

The role of temporal cues in word identification by younger and older adults: Effects of sentence context

Sandra Gordon-Salant and Grace Yeni-Komshian

Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

Peter Fitzgibbons

Department of Hearing, Speech, and Language Sciences, Gallaudet University, Washington, DC 20002

(Received 6 August 2007; revised 8 August 2008; accepted 17 August 2008)

Prior investigations, using isolated words as stimuli, have shown that older listeners tend to require longer temporal cues than younger listeners to switch their percept from one word to its phonetically contrasting counterpart. The extent to which this age effect occurs in sentence contexts is investigated in the present study. The hypothesis was that perception of temporal cues differs for words presented in isolation and a sentence context and that this effect may vary between younger and older listeners. Younger and older listeners with normal-hearing and older listeners with hearing loss identified phonetically contrasting word pairs in natural speech continua that varied by a single temporal cue: voice-onset time, vowel duration, transition duration, and silent interval duration. The words were presented in isolation and in sentences. A context effect was shown for most continua, in which listeners required longer temporal cues in sentences than in isolated words. Additionally, older listeners required longer cues at the crossover points than younger listeners for most but not all continua. In general, the findings support the conclusion that older listeners tend to require longer target temporal cues than younger normal-hearing listeners in identifying phonetically contrasting word pairs in isolation and sentence contexts.

© 2008 Acoustical Society of America. [DOI: 10.1121/1.2982409]

PACS number(s): 43.71.Lz, 43.71.Es, 43.66.Mk, 43.66.Sr [RLF]

Pages: 3249–3260

I. INTRODUCTION

The overall objective of this study was to determine the effect of stimulus context on perception of target temporal cues in speech by younger and older listeners. Although investigations of perception of temporal cues in speech continua have often demonstrated performance differences in younger and older listeners [e.g., see [Strouse et al. \(1998\)](#) and [Gordon-Salant et al. \(2006\)](#)], these studies have primarily presented isolated words containing the targeted temporal cues. It may be that perception of target cues in speech stimuli presented in isolation may not predict perception of these same cues when embedded in a sentence context.

The assessment of perception of temporal cues in speech by younger and older listeners emerges from a larger body of evidence showing that older human listeners exhibit slowed processing of acoustic signals. This conclusion derives from experiments employing nonspeech signals, as well as those employing speech sounds. For example, many investigations have demonstrated an age effect for detection and discrimination of simple nonspeech sounds presented in isolation. Older listeners show longer duration difference limens (DLs) for discrimination of reference stimuli defined as pure tones or silent intervals ([Fitzgibbons and Gordon-Salant, 1994](#); [Bergeson et al., 2001](#); [Grose et al., 2006](#)) and longer gap detection thresholds for gaps inserted between tonal markers or noise bands ([He et al., 1999](#); [Snell and Hu, 1999](#); [Lister et al., 2002](#)).

Age-related slowing of perceptual processing is evident also in experiments measuring identification of isolated words, which varied on a single temporal cue along a con-

tinuum. In one study, older listeners showed more gradual slopes than younger listeners in identification functions derived from judgments of synthetic speech stimuli varying along a voice-onset time (VOT) continuum ([Strouse et al., 1998](#)). In another study, [Gordon-Salant et al. \(2006\)](#) used natural speech stimuli and found that older listeners required a longer acoustic cue to shift their perception from one phoneme category to another along a stimulus continuum. For example, in a continuum that varied the formant transition duration to cue a stop-glide distinction, younger listeners required a transition duration of about 29 ms to shift their perception from the stop to the glide, whereas older listeners required a transition duration of about 37 ms to shift their perception from the stop to the glide. Thus, for this one continuum, older listeners required the target cue to be approximately 30% longer than the younger listeners to differentiate the contrasting phoneme with the longer cue from the contrasting phoneme with the shorter cue. The age effect in the study of [Gordon-Salant et al. \(2006\)](#) was observed for the silent interval cue for a sibilant-affricate distinction (as in *dish* versus *ditch*), in addition to the transition duration cue for a stop-glide distinction described above (as in *beat* versus *wheat*). However, in this previous study, the age effect was not observed consistently across different speech continua, such as those varying VOT (*buy* versus *pie*) and vowel duration (*wheat* versus *weed*).

The temporal cues discussed above are important for distinguishing certain speech sounds in English as well as in other languages ([Lisker and Abramson, 1964](#); [Chen, 1970](#)). The critical values of these temporal cues are, however, lan-

guage specific. As a result, non-native speakers of English often produce speech with temporal characteristics that correspond more to their own language than to English. In other words, the accented speech of non-native speakers contains alterations of numerous acoustic characteristics of English, particularly those that involve the temporal aspects of speech, such as VOT (Flege and Eefting, 1988; MacKay *et al.*, 2000) and vowel duration (Chen, 1970; Fox *et al.*, 1995).

These distinctions between native and accented speech, coupled with the age-related differences in perception of temporal acoustic cues, have potential implications for identifying the source of older listeners' difficulty in understanding accented English. At least one report suggests that older listeners have considerable difficulty in recognizing words and sentences spoken by non-native speakers of English (Burda *et al.*, 2003). One possible reason for older listeners' problems in understanding accented English is difficulty distinguishing key words that are cued by temporal contrasts. The prior investigation of perception of temporal cues in isolated words (Gordon-Salant *et al.*, 2006) revealed age-related changes in perception of two cues (silent interval duration and transition duration), but these are not the temporal cues that are cited as being altered in accented English (i.e., VOT and vowel duration). However, it is possible that perception of VOT and vowel duration in a sentence context may be affected differentially by age. The effect of sentence context on perception of temporally mediated target cues is currently unknown.

In psychoacoustic studies, placement of a nonspeech target stimulus in a sequential context is known to alter perception of the target, particularly by older listeners. Fitzgibbons and Gordon-Salant (1995) measured duration DLs for tones and silent intervals, each with a reference duration of 250 ms, presented in isolation and embedded in a sequence of five 250 ms components. The 250 ms standard component duration was chosen to mimic the average duration of a word. Differential effects of sequential context for younger and older listeners varied with the two types of stimuli (tones and silent intervals). Whereas the young listeners showed a constant DL for tones in isolation and in the sequence, older listeners showed significantly larger DLs for the tones in the five-component sequence compared to the DLs for the same tones measured in isolation. However, for reference silent intervals, both younger and older listeners showed larger DLs in the sequence of five components, but the sequential context effects were still larger for the older listeners. The source of the pronounced age-related deficit for sequential stimuli is unknown but may be associated with age-related decline in global cognitive abilities, coupled with a slowing in component processes (Salthouse, 1996). Deficits in working memory, divided attention, and search capacity, as well as limited processing resources for multiple mental, perceptual tasks have all been reported in older people [e.g., see Baddely (1986), Fisk and Rogers (1991), Tun (1998), and Humes *et al.* (2006)]. It is possible that any one of these attributes of cognitive decline, or a combination of these factors, is implicated in an older listener's reduced ability to make judgments about temporally mediated stimuli in a se-

quential context. Nevertheless, older listeners are able to make excellent use of contextual meaningful information for understanding speech materials (Pichora-Fuller *et al.*, 1995; Gordon-Salant and Fitzgibbons, 1997), particularly in difficult listening situations. If the nature of the contextual information is semantically limited, however, it may be predicted that age-related differences in perception of temporal cues in speech will be magnified in a sentence context.

The purpose of this study was to investigate the hypothesis that listener perception of target temporal cues to identify words depends on the stimulus context (isolated words versus semantically limited sentences). Placing a cue in a sentence context may influence listener perception of that cue because of the mediating effects of phonemes that precede it, the increased memory load, and/or the increased attention required to locate the cue in a surround. A secondary hypothesis was that older listeners require longer acoustic cues than younger listeners to shift their percept from a phoneme with a shorter reference cue to its contrasting phoneme and that these age-related performance differences may interact with stimulus context. The age-related effects could be manifested primarily in the form of longer temporal cues required by older listeners at crossover points and secondarily in more gradual slopes, consistent with previous studies [e.g., see Strouse *et al.* (1998)]. The interaction effect, if observed, could be seen as a larger difference in performance between younger and older listeners in the sentence context because of increased cognitive demand that may have a greater impact on the older listeners.

A final issue in any examination of age-related performance differences on listening tasks concerns differentiating the effects of age from those attributed to hearing loss, particularly because hearing loss has a high prevalence rate among older people [e.g., see Cruickshanks *et al.* (1998)]. In the present study, the effects of age were investigated through a comparison of performance by older listeners with normal hearing to performance by young noise-masked listeners. The detection thresholds of older listeners with normal hearing are often poorer than those of younger listeners with normal hearing. To accommodate this potential confounding effect, a low-level masking noise was presented to a group of young normal-hearing listeners to shift their thresholds to approximate those of the older normal-hearing listeners. An additional, separate group of young normal-hearing listeners without noise masking served as a control group to validate the continua, as well as to determine the extent to which the noise masking affected performance. Finally, the performance of older hearing-impaired listeners was tested as well to assess the effects of hearing impairment on these tasks. The specific effects of hearing impairment are important to evaluate because previous investigations have shown that hearing loss affects the identification of speech stimuli varying in a single temporal acoustic dimension [e.g., see Gordon-Salant *et al.* (2006)].

II. METHOD

A. Participants

Four groups of listeners ($n=15$ /group) participated in these experiments. The first three groups included a control

TABLE I. Pure tone thresholds in dB HL (re: ANSI 2004) in the test ear of the four listener groups ($n=15/\text{group}$). Data shown are group means, with standard deviations in parentheses

Group	Frequency (Hz)				
	250	500	1000	2000	4000
Young normal	7.6(4.9)	7.0(4.9)	5.0(4.2)	3.3(8.2)	1.0(6.6)
Young noise masked	11.0(4.3)	13.3(3.1)	15.3(4.0)	18.7(4.0)	20.3(4.0)
Older normal	15.3(5.7)	15.3(5.3)	14.6(6.3)	17.8(8.0)	18.9(6.2)
Older hearing impaired	25.0(9.6)	24.7(10.3)	27.7(10.7)	40.7(10.0)	51.7(7.7)

group of young listeners (Yng Norm; ages 18–29 years, mean age=22.26) with normal hearing (pure tone air conduction thresholds <20 dB HL re: ANSI (2004), 250–4000 Hz), a group of older listeners (Older Norm; ages 65–79 years, mean age=70.93) with normal hearing, and a group of older listeners (Older Hrg Imp; 66–80 years, mean age=72.86) with mild-to-moderately severe sloping sensorineural hearing loss. Because pure tone detection thresholds are often different between younger and older normal-hearing listeners, spectrally shaped noise masking was presented to a separate fourth group consisting of young normal-hearing listeners (Yng Noise Mask; $n=15$, ages 18–25 years, mean age=21.5) to equate their masked thresholds to the unmasked thresholds of the older normal-hearing group. This masking noise was added to the stimuli presented to this group during all testing conditions. Comparisons of performance between the young noise-masked and older normal-hearing groups were used to assist in the interpretation of the effects of age (Hargus and Gordon-Salant, 1995; Gordon-Salant *et al.*, 2006). The low-level masking noise was created using COOL EDIT PRO (Version 2, Syntrium Software) to shift the thresholds of these young listeners to be equivalent (within 5 dB) to the average thresholds at each frequency (250–4000 Hz) of the older group with normal hearing. None of the listeners in the young noise-masked group served in the young normal-hearing control group (unmasked). Mean thresholds of the four listener groups are shown in Table I.

All listeners were required to have good to excellent recognition scores for monosyllabic words (Northwestern University Test No. 6; Tillman and Carhart, 1966) presented under earphones, tympanograms demonstrating normal peak admittance, peak pressure, and equivalent volume (Roup *et al.*, 1998), acoustic reflexes elicited at levels that were within the 90th percentile for individuals with normal hearing or a cochlear hearing loss (Gelfand *et al.*, 1990), and transient-evoked otoacoustic emissions that were present in the listeners with normal hearing and absent in the listeners with hearing impairment. These criteria were intended to ensure that listeners with normal hearing did not have significant peripheral auditory dysfunction and that listeners with hearing loss had a primarily cochlear site of lesion. In addition, American English was the native language of all participants. Finally, all listeners were screened for general cognitive awareness (Short Portable Mental Status Questionnaire, Pfeiffer, 1977) and were required to score at least 9 out of 10 on this test.

B. Stimuli

The stimuli were phonetically contrasting word pairs that varied in a single temporal cue. These words were presented either in isolation or in a sentence context. Eight different speech continua were developed for this purpose. The stimuli on each continuum varied in a single temporal cue in seven equal-interval steps. Three of the continua, consisting of nouns (*buy/pie*, *dish/ditch*, and *beat/wheat*), were used in a previous experiment (Gordon-Salant *et al.*, 2006), and five additional continua, all consisting of names (*Greg/Creig*, *Sue/Stu*, *Bill/Will*, *Alisa/Aliza*, and *Mr. Rite/Mr. Ride*), were developed specifically for this experiment. The name continua were implemented in order to permit the placement of the stimulus word at the beginning or end locations in a sentence. In most instances, each temporal cue was examined twice in order to determine whether or not patterns of performance were consistent across different continua varying on the same acoustic cue.

A male talker (age 19 years) with a general American dialect recorded all of the speech stimuli with the names and nouns produced in isolation and in the context of a sentence. The names and nouns spoken in isolation were used to create the tokens for the speech continua, as described below. These tokens were then presented either in isolation or inserted into carrier sentences (replacing the recorded unprocessed word at the beginning or end of a sentence). The carrier sentences were taken from or modeled after the low probability Revised Speech Perception in Noise (R-SPIN) lists (Bilger *et al.*, 1984), which do not provide semantic information to cue the test word. Stimuli were recorded directly onto a laboratory PC using a microphone, a preamplifier, and a 16 bit sound card (Audigy Sound Blaster 2ZS). The name stimuli were recorded in the context of multiple sentences, with the names occurring at or near the beginning of a sentence, as well as at the end of a sentence. The noun stimuli were recorded at the end of different carrier sentences and were usually preceded by the word “the.” Altogether, there were six different carrier sentences for the name stimuli and 70 different carrier sentences for the noun stimuli. The numbers of carrier sentences for noun and name stimuli were unequal, reflecting the available corpus of appropriate sentences from the R-SPIN lists. Given the nature of the two-alternative forced choice identification task, the lack of semantic information in the carrier sentences, and the repetition of all of the carrier sentences multiple times throughout the experiment,

it was assumed that the different number of carrier sentences for a given type of stimulus had only minimal influence on results.

The digital recordings were edited using **COOL EDIT PRO** and **WEDW** (2008). First, the stimuli in each contrasting word pair recorded in isolation were extracted from the recordings. These isolated stimuli, rather than the stimuli recorded in the sentence contexts, were chosen for processing for two reasons: (1) to facilitate comparison with a previous experiment (Gordon-Salant *et al.*, 2006) that also used stimuli recorded in isolation and (2) to permit a direct comparison of perception of the exact same words presented in isolation and in sentence contexts. The words in each pair were analyzed acoustically and modified to create a continuum of seven stimuli in which the single temporal cue of interest was varied in equal steps. It should be noted that in natural speech, more than one acoustic cue may vary between two phonetically contrasting words. Nevertheless, the primary acoustic cue that contributes to the perception of these stimuli is the target temporal cue modified in each of the continua. Four acoustic cues were sampled in a total of eight separate continua: VOT to signal stop-consonant voicing, silent interval duration to distinguish either fricative versus affricate (/ʃ/ versus /tʃ/) or the fricative /s/ from the fricative-stop cluster /st/, formant transition duration as a cue for stop versus semi-vowel, and vowel duration as a cue for voicing in postvocalic consonants.

A complete description of the creation of the noun continua (*buy/pie*, *dish/ditch*, and *beat/wheat*) can be found in a previous publication (Gordon-Salant *et al.*, 2006). They will be described only briefly below. Stimulus development for the proper name stimuli was comparable to that used for the noun stimuli. Detailed descriptions of the creation of these name stimuli (*Greg/Creig*, *Sue/Stu*, *Bill/Will*, *Alisa/Aliza*, and *Rite/Ride*) are also presented.

Voice-onset time. VOT was altered in the *buy/pie* continuum by varying the duration of the aspiration in the natural token, *pie*, from 0 ms (*buy*) to 60 ms (*pie*) in 10 ms steps. The *Greg/Creig* continuum also manipulated VOT. The endpoint stimulus, *Creig*, was modified from the original recording by first removing its burst and 100 ms of aspiration, and then inserting a burst from a natural *Greg*. Stimuli on the seven-point continuum were created by adjusting the amount of aspiration in 10 ms steps, from 20 ms (*Greg*) to 80 ms (*Creig*).

Silent interval duration. The target cue in the *dish/ditch* continuum was the silent interval duration between the initial consonant-vowel and the final consonant, /ʃ/, from 0 ms (*dish*) to 60 ms (*ditch*) in 10 ms steps. The target cue for the *Sue/Stu* continuum was the silent interval duration between /s/ and the final vowel. The endpoint stimulus, *Sue*, was a hybrid; it was created by modifying the original *Stu* stimulus as follows. First, the /s/ from *Stu* was removed and replaced with the natural /s/ from *Sue*. Next, the burst from *Stu* was attenuated by 12 dB, and the 92 ms silent period in *Stu* was removed. This produced the endpoint stimulus perceived as *Sue*, with a 0 ms silent interval. The remaining stimuli on the continuum were created by inserting silent segments, in 5 ms steps, following the fricative /s/ in the endpoint *Sue* stimulus.

The endpoint stimulus, *Stu*, had a silent interval of 30 ms.

Transition duration. For the *beat/wheat* continuum, the transition duration of the initial consonant was varied from 7 ms (*beat*) to 51 ms (*wheat*) in 7–8 ms steps (i.e., a single glottal pulse). The *Bill/Will* continuum was created by reducing the natural transition duration of the original stimulus, *Will*, down to 60 ms from 160 ms. This endpoint stimulus was perceived as *Will*. Single pulses of the glide (approximately 10 ms/pulse) were removed to create the remaining stimuli on the continuum, with the endpoint, *Bill*, having a transition duration of 0 ms. The endpoint stimuli *Bill* and *Beat* did not contain a burst, although they were perceived as the intended phoneme (/b/).

Vowel duration. The target cue in the *Alisa/Aliza* continuum was vowel duration as a cue for voicing contrasts in post-vocalic fricatives (Denes, 1955). The endpoint stimulus, *Aliza*, was the natural stimulus whose vowel duration was 356 ms. The duration of the vowel, /i/, was cut from this endpoint stimulus by two pulses (approximately 19–20 ms) for each successive stimulus, such that the vowel of the endpoint stimulus, *Alisa*, was 242 ms. The *Rite/Ride* continuum varied vowel duration as a cue to final consonant voicing. (Note: these stimuli were preceded by *Mr.* or *Ms.*, as in *Mr. Ride*, to make these stimuli proper nouns.) The endpoint stimulus, *Ride*, was a hybrid of the original recording of *Ride*, with the final consonant release of *Ride* removed and replaced with the release of *Rite*. The duration of the consonant, /r/, and the closure duration for the final consonant were held constant, but the steady-state vowel portion was modified to produce the different stimuli in this continuum. The steady-state vowel portion in the endpoint stimulus, *Ride*, was 294 ms. Additional stimuli in the continuum were created by removing two pulses per step (approximately 19–20 ms) from the steady-state portion of the vowel, such that the vowel duration of the endpoint stimulus, *Rite*, was 174 ms.

The isolated word stimuli from each continuum were equated in rms level. As noted earlier, carrier sentences were recorded into which the processed stimuli from a continuum were inserted. The noun stimuli were inserted at the end of a sentence. There were 70 carrier sentences used for the noun stimuli in each continuum (seven steps/continuum and ten repetitions of each token). The name stimuli were inserted at the beginning of a carrier sentence (“*Sue* heard Tom called about the coach.”) and at the end of another carrier sentence (“Ruth could not have known about *Sue*.”). This pattern was used for the rest of the contrasting name stimuli. The natural pauses in the original sentences were preserved when creating these sentences. The sentence stimuli across all continua were analyzed and equated in rms level. Two separate calibration tones were created to be equivalent to the rms levels of the isolated word stimuli and the sentence stimuli, respectively. The isolated word and sentence stimuli, as well as the calibration tones, were stored in separate waveform files on the PC.

C. Procedure

Each of the eight speech continua were presented both as isolated words (eight conditions corresponding to the

three noun and five name continua) and as words in a sentence (three conditions corresponding to the noun stimuli placed at the end of sentences and ten conditions corresponding to the five name stimuli, which were placed at the beginning and at the end of sentences) for a total of 21 (8+3+10) conditions. Half of the participants were tested on the isolated words first, and half of the participants were tested on the sentences first. The order of presentation of the specific speech continua was randomized separately for the isolated words and the sentences, and this order also varied across participants.

During testing, the listener was seated in a double-walled sound attenuating chamber. ECOS2 software (Avaaz Innovations) controlled stimulus presentation and timing, presented visual displays to the listener, and tracked the listener's responses. For each condition, the seven stimuli on the target continuum were presented ten times in randomized order, with each randomization occurring within a block of the seven stimuli (i.e., ten randomized blocks were presented). Speech stimuli were played out of a sound card (Audigy Sound Blaster 2ZS), amplified (Crown D-75A), mixed (Coulbourn S-24 audiomixer amplifier), and delivered to an insert earphone (Etymotic ER-3A). The signal level was 85 dB SPL. For the young noise-masked listeners, a CD of the masking noise was played out of a CD player (Tascam model CD-RW402), amplified (Crown D-75A), attenuated (Hewlett-Packard 350D), and mixed with the target speech signal (Coulbourn audiomixer amplifier). The level of the spectrally shaped noise was adjusted to produce 50 dB SPL in the insert earphone, as measured in a 2 cm³ coupler. The test ear was the ear with better hearing sensitivity for the hearing-impaired listeners and was randomized for the normal-hearing listeners.

Following the stimulus presentation, a computer monitor presented a text display of the two possible response choices, corresponding to the two endpoint stimuli on the continuum (e.g., *buy* and *pie*). The listener's task was to select the stimulus perceived using a computer mouse click. Listener responses were self-paced. Prior to commencing each condition, instructions were given orally and in a visual display on the computer monitor. Before the experimental conditions began, listeners were given practice using another speech continuum that was created exclusively for practice purposes.

The entire experiment, including preliminary audiometric testing and experimental testing, was concluded in approximately 6 h, scheduled in several 2 h visits. Listeners were provided frequent breaks to minimize fatigue and were reimbursed for their participation in the experiment. The protocol for this experiment was approved by the University of Maryland Institutional Review Board for Human Subjects Research.

D. Analysis

Two values were calculated from each participant's identification function for each continuum: crossover point and slope. In addition, the endpoint accuracy was calculated for the stimuli presented in isolation, as an indicator of how

well the listeners identified the exemplar stimuli (stimuli 1 and 7 on each continuum); lower accuracy indicates that the listeners had a relatively unstable perception of the target stimulus (Munson and Nelson, 2005).

The slope and 50% crossover point of each individual's identification function for each stimulus continuum were estimated using a fitting procedure described by Wichman and Hill (2001a, 2001b) and the accompanying software (PSIGNIFIT, available from <http://bootstrap-software.org/psignifit/>). The logistic option was used to characterize the shape of the underlying psychometric distribution of the measured identification functions. The estimated crossover point for each function represents the stimulus value at which a listener changes his or her perception from one stimulus to the contrasting stimulus on the continuum. The slope of the function indicates how abrupt and consistent the shift is in perception from one phoneme category to another. Steep slopes may suggest that the stimuli on the continuum were perceived with certainty, whereas a gradual slope reflects the uncertainty in perception of the stimuli encompassed by the slope.

Some participants were random in their responses. Data from these participants were removed from the analysis for a particular condition if the identification scores of both endpoint stimuli did not exceed 50%, suggesting that the listener did not perceive the two endpoint stimuli as distinct. In these cases, a crossover point and slope are not meaningful. A subset of these participants perceived all of the stimuli along the continuum as one of the endpoint stimuli (e.g., flat slope). The PSIGNIFIT program could not derive crossover points and slopes for these flat responses. In all, there were several participants who showed random performance for a particular condition, although the specific participants who exhibited random identification performance on one condition were not necessarily those who showed random performance on other conditions. Participants with random endpoint accuracy on a particular condition came from all listener groups. The proportion of data dropped from each listener group, across all 21 conditions, was 5.7% for young normal controls, 6.0% for young noise-masked listeners, 9.2% for older normal-hearing listeners, and 10.8% for older hearing-impaired listeners. The relatively small proportion of responses that were dropped from the total data set of each group suggests that these stimuli were generally perceived as the intended words and hence are reasonably natural and valid. Goodness of fit for each individual function was determined by calculating the correlation coefficients between the PSIGNIFIT predicted data points and the observed data points. The means of the correlation coefficients for the functions associated with the retained data for isolated word and sentence stimuli were 0.98 and 0.97 ($R^2=0.96$ and 0.94), respectively, while those associated with the rejected data for isolated word and sentence stimuli were 0.42 and 0.56 ($R^2=0.18$ and 0.31), respectively.

III. RESULTS AND DISCUSSION

Participants in the young normal-hearing group served as the control group to verify the validity of the stimuli. Figure 1 plots the percentage of "buy" judgments as a func-

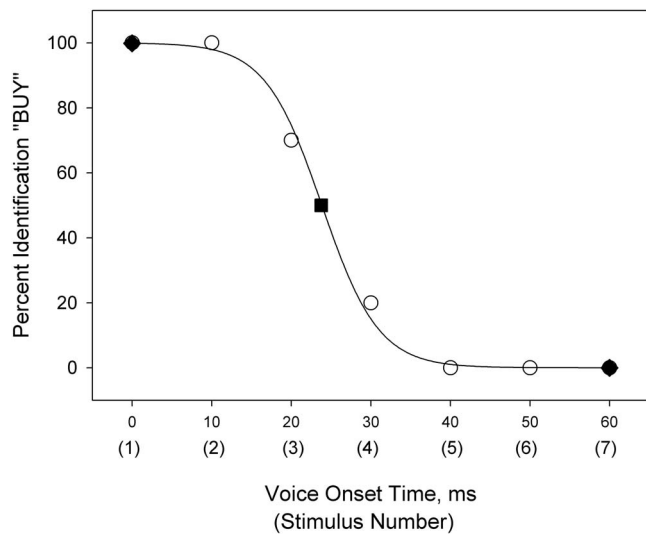


FIG. 1. Sample identification function for the *buy/pie* continuum of one listener from the young normal-hearing (Yng Norm control group). The x-axis is labeled with the actual VOT of stimuli on the continuum as well as the stimulus number corresponding to each VOT value (in parentheses). Open circles are the raw data points, the line is the PSIGNIFIT fitted function to these data, the filled square is the PSIGNIFIT derived crossover point, and the filled diamonds are the PSIGNIFIT derived endpoints.

tion of VOT for a representative young normal-hearing participant. Actual data points are shown (open circles) as well as the PSIGNIFIT fitted function for these data and the derived crossover point (filled square) and endpoints (filled diamonds). The figure shows that this listener categorized the shortest VOT (stimulus 1) as “buy” and the longest VOT (stimulus 7) as “pie.” The performance of this listener on this stimulus pair was typical of the performance of all of the listeners on most of the stimulus pairs, indicating that participants heard the endpoint stimuli as intended. Typically, the crossover point for all listeners occurred in the region defined by stimuli 3, 4, and 5. The PSIGNIFIT curve also demonstrates the accuracy of the fitted function for data of this form.

A. Endpoint accuracy for isolated stimuli

Percent identification scores were analyzed for the two endpoint stimuli presented in isolation because these stimuli served as the baseline stimuli. Analyses were conducted separately for stimuli from each of the eight continua, following arc-sine transformations of the identification scores. In general, the level of performance of all groups was high for these endpoint stimuli on each continuum (range = 87%–100%; mean = 97.01%). This is not unexpected, given that responses on continua demonstrating random performance were eliminated from the analyses. Separate analyses of variance (ANOVAs) were conducted for each continuum using a split-plot factorial design with one between-subjects factor (group: four levels) and one within-subjects factors (stimulus: two endpoints). In this and all subsequent analyses, interaction effects were examined with simple main effect analyses and multiple comparison testing (Bonferroni method) as appropriate.

Overall, effects of group were few in these analyses, indicating that the four listener groups were reasonably well equated for recognition of the endpoint stimuli in each continuum presented in isolation (the baseline condition). A significant main effect of group [$F(3, 49) = 3.60, p < 0.05$] was observed for one continuum in the domain of vowel duration (*Alisa/Aliza*), but not in the other continuum for vowel duration (*Rite/Ride*). The significant effect reflected higher performance by the two younger groups compared to the older normal-hearing group. A significant stimulus by group interaction was observed for one VOT continuum (*buy/pie*) [$F(3, 54) = 3.57, p < 0.05$] but not for the other VOT continuum (*Greg/Creig*). The source of the significant interaction was higher recognition scores for *pie* (100%) than for *buy* (96%) for the young noise-masked listeners only.

The only significant main effects of stimulus were observed for one VOT continuum [*Greg/Creig*: $F(1, 51) = 11.28, p < 0.01$] and one continuum varying formant transition duration [*Bill/Will*: $F(1, 51) = 8.45, p < 0.01$]. Significant effects were not observed for the other VOT continuum (*buy/pie*) or for the other formant transition duration continuum (*beat/wheat*). Listeners, on average, showed higher recognition scores for the endpoint stimulus *Greg* (97%) compared to the endpoint stimulus *Creig* (92%). Although the endpoint stimulus *Will* was identified with higher accuracy than *Bill*, it is noteworthy that the mean performance scores exceeded 90% for this continuum. In summary, all target stimuli were perceived highly accurately in the baseline condition in all listener groups.

B. Crossover points

Each continuum was structured such that the endpoint stimulus with the shorter acoustic cue served as the reference stimulus. Crossover points that were closer to this reference stimulus were expressed as shorter values (in milliseconds) than those that were closer to the alternate endpoint. Thus, a longer crossover point would indicate that a given listener (or group) requires a longer acoustic cue, e.g., VOT, to shift their percept from voiced to voiceless.

1. Voice-onset time

The mean crossover points for the four listener groups for the two continua that varied VOT (*buy/pie*, *Greg/Creig*), presented in isolation and in the sentence contexts, are shown in Fig. 2. For the *buy/pie* VOT continuum, it is apparent that these crossover points vary considerably for stimuli presented in isolation and in the sentence context and between listener groups. For a few participants, the estimated crossover point was beyond the range of stimulus values; nevertheless, data of these participants were included in the analysis. This outcome is due to the Psignifit fitting function estimates in cases where the listener’s identification responses were predominantly for one stimulus on the continuum, with an abrupt shift to the percept of the contrasting phoneme for the other endpoint stimulus. ANOVAs conducted on the crossover points revealed significant main effects of context [$F(1, 53) = 65.34, p < 0.01$] and group [$F(3, 53) = 23.46, p < 0.01$] and a significant interaction be-

Cue Domain: Voice Onset Time

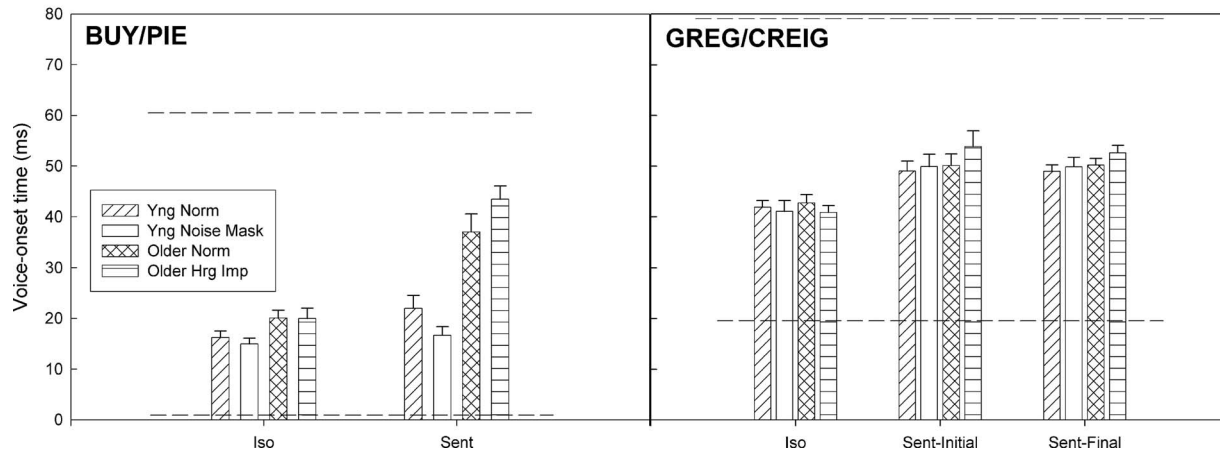


FIG. 2. Mean crossover points (in ms) for the VOT cue for the four listener groups measured in two continua (*buy/pie* and *Greg/Creig*) presented as isolated words (Iso) and in a sentence context (Sent) for the *buy/pie* continuum, and as isolated words (Iso) and in a sentence-initial position (Sent-Initial) and in a sentence-final position (Sent-Final) for the *Greg/Creig* continuum. Error bars show 1 standard error of the mean. The dashed horizontal lines indicate the minimum and maximum VOT values for the endpoint stimuli on each continuum.

tween the two [$F(3, 53) = 11.80, p < 0.01$]. *Post hoc* testing showed that the group effect was significant for the sentence context only ($p < 0.01$), with the two older groups exhibiting longer crossover points than the two younger groups.

The ANOVA on the crossover points for the *Greg/Creig* continuum also assessed the main effects of context and group, but the context effect had three levels (the design for all of the “name” continua): words in isolation, words at the initial position of sentences, and words at the final position of sentences. A significant main effect of context [$F(2, 76) = 59.57, p < 0.01$] was observed. The main effect of group and the interaction effect were not significant. A *post hoc* analysis of the context effect indicated that crossover points were longer for the words presented in sentences (either at the beginning or at the end of the sentence) than for isolated words. There were no significant differences in crossover points for words occurring in the two-sentence locations.

One hypothesis was that age-related differences in perception of target temporal cues would be altered when the target was presented in the context of a sentence. Thus, a group by context interaction effect was predicted. This predicted interaction effect was observed for the crossover points in the *buy/pie* continuum. For this continuum, there were no group differences in the crossover points for the words presented in isolation, consistent with previous findings (Gordon-Salant *et al.*, 2006). However, for words presented in the sentence context, there was a significant group effect in which the crossover points for the two older groups were longer than those for the two younger groups. This result indicates that older listeners required a longer VOT to perceive an ambiguous token on the *buy/pie* continuum as *pie* in a sentence context compared to younger listeners. This is similar to psychoacoustic findings indicating that older listeners show longer DLs for tones presented in a sequence compared to tones presented in isolation, but younger listeners show stable DLs in these two conditions (Fitzgibbons and Gordon-Salant, 1995).

For the *Greg/Creig* continuum, the only significant effect was for context, in which all listener groups required

longer VOTs to switch their perception from the voiced *Greg* to the voiceless *Creig* in the sentence context. The divergent findings across the two VOT continua may be related to differences in the specific nature of the VOT cue in these two sets of stimuli. Although both continua varied the duration of the aspiration following a burst as a cue for initial stop voicing, the endpoint stimulus *buy* had a VOT of 0 ms (i.e., no aspiration), whereas the endpoint stimulus *Greg* had a VOT of 20 ms (i.e., a 20 ms aspiration). This is characteristic of VOT values for word initial velar stops reported by Lisker and Abramson (1964). For listeners to shift their percept from *buy* to *pie* may therefore require the detection of the presence of a brief segment of aspiration, whereas for listeners to shift their percept from *Greg* to *Creig* may require the discrimination of a longer segment of aspiration from a shorter one. Clearly, the young listeners (with normal hearing and noise masking) were good at detecting the presence of the aspiration in the *buy/pie* continuum, regardless of the stimulus context (isolation or in sentences), and the older listeners were also reasonably good at this task for the isolated word stimuli. When the stimuli were embedded in the sentence context (final position), the detection of the aspiration segment was more difficult for the older listeners and may be analogous to age effects observed in temporal resolution tasks when the relative location of a brief interruption in a sound is shifted toward the temporal boundary of the stimulus (He *et al.*, 1999). In contrast, the *Greg/Creig* continuum may be more similar to a duration discrimination task for stimulus intervals defined by aspiration. The sentence context effects observed with the *Greg/Creig* continuum appear to parallel the sequential context effects observed in psychoacoustic investigations of duration discrimination of silent intervals by younger and older listeners (Fitzgibbons and Gordon-Salant, 1995), in which both listener groups showed larger DLs for intervals presented in sequences than in isolation. These observations underscore the importance of interpreting context effects in relation to the specific temporal cue under investigation. However, these effects need to be

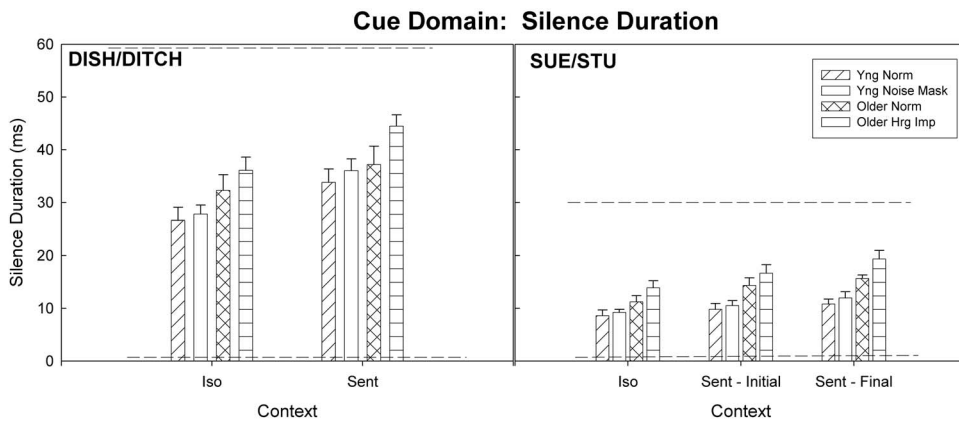


FIG. 3. Mean crossover points (in ms) for the silence duration cue for the four listener groups measured in two continua (*dish/ditch* and *Sue/Stu*) presented as isolated words (Iso) and in a sentence context (Sent) for the *dish/ditch* continuum, and as isolated words (Iso) and in a sentence-initial position (Sent-Initial) and in a sentence-final position (Sent-Final) for the *Sue/Stu* continuum. Error bars show 1 standard error of the mean. The dashed horizontal lines indicate the minimum and maximum silence duration values for the endpoint stimuli on each continuum.

replicated with different tokens of the same target cue, spoken by either the same talker or different talkers.

2. Silent interval

The two continua that varied the duration of a silent interval signaled a shift from a fricative to an affricate (*dish/ditch*) or a shift from a fricative to a consonant cluster (*Sue/Stu*). Mean crossover points derived for these two continua in isolation and sentence contexts are presented in Fig. 3. For the *dish/ditch* continuum, the analysis of the data indicated significant main effects of context [$F(1,54)=34.02$, $p < 0.01$] and group [$F(3,54)=4.00$, $p < 0.05$] but no significant interaction. The main effect of context reflected the longer crossover points observed for the sentence context than for isolated words. Multiple comparison testing revealed that the older hearing-impaired group had longer crossover points than the two young listener groups.

The target cue in the *Sue/Stu* continuum (Fig. 3) was a silent interval duration preceding the final vowel. The ANOVA revealed significant main effects of context [$F(2,92)=27.01$, $p < 0.01$] and group [$F(3,46)=10.68$, $p < 0.01$] and no significant interaction. All listeners exhibited longer crossover points for words in sentences compared to isolated words and longer crossover points for words occurring at the end of the sentence than at the beginning of the sentence. The group effect was attributed to longer crossover points for the older hearing-impaired group compared to the two younger groups. Additionally, the older listeners with normal hearing showed longer crossover points than young listeners with normal hearing.

Overall, similar significant main effects of context and group were observed for the two continua that varied the silence duration cue. The context effect for the *dish/ditch* continuum was attributed to longer crossover points for words in sentences than in isolation. In the *Sue/Stu* continuum, all listeners showed longer crossover points for words in the final sentence position than words in isolation. The final sentence position for these name stimuli is comparable to the location of the target word in sentences for the noun stimuli. Thus, the context effects for these two continua were quite similar.

The group effects in these two continua were also highly comparable. *Post hoc* analyses revealed that for both the *dish/ditch* and *Sue/Stu* continua, the older hearing-impaired

listeners demonstrated longer crossover points than the two groups of younger listeners. Both of these continua varied the silent interval duration as a cue for the distinction /f/ versus /tʃ/ or the distinction /s/ versus /st/; these consonants are classified as sibilants. This effect of hearing impairment was observed previously for the *dish/ditch* continuum presented as isolated words (Gordon-Salant *et al.*, 2006). It is noteworthy that the range in duration of the silent intervals was twice as long for the *dish/ditch* continuum (0–60 ms) as for the *Sue/Stu* continuum (0–30 ms), yet the same group effect was observed. The listeners in the older hearing-impaired group had moderate hearing losses in the high frequencies, and it is likely that the high frequency cues for the sibilant consonants in the target words were less audible in this group compared to the two young groups with normal hearing, even at the high presentation level used in the study. As a result, hearing-impaired listeners may have experienced difficulty in coding the onset and offset of sibilance and in detecting the relatively brief silent interval that is contiguous with these sibilant consonants. Several studies have shown that hearing-impaired listeners have larger gap detection thresholds than normal-hearing listeners when the gap is bounded by a broadband noise burst (Irwin *et al.*, 1981; Florentine and Buus, 1984) or when the gap is bounded by two different markers, such as a broadband noise burst and a tone (Grose *et al.*, 1989; Lister *et al.*, 2000). Thus, reduced ability to detect a silent gap between two markers may contribute to the hearing-impaired listeners' use of the silent interval cue, although this was not assessed directly in the present investigation.

3. Transition duration

The temporal cue for the *beat/wheat* and *Bill/Will* continua was transition duration; the mean crossover points of the four listener groups are shown in Fig. 4. Results of statistical analysis for the *beat/wheat* continuum revealed significant main effects of context [$F(1,44)=48.86$, $p < 0.01$] and group [$F(3,44)=6.89$, $p < 0.01$] and no significant interaction. For this continuum, crossover points were longer for isolated words than for the stimuli in the sentence context. The group effect was attributed to longer crossover points for the two older groups compared to the young normal-hearing group.

Cue Domain: Transition Duration

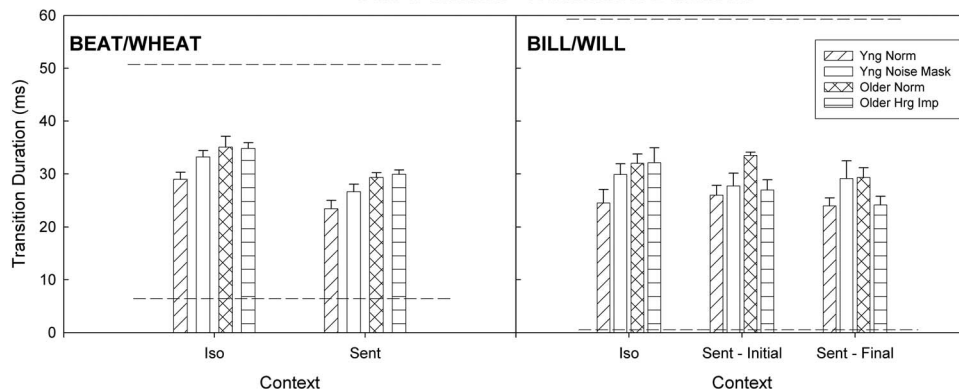


FIG. 4. Mean crossover points (in ms) for the transition duration cue for the four listener groups measured in two continua (*beat/wheat* and *Bill/Will*) presented as isolated words (Iso) and in a sentence context (Sent) for the *beat/wheat* continuum, and as isolated words (Iso) and in a sentence-initial position (Sent-Initial) and in a sentence-final position (Sent-Final) for the *Bill/Will* continuum. Error bars show 1 standard error of the mean. The dashed horizontal lines indicate the minimum and maximum transition duration values for the endpoint stimuli on each continuum.

Somewhat similar results were observed for the *Bill/Will* continuum. Statistical analyses revealed a significant main effect of context only [$F(2,90)=5.79, p<0.01$]. Multiple comparison testing indicated that crossover points were longer for isolated words than for words in the sentence-final position.

Taken together, the two continua that varied transition duration as a cue to the consonant manner (stop/semivowel) exhibited context effects, based on an analysis of the crossover points, that were highly similar and quite different from those observed for the other continua. Specifically, listeners from all groups showed longer crossover points for the isolated words than for the words in the final position of a sentence. One possible explanation for this observation is that listeners relied on the rapid transition to cue the initial stop, /b/, since these stimuli did not contain a burst. For words presented in isolation, listeners perceived transition durations of 0–20 ms (on average) as /b/, which is consistent with the expected duration of this cue for this phoneme (Lieberman *et al.*, 1956; Lisker and Abramson, 1964). When the target noun was embedded in the sentence-final position, it was usually preceded by the vowel /ə/. This may have produced a continuous transition from the vowel /ə/ to the onset of the target stimulus, which in turn may have enhanced the percept of the semivowel /w/ with shorter transition durations. Listeners required an extremely brief transition to signal the onset of the target word as *beat* following this vowel. The same argument does not apply for the name continuum, however, because the phoneme preceding the target was a voiceless stop.

Group effects were observed for the *beat/wheat* continuum. Younger listeners showed shorter crossover points than the two older listener groups, and this finding is similar to a previous report for the *beat/wheat* continuum presented as isolated words (Gordon-Salant *et al.*, 2006). For the *Bill/Will* continuum, there was considerable variability in the crossover points, which likely precluded the observation of any age effects. It is not clear why there was a large variability in the crossover points for the young listeners with normal hearing, especially in light of the fact that the endpoint accuracy for these stimuli was quite high for all listener groups (>94%).

4. Vowel duration

The last two continua, *Alisa/Aliza* and *Rite/Ride*, varied the vowel duration as the target cue for postvocalic consonant voicing. An inspection of the crossover distribution for *Alisa/Aliza* (Fig. 5) suggests that younger listeners perform differently than the older listeners, and the ANOVA confirmed this impression. A significant main effect was observed for listener group [$F(3,42)=16.18, p<0.01$]. *Post hoc* testing revealed that the two older groups had longer crossover points than the two younger groups. None of the other effects were significant. For the *Rite/Ride* continuum (Fig. 5), the results showed a significant main effect of context [$F(2,108)=5.37, p<0.01$] and a significant main effect of group [$F(3,54)=4.24, p<0.01$], with no significant interaction between these two variables. A *post hoc* analysis of the group effect showed a significant difference in mean per-

Cue Domain: Vowel Duration

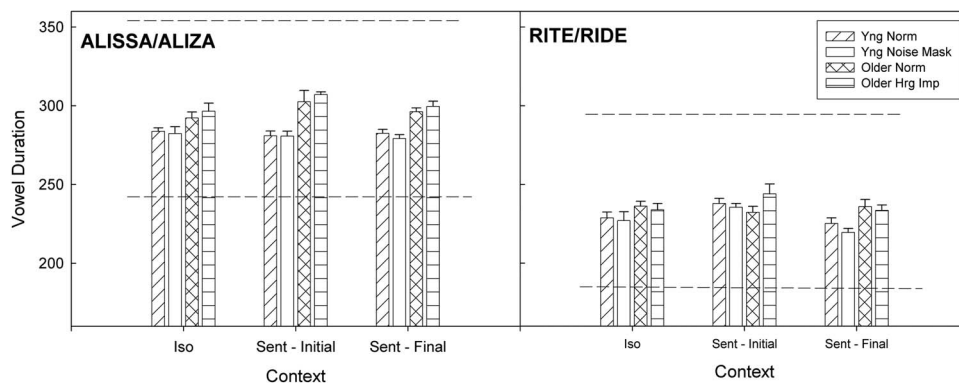


FIG. 5. Mean crossover points (in ms) for the vowel duration cue for the four listener groups measured in two continua (*Alisa/Aliza* and *Rite/Ride*) presented as isolated words (Iso) and in a sentence-initial position (Sent-Initial) and in a sentence-final position (Sent-Final). Error bars show 1 standard error of the mean. The dashed horizontal lines indicate the minimum and maximum vowel duration values for the endpoint stimuli on each continuum.

formance between the young noise-masked listeners and the two older listener groups, which was attributed to longer crossover points for the older groups. Crossover points for words in the sentence-initial position were overall longer than for words in the sentence-final position as well as for words in isolation.

Observed group effects were similar for these two continua. Analyses of crossover points for the *Alisa/Aliza* continuum revealed that the two groups of younger listeners demonstrated shorter crossover points than the two groups of older listeners. In the *Rite/Ride* continuum, the young noise-masked listeners also showed shorter crossover points than the two groups of older listeners. This group difference in crossover points was found consistently for stimuli presented in isolation and in the sentence context for both continua. For both continua, then, older listeners had longer crossover points than younger listeners. Because vowel duration is a well-established cue for postvocalic voicing in stops (Peter-son and Lehiste, 1960) and fricatives (Denes, 1955), the current findings indicate that the older listeners required longer vowel durations than younger listeners to alter their percept of postvocalic consonantal voicing.

A context effect was also shown for the *Rite/Ride* continuum. All listeners showed longer crossover points for words presented in the sentence-initial position than for words presented either in isolation or in the sentence-final position. This is the only continuum in which this pattern of context effects was observed. The presentation of the target item in the initial position of the sentence was somewhat different in this continuum than in all other continua because the target was embedded in the phrase “Mr. Rite” or “Mr. Ride.” As a result, the target cue was actually embedded in the second word in the sentence, rather than the first word. This may have made the target cue less perceptible or less accessible to the listeners in this location compared to the words in isolation or in the sentence-final position.

C. Slopes

Similar statistical analyses to those reported above were conducted for the slope values derived from the PSIGNFIT analysis of each listener’s identification function for each condition. In general, the analyses of the slope values did not reveal significant effects. Of the 24 effects analyzed (8 continua \times 3 effects/continuum), only three were significant. The effects of context only were significant for one VOT continuum [*buy/pie*; $F(1, 54) = 20.46$, $p < 0.01$] and one transition duration continuum [*beat/wheat* $F(1, 44) = 12.15$, $p < 0.01$]. Slope values were higher for the isolated word stimuli compared to the words in sentences for both of these continua. It should be noted that significant effects of context on slope values were not observed for the other VOT continuum (*Greg/Creig*) or the other transition duration continuum (*Bill/Will*). A significant main effect of group was observed for one continuum varying silent interval duration [*Sue/Stu*: $F(3, 46) = 4.53$, $p < 0.01$]. Multiple comparison testing showed that the young noise-masked listeners exhibited steeper slopes than the two older listener groups on this continuum. Group effects were not observed for the other

continuum varying silent interval duration (*dish/ditch*). In summary, the identification function slopes varied minimally between groups and contexts for the different continua.

IV. GENERAL DISCUSSION

The principal hypothesis concerning differential performance by all listener groups in the two stimulus contexts was observed for many temporal cues. In most cases, the sentence context effect (when present) resulted in longer crossover points for sentences than for individual words. The older listener groups also perceived the target temporal cues differently from the younger listener groups for many stimulus continua.

Age-related differences in the perception of temporal cues in natural speech continua were observed in this study in most, but not all, continua. In all cases where age effects were observed, older listeners required a longer temporal cue to alter their percept from the reference stimulus (that containing the shorter cue) to the alternate stimulus. The strongest age-related effects were observed in the two continua that varied the vowel duration as a cue to postvocalic voicing (*Alisa/Aliza* and *Rite/Ride*), and these age effects were seen for all stimulus contexts. Age effects were also observed in one continuum that varied the transition duration (*beat/wheat*) for stimuli presented in both word and sentence contexts. A modest age effect was observed for the *Sue/Stu* continuum as well. Finally, strong age effects were observed for one VOT continuum (*buy/pie*) in the sentence context only. These observations are consistent with the notion that for many stimuli, older listeners have difficulty processing relatively brief changes in duration compared to younger listeners. The present findings suggest, however, that the problems older people have in the perception of brief target temporal cues in speech are cue specific. The duration of temporal cues in natural speech is highly variable and dependent on the talker, the stimulus context, and the specific target phonemes. The current results show that a different pattern of age-related effects can emerge for the same type of temporal cue when it appears in different phonemes. This may be one contributing factor to the common observation of wide variability in speech recognition performance of older listeners.

These findings may have implications for the perception of accented speech by elderly listeners. For example, stimuli in the middle of a speech continuum may be similar to those produced in accented English. A native Spanish speaker of English will produce the intended word, *pie*, with a shorter VOT than a native English speaker, resulting in the perception of this word as *buy* by a native English listener. Because older listeners require longer VOTs than younger listeners to perceive *pie* in sentence contexts, it is even more likely that they will misperceive the intended *pie* produced by native Spanish speakers. A similar phenomenon may occur as a result of altered vowel duration in accented speech. Non-native speakers of English often produce shorter vowels than native speakers, resulting in a greatly reduced distinction in vowel duration as a cue to postvocalic consonant voicing (Chen, 1970). This could result in the perception of a postvocalic consonant as voiceless rather than voiced, particularly

by older listeners who require longer vowel durations in the region of uncertainty to shift their perception to the voiced cognate. In the current experiment, the ambiguous words in the middle of the continuum were presented in the context of unaccented English. At present, it is unknown if embedding these same words in the context of accented English would alter the listener perception of the ambiguous target words. It is possible that listeners would have less difficulty because of perceptual normalization, but, alternatively, listeners may have more difficulty because contextual cues may not be well understood. At least one report (Burda *et al.*, 2003) has shown that older listeners recognize accented English more poorly than young adult listeners. The current findings are useful for explaining some possible underlying mechanisms for this observation.

Age-related differences in identification function slopes were not observed for most continua. It was anticipated that older listeners, who often exhibit poorer temporal resolution than younger listeners, would show more gradual identification function slopes on speech continua that varied temporal cues compared to younger listeners. For example, Strouse *et al.* (1998) reported significantly different slope values between younger and older listener groups for the perception of a synthetic speech continuum varying from /ba/ to /pa/. There are some procedural differences between the study of Strouse *et al.* (1998) and the present study that might account for differing outcomes regarding slope values. First, the stimuli in the present study were natural speech tokens, rather than synthetic speech tokens as used in the study of Strouse *et al.* (1998). With real speech, there are multiple cues in addition to the temporal cue that was manipulated in the present study. Listeners may have made use of these cues. Second, the current investigation used the PSIGNIFIT analysis (a maximum likelihood method with the selection of a logistic form of the psychometric function) to estimate the crossover points and slopes from the identification data, whereas the study of Strouse *et al.* (1998) used a linear regression analysis to derive crossover points and slopes. These methods for estimating the slopes are quite different and possibly contributed to different outcomes. Nevertheless, the current analysis of the data revealed clear differences in crossover durations required by younger and older listeners in many of the speech continua. Taken together, these findings suggest that for many of the temporally mediated continua, the entire function appears to be shifted to longer stimulus values for the older listeners, but the shape of the function is not notably different for younger and older listeners.

Hearing loss effects were seen consistently in the continua that varied the silent interval duration. Hearing-impaired listeners required a longer period of silence to perceive a stop adjacent to a sibilant, as in *Stu* or *ditch*, compared to listeners with normal hearing. This finding may be related to reduced audibility of the sibilance and/or poorer gap detection ability. These results may extend as well to the perception of accented English by listeners with hearing loss. Non-native speakers of English whose first language is Spanish frequently shorten the silent interval preceding /ʃ/ when saying /tʃ/. The implication of the current findings is that

listeners with hearing loss may misperceive affricate phonemes in English.

Low-level noise masking was employed in this study in an effort to equate the audibility of speech signals between the young normal-hearing (noise-masked) listeners and the older normal-hearing listeners. Differences in performance between the older normal-hearing listeners and young noise-masked listeners were observed for some of the measures (e.g., crossover points for one VOT continuum in the sentence context) and underscored the presence of age effects that were not confounded by possible differences in high-frequency sensitivity between the younger and older listeners with clinically normal hearing. However, a second group of young normal-hearing listeners without noise masking was also employed to serve as a baseline referent group. Overall, the results indicate that the two young listener groups (with and without noise masking) performed similarly for all conditions. This finding suggests that the use of a low-level noise masker to equate the thresholds of younger and older listeners with normal hearing may not be necessary for investigations of age-related differences in perceptual judgments involving speech continua presented at high supra-threshold levels.

Context effects were observed for most of the speech continua. For several continua, listeners required longer target cues in the sentence context than in isolation, as in *Greg/Creig*, *dish/ditch*, *Sue/Stu*, *Rite/Ride*, and, for older listeners, *buy/pie*. However, for the continua that varied the transition duration (*beat/wheat* and *Bill/Will*), the context effect was attributed to shorter crossover points for words presented in sentence contexts than in isolation. These divergent results are likely due to the nature of the exact sentence context for these stimuli rather than to differential effects of memory. If memory effects played a significant role in these findings, the context effects would be consistent for all stimulus continua, with age-related deficits being exaggerated in all sentence contexts. Rather, the current findings suggest that context effects, when observed, generally show different crossover points for words in the sentence-final position than either isolated words or words in the sentence-initial position, suggesting that effects of specific phonetic sequences play a role in these results.

Overall, the results tentatively suggest that when older listeners with or without hearing loss require longer target temporal cues than younger normal-hearing listeners for words presented in isolation, they are likely to require a similar lengthening of these target cues in sentence contexts. This was observed in five of the eight continua. Moreover, the finding of context effects in seven of the eight continua suggests that patterns of performance observed for the perception of isolated words do not necessarily predict speech perception performance for these same words in the context of a sentence. The current study demonstrates the importance of evaluating the perception of target speech cues in sentence contexts.

ACKNOWLEDGMENTS

This research was supported by a MERIT award from the National Institute on Aging (R37AG09191). The authors

are grateful to Michele Spencer, Jessica Barrett, and Maureen D'Antuono for their assistance with data collection and analysis, and to Ed Smith for implementing the PSIGNIFIT analyses. The authors also appreciate the valuable comments of three anonymous reviewers.

- ANSI (2004). ANSI S3.6-2004, "American National Standard Specification for Audiometers (Revision of ANSI S3.6-1996), (American National Standards Institute, New York).
- Baddely, A. D. (1986). *Working Memory* (Oxford University Press, Oxford, England).
- Bergeson, T. A., Schneider, B. A., and Hamstra, S. J. (2001). "Duration discrimination in younger and older adults," *Can. Acoust.* **29**, 3–9.
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (1984). "Standardization of a test of speech perception in noise," *J. Speech Hear. Res.* **27**, 32–48.
- Burda, A. N., Scherz, J. A., Hageman, C. F., and Edwards, H. T. (2003). "Age and understanding speakers with Spanish or Taiwanese accents," *Percept. Mot. Skills* **97**, 11–20.
- Chen, M. (1970). "Vowel length variation as a function of the voicing of the consonant environment," *Phonetica* **22**, 129–159.
- Cruikshanks, K. L., Wiley, T. O., Tweed, T. S., Klein, B. E. K., Klein, R., Mares-Perlman, J. A., and Nondahl, D. M. (1998). "Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin: The epidemiology of hearing loss study," *Am. J. Epidemiol.* **148**, 879–886.
- Denes, P. (1955). "Effect of duration on the perception of voicing," *J. Acoust. Soc. Am.* **27**, 761–764.
- Fisk, A. D., and Rogers, W. (1991). "Toward an understanding of age-related memory and visual search effects," *J. Exp. Psychol. Gen.* **120**, 131–149.
- Fitzgibbons, P., and Gordon-Salant, S. (1994). "Age effects on measures of auditory duration discrimination," *J. Speech Hear. Res.* **37**, 662–670.
- Fitzgibbons, P. J., and Gordon-Salant, S. (1995). "Duration discrimination with simple and complex stimuli: Effects of age and hearing sensitivity," *J. Acoust. Soc. Am.* **98**, 3140–3145.
- Flege, J. E., and Eefting, W. (1988). "Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation," *J. Acoust. Soc. Am.* **83**, 729–740.
- Florentine, M., and Buus, S. (1984). "Temporal gap detection in sensorineural and simulated hearing impairments," *J. Speech Hear. Res.* **27**, 449–455.
- Fox, R. A., Flege, J. E., and Munro, J. (1995). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," *J. Acoust. Soc. Am.* **97**, 2540–2551.
- Gelfand, S., Schwander, T., and Silman, S. (1990). "Acoustic reflex thresholds in normal and cochlear-impaired ears: Effect of no-response rates on 90th percentiles in a large sample," *J. Speech Hear. Disord.* **55**, 198–205.
- Gordon-Salant, S., and Fitzgibbons, P. (1997). "Selected cognitive factors and speech recognition performance among young and elderly listeners," *J. Speech Lang. Hear. Res.* **40**, 423–431.
- Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Barrett, J. (2006). "Age-related differences in identification and discrimination of temporal cues in speech segments," *J. Acoust. Soc. Am.* **119**, 2455–2466.
- Grose, J. H., Eddins, D. A., and Hall, J. W., III (1989). "Gap detection as a function of stimulus bandwidth with fixed high frequency cut-off in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **86**, 1747–1755.
- Grose, J. H., Hall, J. W., III, and Buss, E. (2006). "Temporal processing deficits in the pre-senescent auditory system," *J. Acoust. Soc. Am.* **119**, 2305–2315.
- Hargus, S. E., and Gordon-Salant, S. (1995). "Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment," *J. Speech Hear. Res.* **38**, 234–243.
- He, N.-J., Horwitz, A. R., Dubno, J. R., and Mills, J. (1999). "Psychometric functions for gap detection in noise measured from young and aged subjects," *J. Acoust. Soc. Am.* **106**, 966–978.
- Humes, L. E., Lee, J. H., and Coughlin, M. P. (2006). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," *J. Acoust. Soc. Am.* **120**, 2926–2937.
- Irwin, R. J., Hinchcliff, L. K., and Kemp, S. (1981). "Temporal acuity in normal and hearing-impaired listeners," *Audiology* **20**, 234–243.
- Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (1956). "Tempo of frequency change is a cue for distinguishing classes of speech sounds," *J. Exp. Psychol.* **52**, 127–137.
- Lisker, L., and Abramson, A. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," *Word* **20**, 384–422.
- Lister, J., Koehnke, J., and Besing, J. (2000). "Binaural gap duration discrimination in listeners with impaired hearing and normal hearing," *Ear Hear.* **21**, 141–150.
- MacKay, I. R. A., Flege, J. E., and Piske, T. (2000). "Persistent errors in the perception and production of word-initial English stop consonants by native speakers of Italian," *J. Acoust. Soc. Am.* **107**, 2802.
- Munson, B., and Nelson, P. B. (2005). "Phonetic identification in quiet and in noise by listeners with cochlear implants," *J. Acoust. Soc. Am.* **118**, 2607–2617.
- Peterson, G. E., and Lehiste, I. (1960). "Duration of syllable nuclei in English," *J. Acoust. Soc. Am.* **32**, 693–703.
- Pfeiffer, E. (1977). "A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients," *J. Am. Geriatr. Soc.* **23**, 441–443.
- Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (1995). "How young and old adults listen to and remember speech in noise," *J. Acoust. Soc. Am.* **97**, 593–608.
- Roup, C. M., Wiley, T. L., Safady, S. H., and Stoppenbach, D. T. (1998). "Tympanometric screening norms for adults," *Am. J. Audiology* **7**, 55–60.
- Salthouse, T. A. (1996). "The processing-speed theory of adult age differences in cognition," *Psychol. Rev.* **103**, 403–428.
- Snell, K. B., and Hu, H.-L. (1999). "The effect of temporal placement on gap detectability," *J. Acoust. Soc. Am.* **106**, 3571–3577.
- Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (1998). "Temporal processing in the aging auditory system," *J. Acoust. Soc. Am.* **104**, 2385–2399.
- Tillman, T. W., and Carhart, R. C. (1966). "An expanded test for speech discrimination utilizing CNC monosyllabic words: N. U. Auditory Test No. 6," Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, pp. 1–12.
- Tun, P. (1998). "Fast noisy speech: Age differences in processing rapid speech with background noise," *Psychol. Aging* **13**, 424–434.
- WEDW (waveform display, labeling and editing software), http://www.asel.udel.edu/speech/Spch_proc/software.html (Last viewed January 8, 2008).
- Wichmann, F. A., and Hill, N. J. (2001a). "The psychometric function: I. Fitting, sampling, and goodness of fit," *Percept. Psychophys.* **63**, 1293–1313.
- Wichman, F. A., and Hill, N. J. (2001b). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," *Percept. Psychophys.* **63**, 1314–1329.