Published in final edited form as:

Ear Hear. 2008 October ; 29(5): 725–745. doi:10.1097/AUD.0b013e31817bdd42.

Classification and cue weighting of multidimensional stimuli with speech-like cues for young normal-hearing and elderly hearing-impaired listeners

Xin Wang and Larry E. Humes

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

Abstract

Purpose—The purpose of this study was to investigate how young normal-hearing (YNH) and elderly hearing-impaired (EHI) listeners make use of redundant speech-like cues when classifying non-speech sounds having multiple stimulus dimensions.

Design—A total of four experiments were conducted with 10–12 listeners per group in each experiment. There were 27 stimuli, making use of all possible combinations of 3 stimulus values along each of three cue dimensions. Stimuli were comprised of two brief sequential noise bursts separated by a temporal gap. Stimulus dimensions were: 1) the center frequency of the noise bursts; 2) the duration of the temporal gap separating the noise bursts; and 3) the direction of a frequency transition in the second noise burst.

Results—Experiment 1 verified that the stimulus values selected resulted in adjacent steps along each stimulus being easily discriminable [$P(c) \geq 90\%$]. In Experiment 2, similarity judgments were obtained for all possible pairs of the 27 stimuli. Multidimensional scaling confirmed that the three acoustic dimensions existed as separate dimensions perceptually. In Experiment 3, listeners were then trained to classify three exemplar stimuli. Following the training, they were required to classify all 27 stimuli and these results led to the derivation of attentional weights for each stimulus dimension. Both groups focused their attention on the frequency-transition dimension during the classification task. Finally, Experiment 4 demonstrated that the attentional weights derived in Experiment 3 were reliable and that both EHI and YNH participants could be trained to shift their attention to a cue dimension (temporal-gap) not preferred in Experiment 3, although older adults required much more training to achieve this shift in attention.

Conclusion—For the speech-like, multidimensional acoustic stimuli employed here, YNH and EHI listeners attended to the same dimensions of the stimuli when classifying them. In general, the EHI listeners required more time to acquire the ability to categorize the stimuli, as well as to change their focus to alternate stimulus dimensions.

Many sounds a listener encounters in everyday listening contain multiple, often redundant, acoustic cues. In the past half century, for instance, researchers have discovered that many acoustic cues, including static and dynamic frequency and intensity cues, contribute to the perception of different classes of speech sounds (e.g., Liberman et al., 1954; Stevens and Blumstein, 1978; Lisker, 1986; Nearey, 1989; Dorman and Loizou, 1996). Likewise, other researchers have identified multiple cues that enable the identification of other complex, acoustic stimuli, such as environmental sounds (e.g., Ballas, 1993; Gygi, Kidd & Watson, 2004). The way in which the listener makes use of multiple acoustic cues when processing many naturally occurring sounds can be considered a fundamental aspect of auditory processing. For example, Francis, Baldwin and Nusbaum (2000) demonstrated that, even though redundant acoustic cues may be available acoustically for the categorization of stop

consonants, listeners tended to base their categorization responses primarily on one (and the same) acoustic cue.

In the context of speech perception, the phonetic trading relation theory proposed by Repp (1982) indicated that, when multiple cues signal the same percept, strengthening one of them could offset the weakening of another. As people age, it is well known that many of them develop a high-frequency sensorineural hearing loss (e.g., Corso, 1963; Spoor, 1967; Moscicki et al., 1985; Wiley et al., 1998), which essentially serves to low-pass filter conversational-level speech stimuli for the older listeners. Age-related hearing loss progresses over many years and often 10–15 years pass before an older person seeks help for the hearing loss (Gianopoulos and Stephens, 2005). As a result, it is reasonable to ask whether long-term high-frequency hearing loss will lead older listeners to weight certain cues in complex, multidimensional acoustic stimuli differently from young listeners with normal hearing when the audibility of the higher frequencies has been restored.

For speech, there is some evidence to support the notion that older listeners with impaired hearing focus on cues which differ from those preferred by listeners with normal hearing. Lindholm and colleagues (1988), for example, investigated how young normal-hearing listeners and older listeners with mild-to-moderate sensorineural hearing loss identified stop-consonant place of articulation by using 18 synthesized CV syllables which contained three acoustic cues (formant transition, spectral tilt, and abruptness of frequency change). They found that compared to normal-hearing young adults, older hearing-impaired listeners relied more on spectral tilt and abruptness of frequency change than younger normal-hearing listeners, which suggests that temporal characteristics and gross spectral shape provided more information to the older hearing-impaired listeners in the perception of stop consonants. On the other hand, young normal-hearing listeners primarily relied on the frequency transition cue.

Elderly hearing-impaired listeners have also been found to place more weight on temporal cues rather than spectral cues in stop consonant place of articulation identification in both unaided and aided situations. Hedrick and Younger (2001), for example, examined how young normal-hearing listeners and older listeners with mild-to-moderate sensorineural hearing loss weighted two available consonant cues onset frequency of the second and third formant (F2/F3) transition and the consonant-to-vowel ratio (CVR) measured between the amplitude of the consonant burst and that of the following vowel in the F4/F5 frequency region. These cues were contained both in unaided synthetic /CV/ syllables and in the same stimuli processed by a K-amp amplitude-compression circuit. They found that normal-hearing listeners placed more weight on the frequency transition cue for processed stimuli than for unprocessed stimuli. However, older hearing-impaired listeners, who originally weighted the CVR cue more than the frequency transition cue with unprocessed stimuli, weighted the CVR cue even more heavily with processed stimuli. The results suggested that the two populations weighted the available cues differently, especially in the aided situation. Relative to the frequency transition cue, CVR may be a more important cue for older hearing-impaired listeners for a stop consonant place of articulation identification test. A more recent study completed by Hedrick and Younger (2007) confirmed their previous finding. They further discovered that the weightings of those two cues for voiceless stop consonants depend on factors such as listening conditions (quiet, noise and reverberation), hearing loss and age. In summary, these research findings on a limited set of CV speech stimuli suggest that older hearing-impaired listeners weight some of the multiple cues available for syllable identification differently than younger adults with normal hearing.

In spite of the wide application of synthetic speech in the study of acoustic cues, use of non-meaningful, complex, multidimensional stimuli provides unique advantages compared to

speech. It not only enables the construction of novel stimuli having unique stimulus dimensions, but also enables the study of the acquisition of learned categorization or classification by younger and older adults (Christensen & Humes, 1997; Holt & Lotto, 2006). For example, Christensen and Humes (1997) examined how young normal-hearing listeners made use of speech-like cues in stimuli with multiple acoustic dimensions. Stimuli were comprised of a sequence of two brief noise bursts having three stimulus dimensions or cues including a static spectral cue (center frequency of a frication noise), a dynamic spectral cue (the extent of frequency transition in a formant resonance running through the second of two noise bursts), and a temporal cue (the duration of a silent gap between the two noise bursts). They found that the pattern of classification responses for most of the young normal-hearing listeners reflected a greater weighting of the frequency-transition dimension. Others have also constructed artificial, complex, multidimensional sounds to study the contributions of each dimension to the ability of young normal-hearing listeners to classify or categorize the stimuli (Holt & Lotto, 2006). Christensen and Humes (1997) also demonstrated that it was quite easy for young normal-hearing listeners to shift their attentional weights to other, non-preferred, redundant stimulus dimensions with minimal amounts of training.

As a follow-up to the Christensen and Humes (1997) study, in this study we examined how elderly hearing-impaired listeners made use of multiple speech-like cues in stimuli similar to those used in Christensen and Humes (1997) study. Due to their hearing loss, typically developing over a period of several years, elderly hearing-impaired listeners might weight some stimulus dimensions differently than young normal-hearing listeners. Specifically, due to the degradation or loss of the high-frequency spectral cue from the high-frequency sensorineural hearing loss, the elderly hearing-impaired listeners might give primacy to other cues, such as temporal cues (e.g., Repp, 1982). If this is true, then we would expect to see different cue weighting strategies between the two populations.

In addition, the plasticity of the aged brain will be examined in this study by training older listeners to shift their attention from one stimulus dimension to another. If the older adults are trainable, they should be able to shift their attention from one cue dimension to another, the same as the young normal-hearing listeners (Christensen and Humes, 1997). Furthermore, because elderly people, regardless their hearing status, often experience general cognitive declines (e.g., Sommers, 2005), we may see a difference in terms of the time course of initial category learning and in subsequently learning to shift weighting from one dimension of the stimuli to another.

Finally, spectral shaping, similar to the shaping provided by a well-fit hearing aid, will be applied to stimuli to ensure full audibility of each dimension. The questions of interest here were, once audibility has been restored, are there differences in cue weighting between young normal-hearing adults and older hearing-impaired adults, as well as in the time required to learn these categories? There was little doubt that inaudible cues would receive no attention from the elderly listeners. We were most interested in how the groups would differ once audibility had been restored for the older adults with impaired hearing.

There were a total of four experiments in this study. The first two experiments were necessitated by the fact that the stimuli used here differed somewhat from those used previously by Christensen and Humes (1997) due, in part, to the need to ensure audibility of the high-frequency portions of the stimuli. The first experiment measured the ability of the participants to discriminate stimuli differing by one step along each of the three stimulus dimensions. If participants could not perform this discrimination task readily, following a minimum amount of training, they could not continue into the remaining experiments. The second experiment made use of similarity ratings and multidimensional scaling to establish

that the three acoustic dimensions were mapped to three largely independent perceptual dimensions. In the third experiment, the participants learned to classify three exemplar stimuli comprised of three redundant cue dimensions and, once capable of classifying these exemplars with 90% accuracy, were asked to classify the full set of 27 stimuli generated by all combinations of the three stimulus values along each of the three stimulus dimensions. It is in this experiment that the attentional weights given to each stimulus dimension by the listener are generated. Finally, in the fourth experiment, once the listener's weights have been derived in Experiment 3, the ability to shift the attentional weights to a new stimulus dimension is examined. Additional details for each experiment are provided below. We begin with a description of general methods applicable to the entire study and then present the methods and results for each experiment in turn.

GENERAL METHODS

Participants

Two groups of native English speakers who differed in age and hearing loss participated in this study: 1) young normal-hearing listeners (YNH); and 2) elderly hearing-impaired listeners (EHI). There were 11 YNH listeners between the ages of 19 and 27 y ($M = 22$ y) with 1 male and 10 females. A total of 10 YNH participants finished the whole experiment and partial data are available from one participant who withdrew from the study after completing about 95% of the study. There were 12 EHI participants between the ages of 66 and 79 y ($M = 72$ y) with 6 males and 6 females. Three of them, two males and one female, had been using hearing aids for both ears for about one to three years prior to entering the study. A total of 16 EHI participants had enrolled in this study, but three could not achieve the criterion performance level for the initial discrimination task and one dropped out this study for personal reasons prior to completing the initial set of discrimination measurements.

All YNH participants had pure-tone air-conduction thresholds of less than or equal to 20 dB HL at octave frequencies from 250 to 8000 Hz (ANSI, 1996) bilaterally. All EHI participants had bilateral sloping sensorineural hearing loss of less than 65 dB HL at octave frequencies from 500 to 3000 Hz in order to ensure a ≥ 10 dB sensation level for the stimuli at each octave frequency. The right ear was chosen as the test ear. The audiogram for the test ear of each EHI participant is illustrated in Figure.1.

Stimuli

Twenty-seven stimuli, representing all possible combinations of three stimulus values along three cue dimensions, were developed using the Klatt speech-synthesis software (Klatt and Klatt, 1990). Stimuli were patterned after those used by Christensen and Humes (1997), but several parameters were varied. First, the specific version of the Klatt synthesizer used in this study differed from that used previously by Christensen and Humes (1997) and some of the parameters available differed slightly across versions. Second, fricative noise center frequency was increased from around 1650 Hz to 2250 Hz (600 Hz on average) in order to move this cue to a frequency region for which most of the EHI participants had sensorineural hearing loss (See Figure 1). We assumed that it might be less likely for participants to pay attention to spectral cues within this frequency range due to the inaudibility of sounds in this frequency region over a period of many years. Third, pilot data showed that, EHI participants could not easily [Percent correct, $P(c) \geq 90\%$] discriminate the temporal-gap and frequency-transition cues when the same set of the stimuli in Christensen and Humes (1997) was used. Therefore, the temporal gap cue was increased slightly and endpoint of the relative frequency transition was enhanced considerably to make these cue dimensions more salient and easier for EHI participants to discriminate. Finally, the duration of the second noise burst in this study was doubled (from 100 to 200 ms) for the

same reasons. The synthesizer parameter values for the final stimuli are shown in Table 1. It should be noted that there was no attempt to make the stimuli too speech-like in values along each cue dimension or it would begin to sound like speech. The duration of the temporal gap, the extent of the frequency transition and the center frequencies of the frication noise were selected to generally approximate those values encountered in speech, but did not mimic the acoustic properties of any specific speech sounds closely.

The stimuli each contained one 100-ms fricative noise burst which was separated by a temporal gap (silent interval) from a second fricative noise burst that was 200-ms in duration. In addition, the second burst differed from the first one by having a formant-frequency transition throughout. A total of three cue dimensions were available in each stimulus. These dimensions were: 1) the center frequency of the fricative noise; 2) the temporal gap between the two fricative noise bursts; and 3) endpoint of the frequency transition in the second fricative noise burst. The three values (small, medium and large) along each of the three stimulus dimensions include were 2000, 2250, and 2500 Hz for center frequency of the fricative noise, 10, 70 and 130 ms for temporal gap, and 1500, 2250 and 3000 Hz for endpoint of the frequency transition, respectively. Each dimension could take one of the three possible values for a given stimulus. Another low-frequency bandpass-filtered noise (500 to 1000 Hz), which only represented the temporal gap cue, was created and added to each stimulus in order to ensure that the temporal gap cue would be available in both low- and high-frequency regions. The reason for adding this low-frequency temporal gap cue was to make this cue a broad-band temporal cue that was available to the listener in both the near-normal-hearing low-frequency region as well as the higher frequency region with cochlear pathology.

Three exemplar stimuli using the smallest, middle, and largest values in each of the three cue dimensions are illustrated schematically in Figure 2. The top panel represents the stimulus with the lowest values on the center frequency of the 1000-Hz wide fricative noise burst (center frequency = 2000 Hz), lowest values on the endpoint of the frequency transition in the second fricative noise burst (frequency transition = 1500-Hz endpoint, or -500 Hz frequency transition), and smallest temporal gap between the first and second noise burst (temporal gap = 10 ms). The middle panel represents the stimulus with the medium values on the three dimensions (center frequency = 2250 Hz, frequency transition = 2250-Hz endpoint, or 0 Hz frequency transition, and temporal gap = 70 ms.). The bottom panel represents the stimulus with the highest or largest values on all three dimensions (center frequency = 2500 Hz, frequency transition = 3000-Hz endpoint, or +500 Hz frequency transition, temporal gap = 130 ms).

It should be noted that the labels for each stimulus dimension (noise center frequency, frequency transition endpoint, and temporal gap duration) are the nominal labels for each of the three dimensions. For example, since the durations of the first and second fricative noise bursts were fixed at 100 and 200 ms, respectively, increasing temporal gap size also increased overall duration from stimulus onset to offset. In either case, however, this is a temporal cue. Similarly, endpoint frequency for the transition could also be defined as the direction and extent of the relative frequency transition (-500, 0 and +500 Hz, respectively). Regardless, this comprises a dynamic spectral cue. Finally, since the bandwidth of the frication noise was fixed at 1000 Hz throughout, the nominal stimulus-dimension label of noise center frequency could easily be labeled as the upper or lower frequency of the noise band. Regardless, this cue represents a static spectral cue confined to the higher frequencies.

All stimuli were equated for rms amplitude following synthesis. Then, spectral shaping was applied to ensure the audibility of the high-frequency part of the stimuli for EHI participants with varying degrees of sloping high-frequency sensorineural hearing loss (see Figure 1).

The amplitude at 1250, 1600, 2000, 2500, and 3150 Hz was increased by 4, 8, 12, 16 and 20 dB, respectively, approximating the gain that would be provided by various hearing-aid gain-prescription formulas (Humes, 1991). After the spectral shaping, all 27 stimuli were re-equalized in amplitude and were bandpass-filtered from 1500 to 3000 Hz, the frequency region in which the frequency transition and center frequency cues were located. Finally, to make the temporal-gap cue available over a broader range of frequencies, low-frequency bandpass-filtered noises (500 to 1000 Hz), which only contained the temporal gap cue, were created and added to the stimuli. The amplitude of the low-frequency bandpass-filtered noise was about 15 to 18 dB lower than the peak value of the stimuli to minimize upward spread of masking. Therefore, in the final stimuli, the frequency transition cue and the center frequency cue were available only in the high frequencies. The temporal gap cue, however, was available at both low and high frequencies.

In order to calibrate the sound pressure level of the 27 stimuli, we created one 2-s calibration noise CalMin, which matched the minimum amplitudes of the FFT spectra across all 27 stimuli, using Adobe Audition 1.5. After the calibration stimulus was created, the overall (linear) and one-third octave band levels in dB SPL were measured, the latter at center frequencies from 125 through 8000 Hz, using a Larson-Davis model 800 sound level meter. The calibration sounds were measured in a 2-cm³ coupler, the same reference point used for the calibration of dB HL for the threshold measurements. For the entire set of 27 stimuli, the maximum and minimum overall levels for the stimuli were 106 and 91.4 dB SPL, respectively. The 1/3-octave-band sound pressure levels of the calibration noise and the average hearing threshold for both YNH and EHI participants can be seen in Figure 3. The latter are based on the assumption that thresholds for pure tones and for 1/3-octave-band noises in hearing-impaired listeners are equivalent (Cox & McDaniel, 1986). Although the stimuli were equated for rms amplitude after spectral shaping and filtering, the shaping needed to ensure audibility had differential effects on the amplitudes of the 27 stimuli. Basically, those stimuli with higher center frequencies and higher frequency transition endpoints had greater amplitudes than those with lower values on each of these dimensions. To minimize the use of this covarying amplitude cue, attenuation varying from 0 to 10 dB was randomly added to all stimuli across intervals of a trial (i.e., intensity was roved). Therefore, from interval to interval, the overall sound level could vary from a low of 81.4 dB SPL to a high of 106 dB SPL, depending on the specific stimulus selected and the random attenuation setting. From Figure 3, we can see that the minimum amplitudes across the 27 stimuli (CalMin) were at least 30 dB sensation level for the average EHI participants threshold from 500 to 3000 Hz. This meant that, even with the extreme amount of amplitude roving possible (−10 dB), the minimum amplitudes across the 27 stimuli would still be at least 15–20 dB above threshold for the EHI listeners. Similarly, all stimuli were at or above threshold even for the EHI participants with the worst thresholds (Maximum EHI threshold) from 500 to 3000 Hz and the extreme amount of amplitude roving (−10 dB).

Procedures

All EHI participants passed the Mini-Mental State Exam, a screening test of cognitive function, with a MMSE score of at least 27 of 30 (Folstein et al., 1975). All testing was completed in a single-walled sound booth having attenuation characteristics compatible with ears-covered threshold measurements (ANSI, 1991). Participants were seated in front of the computer monitors and oral instructions were given before the test started. The stimuli were delivered through ER-3A insert earphones.

Listeners participated in several 1- to 2-hour sessions each week for 4 to 8 weeks. On average, YNH subjects needed about 11 sessions to complete the whole project. EHI subjects, on the other hand, showed a much wider variation in terms of the sessions needed (11 to 33). Participants were paid for their participation.

EXPERIMENT 1: DISCRIMINATION OF ADJACENT STIMULUS PAIRS

Method

Discrimination of stimuli along each stimulus dimension was measured using a modified two-alternative-forced-choice paradigm (2AFC). Three stimuli were presented per trial with the first sound as the standard and the following sounds being either the same or different from the standard sound. Participants were required to listen to all three stimuli then indicate which of the two sounds following the standard differed from the standard sound by clicking on the “First” or “Second” button on the computer screen. All of the 27 stimuli were paired by the target dimension, the dimension that participants were required to discriminate in each session, and the adjacent values (i.e., smallest to middle, or middle to largest) along each target dimension. For instance, gp12 means the smallest and middle value along temporal gap dimension were compared. The adjacent values along each dimension were compared while the context (other two dimensions, each taking one of the three possible values) was varied randomly from trial to trial. To balance the order, the adjacent values were compared twice. For example, for discrimination of the stimuli differing in temporal gap, the stimuli with the smallest temporal gap served as the standard stimulus and the comparison stimulus included those stimuli with the middle temporal gap. In another block involving this pairing of the smallest and medium temporal gaps, stimuli with the medium gap served as the standard and the comparison stimulus had the smallest temporal gap. In total, there were 12 pairs of discrimination measurements (3 dimensions $\times$ 2 values $\times$ 2 orders), namely, fn12, fn21, fn23, fn32, gp12, gp21, gp23, gp32, tr12, tr21, tr23 and tr32, in which “fn” stood for center frequency of the frication noise, “gp” stood for temporal gap and “tr” stood for frequency transition, respectively. In each pair of these discrimination measurements, there were a total of nine contexts (for example, 3 center frequencies $\times$ 3 frequency transitions = 9 contexts in the case of temporal-gap comparisons,) and each was presented eight times in random order for a total of 72 trials in a given block of discrimination trials. On a given discrimination trial, however, the context was fixed. Prior to measurement of discrimination performance, each participant received 27 trials of each discrimination pair as practice. Trial-by-trial feedback was provided during practice and the test sessions for the discrimination measurements. Participants were required to get 90% accuracy for each stimulus contrast before proceeding to Experiment 2. If a participant failed to reach 90% accuracy for specific stimuli, the corresponding 27-trial training sessions were repeated, followed by another set of discrimination measurements.

Results and Discussion

All YNH and most of the EHI participants passed the discrimination test with an average accuracy of $\geq 90\%$ for each stimulus pair after the training. Due to the limited number of prospective older participants and the relatively high difficulty of the task, two EHI participants who had minor difficulty (accuracy scores around 80%) in discriminating one or two (of the 12) stimulus pairs were retained as participants for this experiment. In general, YNH listeners required 2 to 6 sessions ($M = 5$ hours; $SD = 2$ hours) to complete the training for the discrimination task. EHI participants required 2 to 8 sessions ($M = 6.7$ hours; $SD = 2.6$ hours) to complete the training. Independent-sample *t*-tests showed there was no significant difference between YNH and EHI participants in terms of the training time required to reach the criterion performance level ($t = 1.87$, $p = 0.08$).

Percent-correct scores were transformed into rationalized arcsine units (Studebaker, 1985) to stabilize the error variance prior to statistical analysis. The GLM analysis showed significant main effects of cue type [$F(2, 42) = 24.85$, $p = 0.00$] and participant group [$F(1, 21) = 4.96$, $p = 0.04$]. However, there was not a significant interaction between cue type and participant group [$F(2, 42) = 0.60$, $p = 0.55$]. The YNH group performed significantly better than EHI

group in discriminating all three cue dimensions (see Figure 4). In order to examine the differences among the 3 cue types, a series of paired-sample t-tests were performed. Results showed that the frequency-transition cue was discriminated significantly better than temporal-gap cue ($t = -4.097$, $p = 0.00$); and temporal-gap cue was discriminated significantly better than center-frequency cue ($t = -2.261$, $p = 0.03$). Nonetheless, the data in Figure 4 indicate that the three cue dimensions could be easily discriminated ($\geq 90\%$) for both groups after a brief amount of training. Table 2 shows participants' final performance on the discrimination test. It is clear that the adjacent values in each of the three stimulus dimensions could be easily discriminated by both groups after a brief amount of training ($\geq 90\%$). Of course, if adjacent steps are easily discriminated, then the larger steps (between minimum and maximum values) can be assumed to be even more readily discriminated.

EXPERIMENT 2: SIMILARITY JUDGMENTS AND MULTIDIMENSIONAL SCALING

Method

Following the discrimination test and training, participants were required to make similarity judgments for all 351 possible pairings of the 27 stimuli. Two sets of measurements, with and without intensity rove, were obtained in order to evaluate the contribution of a potential intensity cue associated with the combination of spectral shaping and either of the two frequency-based cues (center frequency spectrum and frequency transition). Within each set, both orders of the comparison stimuli were used. No practice session was given. The pairs of stimuli were presented in random order. There was a 1-second silent interval between the two stimuli comprising a given paired comparison. Participants were required to judge the similarity of the two sounds based on a nine-point scale. A rating of one indicated that the two sounds were very similar; a rating of nine indicated that the two sounds were very different. Participants were instructed to listen to the sounds and make the similarity judgment while ignoring the intensity differences between the two stimuli. The response consisted of a mouse click on a number (1–9) displayed on the computer monitor, followed by pressing the Enter key on the keyboard. Another pair of stimuli was then presented 1.5 seconds after the response.

Results and Discussion

Multidimensional scaling (MDS; Kruskal and Wish, 1989) was used to analyze the similarity data from both YNH and EHI groups. MDS is a class of mathematical techniques which uncover the hidden structure of the raw data. It takes proximities, numbers indicating how similar any two objects are, as input, calculates the distances between those objects through certain mathematical algorithms (e.g. Euclidean, etc.), and then generates a spatial configuration or representation of the proximities as the output. Therefore, the more dissimilar the two objects are, as shown by their proximity value, the further apart they will be in the spatial representation and vice versa.

Recall that two sets of similarity ratings were obtained from all participants. One set without amplitude roving imposed and one set with amplitude roving across intervals. Therefore, two proximity matrixes were generated for the data sets, one with and one without amplitude rove for each group. The proximity matrices were analyzed using the Alternating Least Squares Scaling (ALSCAL) MDS algorithm (Kruskal and Wish, 1989). For both groups, a three-dimensional ALSCAL MDS solution provided the best fit for the data and their ALSCAL results for the two sets of similarity ratings were very similar (stress = 0.11 to 0.12, $r^2 = 0.87$ to 0.90) for rove and non-rove stimulus sets. Stress and r^2 are goodness of fit statistics which MDS attempts to optimize. The smaller the stress value, or the bigger the r^2 , the better the MDS fit. Since the results from both stimulus sets were very similar, the two

data sets were pooled for analysis. Therefore, one proximity matrix was generated from the pooled similarity judgments for each group. Subsequent MDS analyses showed a three-dimensional ALSCAL MDS solution provided the best fit for the data. The corresponding stress and r^2 values were similar for both groups (YNH: stress = 0.09, $r^2 = 0.91$; EHI: stress = 0.07, $r^2 = 0.95$) and the three perceptual spaces for the two groups were nearly identical. As a result, the two proximity matrices from the YNH and EHI participant groups were pooled together to generate one average proximity matrix, representing the average similarity judgments from both groups, which was again analyzed using the ALSCAL algorithm. Results showed that a three-dimensional ALSCAL MDS solution provided the best fit for the data with a stress value of 0.08 and r^2 value of 0.94. This result suggests that three separate stimulus dimensions exist not only physically, but also perceptually. Therefore, MDS analysis confirmed that the acoustical dimensions in the stimuli mapped to separate perceptual dimensions internally.

To better interpret the data, the original axes were rotated while the distances between points in the three-dimensional spaces remained unchanged. Three two-dimensional plots were generated following the rotation and these are shown in Figure 5. From Figure 5 we can see that the perceptual dimensions discovered from the similarity judgments match the acoustical dimensions. For instance, from the top panel of Figure 5, a plot of the temporal-gap dimension by the frequency-transition dimension, we can see that on the Temporal Gap axis, stimuli S1 to S9, which all have the smallest temporal gap (10 ms), are grouped together at the lower end of the axis (left column), while stimuli S10 to S18, which have the medium temporal gap (70 ms), are grouped in the middle column, and stimuli S19 to S27, which have the largest temporal gap (130 ms), are grouped at the high end (right column) of the axis. The same consistent mapping of acoustic dimensions to perceptual dimensions can be seen in the Frequency Transition axis for the top panel, as well as in the other two two-dimensional plots in Figure 5 (middle and bottom panels).

Individual MDS weights, which participants assigned to each cue dimension when they did the similarity judgments, were calculated by using the individual scaling (INDSCAL) MDS solution (Kruskal and Wish, 1989) for all participants. In Figure 6, the distribution of individual weights is graphed for YNH and EHI participants for all three cue dimensions. For the YNH group, all participants, except Participants 1, 2 and 3, distributed their weights across the three dimensions (INDSCAL weights are from 0.28 to 0.66). Participant 1, 2 and 3, however, allocated most of their attention (INDSCAL weights ≥ 0.7) to the center-frequency dimension (INDSCAL weights are from 0.76 to 0.91). Similarly, for the EHI group, all participants except Participants 1 and 2 distributed their weights across the three dimensions (INDSCAL weights are from 0.32 to 0.65). Participants 1 and 2 directed most of their attention (INDSCAL weights are ≥ 0.7) to the center-frequency dimension. These results suggest that, at this stage in the testing, except for three YNH and two EHI participants who allocated more attention to center frequency, the bulk of the listeners in each group are distributing their attention across all three perceptual dimensions when judging the similarity of stimulus pairs. That is, except for those few participants noted in each group, no particular stimulus dimension dominates perceptually. This is true even though slight, but statistically significant, differences in discrimination performance were observed in the first experiment with the highest discrimination performance observed for the frequency-transition cue.

EXPERIMENT 3: IDENTIFICATION AND CLASSIFICATION OF STIMULI

Method

In this experiment, participants were trained to identify and classify stimuli as circle (exemplar stimuli with lowest or smallest values along each of the three stimulus

dimensions), triangle (exemplar stimuli with middle values along all three dimensions) and square (exemplar stimuli with highest or largest values along all three dimensions). First, a familiarization session was provided to make sure that participants understood what they were expected to do in this task. After one of the three exemplar stimuli was presented randomly, three buttons labeled as circle, triangle and square were displayed on the computer screen. Participants were required to name the stimulus they just heard by clicking on one of the three buttons using the computer mouse. Immediate feedback was provided on each trial. After a participant clicked one of the three response buttons, a message appeared on the computer screen indicating “Your answer is correct”, if the participant chose the correct answer or, if incorrect, “Your answer is incorrect. The correct answer is xxx”. Following the familiarization session, additional exemplar training was provided using the same identification task with feedback. Exemplar training continued until participants reached the criterion performance level (i.e., participants correctly identified all three exemplars with $\geq 90\%$ accuracy). After participants reached the criterion performance level in the exemplar training, they were required to classify all 27 stimuli as circle, triangle, or square, using the rules for stimulus classification they learned in the exemplar training task, without any feedback. Brief refresher training was provided every time participants returned for a new test session to make sure that the criterion performance level for identification was maintained over time. Participants were required to classify a total of 80 repetitions for each of the 27 stimuli (a total of 2,160 stimulus presentations) in random order. As in the previous tests, participants were instructed to ignore the intensity differences when they tried to identify and classify the stimuli.

Results and Discussion

All YNH and EHI participants reached 90% identification accuracy after the exemplar training. In general, YNH participants required one training session (75 trials), which took about 10 minutes, to reach 90% accuracy. EHI participants, especially certain individuals, required much more training in order to identify the stimuli correctly 90% of the time. On average, EHI listeners required 1.2 hours (SD = 1.3 hours) to reach 90% accuracy in the identification of the exemplars. An independent-sample t-test showed that this was a significant difference between YNH and EHI participants in terms of the training time required to reach 90% identification accuracy for the exemplar stimuli. The EHI participants needed significantly more training time than the YNH participants ($t = 2.75$, $p = 0.02$) to achieve this level of accuracy when identifying the three exemplar stimuli.

After learning to successfully identify ($\geq 90\%$ accuracy) the exemplar stimuli, the listeners then proceeded to the classification or identification of all 27 stimuli. The generalized context model (GCM; Nosofsky, 1986) was used to predict performance in the categorization experiment. The GCM model is an exemplar model which assumes that exemplars of categories are stored in memory to represent these categories and the category decisions are made based on the similarity calculation between a probe stimulus and stored exemplars. The GCM, combined with the MDS solution, was used to calculate the attention weights participants placed in each of the cue dimensions in the classification task. The distances between different stimuli in the MDS space determines the similarity among those stimuli, which can be further modified by the selective attention a listener places on each of three cue dimensions. According to GCM, the likelihood that stimulus i will be placed in category J is found by summing the (weighted) similarity of stimulus i to all of exemplars of category J , multiplied by the response bias for the category. This likelihood is then divided by the sum of likelihoods for all categories to determine the conditional probability (GCM weight) that stimulus i will be classified in category J . The GCM weights range from 0 to 1 and must sum to 1.0. [For a full review of the GCM see Nosofsky (1986).]

The GCM weights calculated for each participant across all 3 cue dimensions are shown in Figure 7. In the GCM model, the percentage of variance accounted for by the best-fitting model (pcvar) is one of the values used to evaluate the goodness of fit. The higher the pcvar value, the better the fit. The GCM weights for 9 of the 10 YNH (8 participants' pcvar $\geq$ 95%; 1, pcvar = 86%) and 10 of the 12 EHI participants (7 participants' pcvar $\geq$ 88%, 10 participants' pcvar $\geq$ 78%) were good fits to their classification data. Among these participants, as we can see from Figure 7, seven YNH participants (subject 1–3, 6–7, 9–10) were frequency transition users (majority of their attention was allocated in the frequency-transition dimension) and two YNH participants (subject 4 and 5) were temporal gap users (majority of their attention was allocated in the temporal gap dimension). Similarly, eight EHI participants were frequency-transition users (subject 1–6, 8–9) and 1 EHI participant (subject 7) allocated attention to temporal gap and frequency transition dimensions. One EHI participant (participant 12), whose GCM weight was not as good as the nine EHI participants (pcvar = 78%), was a temporal gap user. For 1 of the 10 YNH (subject 8, pcvar = 69%) and 2 of the 12 EHI (subject 10, 11, pcvar = 39%, 28%) participants, however, they had worse GCM fits. For those participants, the YNH participant (subject 8) allocated attention between the frequency-transition and temporal gap dimensions; 1 EHI participant (subject 10) was a frequency-transition user, and another (subject 11) allocated his attention to the frequency-transition and temporal-gap dimensions. In general, center frequency is the dimension to which the listeners in each group paid the least attention. Most individuals in each group focused their attention on the frequency-transition cue during the classification task. The salience of the formant transition is most likely caused by its importance in speech perception, although some other factors, such as the later temporal location (the second half of the stimulus) and longer duration (200 ms), might also have influence on the weightings (Whalen, 1991).

To illustrate different cue weighting strategies used by listeners in the classification test, the classification results for two typical cue users from the EHI group are plotted in Figure 8. When the 27 stimuli are ordered by frequency transition value along the x-axis, subject 2's classification performance (top panel) reveals that classification decisions were determined almost exclusively by the frequency-transition cue, regardless of the value of the temporal-gap or center-frequency cues. On the other hand, the classification data from subject 12 (middle panel) do not have the same pattern. The classification results are more orderly for this participant when the 27 stimuli are ordered by temporal gap value along the x-axis (bottom panel) than when ordered according to frequency transition (middle panel). To the right of each panel in Figure 8, we can see that subject 2 paid most (almost exclusive) attention (GCM weight = 0.99) to the frequency-transition dimension whereas subject 12 paid most attention (GCM weight = 0.86) to the temporal-gap dimension. Therefore, subject 2 was a typical frequency-transition user while subject 12 was a temporal-gap user.

EXPERIMENT 4: TEMPORAL-GAP TRAINING

Method

In this experiment, participants were trained to attend to the temporal-gap dimension. Five YNH and 11 EHI participants from the previous experiments completed this experiment. Another 7 YNH participants were recruited for this experiment using the same criteria as other YNH participants from previous experiments. Therefore, a total of 12 YNH and 11 EHI participants participated in this experiment.

Since about five months had passed between the completion of the first three experiments and the beginning of this experiment, all returning participants from the previous experiments were required to redo the classification task from Experiment 3 as a refresher. To expedite this refresher, fewer stimulus presentations (half, or 1,080 stimulus

presentations) and only an 80% exemplar-identification criterion (versus 90% in Experiment 3) were employed. However, all the new participants (7 YNH listeners) had to do the full-set of the identification and classification tasks as described previously in Experiment 3 prior to completing this experiment.

After the refresher or the full versions of the classification task, all participants were trained on a subset of the 27 stimuli. A group of stimuli were again presented to participants as circle, triangle or square. Here, however, participants were trained to label stimuli 1, 2, 3, 6, 7 as circle (10-ms gap); 10, 13, 14, 15, 17 as triangle (70-ms gap) and 20, 22, 25, 26, 27 as square (130-ms gap). The training stimuli were chosen based on their perceptual distance in the MDS space derived in Experiment 2, as well as different combinations of values on the center-frequency and frequency-transition dimensions. The training stimuli were distributed evenly in the MDS space and contained all possible combinations of values from other two cue dimensions. As a result, the only common denominator for each stimulus in each of the three categories for the exemplar stimuli was the temporal-gap dimension. Basing categorization responses on this stimulus dimension will optimize classification accuracy. The remaining 12 stimuli not used in training were used as test stimuli to assess the generalization of training to non-trained stimuli.

Training procedures were the same as those of the previous classification task in Experiment 3. First, participants were familiarized with the stimuli and the exemplar training task in a familiarization session (3 repetitions $\times$ 15 training exemplars = 45 trials). Following the familiarization, the exemplar training with trial-by-trial feedback was provided repeatedly until participants reached 80% identification accuracy. Unlike the previous classification task (Experiment 3), however, participants needed to learn to label a group of stimuli, instead of one stimulus, as circle, triangle or square in the exemplar training experiment. Again, they could learn to do so by directing their attention to the temporal-gap dimension, although they were not told explicitly to do this. Rather, they needed to learn to attend to this stimulus dimension through the trial-by-trial feedback received during exemplar training. Each participant had up to 2,400 trials (8 repetitions $\times$ 15 training exemplars $\times$ 20 blocks) to achieve the criterion of 80% accuracy. If participants still could not reach that performance criterion after that, they would not be able to continue with the classification task. Feedback was provided in both familiarization and exemplar training tasks. Once participants reached the exemplar-identification criterion, they would continue with the classification task. In the classification task, participants were presented with all 27 stimuli in random order and required to classify each of them as circle, triangle or square. No trial-by-trial feedback was presented. One exemplar training block was given at the beginning of each new test session as a refresher. A total of 2,160 classification trials (4 repetitions $\times$ 27 stimuli $\times$ 20 blocks) were presented.

Results and Discussion

Since about 5–6 months had passed between completion of Experiment 3 and the initiation of Experiment 4, as noted, “refresher” classification of the exemplar stimuli and full stimulus set was performed by the 11 EHI and 5 YNH listeners from Experiment 3 who participated in Experiment 4. This afforded an opportunity to examine the reliability of the GCM weights derived for 15 out of the total 16 participants (one EHI subject dropped out before completing the “refresher” classification test in Experiment 4). Two paired-sample *t*-tests were computed for the GCM weights for each of the three stimulus dimensions obtained in Experiment 3 and at the beginning of Experiment 4. There were no significant differences in GCM weights on any of the three dimensions ($p > 0.05$) for either group. Further, for both groups, the test-retest correlations were $r > 0.90$ for all three cue dimensions. Thus, these results suggest that the GCM weights derived in Experiments 3 and 4 were reliable.

Most of the participants were successfully trained to allocate their attention to the temporal-gap dimension. One YNH and three EHI participants could not reach 80% accuracy during the exemplar training task after the maximum number of training trials (2,400) and did not continue with the classification task. Another EHI participant dropped out of the project in the temporal-gap training phase. Therefore, a total of 7 EHI and 11 YNH participants completed and passed the temporal-gap training. For the 7 EHI participants, originally 6 were frequency-transition cue users and 1 was a mixed cue user. For the 11 YNH participants, originally 8 were frequency-transition users and 3 were temporal-gap users.

The new GCM weights from the classification task for both YNH and EHI participant groups following the temporal-gap training were calculated and are shown in Figure 9. Recall that only 7 out of 10 EHI participants could actually be trained to shift their attention to temporal gap dimension and therefore completed the classification experiment. As we can see from Figure 9, temporal gap is the dimension that was most heavily weighted by all participants. Comparing this figure to Figure 7, we can clearly see the perceptual weighting shift from the frequency-transition cue to the temporal-gap cue for most of the participants. To verify the finding, two paired-samples t-tests were performed, one for each group, on the GCM weights for the temporal-gap dimension obtained before and after temporal-gap training for those participants completing Experiment-3 measurements (11 YNH and 7 EHI participants). Results showed that the weight assigned to the temporal-gap dimension following the training was significantly higher than the weight obtained before training for both YNH ($t = -5.33$, $p = 0.00$) and EHI ($t = -6.87$, $p = 0.00$) participant groups. Likewise, there was a corresponding significant ($p < 0.01$) decline in the GCM weight for the frequency-transition cue in each group.

The number of blocks needed to achieve the training criterion ($\geq 80\%$ classification) for the YNH and the EHI participants can be seen in Figure 10. Note that although several older adults required an order of magnitude more training than younger adults, some still not achieving the performance criterion then, there were also some older adults who performed well within the range typical of YNH listeners. The median number of training blocks needed by the YNH participants was 1 (interquartile range of 1 to 1) and 17 (3 to 21) for the EHI group. Given the bimodal distribution of values for the EHI listeners in Figure 10, the nonparametric Mann-Whitney U test was used to compare the values from the two groups. The Mann-Whitney U test revealed that there was a significant difference in the average number of blocks needed to reach training criterion between groups ($U = 17.0$, $p = 0.00$). Although there were differences between the two groups in the amount of training time needed to shift attention to the temporal-gap dimension, the distribution of training times was very wide for the EHI participants.

GENERAL DISCUSSION

The purpose of the present study was to investigate how YNH listeners and EHI listeners classify acoustic stimuli containing multiple speech-like stimulus dimensions. Results from Experiment 3 showed that both YNH and EHI participants allocated the majority of their attention to the frequency-transition dimension when classifying these stimuli. The results from the YNH group for the first three experiments are consistent with those of Christensen and Humes (1997), although the stimuli used in present study differ in some details from those used by Christensen and Humes (1997). In addition, Experiment 4 demonstrated that the GCM weights derived were reliable and that both YNH and EHI participant groups could be trained to focus their attention on a non-preferred dimension (temporal gap). This latter finding also had been observed previously in YNH listeners by Christensen and Humes (1997). In general, however, EHI participants needed significantly longer training

time to shift their attention to a non-preferred cue dimension with several remaining unable to do so after the maximum amount of training in this study (2,400 stimulus presentations).

The classification results in Experiment 3 suggest that EHI listeners do not weight the cues in these multidimensional acoustic stimuli differently than YNH participants. The majority of YNH and EHI participants alike allocated most of their attention to the frequency-transition dimension. Francis et al. (2000) found that the frequency-transition cue was also the preferred stimulus cue for YNH listeners for the identification of stop-consonants which were also differentiated by redundant spectral-tilt cues in the initial burst. Similar findings were obtained by Hedrick and Younger (2001) for YNH and spectrally shaped speech stimuli. However, as noted in the introduction, Lindholm et al. (1988) and Hedrick and Younger (2001, 2007) all reported differences in the weighting of cues between YNH and EHI listeners when using speech stimuli. The failure to observe group differences in cue weighting in the present study could be due to the use of non-meaningful, non-speech stimuli in this study. In addition, both Lindholm et al. (1988) and Hedrick and Younger (2001, 2007) manipulated acoustic cues which signaled the consonant portion of the CV syllables and listeners were asked to identify the consonant only. In the present study, however, participants were required to identify the stimuli as a whole using the three acoustic cues. This might be another reason that we failed to see group differences.

Recall, however, that, given the nature of the GCM classification task, a prerequisite was that listeners in this study had to be able to easily and roughly equally ($\geq 90\%$ accuracy) discriminate stimuli differing by one value along each of the three stimulus dimensions. Therefore, it is possible that those EHI listeners who could not achieve the 90% criterion for discrimination of the frequency-transition cue (Experiment 1), and were subsequently excluded from this study were, in fact, participants who had shifted their focus to other dimensions already and couldn't learn to discriminate the differences in frequency transition. As a matter of fact, as is shown in Figure 11, at least one of the three excluded participants (Subject b) had more difficulty on average with the spectrally based cues (center frequency and frequency transition) than those 12 older adults who completed that portion of the study. In general, though, it does not appear that the three EHI participants who could not continue beyond Experiment 1 were primarily temporal-gap users, with one possible exception (Subject b). On the other hand, for the two EHI participants who had minor difficulty (accuracy scores of about 80% instead of 90%) in discriminating one or two stimulus pairs and were retained as participants for the subsequent experiments, one of them was the only gap user among all the EHI participants (temporal gap weight = 0.87) and another distributed her attention to all three dimensions while weighting the temporal gap dimension slightly more than others (temporal gap weight = 0.61).

Holt and Lotto (2006) examined the cue weighting of YNH participants for nonspeech stimuli varying in two acoustic dimensions: center frequency and modulation frequency. Their results suggested that, by manipulating the distribution statistics of the stimuli along the two acoustic dimensions, participants tended to change their weighting from a less informative dimension to a more informative dimension. This result is in accord with the phonetic trading theory of Repp (1982). In the case of hearing loss, the distribution of acoustic cues available to hearing-impaired listeners might vary according to listeners' degree of hearing loss and the time course over which they have experienced the hearing loss. It is possible for elderly hearing-impaired listeners to weight a cue dimension which is more informative to them (e.g. a cue dimension which is still audible given their hearing impairment) more heavily and give less weight to a less- or non-informative cue dimension (e.g., a cue dimension which is no longer fully audible because of the hearing loss). However, the results of our study do not support such an interpretation. For multidimensional acoustic stimuli with stimulus dimensions which are easily (and roughly

equally) discriminable and fully audible, differences in the distribution of attention across stimulus dimensions between the YNH and EHI participants were minimal. This perceptual equivalence of the two groups is also supported by the overlap of the MDS spaces of the two groups and the distribution of INDSCAL weights for both groups in Experiment 2. It needs to be emphasized that this kind of perceptual equivalence is conditional and based on the premise that both YNH and EHI groups could easily [$P(c) \geq 90\%$] and roughly equally discriminate all the stimuli. Indeed, our pilot data suggested that the EHI group, most likely due to their hearing loss, needed more salient cues to achieve good performance [$P(c) \geq 90\%$] in the discrimination task than the YNH group. In this study, we wanted to examine how older hearing-impaired adults would make use of multiple cues after years' of hearing loss when all the cues were roughly equally salient. This is a requirement for the eventual analysis of the results with the GCM. Therefore, to achieve this goal, it was necessary to equalize the performance in the discrimination task for both YNH and EHI groups as a reference point. This is a different question from asking whether use of multiple cues varies when they are not equally salient, as may be the case for naturally produced speech.

It is possible that the initial discrimination training enabled the EHI participants in this study to make better use of previously inaudible cues. To examine this possibility, we looked at the discrimination performance for both groups of listeners after the initial test and these data appear in Figure 12. From Figure 12, we can see that although the YNH group performed better than the EHI group in general, they significantly outperformed the EHI group in only one of the 12 discrimination conditions (t_{12}). Thus, there are no notable differences between the YNH and EHI groups in initial discrimination of adjacent steps along any of these three stimulus dimensions. It does not appear that there were initial performance differences, owing to the presence of high-frequency hearing loss in the EHI listeners, that were trained away to enable the EHI listeners to reach the criterion level of performance for this task. Furthermore, the discrimination test results showed that the frequency-transition cue was significantly better discriminated by both YNH and EHI groups. Given the fact that the frequency-transition dimension was the dimension that both participant groups attended to most, the correlations between participants' discrimination performance and their attention weights were examined. The results showed that there were moderate, but not significant, correlations ($r = -0.51$ to 0.46 ; $p > 0.05$) between participants' discrimination performance in Experiment 1 and their attention weights derived in Experiment 3 for the three cue dimensions.

We further examined the ability of both groups to maintain their identification scores for the three exemplar stimuli during the classification task. Results showed that, on average, both groups maintained good identification for the three exemplars during the classification task (performance remained about 90% correct). While the YNH group almost uniformly performed well (except one participant), there was relatively larger variation in classification performance for the three exemplars among the EHI participants. This same group of participants also showed considerable variation in the amount of training needed with the exemplars to achieve the criterion of 90% correct identification. Perhaps those EHI listeners who required the greatest amount of exemplar training to reach criterion at the beginning of Experiment 3 were also the ones who had the most difficulty identifying these exemplars during the classification of the full set of 27 stimuli at the end of Experiment 3. To examine this possibility, correlations were computed. Correlation analysis showed a strong and significant correlation ($r = -0.82$; $p = 0.00$) between exemplar training time needed to reach criterion and the accuracy with which the exemplars were identified among the full set of 27 stimuli. That is to say, the EHI participants who needed less training time to reach criterion performance at the beginning of Experiment 3 also tended to better classify the three exemplars during the classification task; that is, they could better recognize the three

exemplar stimuli among all 27 stimuli when no feedback was provided in the classification test.

Many EHI participants and most YNH participants could be trained to shift their attention to a generally non-preferred cue dimension (temporal gap) in Experiment 4. Specifically, 5 EHI participants, who originally were frequency-transition cue users, were trained to shift their attention or listener weights to the temporal-gap dimension ($pcvar \geq 91.80$). Similarly, 7 YNH participants, who originally were also frequency transition users, were trained to shift their attention to temporal gap dimension in Experiment 4 ($pcvar \geq 88.86$). Among the YNH listeners in that experiment, three were natural temporal gap users and, as expected, their attention weights on temporal gap dimension increased only slightly after the training (e.g. increased from 0.97 to 1.0, $pcvar \geq 86.07$). The fact that several of EHI participants could be trained to shift their attention to a different cue dimension is noteworthy, even if substantially more time to do so was typically required. This suggests that auditory training could shape the way that EHI participants make use of multiple speech-like cues. In the case of speech, if EHI participants could learn to reallocate their attention to an available redundant cue uncompromised by aging or hearing loss, the loss of information caused by their hearing loss might be minimized. Therefore, if this kind of attention shift through training could also be demonstrated in real speech, it would suggest a possible approach to rehabilitation for the EHI population. Efforts by Francis et al. (2000) to train YNH listeners to focus their attention on a redundant spectral-tilt cue and away from the preferred frequency-transition cue when categorizing stop consonants, however, proved to be unsuccessful. This suggests that, changing an older hearing-impaired listener's cue-weighting strategy may have limited utility as an approach to rehabilitation with natural speech.

It should be noted that all of our older listeners had impaired hearing. Regardless of the substantial variation in the thresholds of different EHI listeners, the performance on the discrimination test and the GCM weights obtained after the exemplar training were not significantly ($r < -0.54$; $p > 0.05$) correlated to listeners' high-frequency pure-tone average (mean of pure-tone thresholds at 1000, 2000 and 4000 Hz). This is somewhat expected given the fact that all 27 stimuli were spectrally shaped to ensure sufficient audibility from 500 to 3000 Hz for all the EHI subjects. Our primary focus is on the use of cues once audibility had been restored for such listeners. As a result, however, it is not possible to attribute the group differences between young and older adults in this study exclusively to age, hearing loss, or the interaction of these two factors. Additional research with elderly normal-hearing and young hearing-impaired listeners would be needed to better identify the factors underlying the observed differences in performance.

There is little question, however, that had this study been conducted without sufficient spectral shaping of the stimuli to ensure audibility then the attention directed to spectral cues would have been diminished. Clearly, unheard acoustic cues can't capture the listener's attention. As a result, the stimuli in this study approximate aided listening conditions in which the audibility of the stimuli has been assured. Under these conditions, both groups of listeners directed most of their attention to the frequency-transition dimension in the classification task.

Acknowledgments

This work supported by a research grant awarded to the second author from the National Institute on Aging (R01 AG 08293). We thank Rob Nosofsky for his assistance with the application of the GCM model. In addition, we thank Gary Kidd and Noah Haskell Silbert for their thoughtful comments on the previous version of this manuscript.

REFERENCES

- American National Standards Institute. Maximum permissible ambient noise levels for audiometric test rooms. (ANSI S3.1-1991). New York: ANSI; 1991.
- American National Standards Institute. Specifications for audiometers. (ANSI S3.6-1996). New York: ANSI; 1996.
- Ballas JA. Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*. 1993; 19:250–267. [PubMed: 8473838]
- Christensen LA, Humes LE. Identification of multidimensional stimuli containing speech cues and the effects of training. *Journal of the Acoustical Society of America*. 1997; 102:2297–2310. [PubMed: 9348688]
- Corso JF. Age and sex differences in pure tone thresholds: survey of hearing levels from 18 to 65 years. *Archives of Otolaryngology*. 1963; 77:385–405. [PubMed: 14041404]
- Cox RM, McDaniel DM. Reference equivalent threshold levels for pure tones and 1/3-octave noise bands: Insert earphone and TDH-49 earphone. *Journal of the Acoustical Society of America*. 1986; 79:443–446. [PubMed: 3950196]
- Dorman MF, Loizou PC. Relative spectral change and formant transitions as cues to labial and alveolar place of articulation. *Journal of the Acoustical Society of America*. 1996; 100:3825–3830. [PubMed: 8969483]
- Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198. [PubMed: 1202204]
- Francis AL, Baldwin K, Nusbaum HC. Effects of training on attention to acoustic cues. *Perception and Psychophysics*. 2000; 62:1668–1680. [PubMed: 11140187]
- Gelfand SA, Porrazzo J, Silman S. Age effects on gap detection thresholds among normal and hearing impaired subjects. *Journal of the Acoustical Society of America*. 1988; 83:75.
- Gianopoulos I, Stephens D. General considerations about screening and their relevance to adult hearing screening. *Audiological Medicine*. 2005; 3:165–174.
- Gordon-Salant S, Fitzgibbons PJ. Selected cognitive factors and speech recognition performance among young and elderly listeners. *Journal of Speech, Language and Hearing Research*. 1997; 40:423–431.
- Grose JH, Hall III JW, Buss E. Temporal processing deficits in the pre-senescent auditory system. *Journal of the Acoustical Society of America*. 2006; 119:2305–2315. [PubMed: 16642844]
- Gygi B, Kidd GR, Watson CS. Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America*. 2004; 115:1252–1265. [PubMed: 15058346]
- Hedrick MS, Schulte L, Jesteadt W. Effect of relative and overall amplitude on perception of voiceless stop consonants by listeners with normal and impaired hearing. *Journal of the Acoustical Society of America*. 1995; 98:1292–1303. [PubMed: 7560503]
- Hedrick M, Younger MS. Perceptual weighting of relative amplitude and formant transition cues in aided CV syllables. *Journal of Speech, Language and Hearing Research*. 2001; 44:964–974.
- Hedrick M, Younger MS. Perceptual weighting of stop consonant cues by normal and impaired listeners in reverberation versus noise. *Journal of Speech, Language and Hearing Research*. 2007; 50:254–269.
- Holt LL, Lotto AJ. Cue weight in auditory categorization: Implication for first and second language acquisition. *Journal of the Acoustical Society of America*. 2006; 119:3059–3071. [PubMed: 16708961]
- Humes LE. Prescribing gain characteristics of linear hearing aids. In: Studebaker, GA.; Bess, FH.; Beck, L., editors. *The Second Vanderbilt/VA Hearing Aid Report*. Parkton, Maryland: York Press; 1991.
- Klatt DH. Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*. 1980; 67:971–995.

- Klatt DH, Klatt LC. Analysis synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*. 1990; 87:820–857. [PubMed: 2137837]
- Kruskal, JB.; Wish, M. *Multidimensional Scaling*. Newbury Park: Sage; 1989.
- Lieberman AM, Delattre PC, Cooper FS. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*. 1958; 1:153–167.
- Lieberman AM, Delattre PC, Cooper FS, Gerstman LJ. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*. 1954; 68(8) (Whole No. 379).
- Lindholm JM, Dorman M, Taylor BE, Hannley MT. Stimulus factors influencing the identification of voiced stop consonants by normal-hearing and hearing-impaired adults. *Journal of the Acoustical Society of America*. 1988; 83:1608–1614. [PubMed: 3372871]
- Lisker L, Abramson AS. Some effects of context on voice onset time in English stops. *Language and Speech*. 1967; 10:1–28. [PubMed: 6044530]
- Lisker L. Voicing in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*. 1986; 29:3–11. [PubMed: 3657346]
- Moscicki E, Elkins E, Baum H, McNamara P. Hearing loss in the elderly: An epidemiologic study of the Framingham heart study cohort. *Ear and Hearing*. 1985; 6:184–190. [PubMed: 4043571]
- Nearey TM. Static, dynamic and relational properties in vowel perception. *Journal of the Acoustical Society of America*. 1989; 85:2088–2113. [PubMed: 2659638]
- Nosofsky RM. Attention, similarity, and the identification categorization relationship. *Journal of Experimental Psychology*. 1986; 115:39–57. [PubMed: 2937873]
- Ohde RN, Stevens KN. Effect of burst amplitude on the perception of stop consonant place of articulation. *Journal of the Acoustical Society of America*. 1983; 74:706–714. [PubMed: 6630726]
- Raphael, LJ. Acoustic cues to the perception of segmental phonemes. In: Pisoni, DB.; Remez, RE., editors. *The Handbook of Speech Perception*. Oxford: Blackwell; 2005. p. 182–206.
- Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. *Science*. 1981; 212:947–950. [PubMed: 7233191]
- Repp BH. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychology. Bull.* 1982; 92:81–110.
- Smith EE, Patalano AL, Jonides J. Alternative strategies of categorization. *Cognition*. 1998; 65:167–196. [PubMed: 9557382]
- Sommers, MS. Age-related changes in Spoken Word Recognition, *The Handbook of Speech Perception*. Blackwell Publishing Ltd; 2005. p. 469–493.
- Spoor A. Presbycusis values in relation to noise-induced hearing loss. *International Audiology*. 1967; 6:48–57.
- Stevens KN, Blumstein SE. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*. 1978; 64:1358–1368. [PubMed: 744836]
- Strouse A, Ashmead DH, Ohde RN, Grantham DW. Temporal processing in the aging auditory system. *Journal of the Acoustical Society of America*. 1998; 104:2385–2399. [PubMed: 10491702]
- Studebaker GA. A rationalized arcsine transform. *Journal of Speech, Language and Hearing Research*. 1985; 28:455–462.
- Whalen DH. Perception of the English /s/-/ʃ/ distinction relies on rictive noises and transitions, not on brief spectral slices. *Journal of the Acoustical Society of America*. 1991; 90:1776–1785. [PubMed: 1960273]
- Wiley TL, Cruickshanks KJ, Nondahl DM, Tweed TS, Klein R. Aging and high-frequency hearing sensitivity. *Journal of Speech, Language and Hearing Research*. 1998; 41:1061–72.
- Wright, R. A review of perceptual cues and cue robustness, in *Phonetically Based Phonology*. Hayes, B.; Kirchner, R.; Steriade, D., editors. Cambridge: Cambridge University Press; 2004. p. 34–57.

FIGURE 1.

Pure tone air-conduction hearing threshold for 12 elderly hearing-impaired (EHI) subjects (right ear). The bolded line and circles represent the mean hearing threshold.

FIGURE 2.

Schematics of three exemplar stimuli, which take the small, median and large values in each of the three cue dimensions, from top to bottom panel. The top graph represents the stimulus containing one 100-ms fricative noise burst (represented by the 1st shaded box), which is separated by a 10-ms temporal gap with another 200ms fricative noise burst (represented by the 2nd shaded box). Both bursts center at 2000 Hz (represented by the bold line) and the 2nd burst has an endpoint of the frequency transition of 1500 Hz or a -500 Hz frequency transition (represented by the bold line). The middle graph represents a stimulus with a center frequency at 2250 Hz, a 70-ms temporal gap and an endpoint of the frequency transition of 2250 Hz or a 0 Hz frequency transition. The bottom graph represents a stimulus

with a center frequency at 2500 Hz, a 130-ms temporal gap and an endpoint of the frequency transition of 3000 Hz or a 500 Hz frequency transition. The gray boxes at the bottom of each panel represent the band-pass filtered noises with the temporal gap which line up with those of the fricative noise.

FIGURE 3.

Comparison between shaped stimuli and thresholds for both EHI and YNH groups. The x axis presents the 1/3 OCT frequency from 125 Hz to 8000 Hz. The y axis presents the corresponding amplitude in dB SPL in each 1/3 OCT frequency. The top curve, “CalMin” (filled circles), is the calibration noise matching the trough amplitude across all 27 stimuli. The other three curves from top to bottom are “Maximum EHI threshold” (unfilled circles), the worst threshold across all EHI participants; the “Average EHI threshold” (filled triangles) and the “Average YNH threshold” (unfilled triangles), respectively.

FIGURE 4.

Results of the discrimination task for all YNH and EHI participants. The average percent-correct scores for each dimension are plotted for both YNH and EHI groups. The black, light gray and darker gray bars represent the fn (center frequency), gp (temporal gap) and tr (frequency transition) cue dimension, respectively. Error bars represent standard deviations of each group across all three cue dimensions. The dashed line represents 90% correct. The asterisks indicate significant differences between the two cue dimensions in terms of the discrimination scores.

FIGURE 5. Two-dimensional plots of all 27 stimuli derived by the MDS ALSCAL solution for both YNH and EHI participants. Each plot represents a perceptual space by two dimensions with the relative distance of all 27 stimuli graphed. Each plot is partitioned by the stimulus value on each dimension. All stimuli are labeled as S1 to S27 and their corresponding value in each of three dimensions can be seen from the legend on the right side of the graph.

FIGURE 6. Individual MDS weights derived from MDS INDSCAL solution are plotted for all three cue dimension for each YNH (top panel) and EHI (bottom panel) subject.

FIGURE 7.

GCM weights for each dimension are plotted as stacked bars for each YNH (top panel) and EHI (bottom panel) participant following the classification test. The weights allocated in all the three cue dimensions for each participant sum to 1.

FIGURE 8.

Result of the classification test for two EHI participants--Subject 2 (top panel), a typical frequency transition user, and Subject 12 (middle and bottom panels), a typical temporal gap user. The x-axes in the top and middle panels represent the three possible values that each stimulus can take on the frequency transition dimension, -500, 0 and 500 Hz, from left to right. The x-axis in the bottom panel represents the three possible values that each stimulus can take on the temporal gap dimension, 10, 70 and 130 ms. The y-axis represents the percentage that each stimulus was classified as Circle, Triangle and Square for 80 presentations. The corresponding GCM weight on three dimensions for each participant is plotted beside each graph.

FIGURE 9.

GCM weights for each dimension are plotted as stacked bars for each YNH (top panel) and EHI (bottom panel) participant following the temporal-gap training test. The weights allocated in all the three cue dimensions for each participant sum to 1.

FIGURE 10.

Histogram showing distribution of the number of training blocks needed by each YNH and EHI participants to reach the 80%-correct criterion in the exemplar-classification task. The maximum number of training blocks administered in Experiment 4 was 20. One YNH and three EHI who could not reach the 80% criterion after 20 blocks of training, indicated as “> 20” on the graph.

FIGURE 11.

Percent-correct scores for the fixed-discrimination task were plotted across all 12 stimulus pairs for the three excluded EHI participants (subject a, b, c) after the first assessment. The big bold triangles linked by the bold line represent the average performance of all 12 enrolled EHI participants after their first assessment in fixed-discrimination task.

FIGURE 12.

Percent-correct scores across 12 stimulus pairs in the fixed-discrimination task are plotted for both YNH and EHI groups after their initial assessment. Error bars represent the standard deviations of each group across all 12 stimulus pairs. The asterisk indicates significant differences between the YNH and EHI groups for stimulus pair “tr12”.

Table 1

Klattsyn parameter values for final speech-like stimuli

APPENDIX: KLATT SYNTHESIZER PARAMETER VALUES FOR FINAL SPEECH-LIKE STIMULI			
	Parameter	Description	Value
Configuration parameters			
1	sr	sample rate	11,025 samples/sec
2	du	duration of utterance	310-ms–10-ms gap 370-ms–70-ms gap 430-ms–130-ms gap
3	nf	number of cascade formants	5
4	ui	number of formants	5
5	ss	source switch	1
6	rs	random seed	9
7	os	output waveform selector	0
Variable parameters			
8	g0	overall gain control	60
9	f0	fundamental frequency	1000
10	at	amplitude of turbulence	0
11	oq	open quotient	50
12	tl	spectral tilt	0
13	sk	glottal skew	0
14	dF	delta F1 (at open glottis)	0
15	db	delta B1 (at open glottis)	0
16	av	amplitude of voicing	60
17	ah	amplitude of aspiration	0
18	F1	frequency of 1st formant	500
19	F2	frequency of 2nd formant	1500
20	F3	frequency of 3rd formant	1200 Hz-2000 Center 1450 Hz-2250 Center 1700 Hz-2500 Center
21	F4	frequency of 4th formant	2000Hz 2250 Hz 2500 Hz
22	F5	frequency of 5th formant	2800 Hz-2000 Center 3050 Hz-2250 Center 3300 Hz-2500 Center
23	f6	frequency of 6th formant	2000Hz
24	fz	nasal zero frequency	280
25	fp	nasal pole frequency	280
26	b1	bandwidth of the 1st formant	60
27	b2	bandwidth of the 2nd formant	90

APPENDIX: KLATT SYNTHESIZER PARAMETER VALUES FOR FINAL SPEECH-LIKE STIMULI

	Parameter	Description	Value
28	b3	bandwidth of the 3rd formant	450Hz
29	b4	bandwidth of the 4th formant	600Hz
30	b5	bandwidth of the 5th formant	450Hz
31	b6	bandwidth of the 6th formant	100Hz
32	bz	nasal zero bandwidth	90
33	bp	nasal pole bandwidth	90
34	ap	amp of parallel voicing	0
35	af	amplitude of friction	60
36	ab	amplitude of friction bypass	0
37	a1	first formant amplitude	0
38	a2	second formant amplitude	0
39	a3	third formant amplitude	50
40	a4	fourth formant amplitude	50
41	a5	fifth formant amplitude	50
42	a6	sixth formant amplitude	60
43	an	nasal formant amplitude	0
44	P1	first formant bandwidth	80
45	P2	second formant bandwidth	250
46	P3	third formant bandwidth	450
47	P4	fourth formant bandwidth	600
48	P5	fifth formant bandwidth	450

Table 2

Results of the fixed-discrimination task for all young normal-hearing (YNH) and elderly hearing-impaired (EHI) subjects across 12 discriminations. Numbers in the each cell indicate average percent correct and standard deviation (in parenthesis).

	fn12	fn21	fn23	fn32	gp12	gp21	gp23	gp32	tr12	tr21	tr23
YNH	95.20 (3.31)	93.31 (6.11)	93.43 (6.61)	93.43 (8.29)	96.72 (4.23)	94.19 (4.96)	93.94 (8.08)	95.58 (2.90)	98.99 (1.26)	98.36 (3.16)	97.60 (3.98)
EHI	93.20 (4.93)	90.16 (5.35)	91.53 (3.96)	92.95 (2.42)	97.14 (3.18)	94.78 (4.13)	91.48 (4.95)	90.25 (5.65)	96.18 (3.03)	96.76 (3.26)	95.62 (3.50)