the U.K. (Smeeth et al., 2002) and Australia (Ward et al., 1993).

The most common communication complaint of older adults with impaired hearing is that they can hear speech, but can't understand it. This is especially true when there are competing sounds, typically other speech sounds, in the background. Hearing aids, to be effective, must address this common complaint and improve speech understanding, especially in backgrounds of competing speech. The primary function of hearing aids is to provide level-dependent and frequency-dependent gain to the acoustic input arriving at the hearing aid microphone(s). The gain provided has the potential to improve speech communication by restoring the audibility of speech sounds rendered inaudible by hearing loss (e.g., Humes, 1991; Humes & Dubno, in press). For older adults, it is mainly higher frequency lower amplitude consonants (e.g., s, f, t, "sh", and "th") that are made inaudible by hearing loss (e.g., Owens, Benedict & Schubert, 1972).

The hearing loss of older adults is greatest in the frequency region for which the amplitude of speech is the lowest; specifically, frequencies ≥ 2000 Hz (ISO, 2000). As a result, a large amount of hearing-aid gain is required in the higher frequencies to restore full audibility of the speech signal through at least 6000 Hz. This is challenging for the clinician to accomplish, even with the use of amplitude compression. In fact, this difficulty has led to the development of alternatives to conventional amplification, such as frequency compression hearing aids or short-electrode cochlear implants (e.g., Turner & Hurtig, 1999; Simpson, Hersbach & McDermott, 2005; Gantz, Turner & Gfeller, 2006), for adults with *severe or profound* degrees of high-frequency hearing loss. Even for many older adults with less severe high-frequency hearing loss, for whom conventional amplification may be the best alternative, restoration of optimal audibility will not always be possible in the higher frequencies (≥ 2000 Hz; e.g., Humes, 2002). Under such circumstances of less-than-optimal audibility, a better-than-normal speech-to-noise ratio is required to achieve "normal" or near-normal speech-understanding performance; that is, to achieve performance that is equivalent to that of young normal-hearing adults (e.g., Lee & Humes, 1993; Humes, 2008; Humes & Dubno, in press). Nonetheless, even with the less-than-optimal audibility provided by hearing aids fit in clinical settings (e.g., Humes et al., 2000; Humes, 2002; Humes et al., 2004), nonrandomized interventional studies of hearing-aid outcomes in older adults have demonstrated that, on average, clinically-fit bilateral hearing aids provide significant and long-term benefit. This benefit has been demonstrated in quiet and in backgrounds of steady-state noise or multi-talker babble (Humes et al., 1997, 1999, 2001, 2002a, 2002b, 2004, 2009; Larson et al., 2000), although aided speech-recognition performance is seldom restored to "normal". Moreover, in laboratory studies of aided speech understanding *with optimal audibility* achieved with earphones, it has been observed that older adults frequently require a better-than-normal speech-to-noise ratio to perform the same as young normal-hearing adults (Humes, Lee, & Coughlin, 2006; Jin & Nelson, 2006; George et al., 2006, 2007; Humes, 2007; Humes et al., 2007; Amos & Humes, 2007). There are many possible reasons underlying the need for a better-than-normal speech-to-noise ratio by older adults wearing hearing aids, including peripheral, binaural/central auditory, and cognitive factors, but there is mounting evidence that the deficits are cognitive in nature, rather than modality-specific deficits in central-auditory function, especially for fluctuating background sounds, including competing speech (Lunner, 2003; George et al., 2006, 2007; Humes, 2002, 2005, 2007; Humes et al., 2006, 2007; Pichora-Fuller & Singh, 2006; Lunner & Sunderwall-Thoren, 2007; Foo et al., 2007; Ronnberg et al., 2008; Humes & Dubno, in press).

As noted previously, much of everyday communication takes place in backgrounds of competing speech and older adults typically express frustration with this listening situation. Further, hearing aids alone don't address this situation adequately. Although contemporary digital hearing aids implement various noise-reduction strategies or directional-microphone technologies, for the most part, these approaches have had only limited success in improving

the speech-to-noise ratio and the benefit to the hearing-aid wearer (e.g., Bentler, 2005; Walden et al., 2004; Cord et al., 2004; Nordrum et al., 2006). This is especially true when the competing stimulus is comprised of speech produced by one or more talkers; particularly, when the target talker at the moment may become the competing talker a moment later. Thus, there are real limits to how much the speech-to-noise ratio can be improved acoustically by the hearing aids in many everyday circumstances.

To recap, there is considerable evidence that older adults listening to amplified speech require a better-than-normal speech-to-noise ratio to achieve “normal” or near-normal speech understanding. Hearing aids alone, however, have been unable to deliver the required improvements in speech-to-noise ratio. Hearing aids do appear to provide significant and long-term benefits, but aided performance often remains less than ideal, especially in the presence of competing speech. This, in turn, appears to lead to the discouraging statistics cited at the beginning of this article about market penetration, hearing aid usage, and hearing aid satisfaction among the elderly.

Given the foregoing, we decided to explore ways to improve the ability of older adults to understand amplified speech for a given speech-to-noise ratio, rather than attempt to improve the acoustical signal-to-noise ratio. (Of course, there is no reason that these two strategies, one focused on the listener and the other on the hearing aid, can’t work in concert.) Our focus was placed on a novel approach to intervention; specifically, a word-based auditory training method with all training conducted in noise backgrounds. Some of the unique features of this novel approach included that: (1) it was word-based, which placed the focus on meaningful speech rather than on sub-lexical sounds of speech, such as individual phonemes; (2) the training protocol was based on closed-set identification of words under computer control, which enabled automation of presentation, scoring, and feedback; (3) multiple talkers were included in training, which facilitated generalization to novel talkers; (4) all training was conducted in noise, the most problematic listening situation for older adults; and (5) following correct/incorrect feedback on every trial, both auditory and orthographic feedback associated with the correct and incorrect responses were provided to the listener following incorrect responses.

Although the word-based auditory-training method has many novel features that make it unique among auditory-training procedures, it is important to note that auditory training has had a long history in audiology. For the most part, however, the focus historically has been placed on young children with profound or severe-to-profound hearing loss. Limited efforts have been directed to adults with impaired hearing, with still fewer studies directed toward *older* adults with impaired hearing. A recent evidence-based systematic review of the literature on the benefits of auditory training to adult hearing aid wearers (Sweetow & Palmer, 2005) found only six studies that met criteria for valid scientific evidence and subsequent analysis, three of which were conducted with older adults. Following review of these six studies, Sweetow and Palmer (2005) concluded that “...this systematic review provides very little evidence for the effectiveness of individual AT (auditory training). However, there is some evidence for efficacy.”(Pg. 501). The distinction made by these authors is that benefits of auditory training had been observed under optimal conditions in a restricted research environment (efficacy), but not under routine clinical or field conditions (effectiveness), the latter typically requiring some form of clinical-trial research. Sweetow & Palmer (2005) also noted that there was some indication that synthetic approaches, those involving words, phrases, sentences or discourse, were more likely to be efficacious than analytic approaches focusing on sublexical features of speech.

Since the publication of this systematic review in 2005, two developments regarding the efficacy and effectiveness of auditory training in older adults with hearing aids are

noteworthy. First, Sweetow and Sabes (2006) published data from a multi-center clinical evaluation of a cognitive-based top-down auditory-training system referred to as Listening and Auditory Communication Enhancement (LACE; Sweetow and Henderson-Sabes, 2004). The short-term benefits of the top-down approach pursued by LACE in older adults were encouraging. Second, a series of articles was published describing the development, evaluation, and efficacy of our automated word-based approach to auditory training (Burk, Humes, Strauser & Amos, 2006; Burk & Humes, 2007, 2008).

Through a series of laboratory studies evaluating this word-based auditory training protocol (Burk et al., 2006; Burk & Humes, 2007, 2008), we have learned the following: (1) older adults with impaired hearing could improve their open-set recognition of words in noise from about 30–40% correct before training to 80–85% correct following training; (2) training generalized to other talkers saying the same trained words, but only slight improvements (7–10%) occurred for new words, whether spoken by the talkers heard during training or other talkers; (3) improvements from training, although diminished somewhat by time, were retained over periods as long as 6 months (maximum retention interval examined to date); (4) similar gains were observed when the feedback was either entirely orthographic (displaying correct and incorrect responses on the computer screen) or a mix of orthographic and auditory (re-hearing the correct and incorrect words following incorrect responses), but not when the feedback was eliminated entirely (Burk et al., 2006); and (5) there was little transfer of the word-based training to the open-set recognition of sentences in noise. The methodological details varied from study to study, but some common features included the use of closed-set identification of vocabularies of 50–150 words spoken by multiple talkers in steady-state noise during training and the assessment of training effectiveness with these same words, as well as additional words and sentences spoken by the same talkers used in training and by novel talkers. In addition, for all the experiments with older adults with impaired hearing, the spectra of the speech and noise were shaped and the stimuli delivered via earphones to ensure sufficient audibility of the long-term spectrum of the speech stimulus through *at least* 4000 Hz.

We have interpreted this general pattern of findings, confirmed several times now, to suggest that the training process is primarily lexical in nature and serves to reinforce the link between the degraded encoding of the acoustic input and the intact phonological representation of that input in the listener's lexicon. This link may have been weakened through years of gradual hearing loss or through some other aspect of aging. Although training with multiple talkers enhances generalization of the training to other talkers, the lexical nature of the training limits the generalization of trained words to other novel words. This appeared to be the primary reason for the failure of the word-based training to generalize to sentences. In retrospect, we realized that there was very little overlap (about 7–9%) between the words used in training and the words comprising the sentences used to assess generalization after training. When only those words in the sentences that had been used as training words were examined, much larger improvements in performance were observed as a result of training (Burk & Humes, 2008).

Given the foregoing, our most recent work in the area of word-based auditory training is described in this report and has focused on the most frequently occurring words in spoken (American) English. The rationale for this emphasis is that, if the training is word-specific, then the focus should be placed on words used most frequently in everyday conversation. A relatively small number of words, 400–600, represent 80–90% of the words used most frequently in spoken conversation (e.g., French et al., 1931; Godfrey, Holliman & McDaniel, 1992). Having demonstrated that older adults showed significant improvements in the recognition of spoken words in noise for sets of words as large as 150, we explored the expansion of set size by a factor of four to 600 words. If older adults could improve with

this large set, then the use of top-down contextual processing, an ability that does not appear to diminish with age (e.g., Humes et al., 2007), should help fill in the gaps for the untrained words encountered in everyday conversation. We also incorporated some limited training with frequently occurring phrases as a syntactic bridge between isolated words and sentences. In addition, given emerging evidence regarding the special difficulties of older adults with fluctuating, speech like competition, the two-talker version of ICRA noise (Dreschler et al., 2001) was used as the competition in this experiment. Finally, the words comprising the sentences used to assess generalization of training this time had much more overlap (50–80%) with the training vocabulary.

METHODS

Participants

There were two primary groups of participants in this study: young adults with normal hearing (YNH) and older adults with impaired hearing (OHI). Each of these two groups was further divided into two subgroups based on the particular version of the training protocol each received. The differences in the two training protocols are described below and are referred to here simply as Protocol 1 and Protocol 2. Table 1 summarizes the ages, gender, and test ear for the participants in all subgroups and also provides pure-tone averages for the OHI subgroups. Independent-sample t-tests showed that there were no significant differences ($p > 0.1$) in age between the two YNH or between the two OHI subgroups. The mean pure-tone averages for the two OHI subgroups did not differ significantly ($p > 0.1$). Figure 1 provides the mean air-conduction pure-tone thresholds for the test ear for the two OHI subgroups. All told, there were 20 YNH and 16 OHI participants in this study.

Stimuli and Materials

There were a total of four types of speech materials used in the evaluation and training portions of this study: (1) frequent words; (2) frequent phrases; (3) re-recordings of select sentences from the Veterans Administration Sentence Test (VAST; Bell & Wilson, 2001); and (4) the Auditec recording of CID Everyday Sentences (Davis & Silverman, 1978). Since only the last of these sets of materials were previously developed and commercially available, the other three sets of materials are described in detail here.

The frequent words, frequent phrases, and modified VAST sentences were recorded by the same four talkers (A, B, C and D), all of the Midland North American English dialect. Dialect was established simply by the geographical birthplace and residency of the talkers and their parents. Two talkers (B, D) were female, aged 24 and 34 years, and two talkers (A, C) were male, both 23 years of age. All four talkers had normal hearing and speech and none of them were professional speakers. Materials were recorded in a double-walled sound room with an Audio Technica AT3035 cardioid capacitor microphone positioned on a table top, in a supplied custom shock mount and stand, approximately 9 inches from the talker's mouth. An LCD computer monitor located about 18–24 inches in front of the talker displayed the word, phrase or sentence to be produced prior to each recording and also provided a digital VU meter to monitor the voice level during production. The talker was instructed to produce each item clearly but in a conversational manner. Three successive recordings of each stimulus were produced by the talker and the "Enter" key on the computer keyboard was pressed when the third production was completed. If the recording program detected peak clipping or a poor signal-to-noise ratio, three repetitions of the stimulus item were re-recorded immediately. The output of the microphone was pre-amplified (Symetrix Model 302) prior to digitization by a Tucker-Davis Technologies (TDT) 16-bit analog-to-digital converter at a sampling rate of 44.1 kHz.

Following completion of the recordings, the three tokens of each stimulus were listened to, and the respective waveforms examined, using Adobe Audition 2.0. Based on the subjective listening experience and the visual inspection of the waveforms, a trained research assistant selected the best token among the three or, if no discernible differences were observed among the tokens, selected the middle of the three tokens. The selected token was edited at zero crossings of the waveform to remove silence at the beginning or end of the word, phrase or sentence. Following this editing, a second research assistant or one of the authors examined the waveforms and spectrograms of the final set of recordings and listened to each. In some cases, recordings were designated as being unacceptable by the first auditor, the second auditor, or both. In any of these cases, the recorded items were noted and subsequently re-recorded by the talker. This process of recording by the talker and auditing by research assistants in the laboratory was repeated iteratively as needed until a complete set of materials was available for each of the four talkers. Once the stimuli had been finalized, Adobe Audition 2.0 was used to normalize each stimulus to a peak level of -3 dB, low-pass-filter (18-order, Butterworth) each at 8000 Hz, and then re-sample each stimulus at 48.828 kHz (a rate compatible for playback via the TDT equipment). For each set of materials (words, phrases, and sentences), all stimulus files were concatenated into one long file for each talker and the average RMS amplitude was measured using a 50-ms window in Adobe Audition 2.0. Amplitudes of the stimulus materials were adjusted as needed to equate average RMS amplitudes across talkers.

Regarding the frequent words, a total of 1500 most frequently occurring words in English were recorded by each of the four talkers. In the end, only the top 600 most frequently occurring words were used in this study, with proper first names, such as “John”, “Thomas”, “Jane”, etc., eliminated. Homophones, if occurring among the 600 most frequent words, such as “to”, “too” and “two”, were recorded and tested like all other words among the top 600. Such cases, when they occurred, can be likened to multiple tokens of the same sound pattern or spoken word produced by each talker, but mapped to a unique orthographic representation. There are various corpora of most frequently occurring words in English available as resources, some based on written English and others based on oral English. Lee (2003), however, demonstrated that the word frequencies were very similar in both types of corpora and, as a result, recommended using the largest available. To this end, the corpus of Zeno et al. (1995) based on over 17 million words was used to establish word frequency. Since this corpus is based on written English, comparisons were made between the 600 most frequently occurring words from this corpus (with proper names eliminated) and an old (French, Carter & Koenig, 1931) and more recent (Switchboard Corpus; Godfrey et al., 1992) word-frequency corpus developed from recordings of telephone conversations. There was 60–70% overlap among these three corpora, consistent with the findings of Lee (2003) based on other corpora. Although the majority of words were comprised of one or two syllables, there were 20 three-syllable and 5 four-syllable words among the 600 most frequently occurring words.

Pilot testing was completed for five young normal-hearing listeners for the final versions of the 600 words recorded by four talkers. This open-set word-recognition testing was completed at a 0-dB signal-to-noise ratio with the same noise that was used in this study (see below). Two additional listeners were tested at a -5 dB signal-to-noise ratio to get some idea as to performance levels and shape of the psychometric function. The mean percent-correct scores for the 2400 word stimuli (4 talkers X 600 words) were 46.1% for the two YNH listeners tested at a signal-to-noise ratio of -5 dB and 69.4% for the other five YNH listeners tested at a signal-to-noise ratio of 0 dB, indicating slope of about 5%/dB in the linear portion of the psychometric function for YNH listeners. The larger set of pilot data for the 0-dB signal-to-noise ratio was used to evaluate the relative difficulty of the 600 words. Since each word was produced by four talkers, a word was somewhat arbitrarily tabulated as

being correct for a given listener if that word was correctly recognized for productions by three or more of the talkers. Otherwise, the word was tabulated as incorrect for that listener. Words were then partitioned into six different groups to reflect their general difficulty for the listeners: (1) words tabulated as incorrect for all 5 listeners (most difficult); (2) words tabulated as incorrect for 4 of the 5 listeners; (3) words tabulated as incorrect for 3 of the 5 listeners; (4) words tabulated as incorrect for 2 of the 5 listeners; (5) words tabulated as incorrect for 1 of the 5 listeners; and (6) words tabulated as incorrect for 0 of the 5 listeners (easiest). Of course, the actual number of words falling into each of these categories will vary with the signal-to-noise ratio. The 0-dB signal-to-noise ratio, however, yielded a reasonable distribution of words across these six categories of difficulty. The percentages of words falling into each of these six categories were 5, 10, 18, 17, 25 and 25% progressing from the most difficult to the easiest categories, respectively. A total of 147 of the 600 words fell into the easiest category (i.e., all five listeners correctly recognized the word when spoken by at least three of the four talkers). These 147 words were combined with 53 arbitrarily selected words of the 149 from the next easiest category to form a subset of 200 relatively easy to recognize words. These words subsequently received less training during the training protocol than the 400 remaining, generally more difficult words. It should be noted that the 200 easiest words, as defined here for the analysis of these pilot data, were not simply the 200 most frequent words among this set of 600 most frequent words. When all of the words are high in word frequency, word frequency alone is not sufficient to predict performance.

The frequent phrases were obtained from the compilation of contemporary (post 1980) written and oral British and American English published by Biber et al. (1999). The focus for purposes of this project was placed on the analysis of conversational sources by these authors. These conversational samples were obtained from nearly 4 million words produced by 495 speakers of British English and almost 2.5 million words produced by 491 speakers of American English, with roughly equal distributions of the British and American talkers regarding talker gender and age decades up to 60 year olds. A unique feature of the Biber et al. (1999) analysis is the tabulation of the frequency of occurrence of “lexical bundles”; recurrent sequences of words, such as “I don’t” or “do you want”. The authors note that three-word and four-word bundles are extremely common in conversational English. For example, three-word bundles occur over 80,000 times per million words in conversation and four-word bundles over 8,500 times. This represents nearly 30% of the words in conversation that occur in recurrent lexical bundles or phrases. In developing these materials for this project, we selected the most frequently occurring four-word and five-word lexical bundles from Biber et al. (1999) and, in some cases, modified these to form a more complete phrase by adding frequently occurring words to the beginning or end of the bundle. For example, “going to have a” is among the most frequently occurring four-word lexical bundles in Biber et al. (1999). To this bundle, the word “family”, itself a frequently occurring word, was appended to the end to form a five-word phrase: “going to have a family”. In the end, a total of 94 frequently occurring phrases, 36 four-word and 58 five-word phrases, were selected for use in this study. Eighteen frequently occurring words comprised roughly 70% of the words in the phrases with 84 additional words comprising the remainder for a total of 102 unique words from which the phrases were derived, with all 102 of the words being very frequent in occurrence in isolation. Pilot testing of the phrases with five YNH listeners showed mean open-set recognition of the words comprising the phrases to be 95.4% averaged across all four talkers at a signal-to-noise ratio of -3 dB.

The modified VAST materials represented a subset of the sentences developed with keywords having high frequencies of occurrence. The VAST sentences were developed with three keywords per sentence and the keywords were distributed such that one occurred near the beginning of the sentence, one in the middle of the sentence, and one near the end of the

sentence (Bell & Wilson, 2001). A total of four sets of 120 sentences were developed, with word frequency (high or low) and neighborhood density (sparse or dense) of the keywords varying across subsets. From each of the two subsets comprised of keywords with high frequency of occurrence, the sentences were screened by the first author for semantic redundancy. This was necessary because, to meet the constraints of word frequency, neighborhood density, and keyword sentence position, some of the sentences in the VAST corpus do not have the semantic redundancy common to many other sentence materials. In the end, 25 sentences (having a total of 75 keywords) were selected from each of the two high-frequency subsets based on the semantic redundancy of the sentence. Sentences with high semantic redundancy, such as “The *point* of the *knife* is too *sharp*” or “The *model* wore a *plain* black *dress*” (keywords in italics) were retained, whereas others with less redundancy or “semantic connectedness” among words (e.g., “The *gang* had to *live* in a small *flat*”) were not recorded. This judgment regarding the redundancy among the words comprising each VAST sentence was made by the first author and was entirely subjective in nature. Pilot open-set recognition testing with five YNH listeners for these sentence materials showed a mean percent-correct score of 86.6% averaged across all four talkers at a signal-to-noise ratio of -5 dB.

Lists A, B, C and D of the Auditec recording of the CID Everyday Sentences by a male talker were also used in this study. These sentence materials were chosen because of several constraints imposed during their development (Davis & Silverman, 1978), including the use of words with a high frequency of occurrence in English. Each list is comprised of ten sentences with a total of 50 keywords that are scored. There are approximately two short (2–4 words), two long (10–12 words) and six medium-length (5–9 words) sentences in each list. Further, the sentence form varies in a list such that six are declarative, two are imperative, and two are interrogative (one with falling intonation and one with rising intonation).

ICRA Noise (Dreschler, Verschuure, Ludvigsen & Westermann, 2001) was used as the background noise for all testing and training sessions. In particular, this study made use of the two-talker noise-vocoded competition (Track 6 on the ICRA Noise CD), in which one talker is male and the other female. From the 10-minute recording of this noise on the CD, various noise segments were randomly selected (at zero crossings) for presentation during this study with the length of the segments adjusted based on the particular speech materials used (further details below). This particular fluctuating speech-like, but unintelligible, noise was selected for use in this study because this is exactly the type of competition in which older hearing-impaired listeners have considerable difficulty understanding speech.

Equipment and Calibration

For each set of materials recorded in the laboratory (frequent words, frequent phrases, and the modified VAST sentences) by four talkers, each stimulus was normalized to a peak level of -3 dB, upsampled at a rate of 48.828 kHz, low-pass filtered at 8000 Hz, and then equated in average RMS amplitude within Adobe Audition. To measure the long-term-average amplitude spectra for the final set of stimuli, the wave files for a given type of material produced by a given talker were concatenated to form one long wave file. This wave file was then analyzed within Adobe Audition using 2048-point FFTs (Blackmann window) and the long-term average spectrum computed. Similar processing was applied to the two-talker vocoded noise from the ICRA Noise CD (Track 6). Initially, an arbitrary two-minute segment of the ICRA noise was selected and digitized. From this, 20 2-second non-overlapping samples were excised for use as maskers for the word stimuli. Twenty longer 4-second samples of ICRA noise, needed for the longer phrases and sentences, were generated by randomly selecting two of the 2-second noise files for concatenation and generating 20 such pairs. The RMS amplitude and long-term spectra of the ICRA noise was then adjusted to match that of the speech stimuli. Figure 2 displays the long-term amplitude spectra

measured in Adobe Audition for the words (top panel), phrases (middle panel) and sentences (bottom panel) produced by each of the four talkers, as well as the corresponding long-term spectra for the ICRA noise used as competition with these materials.

A steady-state calibration noise was then generated to match the RMS amplitude and long-term spectrum of the ICRA noise used in this study. This noise was then controlled and presented from the same TDT hardware and Matlab software used in this experiment to establish sound pressure levels for presentation. This noise was presented from the TDT 16-bit digital-to-analog converter at a sampling rate of 48.828 kHz, routed through a TDT PA-4 programmable attenuator and a TDT HB-7 headphone buffer prior to transduction by an Etymotic Research ER-3A insert earphone. The acoustic output of the insert earphone for the calibration noise was measured in an HA-2 2-cm³ coupler using the procedure described in ANSI (2004) and a 1-inch microphone coupled to a Larsen-Davis Model 824 sound level meter set for fast response time. With 0 dB attenuation in the programmable attenuator and the headphone buffer, the overall level of the calibration noise was 114.7 dB SPL. The 1/3-octave-band spectrum of the calibration noise is illustrated in Figure 3.

For the young adults with normal hearing, the headphone buffer was increased to the maximum setting of 27-dB attenuation which resulted in a presentation level for all speech materials and talkers of 87.7 dB SPL. For the older adults with impaired hearing, calculations were made based on the 1/3-octave-band levels in Figure 3, an assumed overall sound pressure level for conversational speech of 68 dB SPL (i.e., attenuating the 1/3-octave-band levels in Figure 3 by about 47 dB), and the participant's pure-tone thresholds from his or her audiogram so that the RMS 1/3-octave spectrum was 20 dB above threshold through 4000 Hz. This simulated a well-fit hearing aid that optimally restored audibility of the speech spectrum through at least 4000 Hz and often resulted in the speech signal being at least 10 dB above threshold at 5000 Hz. The 1/3-octave band levels were realized for each participant by using a combination of headphone-buffer setting and digital filtering within the controlling Matlab software. Additional details regarding this approach to spectral shaping can be found in Burk & Humes (2008). Typically, after spectral shaping, the overall sound pressure level of the steady-state calibration stimulus would be approximately 100–110 dB SPL.

Based on pilot testing with these new speech materials, the following speech-to-noise ratios were used for the young adults with normal hearing: (1) –5 dB for frequent words during assessment and training; (2) –8 dB for frequent phrases during assessment and training; and (3) –10 dB for modified VAST sentences during assessment. For most of the older adults with impaired hearing, the following speech-to-noise ratios were used (exceptions noted in parentheses): (1) –2 dB for frequent words during assessment and training (+2 dB for one person during assessment and –5 dB for two older participants during training); (2) –8 dB for frequent phrases during assessment and training (–4 dB for one older adult); and –8 dB for VAST sentences during assessment (–4 dB for one listener). The exceptions noted were based on observations during baseline assessment or during the first few blocks of training for which the participant was performing considerably above or below the expected range of performance (midway between ceiling and floor). Overall, the speech-to-noise ratios were about 2–3 dB better for the older adults during training and assessment. At first glance, these signal-to-noise ratios appear to be very severe. It should be kept in mind, however, that these speech-to-noise ratios are defined on the basis of the steady-state calibration noise whereas the actual competition was the fluctuating two-talker vocoded ICRA noise.

All testing and training was completed in a sound-treated test booth meeting ANSI (1999) ambient noise standards for threshold testing under earphones. Four listening stations were housed within this sound-treated booth with a 17-inch LCD flat-panel touch-screen monitor

located in each listening station. Although all testing was monaural, both insert earphones were worn by the listener during testing with only the test ear activated. This helped to minimize acoustic distractions generated by other participants in the test booth during the course of the experiment. All responses during closed-set identification training were made via this touch screen. For word-based training, 50 stimulus items were presented on the screen at a time, in alphabetical order from top to bottom and left to right, with a large font that was easily read (and searched) by all participants. Additional features of the training protocol, including orthographic and acoustic feedback on a trial-by-trial basis, have been described in detail elsewhere for the word-based training (Burk & Humes, 2008). With regard to the phrase-based training, closed sets of 11 or 12 phrases were displayed on the computer screen at a time and the listener's task was to select the phrase heard. Otherwise, the training for phrases was identical to that for words.

Procedures

Following initial hearing testing, participants were assigned to one of two training protocols that differed in the way the training materials were grouped for presentation during training. Recall that there were a total of 600 frequently occurring words (lexical items), each spoken by four talkers, for a total of 2400 stimulus items for the word-based portion of the training program. As noted in the Introduction, the 600 most frequently occurring words represent about 90% of the words in spoken conversation, but is also a set size that is four times larger than the largest set used previously in our laboratory in similar training studies. This large set size was selected for use in this study in the hopes of establishing a limit for the tradeoff between set size and training benefit. The repetitions of the 2400 stimulus items were sequenced according to two different training protocols. In one protocol, referred to arbitrarily as Protocol 1, the first presentation of all 2400 items was completed before the second presentation of any of the 2400 stimuli occurred, the second presentation of all 2400 stimuli was completed before a third presentation of any of the 2400 stimuli occurred, and so on. The second approach, referred to as Protocol 2, was designed to break this large set of materials into smaller subsets of 600 stimuli (or 150 lexical items), then present multiple repetitions of these stimuli before proceeding to the next subset of 600 stimuli.

Before contrasting the two training protocols in more detail, there were several procedures common to both training protocols and their evaluation and these are reviewed here. Each protocol was preceded by two sessions of baseline testing and followed by two sessions of post-training evaluation. All of this pre- and post-training assessment was open-set recognition testing, unlike the training itself, which was all closed-set identification. Oral responses were provided by the subjects and recorded for off-line analysis and scoring by trained research assistants. For the pre-training baseline, the testing consisted of the presentation of 20 CID Everyday Sentences (Lists A & B) digitized from the Auditec CD with 100 keywords scored, 50 VAST sentences produced by four talkers for a total of 200 sentences with 600 keywords scored, 94 frequent phrases spoken by four talkers for a total of 434 words scored, followed by 200 of the 2400 word stimuli. The 200 word stimuli used for assessment were selected quasi-randomly such that no lexical items were repeated and each of the four talkers produced 50 of the 200 words. The stimuli and order of presentation for the post-training assessment were identical to pre-training except that an extra set of 20 CID Everyday Sentences (Lists C & D) was also presented. The rationale for including an alternate set of sentences was that there was a slight possibility that the listeners might retain some of the sentences from CID Everyday Sentences 1 in memory from the first exposure several weeks earlier, given their high semantic content and the availability of just one talker for those materials, and do better in post-training assessment because of this, rather than due to the intervening training.

As noted, two different protocols were followed for the training sessions. For Protocol 1, an 8-session training cycle was devised. In each of these 8 training sessions, each about 75–90 minutes in duration and administered on separate days, 8 blocks of 50 lexical items were presented, for a total of 400 lexical items, with 100 presented by each of the four talkers. These are the 400 items that were found in pilot testing to be the 400 more difficult words, as noted above. This was followed by 2 blocks of 50 words from the set of 200 easiest items identified in pilot testing with 25% of these 100 stimulus items spoken by each of the four talkers. Finally, the training session for each day concluded with the presentation of four sets of 11–12 phrases with approximately 25% of each set produced by each of the four talkers. Because the training cycle was composed of 8 daily sessions, at the end of a training cycle there had been two repetitions of the 1600 word stimuli (400 words X 4 talkers) found to be more difficult in pilot testing, one repetition of the 800 word stimuli (200 words X 4 talkers) found to be easiest in pilot testing, and one repetition of the 376 phrase stimuli. Protocol 1 was designed so that each participant would complete three 8-session training cycles between pre-training baseline and post-training evaluation. This total of 24 training sessions was typically completed at a rate of 3 sessions per week for a total of 8 weeks of training, but some participants preferred a slower rate and required about 12 weeks for training.

For Protocol 2, 6-session training cycles were implemented. In a given 6-day training cycle, the first four training sessions were devoted to 400 of the 1600 more difficult word stimuli (100 of the 400 corresponding lexical items) with the same 400 stimuli (100 lexical items) each of the four days. This four-day block was followed by a day devoted to training on 400 of the 800 easier word stimuli (or 100 of the 200 easier lexical items) with a subsequent day devoted to training with half of the phrases (188 of the 376 phrase stimuli, 25% by each talker, or 47 of the 94 lexical representations of the phrases). This six-session training cycle was repeated four times with each cycle covering a new 25% of the stimulus corpus. During the training sessions involving phrases at the end of the second and third training cycles, brief refresher sessions (2–6 blocks of 50 stimuli) were added for the word stimuli from earlier cycles. This again resulted in a total of 24 training sessions and, based on the participant's preferences, typically required a total of 8–12 weeks for completion.

In the end, the amount of stimulus repetition during training, including the refresher blocks for the smaller-subset protocol, was roughly equivalent for both protocols. For those completing either protocol, there were 6 repetitions of the 1600 more difficult word stimuli (or 24 repetitions of the 400 corresponding lexical items), 3 repetitions of the 800 easier word stimuli (or 12 repetitions of the 200 corresponding lexical items), and 3 repetitions of the 376 stimulus phrases (or 12 repetitions of the 94 corresponding lexical phrases).

RESULTS AND DISCUSSION

Of the twenty young adults with normal hearing, 9 completed Protocol 1 and 11 completed Protocol 2. Of the 16 older adults with impaired hearing, 10 completed Protocol 1 (Participants 3–9, 11, 12, & 14) and 6 completed Protocol 2 (Participants 1, 2, 10, 13, 15, & 16). Two separate between-subject General Linear Model analyses were performed, one for each group of participants, to examine the effects of training protocol on training outcome. No significant effects ($F(1,12) = 0.5$, for young adults, and $F(1,14) = 0.1$, for older adults; $p > .1$) of training protocol were observed in either group for any of the post-training performance measures. As a result, data have been combined for both protocols in all subsequent analyses.

Of the 16 older adults in this study, one in Protocol 1 (Participant 8) and two in Protocol 2 (Participants 13 & 15) were unable to complete the entire protocol due to scheduling conflicts or health issues that arose. For one of these individuals (Participant 15), only the

last two training sessions were missed, but for the other two, one completed 75% of the protocol (Participant 13) and the other only 50% of the protocol (Participant 8). To decide whether the post-training data from these three participants should be included in the group analyses of post-training performance, three other older adults from the same protocols who had the closest pre-training baseline performance to the three older adults with partial training were selected for comparison purposes. There are a total of five post-training performance measures: CID Everyday Sentences 1 (Lists A & B, same as baseline), CID Everyday Sentences 2 (Lists C & D), 200 randomly selected frequent-word stimuli (the same randomly selected stimulus set for all listeners with 50 unique words spoken by each of the four talkers), 200 modified VAST sentences (same 50 sentences spoken by each of the four talkers), and 376 frequent phrases (the same 94 phrases spoken by each of the four talkers). When comparing pre-training baseline measures to post-training performance measures, there is a direct comparison available for all measures except CID Everyday Sentences 2. For the analysis of post-training benefits, we have assumed that the baseline score for CID Everyday Sentences 1 (Lists A & B) is representative of the baseline for CID Everyday Sentences 2 (Lists C & D), although this has not been confirmed directly by us or, to the best of our knowledge, by others. When comparisons were made between the mean training benefit received for each of these five performance measures for the subgroup of older adults with abbreviated training protocols and a baseline-matched subgroup of older adults with complete training protocols, only one of the differences in training benefit exceeded 2% between these two subgroups. This was for the sample of 200 frequent words, with the complete-training group experiencing a 21.7% improvement (46.3% baseline to 68% post-training) in the open-set recognition of these items following training whereas the abbreviated-training group experienced only a 13.3% improvement (47.0% baseline to 60.3% post-training). Again, other than this difference, the post-training improvements were within 2% for these two subgroups. Given that the training sessions primarily involve word-based training, it is perhaps not too surprising that the word-based post-training assessment was the one which revealed the largest differences between the two subgroups of older adults. Given the good agreement between these two subgroups for 4 of the 5 measures, however, it was decided that the data from those completing an abbreviated training protocol would be retained rather than discarded.

Figure 4 provides the group data for the young adults with normal hearing (top) and the older adults with impaired hearing (bottom). Means and standard deviations are shown in rationalized arcsine units (Studebaker, 1985) for pre-training baseline (black bars) and post-training performance (grey and white bars). Pair-wise t-tests were conducted for each of the five pre- vs. post-training comparisons for each group and the asterisks above the vertical bars in Figure 4 mark those found to be significant ($p < .05$, Bonferroni adjustment for multiple comparisons). The thin horizontal lines above the vertical bars representing the scores for the CID Everyday Sentences indicate that both of these pre-training to post-training comparisons were significant (i.e., CID 1 pre-training to CID 1 post-training and CID 1 pre-training to CID 2 post-training). For both groups of participants, the improvement in performance following training is largest for frequent words (scores improve about 20–30 RAU), the materials emphasized during training. Smaller, but significant, improvements were observed for all of the other performance measures obtained after training (scores improve about 12–20 RAU).

The individual differences in training benefits among the older adults with impaired hearing were examined next. This group, of course, is the group targeted for the training regimen, with the young normal-hearing listeners included for group purposes and to demonstrate efficacy of the regimen should the older adults fail to reveal significant benefits of training. Figure 5 provides the individual improvements from pre-training baseline to post-training evaluation for the 200 randomly selected frequent words. The solid horizontal line in this

figure illustrates the 95% critical difference in RAU for a test comprised of 200 items (Studebaker, 1985). Fourteen of the 16 older adults revealed significant benefits in open-set word recognition in noise following word-based training. One of the two older adults not achieving significant benefits from training is one of three individuals who had an abbreviated training protocol. This individual was in Protocol 2 and completed 75% of the protocol, which means that training was received for only 450 of the 600 words prior to withdrawal. Some of the words for which no training was received were among the set of 200 randomly selected words used in assessment and this lack of training for those words likely contributed to the lower post-training score on these materials for this individual.

Figure 6 provides an illustration of individual improvements for the 16 older adults on the remaining four assessment measures. Each vertical bar again represents the individual benefits measured in one of the older adults and the horizontal lines represent 95% critical differences. In this figure, however, since some decreases in performance were observed following training, the critical difference line is provided above and below zero to determine whether the observed declines were significant. For the two sets of CID Everyday Sentences (top panels), 10 of 16 older adults showed significant improvements for the CID 1 stimulus materials following training whereas 11 of 16 showed significant improvements for the CID 2 stimulus materials. Because these materials were produced by a talker not used during training saying sentences that were not encountered during training, this demonstrates that about 75% of the older adults were able to generalize word-based training with a set of four talkers to novel sentences produced by a novel talker.

For the modified VAST sentences and the frequent phrases, shown in the bottom panel of Figure 6, Figure 12 of 16 older adults exhibited significant improvements in performance on each measure. Both of these sets of materials were spoken by the same four talkers that the participants heard in training, but only the frequent phrases were actually used in training. Thus, these results provide evidence of training benefits for trained materials (frequent phrases) and generalization to new materials (modified VAST sentences).

Closer inspection of the individual data in Figure 6 reveals that four older adults, Participants 6, 8, 9 and 10, failed to show significant improvements on three of the four performance measures included in this figure. We wondered if these four participants shared some common factors that might be underlying the limited benefits experienced and conducted a series of multiple-regression analyses for the various measures of post-training benefit (Figures 5 and 6) to explore this further. Because the two measures of training benefit for the two versions of the CID Everyday Sentences were strongly correlated ($r=0.87$, $p<.001$), the regression analysis for these materials was conducted only for the CID 1 set of materials; the materials for which both pre- and post-training scores were available. The predictor variables examined included the high-frequency average hearing loss (mean hearing loss at 1000, 2000 and 4000 Hz in the test ear), age, and the pre-training score for the same materials. The last measure was included because we have observed frequently in previous training studies, as have others, that the largest gains are often observed in those with the most to gain (Burk et al., 2006; Burk & Humes, 2007, 2008).

The only significant regression solution that emerged from these four regression analyses was the one with training benefit for the CID 1 materials as the dependent variable. Two predictor variables accounted for 61.6% of the variance in CID 1 improvement: (1) pre-training baseline score on the CID 1 materials (43.5% of the variance); and (2) age (18.1% of the variance). The standardized regression weights for these two variables were both negative indicating that the higher the pre-training baseline score and the older the participant, the smaller the post-training benefit for the CID Everyday Sentences. Figure 7 illustrates these two associations as scatterplots of the individual CID 1 training

improvements (y axes) versus CID pre-training baseline score (top, x axis) and age (bottom, x axis). The data points for the four older adults who showed the least amount of generalization of training (Participants 6, 8, 9 and 10) have been circled in each scatterplot. These four older individuals were among the oldest and also among those having the highest pre-training baseline scores, with both factors contributing to the smaller amounts of training benefit observed in these individuals, at least for the two training-benefit measures based on the CID Everyday Sentences.

Conclusions

In general, this study, like our earlier studies in this same area, supports the efficacy of a word-based approach to auditory training in older adults with impaired hearing. Overall, however, when the results of the present study are compared to our earlier studies using similar word-based training approaches (Burk et al. 2006; Burk & Humes, 2007, 2008), two differences in outcome are striking. First, and perhaps most importantly, generalization of word-based closed-set identification training to improvements in the open-set recognition of sentences has been demonstrated in both the group data and in about 75% of the older individuals. We believe that this was made possible in this study for two primary reasons. First, there was greater overlap between the vocabulary or lexical content of the training words and the subsequent measures of sentence recognition than in our prior studies. This is entirely consistent with the presumed underlying lexical nature of this word-based training paradigm. Second, the addition of closed-set identification of frequent phrases most likely enhanced the generalization to sentences. These phrases include inter-word co-articulatory information as well as prosodic information and intonation. Some have argued that prosodic and intonation information is also stored in the lexicon (e.g., Lindfield, Wingfield and Goodglass, 1999). If so, then the inclusion of frequent phrases in the training protocol could have facilitated the transition from word-based training to sentence recognition.

The second striking difference between these results and our earlier work with similar word-based training programs is that the effects of training, although significant and pervasive, are considerably smaller than those observed previously, especially for trained words. In our previous work, however, the set sizes for the trained words were much smaller, typically 50–75 words. In our most recent study (Burk & Humes, 2008), two sets of 75 words were used as training materials, with training being completed with one set of 75 words before proceeding to the next set of 75 words. The total duration of training in that study was similar to that in this study. In the preceding study, open-set recognition of the trained words improved from scores of about 30–40% correct at pre-training baseline to 80–85% at post-training evaluation, an improvement of 40–50%. In the current study, by comparison, open-set word-recognition scores increased from about 50% at pre-training baseline to about 70% at post-training evaluation, an improvement of 20%. There are several possible explanations for the reduced improvement in the current study, but foremost among these is the reduced amount of training time *per word* in this study compared to the earlier investigation. Although total training time was about the same in both studies (about 24 sessions of 75–90 minutes each), the training set was comprised of a total of 150 words in one study and 600 words in the other, in turn reducing the amount of training time per word or number of repetitions of each training word in the current study to about 25% of that in our earlier work. Whether this is, in fact, the explanation for these differences in word-recognition performance must await further research in which the current word-based protocol is extended in duration.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported, in part, by a research grant to the first author from the National Institute on Aging (R01 AG008293). The authors wish to thank Adam Strauser, Charles Perine, Abby Poyser, Patricia Ray, and Amy Rickards for their assistance with stimulus production and subject testing.

Work supported, in part, by a research grant to the first author from the National Institute on Aging (R01 AG008293).

REFERENCES

- American National Standards Institute. American National Standard Maximum Permissible Noise Levels for Audiometric Test Rooms, ANSI S3.1. New York: American National Standards Institute; 1999.
- American National Standards Institute. American National Standard Specifications for Audiometers, ANSI S3.6. New York: American National Standards Institute; 2004.
- Amos NE, Humes LE. Contribution of high frequencies to speech recognition in quiet and noise in listeners with varying degrees of high-frequency sensorineural hearing loss. *J Speech Lang Hear Res.* 2007; 50:819–834. [PubMed: 17675588]
- Bell TS, Wilson RW. Sentence materials controlling word usage and confusability. *J Am Acad Audiol.* 2001; 12:514–522. [PubMed: 11791938]
- Bentler RA. Effectiveness of directional microphones and noise reduction schemes in hearing aids: A systematic review of evidence. *J Am Acad Audiol.* 2005; 16:477–488.
- Biber, D.; Johansson, S.; Leech, G.; Conrad, S.; Finegan, E. *Longman Grammar of Spoken and Written English.* Harlow: Essex: Pearson Education Ltd; 1999.
- Burk MH, Humes LE. Effects of Training on Speech-Recognition Performance in Noise using Lexically Hard Words. *J Speech Lang Hear Res.* 2007; 50:25–40. [PubMed: 17344546]
- Burk MH, Humes LE. Effects of Long-Term Training on Aided Speech-Recognition Performance in Noise in Older Adults. *J Speech Lang Hear Res.* 2008; 51:759–771. [PubMed: 18506049]
- Burk MH, Humes LE, Amos NE, Strauser LE. Effect of Training on Word-Recognition Performance in Noise for Young Normal-Hearing and Older Hearing-Impaired Listeners. *Ear Hear.* 2006; 27:263–278. [PubMed: 16672795]
- Cord MT, Surr RK, Walden BE, Dyrland O. Relationship between laboratory measures of directional advantage and everyday success with directional microphone hearing aids. *J Am Acad Audiol.* 2004; 15:353–364. [PubMed: 15506497]
- Davis, H.; Silverman, SR. *Hearing and Deafness.* 4th edition. London: Holt, Rinehart and Winston; 1978.
- Dreschler WA, Verschuure H, Ludvigsen C, Westermann S. ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. *International Collegium for Rehabilitative Audiology. Audiology.* 2001; 40:148–157. [PubMed: 11465297]
- Foo C, Rudner M, Ronnberg J, Lunner T. Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and processing capacity. *Journal of the American Academy of Audiology.* 2007; 18:553–566.
- French NR, Carter CW Jr, Koenig W Jr. The words and sounds of telephone conversations. *Bell System Technical Journal.* 1931; 9:290–324.
- Gantz BJ, Turner CW, Gfeller K. Acoustic plus electric speech processing: Preliminary results of a multicenter clinical trial of the Iowa/Nucleus hybrid implant. *Audiol Neuro-Otol.* 2006; 11 Suppl 1:63–68.
- George ELJ, Festen JM, Houtgast T. Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am.* 2006; 120:2295–2311. [PubMed: 17069325]
- George ELJ, Zekveld AA, Kramer SE, Goverts ST, Festen JM, Houtgast T. Auditory and nonauditory factors affecting speech reception in noise by older listeners. *J Acoust Soc Am.* 2007; 121:2362–2375. [PubMed: 17471748]

- Godfrey JJ, Holliman EC, McDaniel J. SWITCHBOARD: Telephone speech corpus for research and development. *Proc IEEE Int Conf Acoust Speech Sig Proc.* 1992;517–520.
- Hetzel L, Smith A. The 65 years and over population: 2000. U.S. Department of Commerce, Economics and Statistics Administration. 2001:1–8.
- Humes LE. Understanding the speech-understanding problems of the hearing impaired. *J Amer Acad Audiol.* 1991; 2:59–70. [PubMed: 1768875]
- Humes LE. Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *J Acoust Soc Am.* 2002; 112:1112–1132. [PubMed: 12243159]
- Humes LE. Do "auditory processing" tests measure auditory processing in the elderly? *Ear Hear.* 2005; 26:109–119. [PubMed: 15809539]
- Humes LE. The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *J Am Acad Audiol.* 2007; 18:590–603. [PubMed: 18236646]
- Humes, LE. Issues in the assessment of auditory processing in older adults. In: Cacace, AT.; McFarland, DJ., editors. *Current Controversies in Central Auditory Processing Disorder (CAPD)*. San Diego: Plural Publishing; 2008. p. 121-150.
- Humes, LE.; Dubno, JR. Factors affecting speech understanding in older adults. In: Gordon-Salant, S.; Frisina, RD.; Popper, AN.; Fay, RR., editors. *The Aging Auditory System: Perceptual Characterization and Neural Bases of Presbycusis*. New York: Springer; (in press)
- Humes LE, Humes LE, Wilson DL. A comparison of single-channel linear amplification and two-channel wide-dynamic-range-compression amplification by means of an independent-group design. *Amer J Audiol.* 2004; 13:39–53. [PubMed: 15248803]
- Humes LE, Lee JH, Coughlin MP. Auditory measures of selective and divided attention in young and older adults using single-talker competition. *J Acoust Soc Am.* 2006; 120:2926–2937. [PubMed: 17139749]
- Humes LE, Christensen LA, Bess FH, Hedley-Williams A. A comparison of the benefit provided by a well-fit linear hearing aid and instruments with automatic reductions of low-frequency gain. *J Speech Hear Res.* 1997; 40:666–685.
- Humes LE, Christensen LA, Bess FH, Hedley-Williams A, Bentler R. A comparison of the aided performance and benefit provided by a linear and a two-channel wide-dynamic-range-compression hearing aid. *J Speech Lang Hear Res.* 1999; 42:65–79. [PubMed: 10025544]
- Humes LE, Barlow NN, Garner CB, Wilson DL. Prescribed clinician-fit versus as-worn coupler gain in a group of elderly hearing aid wearers. *J Speech Lang Hear Res.* 2000; 43:879–892. [PubMed: 11386475]
- Humes LE, Garner CB, Wilson DL, Barlow NN. Hearing-aid outcome measures following one month of hearing aid use by the elderly. *J Speech Lang Hear Res.* 2001; 44:469–486. [PubMed: 11407554]
- Humes LE, Wilson DL, Barlow NN, Garner CB. Measures of hearing-aid benefit following 1 or 2 years of hearing-aid use by older adults. *J Speech Lang Hear Res.* 2002a; 45:772–782. [PubMed: 12199406]
- Humes LE, Wilson DL, Barlow NN, Garner CB, Amos NE. Longitudinal changes in hearing-aid satisfaction and usage in the elderly over a period of one or two years after hearing-aid delivery. *Ear Hear.* 2002b; 23:428–437. [PubMed: 12411776]
- Humes LE, Burk MH, Coughlin MP, Busey TA, Strauser LE. Auditory speech recognition and visual text recognition in younger and older adults: Similarities and differences between modalities and the effects of presentation rate. *J Speech Lang Hear Res.* 2007; 50:283–303. [PubMed: 17463230]
- Humes LE, Ahlstrom JB, Bratt GW, Peek BF. Studies of hearing aid outcome measures in older adults: A comparison of technologies and an examination of individual differences. *Sem Hearing.* 2009; 30:112–128.
- ISO. *Acoustics-Statistical distribution of hearing thresholds as a function of age*. Basel, Switzerland: International Standards Organization; 2000. ISO-7029
- Jin S-H, Nelson PB. Speech perception in gated noise: The effects of temporal resolution. *J Acoust Soc Am.* 2006; 119:3097–3108. [PubMed: 16708964]
- Kochkin S. MarkeTrak III: Why 20 million in U.S. don't use hearing aids for their hearing loss. Part I. *Hearing Journal.* 1993a; 46(1):20–27.

- Kochkin S. MarkeTrak III: Why 20 million in U.S. don't use hearing aids for their hearing loss. Part II. *Hearing Journal*. 1993b; 46(2):26–31.
- Kochkin S. MarkeTrak III: Why 20 million in U.S. don't use hearing aids for their hearing loss. Part III. *Hearing Journal*. 1993c; 46(4):36–37.
- Kochkin S. MarkeTrak V: Consumer satisfaction revisited. *Hearing Journal*. 2000; 53(1):38–55.
- Kochkin S. MarkeTrak VII: Hearing loss population tops 31 million. *Hearing Rev*. 2005; 12(7):16–29.
- Larson VD, Williams DW, Henderson WG, et al. Efficacy of 3 commonly used hearing aid circuits: A crossover trial. *J Am Med Assoc*. 2000; 284:1806–1813.
- Lee CJ. Evidence-based selection of word frequency lists. *Journal of Speech-Language Pathology and Audiology*. 2003; 27:170–173.
- Lunner T, Sundewall-Thorén E. Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance in a two-channel hearing aid. *J Am Acad Audiol*. 2007; 18:539–552. [PubMed: 18236642]
- Lee LW, Humes LE. Evaluating a speech-reception threshold model for hearing-impaired listeners. *J Acoust Soc Am*. 1993; 93:2879–2885. [PubMed: 8315151]
- Lindfield KC, Wingfield A, Goodglass H. The role of prosody in the mental lexicon. *Brain Lang*. 1999; 68:312–317. [PubMed: 10433775]
- Lunner T. Cognitive function in relation to hearing aid use. *International J Audiol*. 2003; 42 Suppl 1:S49–S58.
- Nordrum S, Erler S, Garstecki D, Dhar S. Comparison of performance on the hearing in noise test using directional microphones and digital noise reduction algorithms. *Am J Aud*. 2006; 15:81–91.
- Owens E, Benedict M, Schubert ED. Consonant phonemic errors associated with pure-tone configurations and certain kinds of hearing impairment. *J Speech Hear Res*. 1972; 15:308–322. [PubMed: 5047869]
- Pearsons, KS.; Bennett, RL.; Fidell, S. Speech levels in various noise environments. Washington, DC: U.S. Environmental Protection Agency; 1977. Report No. EPA-600/1-77-025
- Pichora-Fuller KM, Singh G. Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. *Trends Amplif*. 2006; 10:29–59. [PubMed: 16528429]
- Ronnberg J, Rudner M, Foo C, Lunner T. Cognition counts: A working memory system for ease of language understanding (ELU). *International J Audiol*. 2008; 47 Suppl 2:S171–S177.
- Schoenborn CA, Marano M. Current estimates from the National Health Interview Survey. *Vital Health Stat*. 1988; 10(166):1–233.
- Simpson A, Hersbach AA, McDermott HJ. Improvements in speech perception with an experimental nonlinear frequency compression hearing device. *Int J Audiol*. 2005; 44:281–292. [PubMed: 16028791]
- Smeeth L, Fletcher AE, Ng ES, Stirling S, Nunes M, Breeze E, Bulpitt CJ, Jones D, Tulloch A. Reduced hearing ownership, and use of hearing aids in elderly people in the UK—the MRC trial of the assessment and management of older people in the community: A cross-sectional survey. *Lancet*. 2002; 359(9316):1466–1470. [PubMed: 11988245]
- Studebaker GA. A “rationalized” arcsine transform. *J Speech Hear Res*. 1985; 28:455–462. [PubMed: 4046587]
- Sweetow R, Henderson-Sabes J. The case for LACE: Listening and auditory communication enhancement training. *Hear J*. 2004; 57(3):32–40.
- Sweetow R, Palmer C. Efficacy of individual auditory training in adults: A systematic review of the evidence. *J Amer Acad Audiol*. 2005; 16:498–508.
- Sweetow RW, Sabes J. The need for and development of an adaptive listening and communication enhancement (LACE) program. *J Am Acad Audiol*. 2006; 17:538–558. [PubMed: 16999250]
- Turner CW, Hurtig RR. Proportional frequency compression for listeners with sensorineural hearing loss. *J Acoust Soc Am*. 1999; 106:877–886. [PubMed: 10462793]
- Walden B, Surr R, Cord M, Dylrund O. Predicting hearing aid microphone preference in everyday listening. *J Amer Acad Audiol*. 2004; 15:365–396. [PubMed: 15506498]

- Ward JA, Lord SR, Williams P, Anstey K. Hearing impairment and hearing aid use in women over 65 years of age. Cross-sectional study of women in a large urban community. *Med J Aust.* 1993; 159:382–384. [PubMed: 8377688]
- Zeno, SM.; Ivens, SH.; Millard, RT.; Duwuri, R. *The Educator's Word Frequency Guide.* New York, NY: Touchstone Applied Science Associates; 1995.

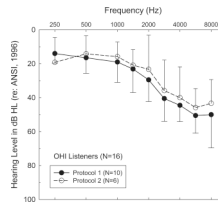


Figure 1. Mean air-conduction pure-tone thresholds for the test ear for the two subgroups of older adults, one received training Protocol 1 (filled circles) and the other received training Protocol 2 (unfilled circles). Thin vertical lines represent one standard deviation above or below the corresponding group mean.

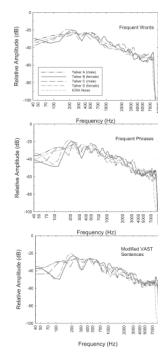


Figure 2. The long-term amplitude spectra measured in Adobe Audition for the words (top panel), phrases (middle panel) and sentences (bottom panel) produced by each of the four talkers, as well as the corresponding long-term spectra for the ICRA noise used as competition with these materials.

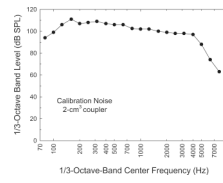


Figure 3. The 1/3-octave-band levels measured in the 2-cm³ coupler for the steady-state calibration noise shaped to match the spectrum of the two-talker ICRA noise.

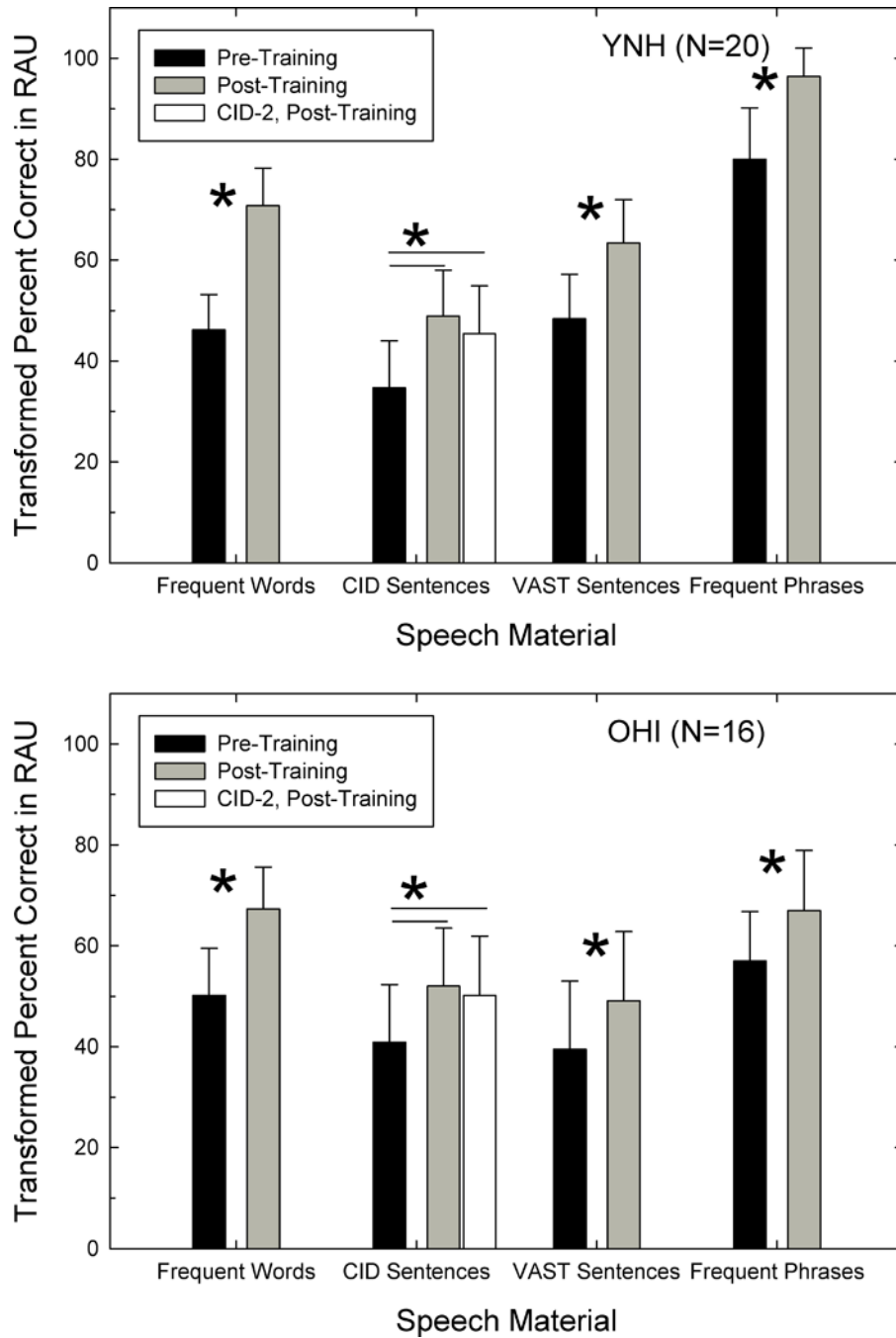


Figure 4. Means and standard deviations are shown in rationalized arcsine units (RAU; Studebaker, 1985) for the young adults with normal hearing (top) and the older adults with impaired hearing (bottom). Black bars represent pre-training baseline performance and grey or white bars represent post-training performance. Thin vertical lines depict corresponding standard deviations.

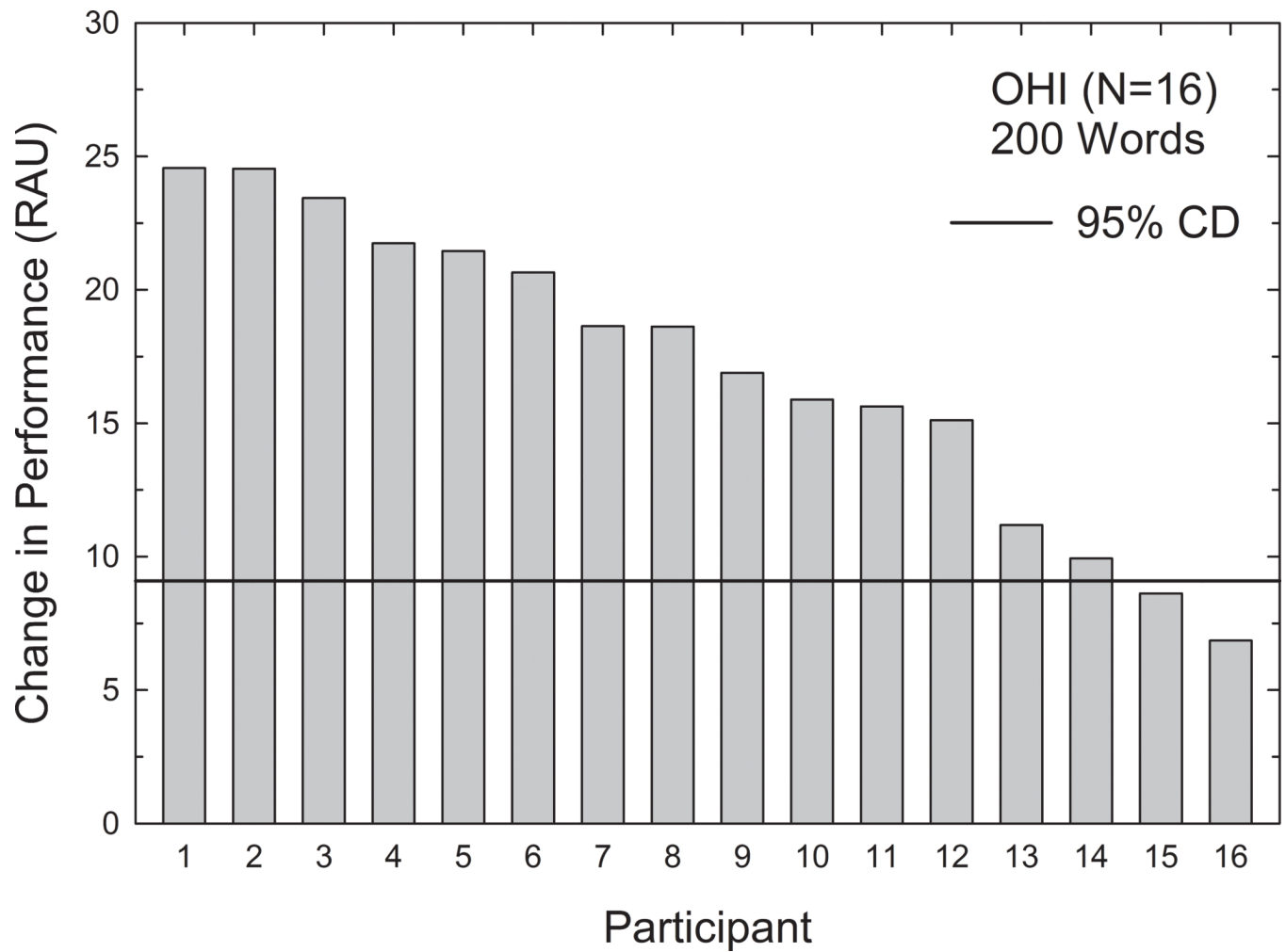


Figure 5. Individual improvements from pre-training baseline to post-training evaluation, training benefit, on the set of 200 randomly selected frequent words for each of the 16 older adults with impaired hearing. The solid horizontal line illustrates the 95% critical difference in RAU for a test comprised of 200 items (Studebaker, 1985).

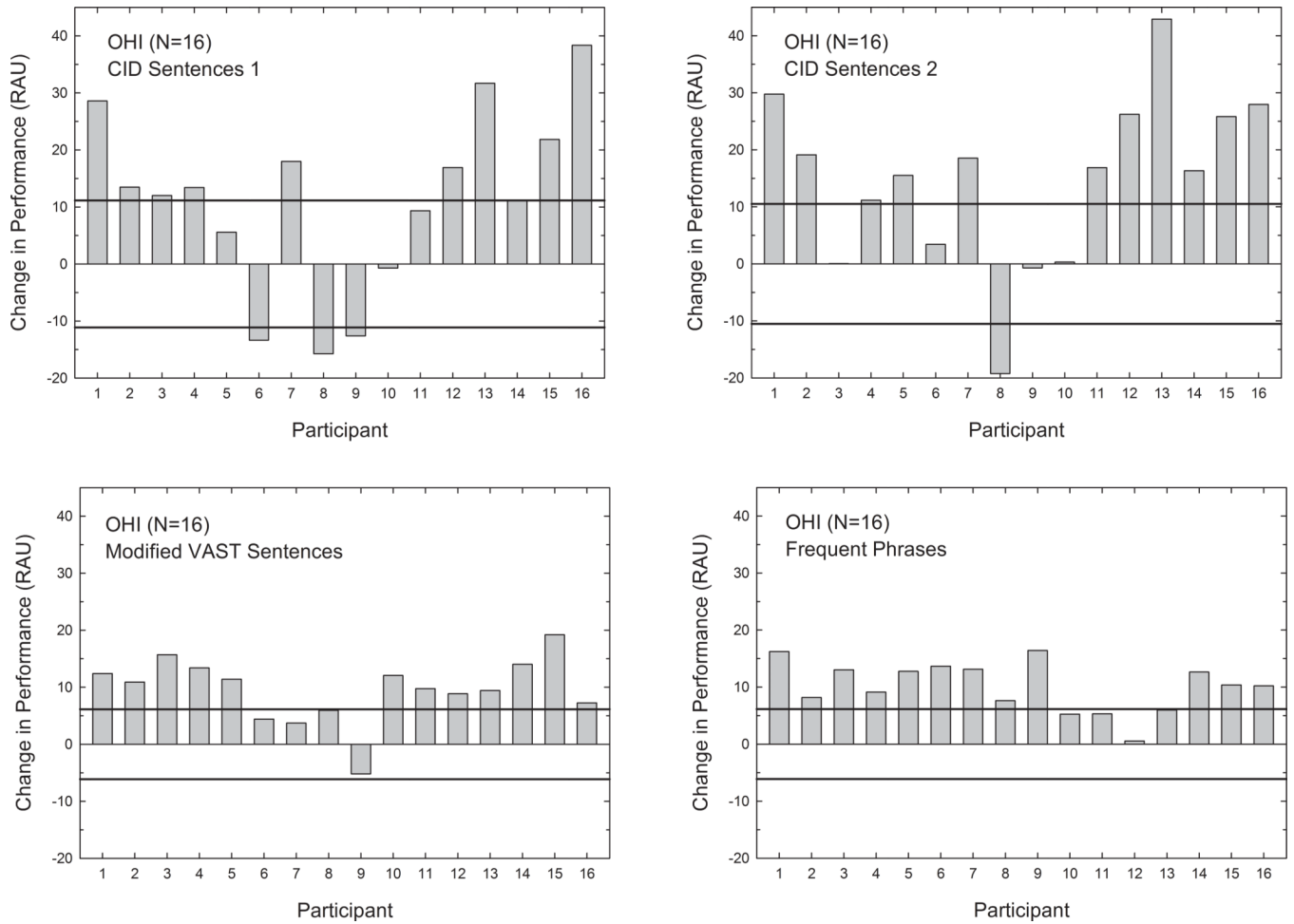


Figure 6. Individual improvements from pre-training baseline to post-training evaluation, training benefit, on for sets of speech materials: (1) CID 1 Everyday Sentences (top left); (2) CID 2 Everyday Sentences (top right); (3) modified VAST sentences (bottom left); and (4) frequent phrases (bottom right). The solid horizontal line in each panel again illustrates the 95% critical difference in RAU for each measure.

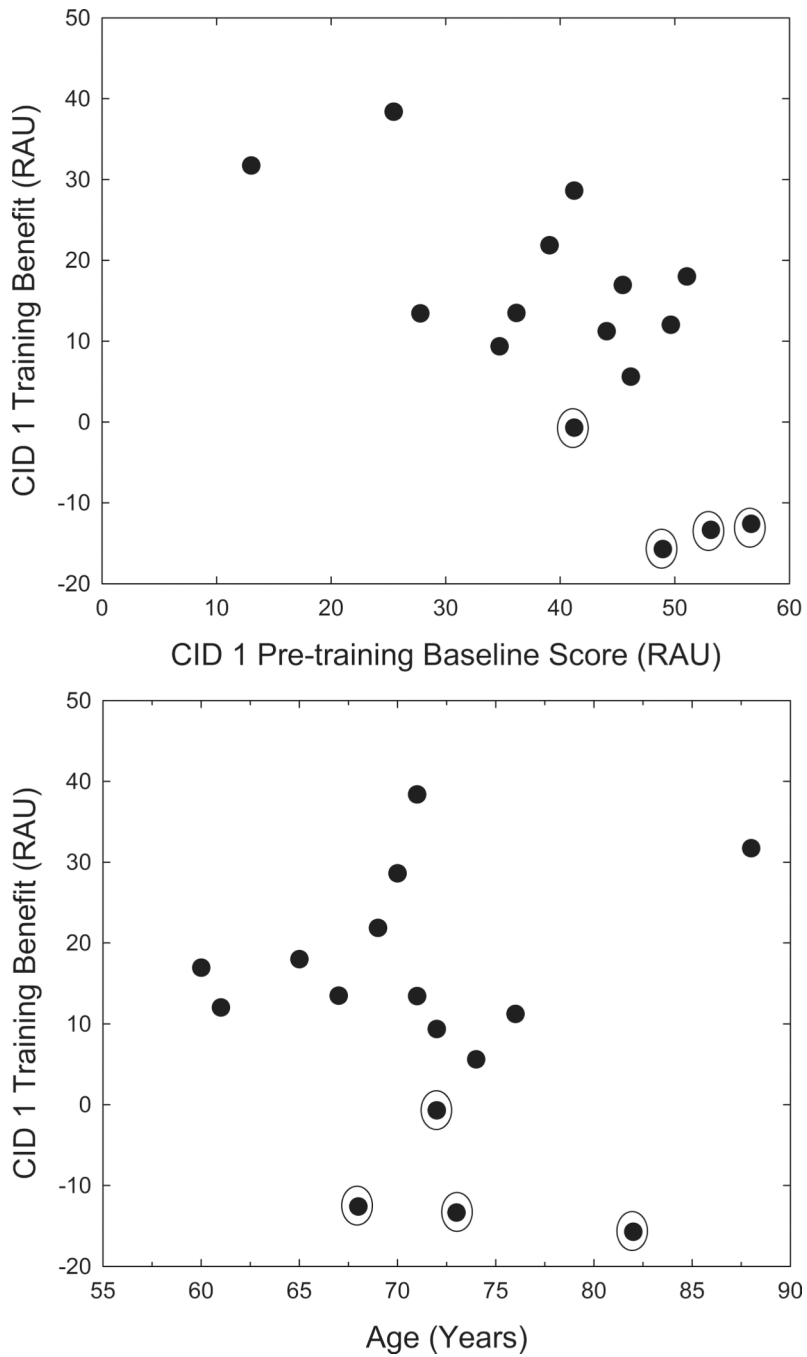


Figure 7. Scatterplots of the individual CID 1 training improvements (y axes) versus CID pre-training baseline score (top, x axis) and age (bottom, x axis). The data points for the four older adults who showed the least amount of generalization of training (Participants 6, 8, 9 and 10) have been circled in each scatterplot.

Table 1

Sample size (N), gender (male, M; female, F), age (mean, M, and standard deviation, SD, in years), ear used in training/testing (right, R; left, L), and high-frequency pure-tone average (HFPTA: 1000, 2000 and 4000 Hz; M, SD) for the participants in each group (YNH = young adults with normal hearing; OHI = older adults with impaired hearing) and training protocol in this study.

Group	Protocol	N	Gender	Age (y)		Test Ear	HFPTA (dB HL)	
				M	SD		M	SD
YNH	1	9	7F, 2M	20.3	3.1	9R, 0L	6.3	3.8
YNH	2	11	7F, 4M	20.8	3.1	11R, 0L	4.5	3.7
OHI	1	10	3F, 7M	70.2	6.8	8R, 2L	31.0	9.6
OHI	2	6	1F, 5M	72.8	7.6	6R, 0L	26.4	14.1