

Vowel identification by younger and older listeners: Relative effectiveness of vowel edges and vowel centers

Gail S. Donaldson,^{a)} Elizabeth K. Talmage, and Catherine L. Rogers

*Department of Communication Sciences and Disorders, University of South Florida, PCD 1017,
4202 E. Fowler Avenue, Tampa, Florida 33620
gdonalds@usf.edu, etalmage@mail.usf.edu, crogers2@usf.edu*

Abstract: Young normal-hearing (YNH) and older normal-hearing (ONH) listeners identified vowels in naturally produced /bVb/ syllables and in modified syllables that consisted of variable portions of the vowel edges (silent-center [SC] stimuli) or vowel center (center-only [CO] stimuli). Listeners achieved high levels of performance for all but the shortest stimuli, indicating that they were able to access vowel cues throughout the syllable. ONH listeners performed similarly to YNH listeners for most stimuli, but performed more poorly for the shortest CO stimuli. SC and CO stimuli were equally effective in supporting vowel identification except when acoustic information was limited to 20 ms.

© 2010 Acoustical Society of America

PACS numbers: 43.71.Es, 43.71.Lz [JH]

Date Received: May 25, 2010 **Date Accepted:** July 6, 2010

1. Introduction

It is well established that normal-hearing native listeners can identify vowels in CVC syllables on the basis of either the dynamic formant transitions that occur at the edges of the vowels, or the quasi steady-state formants that occur at the centers of the vowels, at least when the entire transitions or centers are provided (Strange *et al.*, 1983; Jenkins *et al.*, 1983; Andruski and Neary, 1992). However, less is known about how well these listeners can identify vowels when only a *portion* of the formant transition or vowel center is available. Furthermore, it is not clear whether the two types of cues are equally effective in supporting vowel identification (i.e., whether similar identification performance is achieved when listeners are provided with equal durations of acoustic information taken from the vowel edges or the vowel center). Accordingly, the primary purpose of this study was to evaluate vowel identification in a group of young normal-hearing listeners using /bVb/ syllables that were modified to preserve varying durations of the two types of cues.

A second purpose of the study was to provide preliminary data concerning the possible effects of aging on the relative effectiveness of vowel edges and vowel centers for vowel identification. Several studies have reported that aging preferentially reduces the ability to process rapidly-changing cues to vowel identification, suggesting that older listeners may rely more strongly on quasi steady-state cues located in the vowel centers than on formant transitions located in the vowel edges (Dorman *et al.*, 1985; Elliott *et al.*, 1989; Fox *et al.*, 1992; although cf. Ohde and Abou-Khalil, 2001). The present study sought to further examine this issue by directly comparing older listeners' use of vowel edges versus vowel centers in a vowel identification task that employed naturally produced stimuli.

^{a)} Author to whom correspondence should be addressed.

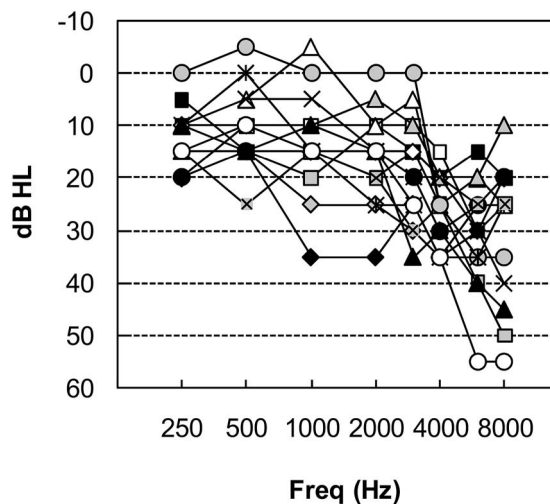


Fig. 1. Better-ear hearing thresholds of the ONH listeners.

2. Subjects and methods

Two groups of subjects were tested. The younger normal-hearing (YNH) group included 12 subjects, aged 18–28 years, with normal hearing sensitivity as defined by pure tone thresholds ≤ 20 dB HL between 250 and 8000 Hz. The older normal-hearing (ONH) group included 15 subjects, aged 56–78 years (mean=64.9 yrs, median=65.5 yrs), with normal or near-normal hearing sensitivity in the better hearing ear as defined by pure tone thresholds ≤ 25 dB HL at 250 Hz and 500 Hz, and ≤ 35 dB HL between 1000 Hz and 4000 Hz. Figure 1 shows individual ONH subjects' better-ear thresholds. All subjects were native speakers of American English.

Study procedures were approved by the University of South Florida Institutional Review Board and each subject provided informed consent prior to participation. Subjects were paid on an hourly basis for their participation.

Stimuli were naturally spoken exemplars of six /bVb/ syllables: “beeb, bib, babe, beb, bab” and “bob.” They were selected from a larger set of stimuli previously recorded and digitized (44.1 kHz sampling rate, 16 bit A/D converter) by Rogers and Lopez (2008). For the present study, three exemplars from each of two male talkers were used. The durations of the vowels in these stimuli ranged from 211 to 353 ms for Talker 1 and from 173 to 381 ms for Talker 2.

The original (FULL) syllables were modified (Syntrillium, Inc., 2000) to create three additional sets of stimuli: (1) GAP20 stimuli. A 20 ms segment of the vowel center was reduced to silence, creating a 20 ms gap at the temporal center of the syllable. This was a control condition to determine whether a brief interruption in the stimulus would degrade performance even when there was minimal loss of acoustic information. (2) Silent-center (SC) stimuli. A variable duration of the vowel center was reduced to silence, preserving 80, 60, 40, 30, 20 or 10 ms of the syllable following the vowel onset and an equal duration of the syllable preceding the final consonant closure. The duration of the silence was lengthened or shortened to equate the overall duration of each SC stimulus to the average duration of the FULL stimuli. Typically, the shorter SC stimuli (those that included 40 ms or less of each edge of the vowel) included only the formant transitions; longer SC stimuli also included the temporal edges of the vowel center. (3) Center-only (CO) stimuli. The consonant-vowel and vowel-consonant transitions were deleted, preserving 100, 80, 60, 40 or 20 ms of the vowel center. Figure 2 shows example stimuli for the syllable “beb,” spoken by a single male talker.

Four pairs of SC and CO stimuli having equal total durations of vowel information were used to assess the relative effectiveness of vowel edges and vowel centers in supporting

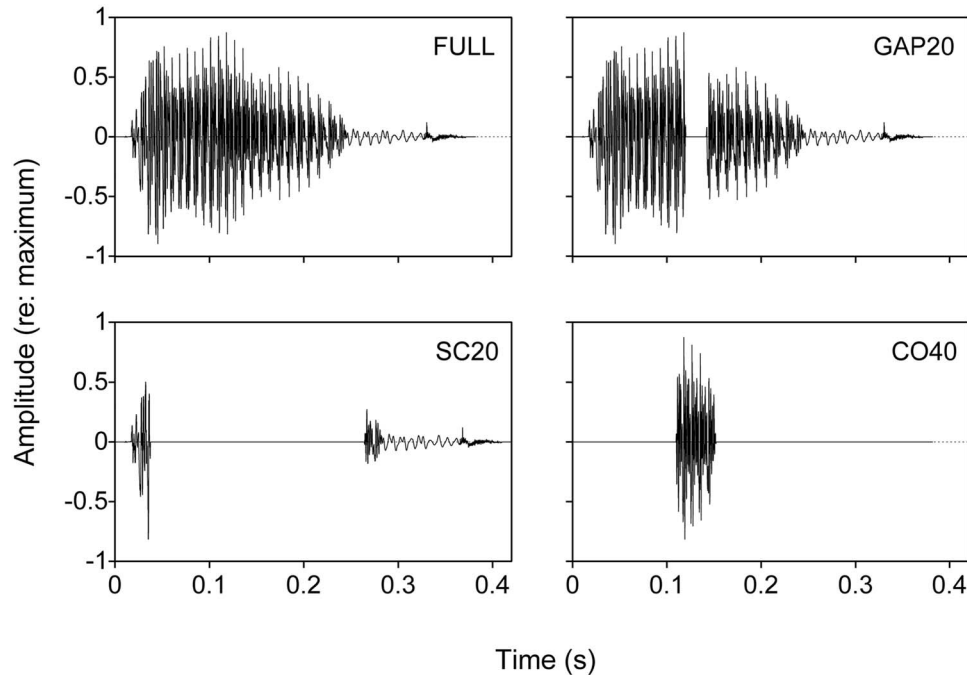


Fig. 2. Example stimulus waveforms for the syllable “bed” spoken by a male talker. The SC20 and CO40 stimuli (bottom row) provide similar total durations of acoustic information (40 ms) from the initial and final vowel edges (SC20) or the vowel center (CO40).

vowel identification: SC10 and CO20; SC20 and CO40; SC30 and CO60; and SC40 and CO80. One such pair (SC20 and CO40) is illustrated in the bottom two panels of Fig. 2.

Vowel identification was tested using a 6AFC vowel-confusion procedure controlled by custom scripts written for EPrime (v1.1, Psychology Software Tools, Inc., 2002) running on a personal computer. Stimuli were played out from a LynxOne sound card, amplified (Crown D75), attenuated (Tucker-Davis Technologies PA-5) and presented through a high-quality loudspeaker (Spendor S3/5se). The subject was seated in a double-walled sound booth approximately 1 m in front of the speaker. A list of the six possible syllables appeared on the computer monitor, and the subject used a computer mouse to make his or her selection after each stimulus presentation. No feedback was provided.

FULL stimuli were presented at an average level of 70 dBA; other stimuli were presented using the same amplification and attenuation settings as the FULL stimuli. This resulted in some stimuli (the SC stimuli and shortest CO stimuli) being perceived as softer than others. For this reason, listening checks were performed with each subject prior to testing to ensure that all stimuli were clearly audible. To prevent subjects from using loudness differences to identify syllables within a given condition, stimulus intensity was selected randomly on each trial from a linear distribution that varied ± 5 dB from the nominal intensity.

Data were collected using a fixed-block design in which a single stimulus condition was tested in each block. This design was selected because the inclusion of multiple stimulus conditions within a given block would increase listener uncertainty, potentially resulting in poorer overall performance (cf., Watson *et al.*, 1976; Liu and Kewley-Port, 2004). Each test block consisted of 36 stimuli (2 speakers \times 3 tokens \times 6 syllables). Prior to the first test block, each subject completed a practice block using FULL stimuli, during which correct-answer feedback was provided. Following this, three complete sets of data were obtained, with each set consisting of one block of each of the 13 stimulus conditions. The 13 conditions were tested in one of the following two orders, with order counterbalanced across subjects: [FULL, GAP20, SC stimuli, CO

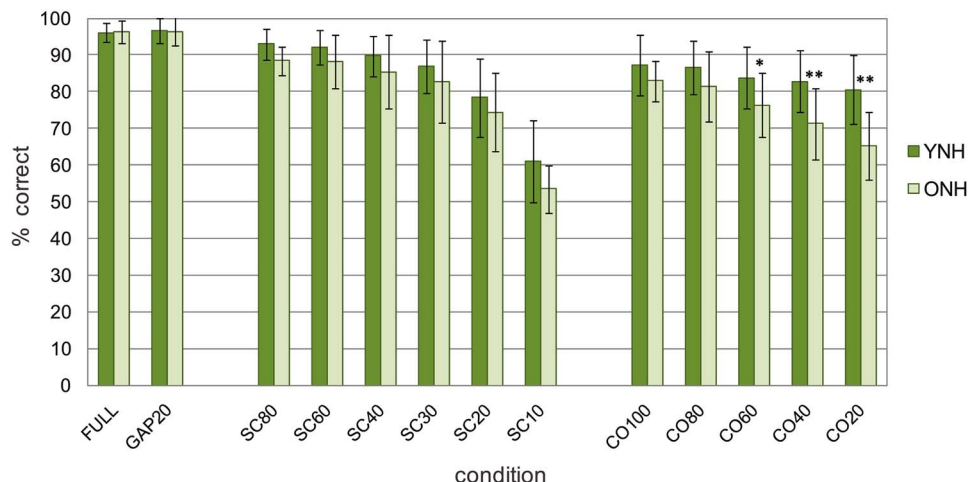


Fig. 3. (Color online) Percent-correct scores for all stimulus conditions. Asterisks indicate stimuli for which ONH listeners scored significantly lower than YNH listeners (*= $p < 0.05$; **= $p < 0.01$).

stimuli] or [FULL, GAP20, CO stimuli, SC stimuli]. In each case, SC and CO stimuli were presented in order of decreasing duration (i.e., easiest to hardest order). Percent-correct scores based on 18 judgments per syllable were converted to rationalized arcsin units (Studebaker, 1985) for purposes of statistical analysis.

3. Results

Overall pattern of performance. Figure 3 shows mean percent-correct scores for all stimulus conditions for both the YNH and ONH listeners. Both groups achieved near perfect performance for the FULL and GAP20 stimuli. Performance was only slightly poorer for the longest-duration SC and CO stimuli, but decreased systematically as the duration of acoustic information in the SC and CO stimuli was reduced.

Effect of age. A two-way (group \times condition) repeated measures ANOVA confirmed a significant main effect of age [$F(1, 300) = 5.9$, $df = 1$, $p < 0.05$], with YNH listeners scoring higher than ONH listeners. However, post-hoc testing indicated that group differences were significant for only the three shortest CO conditions: CO60, CO40 and CO20 (Holm-Sidak pairwise comparisons, $t = 2.4$, $p < 0.05$ for CO60; $t \geq 3.5$, $p < 0.001$ for CO40 and CO20). The largest difference between groups was observed for the CO20 condition, which yielded mean scores of 80.5% for the YNH listeners and 65.1% for the ONH listeners. To determine whether reduced hearing sensitivity could account for ONH listeners' poorer performance for this condition, ONH subjects' vowel identification scores were correlated with hearing thresholds for each of the frequencies represented in Fig. 1 (250 Hz–8 kHz). None of the correlations was significant; however, there was a tendency for higher scores to be associated with better hearing thresholds at 2 kHz ($r = 0.43$, $p = 0.11$). Thus, although it does not appear that hearing sensitivity can account for group differences observed for the shortest CO stimuli, this possibility cannot be ruled out entirely.

Effect of duration. To evaluate the effect of stimulus duration (total duration of acoustic information), separate 2-way (group \times duration) repeated measures ANOVAs were performed for the SC and CO stimuli. For the SC stimuli, there was a significant main effect of duration [$F(5, 125) = 152.9$, $p < 0.001$], but the group \times duration interaction was not significant, indicating that the effect of duration was similar for the YNH and ONH groups. Post-hoc testing (Holm-Sidak pairwise comparisons) indicated that performance did not differ for the three longest-duration stimuli (SC80, SC60 and SC40) and that performance for the SC40 condition was similar to performance for the SC30 condition. Performance levels for the SC20 and SC10 conditions were significantly different from one another and from all longer SC stimuli ($t \geq 6.2$, $p < 0.05$). For the CO stimuli, both the main effect of duration [$F(4, 100) = 35.1$, $p < 0.001$] and the group \times duration

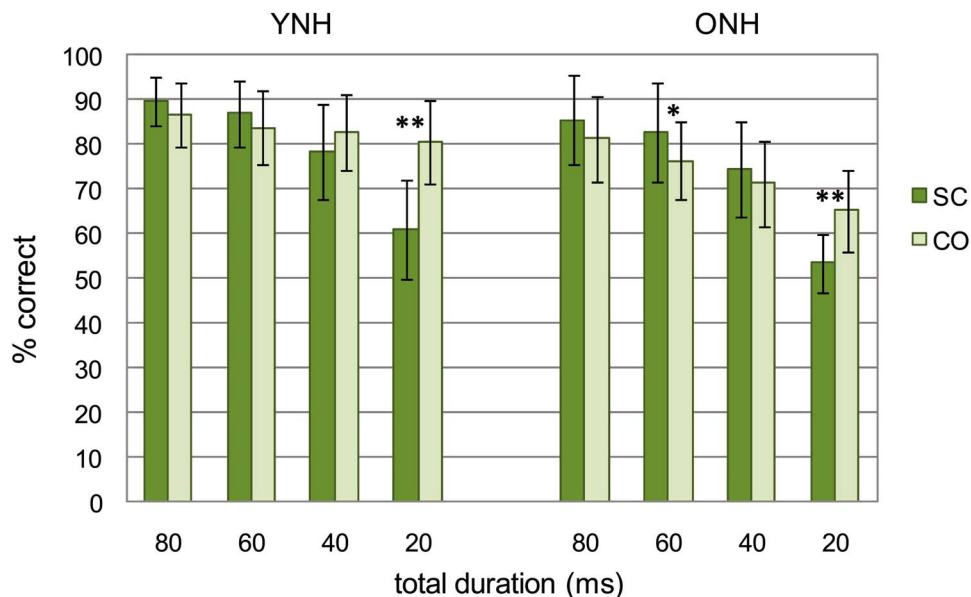


Fig. 4. (Color online) Comparison of SC and CO stimuli having equal total durations. Asterisks indicate pairs of SC and CO stimuli that yielded significantly different scores (*= $p < 0.05$; **= $p < 0.01$).

interaction [$F(4, 100) = 4.8$, $p = .001$] were significant. As is evident from Fig. 3, duration effects were stronger for the ONH subjects than for the YNH subjects. For the ONH subjects, all pairs of CO stimuli except the two longest-duration stimuli (CO100 vs. CO80) differed significantly from each other ($t \geq 5.5$, $p < 0.05$). For the YNH subjects, only three pairs of CO stimuli differed significantly ($t \geq 6.2$, $p < 0.05$): CO100 vs. CO40; CO100 vs. CO20; and CO80 vs. CO20.

Relative effectiveness of vowel edges and vowel centers. Figure 4 compares performance for the four pairs of SC and CO stimuli having equal total durations. For example, the SC40 and CO80 conditions are plotted together because both of these conditions provide a total of 80 ms of acoustic information. Qualitatively, it can be seen that SC and CO stimuli supported similar levels of performance for stimuli having total durations of 40, 60 or 80 ms; however, for stimuli with 20-ms total durations, CO stimuli produced higher scores than SC stimuli. These general findings were confirmed by separate 2-way (duration \times condition) repeated measures ANOVAs for the YNH and ONH subjects. For YNH subjects, both the main effect of condition (SC vs. CO) [$F(1, 33) = 5.9$, $df = 1$, $p < 0.05$] and the duration \times condition interaction [$F(3, 33) = 31.3$, $df = 3$, $p < 0.001$] were significant. Post-hoc testing (Holm-Sidak pairwise comparisons) indicated that CO stimuli produced better performance than SC stimuli for the 20-ms total duration condition only ($t = 8.0$, $p < 0.01$). For ONH subjects, the main effect of condition (SC vs. CO) was not significant, but there was a significant duration \times condition interaction. Post-hoc testing (Holm-Sidak pairwise comparisons) indicated that SC30 produced better performance than CO60 ($t = 2.6$, $p < 0.05$) but that CO20 produced better performance than SC10 ($t = 3.4$, $p < 0.01$).

4. Discussion

Both YNH and ONH listeners maintained substantial levels of vowel identification when listening to silent-center (SC) or center-only (CO) versions of /bVb/ syllables, even when the total duration of acoustic information was as short as 20 ms. This finding implies that acoustic information supporting vowel identification is present throughout the syllable and can be accessed by listeners even during brief glimpses of the edges or centers of the vowel. Except at very short durations (<40 ms of total vowel information), SC and CO segments appear to support equal levels of vowel recognition when equated for overall duration.

ONH listeners performed similarly to YNH listeners for the SC stimuli and the longer-duration CO stimuli, but performed more poorly than YNH listeners for the shorter-duration CO stimuli. Reasons for this are unclear. There remains some possibility that reduced hearing sensitivity among ONH listeners contributed to their relatively poorer performance for shorter CO stimuli; however, correlations between hearing sensitivity and performance did not reach significance in our subject sample.

Fox *et al.* (1992) reported that ONH listeners achieved lower word recognition scores than YNH listeners when vowel centers were removed from CVC monosyllables. However, the performance difference between their groups was only 7.2 percentage points (85.3% for YNH listeners vs. 78.1% for ONH listeners). The mean durations of consonant-vowel and vowel-consonant transitions in Fox *et al.*'s stimuli were 44 ms and 52 ms, respectively; thus, their data are most comparable to our SC40 and SC60 conditions. Average scores for these conditions combined were 90.9% for YNH listeners and 86.8% for ONH listeners, representing a 4.1 percentage point deficit for ONH listeners. This finding is quite similar to the 7.2 percentage point difference observed by Fox *et al.*, especially considering other differences between the studies; however, differences were not statistically significant in the present study, leading to the conclusion that age did not influence results for those conditions.

5. Conclusions

- (1) Normal hearing listeners are able to identify vowels in CVC syllables on the basis of either the vowel edges or the vowel centers, even when only brief portions of edges or centers are provided.
- (2) Older listeners have more difficulty than younger listeners when provided with only brief acoustic information from the vowel center; however, they perform similarly to younger listeners when provided with longer durations of the vowel center, or varying durations of the vowel edges.
- (3) Vowel edges and vowel centers that are matched for total acoustic duration are equally effective in supporting the identification of vowels in CVC syllables, so long as total duration is approximately 40 ms or longer.

Acknowledgment

Partial support for this research was provided by NIH-NIDCD Grant No. 5R03DC005561.

References and links

- Andruski, J. E., and Neary, T. M. (1992). "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables," *J. Acoust. Soc. Am.* **91**, 390–410.
- Dorman, M. F., Marion, K., Hannley, M. I., and Lindholm, J. M. (1985). "Phonetic identification by elderly normal and hearing-impaired listeners," *J. Acoust. Soc. Am.* **77**, 664–670.
- Elliott, L. L., Hammer, M. A., Scholl, M. E., and Wasowicz, J. M. (1989). "Age differences in discrimination of simulated single-formant frequency transitions," *Percept. Psychophys.* **46**, 181–186.
- Fox, R. A., Wall, L. G., and Gokcen, J. (1992). "Age-related differences in processing dynamic information to identify vowel quality," *J. Speech Hear. Res.* **35**, 892–902.
- Jenkins, J. J., Strange, W., and Edman, T. R. (1983). "Identification of vowels in 'vowelless' syllables," *Percept. Psychophys.* **34**, 441–450.
- Liu, C., and Kewley-Port, D. (2004). "Vowel formant discrimination for high-fidelity speech," *J. Acoust. Soc. Am.* **116**, 1224–1233.
- Ohde, R. N., and Abou-Khalil, R. (2001). "Age differences for stop-consonant and vowel perception in adults," *J. Acoust. Soc. Am.* **110**, 2156–2166.
- Rogers, C. L., and Lopez, A. S. (2008). "Perception of silent-center syllables by native and non-native English speakers," *J. Acoust. Soc. Am.* **124**, 1278–1293.
- Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of coarticulated vowels," *J. Acoust. Soc. Am.* **74**, 695–705.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Syntrillium, Inc. (2000). COOLEDT 2000 (Version 1.1).
- Watson, C. S., Kelly, W. J., and Wronton, H. W. (1976). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," *J. Acoust. Soc. Am.* **60**, 1176–1186.