Temporal offset judgments for concurrent vowels by young, middle-aged, and older adults

Daniel Fogerty

*Department of Communication Sciences and Disorders, University of South Carolina, Columbia, South Carolina 29208
fogerty@sc.edu*

Diane Kewley-Port and Larry E. Humes

*Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405
kewley@indiana.edu, humes@indiana.edu*

Abstract: Temporal processing declines with age may reduce the processing of concurrent vowels. For this study, listeners categorized vowel pairs varying in temporal asynchrony as one sound, two overlapping sounds, or two sounds separated by a gap. Two boundaries separating the three response categories, multiplicity and gap-identification, were measured. Compared to young and middle-aged listeners, older listeners required longer temporal offsets for multiplicity. Middle-aged and older listeners also required longer offsets for gap-identification. For older listeners, correlations with various temporal processing tasks indicated that vowel temporal-order thresholds were related to multiplicity, while age and non-speech gap-detection thresholds were related to gap-identification.

© 2012 Acoustical Society of America

PACS numbers: 43.71.Rt, 43.71.Es, 43.66.Sr, 43.71.An [SGS]

Date Received: April 2, 2012 **Date Accepted:** May 10, 2012

1. Introduction

A common complaint that older listeners have is that certain talkers “speak too fast,” making it difficult to understand the intended message. This complaint may in part be justified by accounts of “cognitive slowing” (Salthouse, 1996). In addition, older adults frequently encounter declines in auditory temporal processing (Gordon-Salant and Fitzgibbons, 1993). Recent studies in the Audiology Research Laboratory at Indiana University have documented declines in auditory temporal processing across the lifespan for a host of tasks (Fogerty *et al.*, 2010; Humes *et al.*, 2009, 2010). Consistent with a number of other studies (e.g., Moore *et al.*, 1992; Snell and Frisina, 2000), we have observed large individual differences for a number of our tasks with some older listeners performing within the range of the younger listeners. Of particular interest is the source of difficulty for older adults. Although many of the tasks we have used employed vowels, the ability to identify vowels accurately in isolation does not appear to account for problems with temporally ordering vowels presented in rapid sequence (Fogerty *et al.*, 2012). Therefore, the current study was proposed to investigate another possible locus of difficulty: the categorization of temporal events. For example, it may be that some older listeners have difficulty even identifying whether one or two auditory events occurred during a temporal-order task, let alone the identities of the sounds comprising those events.

Arehart *et al.* (2011) recently explored this ability in their “multiplicity” task using a concurrent vowel paradigm. Listeners were required to report whether they heard one or two vowels presented on a given trial. They found that young listeners correctly categorized the presence of two simultaneous vowels with the same fundamental frequency at 33%, while older listeners performed more poorly, categorizing

two vowels correctly only 8% of the time. With a minimal fundamental frequency separation between the two vowels of one semitone, both age groups categorized the presence of two vowels with over 80% accuracy.

The current investigation explored this ability to judge the presence of multiple vowels using temporal separation, rather than fundamental-frequency separation, as the multiplicity cue. In this study, two vowels were presented with a varied temporal offset (and onset) of the second vowel. This created vowel pairs that listeners could perceive along a continuum as (i) a single vowel sound, (ii) two partially overlapping vowels, or (iii) two vowels separated by a silent gap. These three response alternatives were used to determine two category boundaries: one versus two vowels (multiplicity), and two overlapping vowels versus two vowels separated by a gap (gap-identification). The stimulus onset asynchrony (SOA) that corresponded to the two boundaries between categories (i.e., multiplicity and gap-identification) was measured. Three age groups of listeners were tested, spanning much of the adult lifespan, in order to investigate individual differences on this task and its relation to measures of auditory temporal processing with speech and non-speech stimuli that were reported earlier (see [Humes *et al.*, 2010](#) for an overview of these measures).

2. Methods

We measured SOA boundaries for young, middle-aged, and older adults in making multiplicity and gap-identification judgments for monaural presentation of vowel pairs.

2.1 Listeners

Young ($N = 72$; mean age = 22, 18–30 yr), middle-aged ($N = 31$; mean age = 48, 40–55 yr), and older ($N = 113$; mean age = 71, 60–88 yr) adult listeners participated in this study. Listeners did not have evidence of any cognitive impairment as determined by the Mini Mental State Exam ([Folstein *et al.*, 1975](#)). Maximum hearing thresholds for air conducted pure-tones were not to exceed the following limits in at least one ear: 40 dB HL (hearing level) ([ANSI, 2010](#)) at 250, 500, and 1000 Hz; 50 dB HL at 2 kHz; 65 dB HL at 4 kHz; and 80 dB HL at 6 and 8 kHz. It was also required that there be no evidence of middle ear pathology (air-bone gaps < 10 dB and normal tympanograms). Thresholds were selected to maximize participant eligibility while ensuring sufficient audibility of the experimental stimuli. Although significant variability in hearing sensitivity occurred between listeners and age groups, significant steps were taken to limit the contribution of audibility to performance. This included a high presentation level [i.e., 83 dB SPL (sound pressure level)] and low-pass filtering stimuli at 1800 Hz. In addition, all listeners were able to identify the low-pass filtered test vowels in isolation with at least 90% accuracy. These procedures ensured sufficient audibility for all listeners. Participants were paid for their participation.

2.2 Stimuli

Three vowels /i, ε, a/, spoken by a male Midwestern talker, were spoken rapidly in a /p/-vowel-/t/ context in a carrier phrase and recorded using an Audio-Technica (Stow, OH) AT2035 microphone in a sound-attenuating booth. Vowel productions that had the shortest duration and $F2 < 1800$ Hz were selected for the final stimuli. The words “pit,” “pet,” and “pot” were digitally edited to remove voiceless sounds, leaving only the voiced pitch pulses. Vowels were resynthesized using STRAIGHT ([Kawahara *et al.*, 1999](#)) to be 70-ms long with a steady-state fundamental frequency of 100 Hz. Final stimuli were low-pass filtered in MATLAB at 1800 Hz using a finite impulse response (8th order) filter with 80 dB roll-off to minimize the influence of high-frequency hearing loss on performance, as well as root-mean-square (rms) normalized.

2.3 Procedure

Stimuli were presented via Tucker-Davis Technologies (TDT, Alachua, FL) System III hardware using 16-bit resolution at a sampling frequency of 48828 Hz. The vowel

pairs, shifted by the selected SOA, were added together, output to the TDT digital-to-analog (D/A) converter, and then passed through a programmable attenuator (PA-5), a headphone buffer (HB-7), and finally to an ER-3A insert earphone. The earphone was calibrated using a calibration vowel of equal rms to the test stimuli in a 2-cm³ coupler using a Larson Davis (Depew, NY) model 2800 sound level meter with linear weighting. Two overlapping vowels measured 86 (± 2) dB SPL; isolated, non-overlapping vowels measured 83 dB SPL.

For the categorization task, participants listened to one of two different vowel pairs (/a-ɪ/ or /ɪ-ε/) presented monaurally at different SOAs from 0 ms to 175 ms in 13 steps (step size = 5–20 ms). For non-overlapping vowel pairs, silent gap durations ranged from 5 ms to 105 ms. Participants categorized these vowel pairs as containing one sound, two overlapping sounds, or two sounds separated by a silent gap using three response buttons on a custom designed MATLAB stimulus presentation interface. Thus, listeners were not required to identify the vowels. Listeners completed familiarization trials prior to testing. The categorization task took listeners about 30 min to complete.

Listeners also completed a host of other temporal processing tasks before and after these measurements. Gap detection was assessed for narrow band noise centered at 1000 Hz. Various temporal order abilities were measured using four vowel stimuli (the three used here in addition to /u/) under monaural and dichotic modes of presentation (see Fogerty *et al.*, 2010 for a complete description). Forward and backward masking was also measured using these vowels in the presence of a vowel-like babble masker. Task and group differences on these measures are described elsewhere (Fogerty *et al.*, 2010; Humes *et al.*, 2009, 2010). The focus of this investigation was to determine age group differences in the categorization of overlapping vowels (i.e., multiplicity judgment) and non-overlapping vowels (i.e., gap-identification judgment). The contribution of individual differences in temporal processing abilities to these judgments was also explored.

3. Results and discussion

Of the 216 listeners who participated in this study, 53 listeners only completed testing for one vowel pair (missing data: 18 /a-ɪ/; 35 /ɪ-ε/). An additional 34 listeners only had data for one of the boundary judgments (missing data: 5 multiplicity; 29 gap-identification). Six listeners had data at both boundaries, but for different vowel pairs. Missing boundary data occurred if the listener chose not to use a category. This could occur if the SOA values were not large enough for the participant to identify the silent gap, leading to a missing gap-identification boundary (i.e., floor performance); or the listener was able to use cues other than temporal offset (e.g., spectral cues) to identify the presence of multiple vowels, leading to a missing multiplicity boundary (i.e., ceiling performance). These missing values would also occur if the listener used the category so infrequently that it was never the predominant response at any SOA, resulting in no judgment boundary between categories. Listeners were only excluded in pair-wise analyses for which they had missing data.

3.1 Group differences

Category boundaries were assessed at the intersection of logistic functions fit to each of the three response categories for each individual. Functions fit to the pooled data for each group are displayed in Fig. 1. Median SOA and inter-quartile ranges for the category boundaries are displayed in Fig. 2. Kruskal-Wallis tests demonstrated significant differences between groups at the first category boundary (i.e., multiplicity) (/ɪ-ε/: $X^2(2) = 29.5$, $N = 186$, $p < 0.001$; /a-ɪ/: $X^2(2) = 9.0$, $N = 174$, $p < 0.05$) and at the second category boundary (i.e., gap-identification) (/ɪ-ε/: $X^2(2) = 10.2$, $N = 173$, $p < 0.01$; /a-ɪ/: $X^2(2) = 11.9$, $N = 156$, $p < 0.01$). *Post hoc* Bonferroni-adjusted rank-sum Mann-Whitney tests for the first boundary indicated this small, but significant, difference was isolated to the comparison between young and older listeners (/ɪ-ε/: $U = 1717.0$, $Z = -5.0$, $N = 163$,

Fig. 1. Logistic categorization functions fit to the pooled data for young (black, solid lines), middle-aged (gray, dashed lines), and older (white, dash-dot lines) adult listeners for categorizing the presence of (a) two overlapping vowels, and (b) two separated vowels.

$p < 0.01$, 12.8 ms difference; /a-ɪ/: $U=2000.5$, $Z=-2.5$, $N=149$, $p < 0.012$, 1.8 ms difference) and middle-aged and older listeners for /ɪ-ε/ ($U=605.5$, $Z=-3.4$, $N=121$, $p < 0.01$, 15.1 ms difference). Young listeners also placed the second boundary at a significantly smaller SOA than both older (/ɪ-ε/: $U=2035.5$, $Z=-2.6$, $N=151$, $p < 0.012$, 11.9 ms difference) and middle-aged listeners (/ɪ-ε/: $U=393.0$, $Z=-2.7$, $N=81$, $p < 0.01$, 18.8 ms difference; /a-ɪ/: $U=347.0$, $Z=-3.3$, $N=78$, $p < 0.001$, 41.8 ms difference). Note that while median values for the middle-age group were elevated at the second boundary, this did not result in significantly different performance from the older listeners.

Results suggest that the age-group differences noted previously across temporal-order vowel-identification tasks (Fogerty *et al.*, 2010) may be related to temporal-processing declines that influence the perception of vowel multiplicity, that is, the presence of one versus two overlapping vowels. At very short SOAs, under 20 ms, older adults have more difficulty perceiving the presence of two concurrent vowels. In addition, at longer SOAs, middle-aged and older adults had more difficulty identifying brief silent gaps between vowels. Given that all groups were equally successful at identifying the vowels in these sequences (see Fogerty *et al.*, 2012), and that temporal offset

Fig. 2. Median SOA values for the two boundary judgments using the two vowel pairs. B1= boundary 1, multiplicity; B2= boundary 2, gap-identification. Error bars= inter-quartile range. Asterisks indicate significant differences.

vowel categorization is decreased for older adults, differences in temporal processing, not vowel identification, likely resulted in the temporal-order group differences previously observed (Fogerty *et al.*, 2010). Furthermore, temporal asynchrony provides multiplicity cues at significantly smaller SOAs for all age groups ($p < 0.05$; 53% smaller for middle-aged and older listeners, 17% smaller for young listeners) than required for the identification of monaural asynchronous vowel-pairs for these same listeners (reported earlier by Fogerty *et al.*, 2010). This is in contrast to de Cheveigné (1999), who argued that multiplicity cues are not independent of vowel identity when investigating fundamental frequency differences.

3.2 Individual differences among the older listeners

Given the temporal processing requirements imposed by the categorization task, it is likely that individual differences in temporal processing were related to listener performance. First, the relationship between multiplicity and gap-identification judgments was explored. Correlations across vowel pairs at the same boundary ($r = 0.61$ – 0.67 , $p < 0.01$) were stronger than correlations across boundaries within the same vowel pair ($r = 0.30$ – 0.35 , $p < 0.01$). The moderately strong correlations across vowel pairs suggest generalization of these results beyond one specific vowel pair. However, the relatively weak correlations between boundaries within the same vowel pair suggest two different auditory processes may be involved in these decisions of multiplicity and gap-identification.

Second, of interest here is whether individual differences in performance on a host of auditory processing tasks would predict multiplicity and gap-identification judgments of temporally offset concurrent vowel perception. Analyses of the young and middle-aged groups demonstrated no significant correlations. Therefore, analyses reported here focus only on the older listeners. Correlations were computed between the boundary SOA values and average thresholds obtained for gap detection, monaural and dichotic vowel-identification temporal order, and temporal masking tasks reported earlier (Fogerty *et al.*, 2010; Humes *et al.*, 2010). All measures, except the temporal-masking measures, were correlated in some way with the boundaries measured here. Correlations were slightly more robust for the /i-ε/ vowel pair ($r = 0.29$ – 0.46 , $p < 0.05$ for /i-ε/ versus $r = 0.27$ – 0.35 , $p < 0.05$ for /a-i/), the vowel pair which resulted in the most errors when identifying the vowels presented at threshold SOA values (see Fogerty *et al.*, 2012). Overall, these significant correlations were weak to moderate indicating that, on an individual basis, temporal-processing measures did not account for much variability in the categorization boundaries or vice versa.

The temporal processing measures most representative of associations with boundary categorization were gap detection, monaural temporal order, and dichotic temporal order. A correlational analysis investigated the relationship between these temporal measures, age, and pure-tone averages (0.5, 1, 2 kHz) with boundary thresholds, using SOAs averaged across vowel pair separately for the two boundaries. This analysis indicated that the multiplicity judgment was correlated with monaural temporal order ($r = 0.56$, $p < 0.01$) and dichotic temporal order ($r = 0.31$, $p < 0.01$) while the gap-identification judgment was correlated with gap detection for a narrow band at 1000 Hz ($r = 0.27$, $p < 0.05$) and age ($r = 0.35$, $p < 0.01$). Thus, gap-identification judgments decline with age among the oldest listeners, even after controlling for pure-tone averages ($r = 0.32$, $p < 0.01$). Finally, pure-tone averages were not significantly correlated with either judgment boundary.

In summary, this correlational analysis suggests a moderately strong relationship between temporal order ability and multiplicity judgment, while the relationship between gap detection in noise with the concurrent vowel gap-identification judgment was weak, albeit significant. Because the temporal-order identification task described in Fogerty *et al.* (2010) required identification of vowels in sequence and the vowels could be temporally overlapping, it is reasonable that performance on this task for the older adults would be related to the multiplicity boundary. Likewise, because the gap-

detection task measured the minimum detectable silent period between two bursts of noise it is reasonable that performance on this task for older adults would be related to the gap-identification boundary for these same listeners. It is interesting, however, that no such relationships among tasks were observed for the young or middle-aged listeners. This could be due to the smaller samples of these two age groups or the narrower range of individual differences observed for these measures in these two age groups.

4. Summary and conclusions

Analyses of group differences demonstrated different SOA boundaries for the categorization of concurrent vowels as one versus two vowels (multiplicity) or as two vowels separated by a gap (gap-identification). Non-parametric statistical analysis indicated that older listeners require larger temporal onset asynchronies between concurrent vowels to categorize the number of vowels present, most notably for the vowel pair that had the most identification confusions (see Fogerty *et al.*, 2012). In addition, both middle-aged and older listeners demonstrated a need for increased gap durations between vowels to categorize the presence of the gap, indicating an earlier (i.e., middle-age) onset of gap-identification declines.

Individual differences in temporal processing were only predictive of categorization judgments by the older adults. Temporal processing abilities of young and middle-aged adults were not predictive of concurrent vowel judgment. While aging has been associated with declines in temporal processing tasks (Humes *et al.*, 2009, 2010), the present investigation focused on the underlying sources of these difficulties. For older adults, correlations with a host of temporal processing tasks indicated that multiplicity judgments were most related to temporal-order judgments, while gap-identification judgments for concurrent vowels were related to gap-detection in noise. These results indicate that the age-related declines in temporal processing have two sources. Concurrent vowel judgments for gap-identification appear to involve age-related declines that begin during middle-age and persist among the oldest listeners. In contrast, age-related declines in multiplicity judgments were not observed until after middle-age.

Acknowledgments

The authors would like to thank Dana Kinney and numerous research assistants that assisted in the collection of the data for this project. The study was supported, in part, by the National Institute on Aging Grant No. R01 AG022334 awarded to Larry E. Humes.

References and links

- ANSI. (2010). S3.6–2010, “Specifications for audiometers” (American National Standards Institute, New York).
- Arehart, K. H., Souza, P. E., Muralimanohar, R. K., and Miller, C. W. (2011). “Effects of age on concurrent vowel perception in acoustic and simulated electroacoustic hearing,” *J. Speech Lang. Hear. Res.* **54**, 190–210.
- de Cheveigné, A. (1999). “Vowel-specific effects in concurrent vowel identification,” *J. Acoust. Soc. Am.* **106**, 327–340.
- Fogerty, D., Humes, L. E., and Kewley-Port, D. (2010). “Auditory temporal-order processing of vowel sequences by young and older adults,” *J. Acoust. Soc. Am.* **127**, 2509–2520.
- Fogerty, D., Kewley-Port, D., and Humes, L. E. (2012). “Asynchronous vowel pair identification across the adult lifespan for monaural and dichotic presentations,” *J. Speech Lang. Hear. Res.* **55**, 1–13.
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). “Mini-Mental State: A practical method for grading the cognitive status of patients for the clinician,” *J. Psychiatr. Res.* **12**, 189–198.
- Gordon-Salant, S., and Fitzgibbons, P. (1993). “Temporal factors and speech recognition performance in young and elderly listeners,” *J. Speech Hear. Res.* **36**, 1276–1285.
- Humes, L. E., Busey, T. A., Craig, J. C., and Kewley-Port, D. (2009). “The effects of age on sensory thresholds and temporal gap detection in hearing, vision, and touch,” *Atten. Percep. Psychophys.* **71**, 860–871.
- Humes, L. E., Kewley-Port, D., Fogerty, D., and Kinney, D. (2010). “Measures of hearing threshold and temporal processing across the adult lifespan,” *Hear. Res.* **264**, 30–40.

- Kawahara, H., Masuda-Katase, I., and de Cheveigné, A. (1999). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous frequency-based F0 extraction: Possible role of a repetitive structure in sounds," *Speech Commun.* **27**, 187–207.
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1992). "Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss," *J. Acoust. Soc. Am.* **92**, 1923–1932.
- Salthouse, T. A. (1996). "The processing-speed theory of adult age differences in cognition," *Psychol. Rev.* **103**, 403–428.
- Snell, K. B., and Frisina, D. R. (2000). "Relationships among age-related differences in gap detection and word recognition," *J. Acoust. Soc. Am.* **107**, 1615–1626.