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

Larry E. Humes, Matthew H. Burk, Lauren E. Strauser, and Dana L. Kinney

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: In this study, a repeated-measures group design was used with each of the two groups of participants to evaluate the effects of the word-based training regimen. One group comprised 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 to estimate the magnitude of the benefits that could be achieved in younger listeners.

Results: Significant improvements were observed in the group means for each of five measures of post-training assessment. Prestimulation 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 to 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.

INTRODUCTION

In the 2000 U.S. Census, 35 million Americans older than 65 years were counted, representing 12.4% of the U.S. population (Hetzel & Smith 2001). Approximately 30% of those older than 65 years in the United States, 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 only about 40 to 60% are satisfied with them and use their hearing aids regularly (Kochkin 1993a,b,c, 2000, 2005). Moreover, these figures are typical of those in other countries, such as the United Kingdom (Smeeth et al. 2002) and Australia (Ward et al. 1993).

The most common communication complaint of older adults with impaired hearing (OHI) is that they can hear speech but cannot 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 (Humes 1991; Humes & Dubno Reference Note 1). 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 et al. 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 (Turner & Hurtig 1999; Simpson et al. 2005; Gantz et al. 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; 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 (Lee & Humes 1993; Humes 2008; Humes & Dubno Reference Note 1). Nonetheless, even with the less than optimal audibility provided by hearing aids fit in clinical settings (Humes 2002; Humes et al. 2000, 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 multitalker babble (Humes et al. 1997, 1999, 2001, 2002a,b, 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 et al. 2006, 2007; Jin & Nelson 2006; George et al. 2006, 2007; Humes 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; Humes 2002, 2005, 2007; George et al. 2006, 2007; Humes et al. 2006, 2007; Pichora-Fuller & Singh 2006; Lunner & Sunderwall-Thoren 2007; Foo et al. 2007; Ronnberg et al. 2008; Humes & Dubno Reference Note 1).

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 do not 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 (Walden et al. 2004; Cord et al. 2004; Bentler 2005; 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. However, hearing aids alone have been unable to deliver the required improvements in speech to noise ratio. Hearing aids do seem 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, seems 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 (SNR), rather than attempt to improve the acoustical SNR. (Of course, there is no reason that these two strategies, one focused on the listener and the other on the hearing aid, cannot 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 sublexical 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) after correct/incorrect feedback on every trial, both auditory and orthographic feedback associated with the correct and incorrect responses was provided to the listener after 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 OHI. 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. After review of these six studies, Sweetow and Palmer (2005, p. 501) concluded that “this systematic review provides very little evidence for the effectiveness of individual auditory training. However, there is some evidence for efficacy.” 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 and Palmer 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 multicenter clinical evaluation of a cognitive-based top-down auditory-training system referred to as Listening and Auditory Communication Enhancement (Sweetow & Henderson-Sabes 2004). The short-term benefits of the top-down approach pursued by Listening and Auditory Communication Enhancement in older adults were encouraging. Second, a series of articles were published describing the development, evaluation, and efficacy of our automated word-based approach to auditory training (Burk et al. 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) OHI could improve their open-set recognition of words in noise from about 30 to 40% correct before training to 80 to 85% correct after training; (2) training generalized to other talkers saying the same trained words, but only slight improvements (7 to 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
The recognition of spoken words in noise for sets of words as large that older adults showed significant improvements in the (French et al. 1931; Godfrey et al. 1992). Having demonstrated 90% of the words used most frequently in spoken conversation relatively small number of words, 400 to 600, represent 80 to words used most frequently in everyday conversation. A the training is word specific, then the focus should be placed on (American) English. The rationale for this emphasis is that, if focused on the most frequently occurring words in spoken word-based auditory training is described in this report and has used as training words were examined, much larger improve- tion of the training to other talkers, the lexical nature of the training limits the generalization of trained words to other novel talkers. In addition, for all experiments with OHI, 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, con- firmed 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 seemed 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 to 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 improve- ments 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 to 600, represent 80 to 90% of the words used most frequently in spoken conversation (French et al. 1931; Godfrey et al. 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 4 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 seem to diminish with age (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 International Collegium for Rehabilitative Audiol- ogy (ICRA) noise (Dreschel 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 to 80%) with the training vocab- ulary.

PARTICIPANTS AND METHODS

Participants

There were two primary groups of participants in this study: young adults with normal hearing (YNH) and 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 protocols 1 and 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 of the two OHI subgroups. In total, 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 selected sentences from the Veterans Administration Sentence Test (VAST; Bell & Wil- son 2001); and (4) the Auditec recording of Central Institute for the Deaf (CID) Everyday Sentences (Davis & Silverman 1978). Because 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
recordings were designated as being unacceptable by the first
the final set of recordings and listened to each. In some cases,
of the authors examined the waveforms and spectrograms of
sentence. After this editing, a second research assistant or one
remove silence at the beginning or end of the word, phrase, or
selected token was edited at zero crossings of the waveform to
among the tokens, selected the middle of the three tokens. The
forms, a trained research assistant selected the best token
listening experience and the visual inspection of the wave-
examined, using Adobe Audition 2.0. Based on the subjective
repetitions of the stimulus item were re-recorded immediately.
presses when the third production was completed. If the
talker and the “Enter” key on the computer keyboard was
produced each item clearly but in a conversational manner.
level during production. The talker was instructed to
and also provided a digital volume unit meter to monitor the
talker’s mouth. A liquid crystal display computer monitor
shock mount and stand, approximately 9 inches from the
 editorial microphone positioned on a table top, in a supplied custom
professional speaker. Materials were recorded in a double-walled
sound room with an Audio Technica AT3035 cardioid capaci-
tor microphone positioned in front of the talker displayed the
word, phrase, or sentence to be produced before each recording
also provided a digital volume unit monitor 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 SNR, three
repetitions of the stimulus item were re-recorded immediately.
The output of the microphone was preamplified (Symetrix
Model 302) before digitization by a Tucker-Davis Technolo-
gies (TDT) 16-bit analog-to-digital converter at a sampling rate
of 44.1 kHz.

After completion of the recordings, the three tokens of each
stimulus were listened to, and the respective waveforms were
examined, using Adobe Audition 2.0. Based on the subjective
listening experience and the visual inspection of the wave-
forms, 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. After 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 resample 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-msec win-
dow 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 fre-
quently occurring words in English were recorded by each of
the four talkers. At 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 cor-
pora of most frequently occurring words in English available as
resources, some based on written English and others based on
oral English. However, Lee (2003) 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 more than 17 million
words, was used to establish word frequency. Because 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 et al.
1931) and more recent (Switchboard Corpus; Godfrey et al.
1992) word-frequency corpus developed from recordings of
telephone conversations. There was 60 to 70% overlap among
these three corpora, consistent with the findings of Lee based
on other corpora. Although the majority of words comprised
one or two syllables, there were 20 three-syllable and five
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 com-
pleted at a 0-dB SNR with the same noise that was used in this
study (see below). Two additional listeners were tested at a
−5-dB SNR to get some idea as to performance levels and
shape of the psychometric function. The mean percent-correct
scores for the 2400-word stimuli (four talkers ×600 words)
were 46.1% for the two YNH listeners tested at an SNR of −5
dB and 69.4% for the other five YNH listeners tested at an
SNR of 0 dB, indicating a 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 SNR was used to evaluate
the relative difficulty of the 600 words. Because 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

![Fig. 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. OHI, older adults with impaired hearing.](image-url)
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 five listeners (most difficult); (2) words tabulated as incorrect for four of the five listeners; (3) words tabulated as incorrect for three of the five listeners; (4) words tabulated as incorrect for two of the five listeners; (5) words tabulated as incorrect for one of the five listeners; and (6) words tabulated as incorrect for zero of the five listeners (easiest). Of course, the actual number of words falling into each of these categories will vary with the SNR. The 0-dB SNR, 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. 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 remaining 400, 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). For purposes of this project, the focus 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 years old. A unique feature of the Biber et al. 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 more than 80,000 times per million words in conversation and four-word bundles more than 8500 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. 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. 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 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 an SNR 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 comprising 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 an SNR 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 comprised 10 sentences with a total of 50 keywords that are scored. There are approximately two short (two to four words), two long (10 to 12 words), and six medium-length (five to nine 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 et al. 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 a man and the other a woman. From the 10-min 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 OHI listeners have considerable difficulty in 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 root mean square (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 with Adobe Audition using 2048-point fast Fourier transforms (Blackmann window), and the long-term average spectrum was computed. Similar processing was applied to the two-talker vocoded noise from the ICRA noise CD (Track 6). Initially, an arbitrary 2-min segment of the ICRA noise was selected and digitized. From this, 20 2-sec nonoverlapping samples were excised for use as maskers for the word stimuli. Twenty longer 4-sec samples of ICRA noise, needed for the longer phrases and sentences, were generated by randomly selecting two of the 2-sec noise files for concatenation and generating 20 such pairs. The RMS amplitude and long-term spectra of the ICRA noise were 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. VAST, Veterans Administration Sentence Test.

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 (SPL) 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 before 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-cm$^3$ 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 YNH, 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 OHI, calculations were made based on the 1/3-octave band levels in Figure 3, an assumed overall SPL 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 participants pure-tone thresholds from their 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 the study by Burk and Humes (2008). Typically, after spectral shaping, the overall SPL of the steady-state calibration stimulus would be approximately 100 to 110 dB SPL.

Based on the pilot testing with these new speech materials, the following speech to noise ratios were used for the YNH: (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 OHI, 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 to 3 dB better for the older adults during training and assessment. At first glance, these SNRs seem to be very severe. However, it should be kept in mind 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 were 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 liquid crystal display 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

After initial hearing testing, participants were assigned to one of the 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 section, the 600 most frequently occurring words represent about 90% of the words in spoken conversation and also its set size is four times larger than the largest set used previously in our laboratory in similar training studies. In this study, this large set size was selected for use 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 pretraining baseline, the testing consisted of the presentation of 20 CID Everyday Sentences (Lists A and 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 quasirandomly such that no lexical items were repeated and each of the four talkers produced 50 of the 200 words. The stimuli and the order of presentation for the post-training assessment were identical to pretraining except that an extra set of 20 CID Everyday Sentences (Lists C and 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 because of the intervening training.

As noted, two different protocols were followed for the training sessions. For protocol 1, an eight-session training cycle was devised. In each of these eight training sessions, each about 75 to 90 min duration and administered on separate days, eight 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 earlier. This was followed by two 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 to 12 phrases with approximately 25% of each set produced by each of the four talkers. Because the training cycle was composed of eight daily sessions, at the end of a training cycle there had been two repetitions of the 1600-word stimuli (400 words × four talkers) that were found to be more difficult in pilot testing, one repetition of the 800-word stimuli (200 words × four talkers) that were 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 eight-session training cycles between pretraining baseline and post-training evaluation. This total of 24 training sessions was typically completed at a rate of three 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, six-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 4 days. This 4-day block was followed by a day devoted to training on 400 of the 800 easier word stimuli (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 to 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 to 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 six repetitions of the 1600 more difficult word stimuli (or 24 repetitions of the 400 corresponding lexical items), three repetitions of the 800 easier word stimuli (or 12 repetitions of the 200 corresponding lexical items), and three repetitions of the 376 stimulus phrases (or 12 repetitions of the 94 corresponding lexical phrases).

RESULTS AND DISCUSSION

Of the 20 YNH, nine completed protocol 1, and 11 completed protocol 2. Of the 16 OH, 10 completed protocol 1 (participants 3 to 9, 11, 12, and 14), and six completed protocol 2 (participants 1, 2, 10, 13, 15, and 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 the training outcome. No significant effects (F[1,12] = 0.5 for young adults and F[1,14] = 0.1 for older adults; p > 0.1) 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.

In this study, of the 16 older adults, one in protocol 1 (participant 8) and two in protocol 2 (participants 13 and 15) were unable to complete the entire protocol because of 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 the post-training performance, three other older adults from the same protocols who had the closest pretraining 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 and B, same as baseline), CID Everyday Sentences 2 (Lists C and 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 pretraining baseline measures with 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 and B) is representative of the baseline for CID Everyday Sentences 2 (Lists C and 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 after 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 that revealed the largest differences between the two subgroups of
older adults. Given the good agreement between these two subgroups for four of the five 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 YNH (top) and the OHI (bottom). Means and standard deviations are shown in rationalized arcsine units (Studebaker, 1985) for pretraining baseline (black bars) and post-training performance (gray or white bars) represent post-training performance. Thin vertical lines depict corresponding standard deviations. CID, Central Institute for the Deaf.

Fig. 4. Mean and standard deviation are shown in rationalized arcsine units (RAU; Studebaker, 1985) for the young adults with normal hearing (YNH, top) and the older adults with impaired hearing (OHI, bottom). Black bars represent pretraining baseline performance, and gray or white bars represent post-training performance. Thin vertical lines depict corresponding standard deviations. CID, Central Institute for the Deaf.

The individual differences in training benefits among the OHI were examined next. This group, of course, is the one 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 pretraining 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 comprising 200 items (Studebaker 1985). Fourteen of the 16 older adults revealed significant benefits in open-set word recognition in noise after word-based training. One of the two older adults not achieving significant benefits from the training is one of the three individuals who had an abbreviated training protocol. This individual was in protocol 2 and completed 75% of the protocol, which means that the training was received only for 450 of the 600 words before 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.

Fig. 5. Individual improvements from pretraining 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 composed of 200 items (Studebaker 1985). RAU, rationalized arcsine units; OHI, older adults with impaired hearing; CD, critical difference.
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, 12 of 16 older adults exhibited significant improvements in the performance on each measure. Both of these sets of materials were spoken by the same four talkers that the participants heard in the training, but only the frequent phrases were actually used in the training. Thus, these results provide an evidence of training benefits for trained materials (frequent phrases) and a 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 whether these four participants shared some common factors that might be underlying the limited benefits experienced, and to explore this further, we conducted a series of multiple-regression analyses for various measures of post-training benefit (Figures 5 and 6). Because the two measures of training benefit for the two versions of the CID Everyday Sentences were strongly correlated ($r = 0.87$, $p < 0.001$), the regression analysis for these materials was conducted only for the CID 1 set of 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), the age, and the pretraining 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) pretraining 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 pretraining baseline score and the older the participant, the smaller the post-training benefit for the CID Everyday Sen-
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 OHI. Overall, however, when the results of the present study are compared with 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, a greater overlap between the vocabulary or lexical content of the training words and the subsequent measures of sentence recognition than that 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 interword coarticulatory information as well as prosodic and intonation information. Some have argued that prosodic and intonation information is also stored in the lexicon (Lindfield et al. 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 to 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 to 40% correct at pretraining baseline to 80 to 85% at post-training evaluation, an improvement of 40 to 50%. In this study, by comparison, open-set word-recognition scores increased from about 50% at pretraining baseline to about 70% at post-training evaluation, an improvement of 20%. There are several possible explanations for the reduced improvement in this study, but foremost among these is the reduced amount of training time per word compared with the earlier investigation. Although the total training time was about the same in both the studies (about 24 sessions of 75 to 90 min each), the training set comprised 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 the number of repetitions of each training word in this study to about 25% of that in our earlier work. Whether this is the explanation for these differences in word-recognition performance must await further research in which the current word-based protocol is extended in duration.

Acknowledgments

The authors thank Adam Strauser, Charles Perine, Abby Poyser, Patricia Ray, and Amy Rickards for their assistance with stimulus production and subject testing.

This work was supported, in part, by a research grant from the National Institute on Aging (R01 AG008293) to L.E.H.

Address for correspondence: Larry E. Humes, Department of Speech and Hearing Sciences; Indiana University, Bloomington, IN 47405-7002. E-mail: humes@indiana.edu.

Received October 27, 2008; accepted May 20, 2009.
Appendix A. 600 Words

<table>
<thead>
<tr>
<th>took</th>
<th>upon</th>
<th>went</th>
<th>woman</th>
</tr>
</thead>
<tbody>
<tr>
<td>top</td>
<td>us</td>
<td>were</td>
<td>women</td>
</tr>
<tr>
<td>toward</td>
<td>use</td>
<td>west</td>
<td>word</td>
</tr>
<tr>
<td>town</td>
<td>used</td>
<td>what</td>
<td>words</td>
</tr>
<tr>
<td>tree</td>
<td>using</td>
<td>when</td>
<td>work</td>
</tr>
<tr>
<td>trees</td>
<td>usually</td>
<td>where</td>
<td>worked</td>
</tr>
<tr>
<td>tried</td>
<td>very</td>
<td>whether</td>
<td>working</td>
</tr>
<tr>
<td>true</td>
<td>voice</td>
<td>which</td>
<td>world</td>
</tr>
<tr>
<td>try</td>
<td>walked</td>
<td>while</td>
<td>would</td>
</tr>
<tr>
<td>trying</td>
<td>want</td>
<td>white</td>
<td>write</td>
</tr>
<tr>
<td>turn</td>
<td>wanted</td>
<td>who</td>
<td>writing</td>
</tr>
<tr>
<td>turned</td>
<td>war</td>
<td>whole</td>
<td>year</td>
</tr>
<tr>
<td>two</td>
<td>warm</td>
<td>why</td>
<td>years</td>
</tr>
<tr>
<td>type</td>
<td>was</td>
<td>will</td>
<td>yes</td>
</tr>
<tr>
<td>under</td>
<td>water</td>
<td>wind</td>
<td>yet</td>
</tr>
<tr>
<td>understand</td>
<td>way</td>
<td>window</td>
<td>you</td>
</tr>
<tr>
<td>united</td>
<td>ways</td>
<td>with</td>
<td>young</td>
</tr>
<tr>
<td>until</td>
<td>we</td>
<td>within</td>
<td>your</td>
</tr>
<tr>
<td>up</td>
<td>well</td>
<td>without</td>
<td>you're</td>
</tr>
</tbody>
</table>

Appendix B. 94 Phrases

<table>
<thead>
<tr>
<th>I don't want to go</th>
<th>I'm going to do one</th>
<th>All of a sudden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you want to know?</td>
<td>I'm going to have some</td>
<td>Two and a half years</td>
</tr>
<tr>
<td>Are you going to work?</td>
<td>I've got to go</td>
<td>Do you want to go?</td>
</tr>
<tr>
<td>I don't know how</td>
<td>I haven't got a place</td>
<td>You know what I mean</td>
</tr>
<tr>
<td>I don't know why</td>
<td>I haven't got any</td>
<td></td>
</tr>
<tr>
<td>I don't think so</td>
<td>You're going to have one</td>
<td></td>
</tr>
<tr>
<td>I thought it was</td>
<td>You were going to go</td>
<td></td>
</tr>
<tr>
<td>I think it was</td>
<td>You can have a little</td>
<td></td>
</tr>
<tr>
<td>I'm not going to go</td>
<td>You might as well look</td>
<td></td>
</tr>
<tr>
<td>You know what I did</td>
<td>That's what I mean</td>
<td></td>
</tr>
<tr>
<td>You don't want to know</td>
<td>That's what I said</td>
<td></td>
</tr>
<tr>
<td>You want to go</td>
<td>There's a lot of people</td>
<td></td>
</tr>
<tr>
<td>You don't have to be</td>
<td>I'm going to be</td>
<td></td>
</tr>
<tr>
<td>It's going to be</td>
<td>Go to the bathroom</td>
<td></td>
</tr>
<tr>
<td>Going to have to leave</td>
<td>Put it in the house</td>
<td></td>
</tr>
<tr>
<td>Going to have a family</td>
<td>Going to do it</td>
<td></td>
</tr>
<tr>
<td>Thank you very much</td>
<td>Going to get a house</td>
<td></td>
</tr>
<tr>
<td>Do you know what?</td>
<td>You have to do</td>
<td></td>
</tr>
<tr>
<td>Do you want me to?</td>
<td>Don't worry about it</td>
<td></td>
</tr>
<tr>
<td>What are you doing?</td>
<td>Got a lot of help</td>
<td></td>
</tr>
<tr>
<td>What do you mean?</td>
<td>Can I have a room?</td>
<td></td>
</tr>
<tr>
<td>What do you think?</td>
<td>Have you got a home?</td>
<td></td>
</tr>
<tr>
<td>What do you want?</td>
<td>Have you got any?</td>
<td></td>
</tr>
<tr>
<td>Know what I mean?</td>
<td>Do you want some?</td>
<td></td>
</tr>
<tr>
<td>If you want to see</td>
<td>Do you have to go</td>
<td></td>
</tr>
<tr>
<td>The end of the day</td>
<td>What did you do?</td>
<td></td>
</tr>
<tr>
<td>Or something like that</td>
<td>What did you say?</td>
<td></td>
</tr>
<tr>
<td>But I don't know how</td>
<td>What did he say?</td>
<td></td>
</tr>
<tr>
<td>But I don't think so</td>
<td>What do you call it?</td>
<td></td>
</tr>
<tr>
<td>I don't think he is</td>
<td>What have you got?</td>
<td></td>
</tr>
<tr>
<td>I don't think I did</td>
<td>Don't know what it is</td>
<td></td>
</tr>
<tr>
<td>I don't think it's time</td>
<td>Don't know what to do</td>
<td></td>
</tr>
<tr>
<td>I don't think you are</td>
<td>I know what to do</td>
<td></td>
</tr>
<tr>
<td>No I don't think so</td>
<td>You know what it is</td>
<td></td>
</tr>
<tr>
<td>I thought I would</td>
<td>You don't want to go</td>
<td></td>
</tr>
<tr>
<td>I thought that was it</td>
<td>I want to do it</td>
<td></td>
</tr>
<tr>
<td>I thought you were there</td>
<td>If you've got a home</td>
<td></td>
</tr>
<tr>
<td>I would have thought so</td>
<td>As long as you do</td>
<td></td>
</tr>
<tr>
<td>I want to get one</td>
<td>The back of the house</td>
<td></td>
</tr>
<tr>
<td>I want to do it</td>
<td>The middle of the day</td>
<td></td>
</tr>
<tr>
<td>I want to go</td>
<td>The rest of the day</td>
<td></td>
</tr>
<tr>
<td>I want to see it</td>
<td>Nothing to do with her</td>
<td></td>
</tr>
<tr>
<td>I'll tell you what</td>
<td>In the middle of it</td>
<td></td>
</tr>
<tr>
<td>I mean I don't know</td>
<td>At the same time</td>
<td></td>
</tr>
<tr>
<td>I'm going to do it</td>
<td>For a long time</td>
<td></td>
</tr>
</tbody>
</table>

Appendix C. Modified VAST Sentences

25 Sentences: High-Frequency, Low-Neighborhood Density

1. The point of the knife is too sharp
2. She was sure she had lost a pound
3. The model wore a plain black dress
4. The flock will nest in the wild brush
5. The troop can rest in the warm lodge
6. He had a slim build and his hair was dark
7. The barrel burst when it was too full
8. Use a trap to catch the mouse
9. The breeze helped to clear the fog
10. You will find the van parked on the grass
11. The trip by train was very fast
12. Treat the group to something warm
13. You can grill or boil the beef
14. She felt a rough spot on her skin
15. The couple gathered nuts in the woods
16. His bunk was flat and made of wood
17. The bird left to search for its flock
18. Move over and join us on the couch
19. The little toy has a button you can press
20. This season the grapes will change to purple
21. Use this form to grade his skill
22. I will just try to sleep in the van
23. There was no time to golf last season
24. Drop the reigns and drag the saddle
25. The nest is full of bees that buzz

25 Sentences: High-Frequency, High-Neighborhood Density

1. The rope has been tied in a knot
2. Find a wide pair of shoes for the man
3. Use a fan to keep the room cool
4. I took a long bath in the tub
5. We tried to sink the British fleet
6. The first half of the test was hard
7. Get all the soap off your face
8. He said he would come to the dock
9. The map is on the deck of the ship
10. You can’t park where they load freight
11. A lion’s roar is a sign of fear
12. Set the pole straight so it doesn’t lean
13. The team fought to win the game
14. This season the grapes will change to purple
15. His bunk was flat and made of wood
16. The nest is full of bees that buzz
17. The bird left to search for its flock
18. You can send the bill in the mail
19. You could have been a little
20. The nest is full of bees that buzz
21. Use this form to grade his skill
22. This season the grapes will change to purple
23. There was no time to golf last season
24. Drop the reigns and drag the saddle
25. The nest is full of bees that buzz

REFERENCES


**REFERENCE NOTE**