The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss

Daniel Fogerty,^{a)} Diane Kewley-Port, and Larry E. Humes

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

(Received 25 August 2011; revised 25 June 2012; accepted 26 June 2012)

This study investigated the ability to use cues contained within vowel and consonant segments by older listeners with normal or impaired hearing. Spectral shaping restored audibility for the hearing-impaired group. Word and sentence materials were processed to contain primarily consonants or vowels by replacing segments with low-level speech-shaped noise. The proportion of the total duration of preserved speech was varied by manipulating the amount of transitional information contained within vowel and consonant segments. Older listeners performed more poorly than young listeners on all conditions except when listening to sentences with only the vowels preserved. Results confirmed a greater contribution to intelligibility of vowel segments in sentences, but not in words, for young normal-hearing, older normal-hearing, and older hearing-impaired listeners. Older listeners received a greater benefit than young listeners from vowels presented in a sentence context. Correlation analyses among the older listeners demonstrated an association between consonant and vowel performance in isolated words but not in sentences. In addition, the use of vowel cues in sentences was relatively independent of age and auditory sensitivity when audibility is ensured. Combined, results argue that older listeners are able to use essential cues carried by vowels for sentence intelligibility. © 2012 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4739463>]

PACS number(s): 43.71.Gv, 43.71.Es, 43.71.Lz, 43.71.Ky [PBN]

Pages: 1667–1678

I. INTRODUCTION

Audibility accounts for the majority, but not all, of age differences in speech understanding performance (Humes, 2007). It may be that declines in older listener speech understanding performance are also associated with declines in processing certain informative speech cues. The examination of the relative importance of vowel and consonant information in older adults is of considerable clinical relevance. About two-thirds of the hearing aids sold in the U.S., for example, are purchased by older adults (e.g., Strom, 2006). Once the relative importance of vowel and consonant information in older adults is established, signal-processing strategies may be tailored to emphasize the cues underlying the most important segments.

The current study examined the contribution of acoustic cues present during consonants or vowels to word recognition in isolation and in sentences. Young normal-hearing (YNH), older normal-hearing (ONH), and older hearing-impaired (OHI) listeners who received spectrally-shaped speech were tested to examine the effect of age and cochlear pathology on the use of consonant and vowel cues.

Fundamentally different acoustic cues are carried by consonants and vowels (Ladefoged, 2001; Stevens, 2002). Whereas consonants are characterized by vocal tract constriction, high frequency components, and often aperiodicity, vowels are characterized by sustained voicing, lack of

constriction, and a dominant lower frequency formant structure. Acoustic cues present during these segments are responsible not only for the identity of the segment itself but also carry potential cues regarding the identity of neighboring segments (e.g., Cooper *et al.*, 1952; Strange *et al.*, 1983), suprasegmental cues regarding the entire utterance (Lehiste, 1970), and supralinguistic indexical cues about the talker (Owren and Cardillo, 2006). Acoustic cues present during consonants and vowels likely provide different information for spoken word recognition and are able to carry information beyond segment identity to various degrees.

Greater contributions of vowels compared to consonants have now been supported by a number of studies that have used noise replacement to selectively present either consonants or vowels within the sentence (Cole *et al.*, 1996; Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009; Fogerty and Humes, 2012) but not in isolated words (Owren and Cardillo, 2006; Fogerty and Humes, 2010). These findings are not specific to noise replacement, having also been confirmed using silence or a harmonic complex during the replaced segment (Cole *et al.*, 1996; Owren and Cardillo, 2006; Fogerty and Humes, 2010). The acoustic cues of vowels also appear to be context-dependent, facilitating the recognition of words in sentences more than when presented in words in isolation (Fogerty and Humes, 2010, 2012). Temporal properties of the amplitude envelope appear to be at least partially responsible for this effect (Fogerty and Humes, 2012) and are perceptually weighted more than temporal fine structure in sentences (Fogerty, 2011). Restoring information about the overall amplitude level of each individual vowel in the sentence by dynamically scaling the replacement noise may facilitate sentence recognition (Cole

^{a)}Author to whom correspondence should be addressed. Current address: University of South Carolina, Department of Communication Sciences and Disorders, 1621 Greene St., Columbia, SC 29208. Electronic mail: fogerty@sc.edu

et al., 1996; Stilp and Kluender, 2010). In addition to different acoustic contributions, consonants and vowels appear to have different functional roles in language processing (Caramazza *et al.*, 2000; Nespor *et al.*, 2003; New *et al.*, 2008; Toro *et al.*, 2008; Carreiras and Price, 2008; Carreiras *et al.*, 2009).

Vowel and consonant cues are not discrete but rather are distributed across the consonant-vowel boundary (e.g., Cooper *et al.*, 1952; Strange *et al.*, 1983). Shifting the consonant-vowel boundary to include more or less segmental information has little effect on the relative importance for consonants and vowels in words (Fogerty and Humes, 2010) or sentences (Fogerty and Kewley-Port, 2009), which appears to underscore the somewhat arbitrary, but “convenient,” division the consonant-vowel boundary provides (Ladefoged, 2001). However, shifting the consonant-vowel boundary has the effect of altering the duration of consonants or vowels, which may have significant perceptual implications for older listeners. As aging generally has a negative impact on the ability of listeners to process rapid and brief acoustic signals (e.g., Fitzgibbons and Gordon-Salant, 2004; Shrivastav *et al.*, 2008), it was predicted that older listeners would have reduced word recognition scores for segments with the shortest segment durations, particularly for consonants that are characterized by rapid acoustic changes. In this study, the consonant-vowel boundary was shifted to provide four different proportions of the total duration (PTD) for the words or sentences while equating the total preserved speech duration provided by consonant and vowel conditions.

Finally, correlations were calculated among the measures to determine association between older listeners’ ability to use vowel and consonant information. Thus, this analysis provided a preliminary examination of processing differences between the use of consonant and vowel cues in words and sentences, as well as speech-in-noise recognition more generally.

II. METHODS

A. Listeners

Twenty-four YNH ($M = 21$ yr, range = 18–24 yr), 20 ONH ($M = 71$ yr, range = 62–83 yr), and 20 OHI ($M = 74$ yr, range = 66–84 yr) listeners were paid to participate in the experiment. Older listeners were required to pass (scores ≥ 25) the Mini-Mental State Exam (Folstein *et al.*, 1975). For OHI listeners, maximum hearing thresholds for air conducted pure-tones were not to exceed the following limits in at least one ear: 55 dB hearing level (HL) at 0.25 kHz; 60 dB HL at 0.5, 1, 2, and 4 kHz; and 85 dB HL at 6 and 8 kHz. ONH listeners had hearing thresholds no worse than 20 dB HL for octave frequencies from 0.25–2 kHz and not worse than 30 dB HL at 4 kHz. YNH listeners had hearing thresholds at or below 20 dB HL at all octave frequencies. It was also required that there be no evidence of middle ear pathology (i.e., air-bone gaps < 10 dB and normal tympanograms). Table I lists the average audiometric thresholds (also see Fig. 2) and standard deviations for these three listener groups in the ear that was tested. As can be observed from Table I, OHI listeners had sloping losses, as did ONH listeners in the

TABLE I. Mean (and standard deviation) audiometric thresholds (dB HL) in the test ear for the three listener groups.

Frequency (kHz)	YNH	ONH	OHI
0.25	9 (7)	13 (8)	20 (10)
0.5	11 (7)	13 (6)	21 (9)
1.0	8 (5)	11 (6)	20 (12)
2.0	8 (5)	11 (8)	28 (13)
3.0	6 (4)	14 (9)	40 (12)
4.0	5 (5)	20 (9)	49 (6)
6.0	7 (4)	29 (15)	56 (13)
8.0	6 (5)	34 (19)	59 (16)

highest frequencies tested. The right ear was tested by default unless a better match to the hearing criteria was obtained with the left ear (three ONH and two OHI listeners). Note that, above 3 kHz, the hearing sensitivity of the ONH subjects is considerably poorer than that of the YNH subjects.

YNH and ONH groups listened to acoustically equivalent (i.e., the same) speech presentations. The effect of audibility was controlled in the OHI group by providing these listeners with spectral shaping of the stimuli that restored audibility of the speech signal through 4 kHz. According to the Speech Intelligibility Index (SII; ANSI, 1997), maximum speech intelligibility is obtained when the root-mean-square (rms) speech spectrum is 15 dB above the listener’s hearing threshold, which this spectral shaping procedure ensured (through at least 4 kHz). This shaping is also characteristic of a common clinical prescriptive amplification procedure, the Desired Sensation Level approach (Seewald *et al.*, 1993). Comparisons between YNH and ONH groups allowed for the examination of age differences and comparisons between ONH and OHI allowed for the examination of the effects of cochlear pathology, when audibility has been restored.

B. Stimuli and design

Forty-two sentences were selected from the TIMIT database (Garofolo *et al.*, 1990, www.ldc.upenn.edu). Each sentence was spoken by a different talker from the North Midland dialect region (21 male and 21 female talkers). CVC words ($N = 148$) were selected from recordings of one male talker by Takayanagi *et al.* (2002) reported to be of General American dialect. These sentence and CVC materials have been used in previous noise-replacement studies (Fogerty and Kewley-Port, 2009; Fogerty and Humes, 2010, 2012). All words and sentences were normalized in rms amplitude and presented at a level of 70 dB sound pressure level (SPL) prior to noise replacement and spectral shaping for the OHI group. This ensured that intrinsic level differences between consonants and vowels, present in natural speech, were preserved for the two normal-hearing groups.

Segmental boundaries were previously specified in the TIMIT database by experienced phoneticians and the CVC words used here modeled these boundary marking rules (Zue and Seneff, 1988; Fogerty and Humes, 2010). The full stop closure and burst was assigned to the consonant. Vowels followed by /r/ were treated as single rhoticized vowels. These

segmental boundaries were shifted according to the assigned boundary proportion and adjusted to within 1-ms of the nearest local amplitude minima (i.e., zero-crossing) to minimize the introduction of transients.

Speech materials were segmented at the original C-V boundary (Orig), shifted into the vowel by 15% of the vowel duration (15%VP), shifted into the consonant by 15% of the consonant duration (15%CP), or 30% of the consonant duration (30%CP). These shift conditions allowed for pairings of consonant and vowel conditions at equal preserved speech durations. Figure 1 displays a schematic of these four conditions where the stippled bars show where speech was replaced by noise. These proportional shifts of the boundary provided for the same four values of the PTD of speech presented during the predominant vowel and the predominant consonant segments (PTD = 0.38, 0.46, 0.54, 0.62). These nominal values are averages across all materials. Averaged PTD values for words and sentences were within 0.02. PTD was controlled because other studies of temporally interrupted speech with acoustically rather than linguistically defined interruption parameters have found PTD to be the key parameter for determining word recognition in isolation (e.g., Wang and Humes, 2010) and in sentence contexts (Kidd and Humes, 2012). Similar proportional shifts were used previously (Fogerty and Kewley-Port, 2009; Fogerty and Humes, 2010). Five ONH, five OHI, and six YNH listeners were randomly assigned to each of these four boundary proportions. All listeners were tested with all of the consonants or vowels replaced, using the assigned boundary condition, for isolated word and sentence materials. Thus, this study used a 3 (listener group) $\times$ 4 (boundary proportion) $\times$ 2 (segment replaced) $\times$ 2 (context) design with listener group and boundary proportion as between-subjects factors and the others as repeated-measures factors.

Speech-shaped noise was created, scaled to -16 dB relative to the speech level, and used for replacement of the segments. This noise level was selected to be below the average level for both consonants and vowels and served to assist

in continuity of the speech sample while avoiding phonemic restoration effects. A single noise replacement level was used to avoid providing information regarding the missing segment amplitude. The noise level used in replacement does not influence observed consonant and vowel results for isolated words (Fogerty and Humes, 2010). Furthermore, although Kewley-Port *et al.* (2007) used different noise levels for consonant and vowel replacement, Fogerty and Kewley-Port (2009) obtained similar results using a standard noise level for all replacements: -16 dB relative to the speech level, the same level used in the current study. Cole *et al.* (1996) also found similar results for sentences with silence and noise-replaced conditions.

Two different noise spectra were created, one used for CVC words and one used for the TIMIT sentences. These noises were processed identically, except that one was spectrally shaped to match the long term average spectrum of a concatenation of all CVC words, while the other matched that of the TIMIT sentences. Resulting noise spectra were similarly shaped except for greater intensity of low frequency speech components <200 Hz for the word materials. A unique noise was used for all replacement intervals within a given sentence or word. Consonant conditions preserved all consonants while replacing the vowels with noise. Vowel conditions preserved all vowels and replaced consonant segments with noise. Note that in both cases preserved segments contained some vowel and some consonant cues, the amount associated with the boundary proportion conditions. However, in all cases, these conditions contained predominantly consonant or vowel cues, as indicated in Fig. 1.

All listeners also completed testing on full, unprocessed sentences and words. Novel sentences were used for this full-sentence condition. For the full-word condition, CVC words were repeated in a second presentation after segmental testing. To avoid ceiling performance and obtain a measure of subject variability, these full materials were presented in continuous noise (i.e., the same type of speech-shaped noise used during replacement) at a 0 dB signal-to-noise ratio (SNR). The use of continuous noise for the full words and sentences is in contrast to the segment-replaced words and sentences that were presented in quiet.

C. Procedures

All participants were tested alone in a sound-attenuating booth. Stimuli were presented using Tucker-Davis Technologies System III hardware and passed through a headphone buffer (HB-7) to an ER-3 A insert earphone. Presentation levels were calibrated by presenting the speech-shaped noises matching the CVC and sentence rms and spectra at a sound level of 70 dB SPL using an HA-2 2-cc coupler and a Larson Davis model 2800 sound level meter with linear weighting. Speech materials were presented monaurally. Spectral shaping applied to the stimuli for each OHI listener ensured that the speech spectrum for words and sentences was presented at least 15 dB above each listener's hearing threshold from 100–4000 Hz in 1/3 octave bands. Thus, spectral shaping was applied on an individual basis. The overall presentation level for this group of listeners was estimated at 82 dB SPL

FIG. 1. Schematic of a CVC for consonant and vowel replacement at the four boundary proportions tested. PTD presented is displayed on the left. Black bars = consonant preservation; Gray bars = vowel preservation; Stippled bars = noise replacement.

FIG. 2. Speech spectrum measured in 1/3 octave bands for the TIMIT sentences presented at 70 dB SPL (thin solid line) and after spectral shaping for the average hearing-impaired listener (bold solid line). Dashed lines display mean hearing thresholds in the test ear for the three listener groups.

(SD = 2 dB). Figure 2 displays the speech spectrum level for the TIMIT sentences presented at 70 dB SPL prior to shaping, measured in 1/3 octave bands (see the thin solid line). The solid bold line displays this spectrum after shaping was applied, based on the mean hearing thresholds of the OHI group. Mean hearing thresholds for the three listener groups are also displayed as dashed lines.

For sentence testing, participants were instructed to repeat each sentence as accurately as possible. Audio recordings of responses were made for offline analysis. Sentences in vowel, consonant, and full-utterance conditions were presented fully randomized to the listeners. No sentence was repeated for a given listener. For CVC testing, participants typed what they thought they heard on a PC running a MATLAB open-set response interface. All words were presented to the participants in a random order. Each word was presented only once during the segmental testing. Full-utterance CVCs (i.e., without segmental replacement) were presented at 0 dB SNR in a second block to listeners and were a second presentation of the same words tested previously under segmental replacement. Familiarization trials were provided before sentence and word testing, making use of stimuli not used during testing. No feedback was provided during familiarization or testing. Listeners were encouraged to guess what words they heard when responding.

D. Scoring

Sentence responses were scored offline by two trained raters. Inter-rater agreement was previously established on these same materials at 98%. Sentences were scored based on the number of words correctly repeated. Words were required to be repeated exactly (e.g., no missing or additional suffixes) to be scored correct. Typed CVC word responses were automatically corrected for phonetic misspellings, were manually inspected, and were automatically scored using custom-made software. All word percent-correct scores were transformed to rationalized arcsine units to stabilize the error variance prior to analysis (RAU, [Studebaker, 1985](#)).

III. RESULTS

A. Comparison between listener groups for speech-in-noise perception

The mean performance for full-utterance sentences and words in 0 dB SNR noise for the three listener groups is displayed in Fig. 3. The performance difference across the age groups is evident for both words and sentences. Independent-sample *t*-tests of the full-utterance, unprocessed sentences and words demonstrated significant differences between the YNH listener group and both ONH [sentences: $t(42) = 3.4$, $p < 0.01$; words: $t(41) = 4.4$, $p < 0.01$, Bonferroni correction for multiple comparisons] and OHI listener groups [sentences: $t(42) = 4.1$, $p < 0.01$; words: $t(41) = 2.8$, $p < 0.01$]. No difference between ONH and OHI groups was observed for either words or sentences.

B. Consonant and vowel contributions in words

The results for consonant and vowel conditions in words across PTD for each listener group are displayed in Fig. 4(a). The main effect of PTD was first investigated. As listeners were assigned to separate PTD values for consonant and vowel conditions, a one-way analysis of variance was completed separately for the vowel and consonant conditions. All age groups as a whole performed better with increasing proportion of the word presented for both consonant [$F(3,58) = 13.0$, $p < 0.001$] and vowel [$F(3,58) = 19.3$, $p < 0.001$] conditions. The remaining factors were investigated by means of a mixed-model (2 segment types $\times$ 4 proportional shifts $\times$ 3 listener groups) analysis of variance on the word data. Results demonstrated significant main effects for segment [$F(1,50) = 8.4$, $p < 0.01$] and listener group [$F(2,50) = 28.5$, $p < 0.001$]. Interactions with the segment and listener group [$F(2,50) = 5.5$, $p < 0.01$], segment and shift proportion [$F(3,50) = 103.5$, $p < 0.001$], and three-way interaction of segment, listener group, and shift proportion [$F(6,50) = 2.5$, $p < 0.05$] were found. Therefore, *post hoc* comparisons were performed.

First, comparisons among the three listener groups were investigated at each PTD value tested for words. Similar

FIG. 3. Results for full, unprocessed words and sentences presented in noise at 0 dB SNR. Error bars = standard error of the mean.

a) Words

b) Sentences

FIG. 4. Performance of the three listener groups across the PTD for (a) isolated CVC words and (b) sentences. Bold lines display performance for the older listener groups. Short dashed lines display performance for the normal-hearing groups. Error bars = standard error of the mean.

results were obtained for each PTD and are therefore collapsed for summary here. YNH listeners performed significantly better than ONH listeners for the consonant condition [$t(42) = 4.1$, $p < 0.01$] and better than OHI listeners for both vowel [$t(42) = 3.8$, $p < 0.01$] and consonant [$t(42) = 4.1$, $p < 0.01$] conditions. No difference between ONH and OHI listeners was obtained for any condition.

Second, comparisons between consonant and vowel conditions within each listener group were conducted at each PTD value tested. Results demonstrated no significant difference in the recognition of words for the vowel or consonant presentations at the same PTD values for YNH, ONH, or OHI listeners, with one exception out of the 12 comparisons. OHI listeners at the PTD value of 0.46 performed better when hearing words with the vowel preserved than those with the consonant preserved [$t(3) = 8.5$, $p = 0.001$]. Therefore, overall consonant and vowel contributions in isolated CVC words appear to be near equal across the four PTD values tested and for all three listener groups.

C. Consonant and vowel contributions in sentences

Results for consonant and vowel conditions in sentences across PTD for each listener group are displayed in Fig. 4(b). Separate one-way analyses of variance demonstrated a significant improvement in performance with increasing PTD across all listener groups for consonant [$F(3,60) = 39.4$, $p < 0.001$] and vowel [$F(3,60) = 49.2$, $p < 0.001$] conditions. A mixed-model (2 segment types $\times$ 4 proportional shifts $\times$ 3 listener groups) analysis of variance on the sentence data was performed next. Results demonstrated significant main effects for segment [$F(1,52) = 2637.4$, $p < 0.001$], shift

proportion [$F(3,52) = 12.2$, $p < 0.001$], and listener group [$F(2,52) = 17.7$, $p < 0.001$]. All interactions were also significant ($p < 0.001$). As a result, independent-sample t -tests were conducted at each PTD value to compare the identification of sentences from the vowels versus the consonants separately for each listener group. Results demonstrated that vowels, compared to consonants, resulted in significantly better word recognition scores in sentences at all PTDs tested for all three listener groups ($p < 0.001$). Comparisons between listener groups were also conducted at each PTD. OHI listeners performed as well as ONH listeners for consonant sentences at each PTD. However, for sentences preserving the vowels, no significant difference between any of the listener groups was observed at any PTD. As with words, results were consistent across PTD and were collapsed. YNH subjects performed better than ONH [$t(42) = 2.7$, $p < 0.01$] and OHI [$t(42) = 3.0$, $p < 0.01$] listeners for sentences preserving the consonants.

One possibility for the lack of any age differences for vowels is a performance-level effect. It could be that an effect of age is only observable for more difficult, low-performance conditions (i.e., consonant conditions). However, this explanation is unlikely as age differences were obtained for the full sentence condition in nearly the same performance-level range ($M = 68$ RAU) as the vowel condition ($M = 62$ RAU).

D. Differential contributions of vowel and consonant segments for words versus sentences

A comparison of performance for consonant and vowel segments in the word-only and sentence contexts was

examined. Collapsed across PTD and listener group, vowels in sentences resulted in better performance than vowels in words, $t(61)=22.8$, $p<0.001$. This effect was consistent across PTD values tested and for all three listener groups. In contrast, for consonants, the word context in general resulted in better performance, $t(61)=-5.7$, $p<0.001$. Individual paired-sample t -tests within groups at each PTD probed this finding. Significant differences were confined to 3 cases out of 12 that demonstrated better performance for isolated words [YNH at 0.38, $t(5)=-6.4$, $p<0.01$; 0.46, $t(5)=-4.6$, $p<0.01$; and ONH at 0.46, $t(4)=-5.6$, $p<0.01$]. It is unclear why additional contextual support provided by sentences was associated with poorer performance compared to words for stimuli that preserved the consonants. It may be possible, particularly in the few reduced consonant conditions that reached significance (3 of 12 conditions), that feature cues provided by consonants are more prototypical in these isolated words than in sentence contexts that are characterized by more variable productions. That is, temporally-reduced consonants in isolated words may contain a better representation of local phoneme features than do temporally-reduced consonant productions in sentences.

Figure 5 displays the summary of these four conditions averaged across PTD. In summary, for all listener groups, sentence vowels resulted in the best performance, with the other three conditions resulting in poorer performance. The only significant difference observed among the latter three conditions yielding low scores, pooled across PTD, was that sentence consonants resulted in poorer performance than vowels in words for the ONH group.

The contextual benefit provided by sentences over words was also investigated for full utterances presented at 0dB SNR and for stimuli preserving either the vowels or the consonants. Figure 6 displays the benefit in RAU for sentence minus word performance. As patterns were similar across PTD, results are collapsed across PTD here. No difference between the three listener groups in contextual benefit was observed for the full or consonant presentations. However, ONH [$t(42)=2.9$, $p<0.01$] and OHI [$t(42)=4.1$, $p<0.001$] listeners received more benefit from the added sentence context than YNH listeners when the vowels were

FIG. 5. Average performance of the three listeners groups for the four segmental conditions tested. Error bars = standard error of the mean.

FIG. 6. Benefit of sentence context in RAU over performance for isolated words (sentence-word performance) for full utterances (0dB SNR). Negative benefit indicates that word presentations were identified better than sentence presentations. Error bars = 95% confidence interval.

preserved. No difference was observed between the two older groups.

E. Correlational analysis among the older listeners

As OHI and ONH listeners performed similarly on the conditions in this study, they were pooled into a single group of 40 older listeners. The older listeners were divided into different groups corresponding to the four PTD values tested. Subsequently, older listeners' RAU scores were transformed into z-scores to obtain a standard measure of individual differences within each PTD group.

Correlations were calculated between the individual z-scores on the six measures collected here (full, vowel, and consonant conditions in both words and sentences), as well as for age and high-frequency pure-tone averages (HFPTA; averaged across 1, 2, and 4 kHz). Significant correlations among these measures are listed in Table II.

Correlations demonstrated that listening to speech in noise for full words and full sentences is related ($r=0.54$, $p<0.01$), suggesting a general speech-in-noise ability that is

TABLE II. Correlations for 40 older listeners between z-scores obtained on the six experimental tasks, age, and HFPTA.

	Sentences			Words		
	Full	Vowel	Consonant	Full	Vowel	Consonant
HFPTA			-0.40			
Age	-0.46		-0.49	-0.39		-0.34
<i>Sentences</i>						
Full	—	0.48	0.36	0.54		0.38
Vowel	0.48	—				
Consonant	0.36		—			
<i>Words</i>						
Full	0.54			—	0.59	0.61
Vowel				0.59	—	0.64
Consonant	0.38			0.61	0.64	—

Only correlations significant at alpha=0.05 are displayed. Bold indicates $p<0.01$.

not entirely dependent upon the linguistic context. Full sentence performance in noise was significantly correlated with vowel sentences ($r=0.48$, $p<0.01$), but somewhat more weakly for consonant sentences ($r=0.36$, $p<0.05$). Performance with consonants and vowels in sentences was not significantly correlated ($r=0.29$, $p>0.05$). Full word performance in noise was significantly correlated with both vowel ($r=0.59$, $p<0.01$) and consonant ($r=0.61$, $p<0.01$) words. Performance on consonant and vowel words was also significantly correlated ($r=0.64$, $p<0.01$). Therefore, speech-in-noise perception for sentences may be determined mostly by a listener's ability to process sentence-level vowel cues, but speech-in-noise for words may rely equally on processing vowel and consonant cues. Furthermore, a listener's ability to use consonant and vowel cues is only related in isolated word contexts; different abilities appear to be associated with processing these cues in sentences. This general pattern remained true for separate sub-analyses of ONH and OHI groups.

Analyses with two other measures available from each of the older subjects (i.e., age and HFPTA) demonstrated significant negative correlations of age with full sentence recognition in noise ($r=-0.46$, $p<0.01$). Use of sentence-level consonant cues were also associated with age ($r=-0.49$, $p<0.01$) and HFPTA ($r=-0.40$, $p<0.01$). Neither age nor HFPTA were strongly associated with isolated word recognition abilities for these listeners in the current study. Age and HFPTA were significant, but somewhat more weakly correlated ($r=0.37$, $p<0.05$, not in the table). These correlations suggest that use of sentence-level consonant cues is limited by age/auditory sensitivity, even with audibility ensured via spectral shaping of the stimulus. Use of sentence-level vowel cues was not associated with age or auditory sensitivity declines.

IV. DISCUSSION

The results of this study demonstrate again that vowels contribute more to intelligibility in sentence contexts than in isolated-word contexts. This study has provided further evidence that these essential cues are fully available for older listeners when sufficient audibility is provided. However, the question remains as to why these results are observed. Previous investigations have suggested that consonants may be essential for lexical access (Owren and Cardillo, 2006; Fogerty and Humes, 2010). However, vowels appear to be more involved in carrying suprasegmental and indexical properties of words (e.g., Owren and Cardillo, 2006). For several decades, we have known that suprasegmental cues provide important information in sentence contexts. For example, Cutler and Foss (1977), on the basis of phoneme monitoring reaction times in sentences with normal and non-normal stress patterns, concluded that sentence stress is important for sentence comprehension above and beyond stress patterns that are used to indicate the syntactic function of words. This stress advantage also appears to be independent of the additional perceptual clarity of words, as it is not observed for isolated words (Shields *et al.*, 1974). The stress

pattern of the sentence is aligned to the vowels (Tajima and Port, 2003), which results in vowels providing an informative structure for sentences. Furthermore, this structure aids in the prediction of upcoming stress patterns, perhaps resulting in a predictive advantage in sentence contexts. Recently, we have demonstrated that the vowel amplitude envelope, which includes amplitude and duration properties of sentence stress patterns, may be an essential acoustic component involved in added vowel contributions in sentences (Fogerty and Humes, 2012). The combination of these results appear to converge on the hypothesis that vowels carry the suprasegmental structure of the sentence that results in faster phoneme recognition times and assists in the prediction (or constrains the interpretation) of future sentence patterns or constituents. These findings may also help to explain one possible foundation for the importance of the temporal speech envelope to speech understanding (e.g., Shannon *et al.*, 1995). The envelope provides the amplitude and duration cues in sentences that are essential for conveying these slow-varying suprasegmental cues about the global sentence structure that appear to be predominantly present during the vowels.

The primary purpose of this study was to investigate the effect of age and cochlear pathology on the relative use of consonant and vowel acoustic cues in words and sentences. Results demonstrated that after audibility was accounted for through normal thresholds or spectral shaping, older listeners still performed more poorly than YNH listeners for all conditions except when vowels were presented in sentences. No effect of hearing loss was observed for these listeners after shaping was applied to the stimuli.

The finding here that older listeners perform as well as the young listeners but only for sentences containing predominant vowel cues, adds support to the hypothesis that vowels are carriers of global, sentence-level cues. Cognitive aging is known to cause declines in the processing of new information, yet older adults have relatively sustained semantic abilities (e.g., Wingfield *et al.*, 1994; Pichora-Fuller *et al.*, 1995; Burke and Mackay, 1997). These sustained abilities may reflect the use of preserved semantic processing of global, higher-level vowel cues provided by the vowel envelope. In contrast, the local cues of consonants and vowels in words rapidly vary, the processing of which may require abilities that decline with age. This decline was observed at all PTD values, not just those that provided the shortest glimpse durations as predicted.

In order to provide a preliminary investigation into this possibility, correlations among the older listeners were explored. While performance on vowel and consonant isolated words was significantly correlated, no such correlation was obtained in sentences. Furthermore, age and auditory sensitivity were associated only with consonants in sentences, not vowels. This provides preliminary evidence for shared associations for processing consonants and vowels in words and a dissociation of processing in sentences. Furthermore, this dissociation in sentences may be related to age-related sensory factors limiting the use of consonant cues. Other abilities, preserved with age, appear to underlie the essential ability to use vowel cues.

A. Age differences for processing words and sentences in noise

The examination of the full words and sentences presented in noise at 0 dB SNR allows for the examination of differences in speech-in-noise performance between the three listener groups tested here. For both words and sentences, no effect of hearing loss was found among the older listeners. This demonstrates that restored audibility, via spectral shaping, eliminates any potential performance level differences among ONH and OHI listeners that may have existed. Therefore, cochlear pathology is not involved in differences between ONH and OHI listener performance on the speech tasks assessed here. These findings are consistent with the premise that audibility is the primary predictor of older adult performance, with the remaining variance associated with various issues related to cognitive aging (Humes, 2002, 2007; see also Zurek and Delhorne, 1987).

While no difference between OHI and ONH groups was found, older listeners as a group performed significantly poorer than young listeners for both word and sentence tasks. Gordon-Salant and Fitzgibbons (1997) demonstrated that two factors were involved in older adult performance on speech-perception tasks. First, older listeners, in general, are better able to use contextual cues to predict the final word in the sentence. This is consistent with a number of other studies demonstrating equivalent or better use of supporting linguistic context by older listeners relative to young listeners (e.g., Wingfield *et al.*, 1991, 1994; Pichora-Fuller *et al.*, 1995; Humes *et al.*, 2007). Second, older listeners perform worse when the task demands impose a relatively higher cognitive load (i.e., repeating several elements in a sentence versus repeating a single word). The difference between the word and sentence tasks used here involved both processes. The sentence task had meaningful context available to predict other words in the sentence, but also required a greater memory load for encoding and repeating the entire sentence. Specifically, better contextual processing by the older listeners would predict a greater benefit for the sentence context over the word context, as compared to YNH listeners. Reduced memory load would predict less benefit. Either one or a combination of these factors could be involved in individual performance. However, the lack of a significant difference between the sentence-word benefit of older and younger listeners for the full materials presented at 0 dB SNR (see Fig. 6) possibly suggests a combination of these factors at the group level. This is consistent with Wingfield *et al.* (1994) who found an equal benefit of preceding sentence context on word recognition for young and older adults. In addition, no group differences in benefit from context were observed for materials preserving the consonants. In contrast, older listeners demonstrated a combined benefit of 15 RAU greater than the benefit obtained by the young listeners when listening to materials preserving the vowels (36 versus 50 RAU benefit for young and older listeners, respectively). This result supports the hypothesis that vowels provide higher-order contextual cues to facilitate processing, cues that older listeners are adept at processing.

B. Predictions of relative consonant and vowel contributions from the spectrum

The SII (ANSI, 1997) provides a good prediction of speech intelligibility in noise or with filtering. Frequency importance functions incorporated into SII calculations typically weight the mid frequency bands the most which contain the vowel formant frequencies. The removal of vowels or consonants could be viewed as a filtering process as it alters the long-term average spectrum of the speech sample. Therefore, spectral differences between consonants and vowels using the original consonant-vowel boundary condition were explored to determine if the SII would predict the differences observed here based upon the spectral differences of consonants and vowels in quiet. The speech level in 1/3 octave bands was measured from the long-term average speech spectrum for vowels in sentences, vowels in words, consonants in sentences, and consonants in words. These levels were subsequently used to calculate the SII for these four conditions. As expected, average measurements retained the inherent level differences of vowels and consonants (i.e., consonants were about 9 dB below the vowel level). Frequency importance functions were not available for our materials. Therefore, importance functions for standard speech (ANSI, 1997) were used for both words and sentences. Although the SII is not particularly well designed for predicting the intelligibility of interrupted sentences (see Rhebergen and Versfeld, 2005), the segment replacement procedure used here constrains interruption to provide glimpses of the same spectrum over time, either of consonants or of vowels. Therefore, the method used here to estimate the SII should be a good approximation of expected intelligibility differences between consonant and vowel spectra. Also note that while low-level noise was used during replacement, preserved consonant or vowel spectra were presented to listeners in quiet.

For both words and sentences, similar SII values were obtained for vowels and consonants, although values were slightly better for the consonants (sentences: vowels = 0.85, consonants = 0.89; words: vowels = 0.87, consonants = 0.93). Therefore, SII values did not predict that the vowels in sentences yielded a better performance than consonants in sentences, although they do agree with the behavioral results for performance in isolated words. Consonants had less energy in the low- and mid-frequency range compared to vowels. Therefore, it was somewhat surprising that consonants were associated with similar and even somewhat greater SII values compared to vowels. However, similar SII values were likely obtained because consonant energy in the lower frequency range largely exceeded listener thresholds, likely due to lower frequency energy from semi-vowels and initial vowel transitions. Audibility of lower frequency consonant components was facilitated by the slightly higher conversational level used (70 dB SPL). In addition, as expected, the average consonant level exceeded the vowel level in the higher frequencies.

As the spectral-shaping procedure ensured audibility for consonants and vowels by at least 15 dB for all OHI listeners, SII predictions also could not explain performance

among the OHI listeners when accounting for cochlear filtering. Thus, spectral contributions do not explain the observed differences between consonants and vowels to intelligibility. These results further highlight potential involvement of higher-order cues present in sentence vowels, cues that the SII does not capture.

C. The importance of the vowels for sentence intelligibility

Other studies have demonstrated that PTD is the primary predictor of speech intelligibility during interruption (Miller and Licklider, 1950; Wang and Humes, 2010; Kidd and Humes, 2012). As demonstrated by Fogerty and Kewley-Port (2009), other factors, such as glimpse duration, cannot account for the performance differences between sentences containing predominantly consonants or vowels. Therefore, this study again illustrates the exception to this PTD rule: Vowels in sentences contribute the most to intelligibility. Performance on the other three conditions investigated here (i.e., consonants in sentences, consonants in words, and vowels in words) are all adequately explained by PTD. Vowels in sentences appear to contain additional acoustic cues that facilitate perception beyond that expected from the proportion of the acoustic stimulus presented.

Vowels contain local information about the neighboring consonant (Cooper *et al.*, 1952), and thus are better able to fill in this missing information. However, this is also true for word contexts where vowels do not contribute more than consonants, as observed here and previously (Owren and Cardillo, 2006; Fogerty and Humes, 2010). Furthermore, as is well known, consonants also contain information about the neighboring vowel in the formant transitions (Liberman *et al.*, 1952). Therefore, local vowel and consonant information is distributed across “boundaries” phoneticians assign between segments. In sentences it appears that vowels contain higher-order cues about the global structure of the sentence. Therefore, the apparent contradictory results of Owren and Cardillo (2006), who found poorer discrimination of words based on vowels, and Kewley-Port and colleagues (2007), who found better sentence recognition from vowels, are explained, in part, by the current study. Using the same set of listeners and procedures for word and sentence conditions, results demonstrate that vowel contributions take on added importance only in sentence contexts. That is, vowel contributions are context-dependent. This effect is not due to methodological differences in how words and sentences are processed and tested, but instead due to intrinsic properties of vowels. This statement is supported by recent findings suggesting that this importance is based on the vowel amplitude envelope in sentences (Fogerty and Humes, 2012), which provides information about slow varying changes across the sentence (Rosen, 1992). Indeed, a coarse restoration of some amplitude cues of vowels during their noise replacement reduces, but does not eliminate, the difference observed between consonant and vowel sentences (Stilp and Kluender, 2010). These temporal amplitude cues can be used to predict syllabification and facilitate sentence-level word predictions (Waibel, 1987; Rosen, 1992) and these are clearly higher-order global cues that enhance

the predictability of the entire sentence. Vowels also contain the predominant *F0* contour which also predicts global syntactic structure (Wingfield *et al.*, 1984). Thus, there are several potential sources of sentence-level information carried in vowels that may enhance intelligibility.

D. The importance of the consonants for sentence intelligibility

The focus of this discussion has been on the *relative* contribution of consonants and vowels to speech intelligibility. The large relative benefit of vowels over consonants in sentences is of great significance. However, this discussion is not intended to suggest that consonants make no contribution. First, as discussed in Sec. I, the evidence clearly demonstrates that consonant information is distributed across the sentence and is present during vowels (e.g., Cooper *et al.*, 1952; Strange *et al.*, 1983). Sentences without any consonant information would be highly unintelligible, as is demonstrated in reading (Miller, 1951). Second, the discrete consonants presented here are informative for speech understanding. Even the listener who performed the best for vowel-only sentences at the original boundary only achieved 71% correct. (Incidentally, this was an OHI individual who performed the best. The best YNH listener obtained 66% correct.) As recognition of these sentences in quiet is at 99% correct (Kewley-Port *et al.*, 2007), consonants are responsible for at least a 30 percentage-point gain in intelligibility for quiet presentations, although some of this difference may be related to noise interruption factors, independent of any missing speech. Of course, all studies to date on the relative contribution of consonants versus vowels have been for English, which has more consonants in a sentence than syllable-timed languages such as Spanish or French. Some differences in the relative contribution may change with large phonetic differences in other languages. However, syllable-timed languages also contain a larger proportion of vowel segments than stress-timed languages, such as English (Ramus *et al.*, 1999), and therefore may exaggerate the observed differences between consonants and vowels. It also remains to be seen whether the relative contributions of consonants and vowels are preserved, exaggerated, or reduced in steady-state noise. However, given the low-intensity nature of consonants, the contribution of vowels in noise is likely to be even more important.

A second observation regarding the importance of the consonants involves the difference between word and sentence contexts. While vowels result in better sentence intelligibility, consonants make no additional contributions to intelligibility in sentence versus word contexts.

Consonants appear to be important for providing local feature cues essential for discriminating between lexical word hypotheses (Nespor *et al.*, 2003; Mehler *et al.*, 2006; Fogerty and Humes, 2010). For consonants, this function does not appear to change with the level of linguistic context. Due to their transient nature, consonants are not as well equipped to carry slow-varying dynamic information across the entire sentence as are vowels. However, they still provide local cues for lexical access.

E. Considerations for hearing impairment

For the OHI listeners, consonantal information may take on even more clinical importance. As noted, the performance of the OHI listeners with spectral shaping was equivalent to that of the YNH group for the vowel sentences, but performance for the consonant sentences was considerably lower for the OHI group relative to the YNH group (Fig. 6). Further, for the full unprocessed speech, the OHI listeners perform worse than the YNH group (Fig. 3). Thus, one could argue that signal-processing strategies for hearing aids should not only restore audibility, as for the OHI subjects in this study, but should also maximize the preservation of acoustic cues present during vowels. This vowel information contributes the most to word recognition in sentences, supplies the greatest contextual advantage, and is void of age differences. That is, OHI listeners appear to make full use of this vowel information regardless of what the underlying source is for vowel contributions.

However, this study also lends support for the most common amplification approach of focusing on the restoration of consonant information, the processing of which seems to be the most severely impaired. Therefore, there is an apparent balancing act for applying these results to amplification methods for OHI adults. First, vowel information must be preserved. Second, consonant information must be enhanced, but without an associated cost of reducing the underlying cues provided by the vowels. Attempts to improve consonant cues by increasing the level of the consonant relative to the vowel (i.e., improving the consonant-to-vowel ratio) have demonstrated improvements in consonant recognition for nonsense syllables (Gordon-Salant, 1986, although see Sammeth *et al.*, 1999). Consonant-amplification distorts the overall speech envelope, which provides essential cues for sentence recognition (Shannon *et al.*, 1995). However, for stimuli with limited spectral cues, consonant-amplification improves recognition of some (i.e., voiced stops), although not all, consonants in vowel-consonant-vowel contexts (Freyman *et al.*, 1991). This may be due to the preserved temporal envelope cues carried by the vowels that appear to underlie vowel contributions in sentences (Fogerty and Humes, 2012). However, it is difficult to generalize these selective consonant amplification results in nonsense syllables to sentence contexts.

It appears that sentence-level vowels, compared to consonants, may be more resistant to time compression (Gay, 1978; Max and Caruso, 1997) and reduction of duration (Fogerty and Kewley-Port, 2009). This may be related to normal speech production processes of vowel reduction in natural speech (Max and Caruso, 1997). In contrast, amplitude compression may be more detrimental to vowel contributions, particularly as results from Fogerty and Humes (2012) suggest that the vowel amplitude envelope may be responsible for the greater contributions of vowels in sentences. Whereas phoneme recognition remains relatively intact with compression (e.g., Dreschler, 1989), sentence recognition is clearly reduced (Souza and Kitch, 2001; Jenstad and Souza, 2007). This may, in part, be due to the reduction of vowel amplitude cues that are more informative in sentence contexts. However, declines in sentence intelligibility are

most apparent for the extreme compression conditions not typically implemented clinically (Jenstad and Souza, 2007).

V. CONCLUSIONS AND SUMMARY

Overall, older listeners performed as well as young listeners when sentence-level vowel cues were preserved. YNH listeners performed better than older listeners for sentences and words containing only the consonants or words containing only the vowels. No difference was observed between ONH listeners and OHI listeners who received spectral shaping to ensure audibility through 4 kHz. These results were maintained across four conditions that varied the amount of transitional information contained within the consonants or vowels preserved. Results point to sentence-level vowels as essential carriers of auditory contextual speech information that older listeners are successfully able to use, even when the task demands a relatively greater cognitive load.

This study advanced the hypothesis that vowels carry high-order cues for the processing of sentences by five types of evidence. First, vowels in sentences contribute relatively more to intelligibility than consonants across four proportions of the total sentence duration that manipulated the amount of transitional acoustics at the consonant-vowel boundary. This effect was only observed with sentences where dynamic vowel information might be used to predict the global structure of the sentence. Second, no difference between young and older listeners was obtained for vowel sentences, even though clear performance differences were observed in the other three conditions. Thus, the processing of vowel cues appears to be resilient with aging, which was also evidenced in the correlational analysis among the older listeners. This supports the claim that vowels provide higher-order cues; as such, language processing is preserved with age. Third, the use of consonant and vowel cues among the older listeners was only correlated in isolated word contexts. In sentence contexts, vowel and consonant cues may involve dissociated processes with the use of vowel cues associated the most with full sentence processing. Fourth, older listeners who typically receive a greater benefit from the presence of contextual cues, only did so for materials preserving the vowels. This greater benefit occurred even though greater memory load was involved in the sentence task. No difference in benefit between young and old listeners was obtained for materials containing the full stimulus or just the consonants. Fifth, SII predictions for vowel and consonant contributions to the intelligibility of words and sentences were estimated from the vowel and consonant spectra. No difference, or possibly a slight consonant advantage, was predicted by the SII for both words and sentences. As noted, this prediction was upheld for words, but not sentences. This suggests that spectral differences alone do not explain the observed greater vowel contributions in sentences.

This evidence converges on the possibility that vowels provide contextual cues for the higher-order processing of meaningful sentences. These contextual cues are likely contained in the auditory signal, as evidence suggests that consonants are more important for lexical access (Nespor *et al.*,

2003; Mehler *et al.*, 2006; Fogerty and Humes, 2010). Therefore, supralinguistic auditory cues, such as the vowel amplitude envelope (Fogerty and Humes, 2012), may provide this information about the global sentence structure.

Results strongly suggest that while amplification methods should make concerted efforts to enhance consonantal information, preserving the natural dynamics of vowels is also important. When audibility of the speech materials is ensured, normal-hearing and hearing-impaired older listeners are able to successfully use these sentence-level vowel cues as well as YNH listeners.

ACKNOWLEDGMENTS

This work was supported in part by NIA R01 AG008293 awarded to L.E.H.

- ANSI. (1997). ANSI S3.5-1995, "American national standard methods for the calculation of the speech intelligibility index" (American National Standards Institute, New York).
- Burke, D. M., and Mackay, D. G. (1997). "Memory, language, and ageing," *Philos. Trans. R. Soc. London, Ser. B* **352**, 1845–1856.
- Caramazza, A., Chialant, D., Capasso, R., and Miceli, G. (2000). "Separable processing of consonants and vowels," *Nature* **403**, 428–430.
- Carreiras, M., Gillon-Dowens, M., Vergara, M., and Perea, M. (2009). "Are vowels and consonants processed differently? Event-related potential evidence with a delayed letter paradigm," *J. Cogn. Neurosci.* **21**, 275–288.
- Carreiras, M., and Price, C. (2008). "Brain activation for consonants and vowels," *Cereb. Cortex* **18**, 1727–1735.
- Cole, R., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (1996). "The contribution of consonants versus vowels to word recognition in fluent speech," in *Proceedings of the ICASSP'96*, pp. 853–856.
- Cooper, F., Delattre, P., Liberman, A., Borst, J., and Gerstman, L. (1952). "Some experiments on the perception of synthetic speech sounds," *J. Acoust. Soc. Am.* **24**, 597–606.
- Cutler, A., and Foss, D. J. (1977). "On the role of sentence stress in sentence processing," *Lang Speech* **20**, 1–10.
- Dreschler, W. A. (1989). "Phoneme perception via hearing aids with and without compression and the role of temporal resolution," *Audiology* **28**, 49–60.
- Fitzgibbons, P. J., and Gordon-Salant, S. (2004). "Age effects on discrimination of timing in auditory sequences," *J. Acoust. Soc. Am.* **116**, 1126–1134.
- Fogerty, D. (2011). "Perceptual weighting of individual and concurrent cues for sentence intelligibility: Frequency, envelope, and fine structure," *J. Acoust. Soc. Am.* **129**, 977–988.
- Fogerty, D., and Humes, L. E. (2010). "Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors," *J. Acoust. Soc. Am.* **128**, 3114–3125.
- Fogerty, D., and Humes, L. E. (2012). "The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences," *J. Acoust. Soc. Am.* **131**, 1490–1501.
- Fogerty, D., and Kewley-Port, D. (2009). "Perceptual contributions of the consonant-vowel boundary to sentence intelligibility," *J. Acoust. Soc. Am.* **126**, 847–857.
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," *J. Psychiatr. Res.* **12**, 189–198.
- Freyman, R. L., Nerbonne, G. P., and Cote, H. A. (1991). "Effect of consonant-vowel ratio modification on amplitude envelope cues for consonant recognition," *J. Speech Hear. Res.* **34**, 415–426.
- Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (1990). "DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM," National Institute of Standards and Technology, NTIS Order No. PB91-505065.
- Gay, T. (1978). "Effect of speaking rate on vowel formant movements," *J. Acoust. Soc. Am.* **63**, 223–230.
- Gordon-Salant, S. (1986). "Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing," *J. Acoust. Soc. Am.* **80**, 1599–1607.
- Gordan-Salant, S., and Fitzgibbons, P. J. (1997). "Selected cognitive factors and speech recognition performance among young and elderly listeners," *J. Speech Lang. Hear. Res.* **40**, 423–231.
- Humes, L. E. (2002). "Factors underlying the speech-recognition performance of elderly hearing-aid wearers," *J. Acoust. Soc. Am.* **112**, 1112–1132.
- Humes, L. E. (2007). "The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults," *J. Am. Acad. Audiol* **18**, 609–623.
- Humes, L. E., Burk, M. H., Coughlin, M. P., Busey, T. A., and Strauser, L. E. (2007). "Auditory speech recognition and visual text recognition in younger and older adults: similarities and differences between modalities and the effects of presentation rate," *J. Speech Lang. Hear. Res.* **50**, 283–303.
- Jenstad, L. M., and Souza, P. E. (2007). "Temporal envelope changes of compression and speech rate: Combined effects on recognition for older adults," *J. Speech Lang. Hear. Res.* **50**, 1123–1138.
- Kewley-Port, D., Burkle, Z., and Lee, J. (2007). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," *J. Acoust. Soc. Am.* **122**, 2365–2375.
- Kidd, G. R., and Humes, L. E. (2012). "Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences," *J. Acoust. Soc. Am.* **131**, 1434–1448.
- Ladefoged, P. (2001). *Vowels and Consonants: An Introduction to the Sounds of Languages* (Blackwell, Oxford), pp. 1–191.
- Lehiste, I. (1970). *Suprasegmentals* (MIT Press, Cambridge, MA), pp. 1–194.
- Liberman, A. M., Delattre, P., and Cooper, F. S. (1952). "The role of selected stimulus variables in the perception of unvoiced stop consonants," *Am. J. Psychiatry* **65**, 497–516.
- Max, L., and Caruso, A. J. (1997). "Acoustic measures of temporal intervals across speaking rates: Variability of syllable- and phrase-level relative timing," *J. Speech Hear. Res.* **40**, 1097–1110.
- Mehler, J., Peña, M., Nespor, M., and Bonatti, L. (2006). "The 'soul' of language does not use statistics: Reflections on vowels and consonants," *Cortex* **42**, 846–854.
- Miller, G., and Licklider, J. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**, 167–173.
- Miller, G. A. (1951). *Language and Communication* (McGraw-Hill, New York), pp. 1–298.
- Nespor, M., Peña, M., and Mehler, J. (2003). "On the different roles of vowels and consonants in speech processing and language acquisition," *Lingue e Linguaggio* **2**, 201–227.
- New, B., Araújo, V., and Nazzi, T. (2008). "Differential processing of vowels and consonants in lexical access through reading," *Psychol. Sci.* **19**, 1223–1227.
- Owren, M. J., and Cardillo, G. C. (2006). "The relative roles of vowels and consonants in discriminating talker identity versus word meaning," *J. Acoust. Soc. Am.* **119**, 1727–1739.
- 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.
- Ramus, F., Nespor, M., and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," *Cognition* **73**, 265–292.
- Rhebergen, K. S., and Versfeld, N. J. (2005). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *J. Acoust. Soc. Am.* **117**, 2181–2192.
- Rosen, S. (1992). "Temporal information in speech: Acoustic, auditory, and linguistic aspects," *Philos. Trans. R. Soc. London, Ser. B* **336**, 367–373.
- Sammeth, C. A., Dorman, M. F., and Stearns, C. J. (1999). "The role of consonant-vowel intensity ratio in the recognition of voiceless stop consonants by listeners with hearing impairment," *J. Speech Lang. Hear. Res.* **42**, 42–55.
- Seewald, R. C., Ramji, K. V., Sinclair, S. T., Moodie, K. S., and Jamieson, D. G. (1993). "Computer-assisted implementation of the desired sensation level method for electroacoustic selection and fitting in children: Version 3.1, User's Manual," The University of Western Ontario, London.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Shields, J. L., McHugh, A., and Martin, J. G. (1974). "Reaction time to phoneme targets as a function of rhythmic cues in continuous speech," *J. Exp. Psychol.* **102**, 250–255.

- Shrivastav, M. N., Humes, L. E., and Aylsworth, L. (2008). "Temporal-order discrimination of tonal sequences by younger and older adults: The role of duration and rate," *J. Acoust. Soc. Am.* **124**, 462–471.
- Souza, P. E., and Kitch, V. (2001). "The contribution of amplitude envelope cues to sentence identification in young and aged listeners," *Ear Hear.* **22**, 112–119.
- Stevens, K. N. (2002). "Toward a model for lexical access based on acoustic landmarks and distinctive features," *J. Acoust. Soc. Am.* **111**, 1872–1891.
- Stilp, C., and Kluender, K. (2010). "Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility," *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12387–12392.
- Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of coarticulated vowels," *J. Acoust. Soc. Am.* **74**, 695–705.
- Strom, K. E. (2006). "The HR 2006 dispenser survey," *Hear. Rev.* **13**, 16–39.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Tajima, K., and Port, R. F. (2003). "Speech rhythm in English and Japanese," in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge, UK), pp. 317–334.
- Takayanagi, S., Dirks, D., and Moshfegh, A. (2002). "Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing," *J. Am. Acad. Audiol.* **16**, 494–504.
- Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (2008). "Finding words and rules in a speech stream: Functional differences between vowels and consonants," *Psychol. Sci.* **19**, 137–144.
- Waibel, A. (1987). "Prosodic knowledge sources for word hypothesization in a continuous speech recognition system," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'87*, pp. 856–859.
- Wang, X., and Humes, L. E. (2010). "Factors influencing recognition of interrupted speech," *J. Acoust. Soc. Am.* **128**, 2100–2111.
- Wingfield, A., Aberdeen, J. S., and Stine, E. A. L. (1991). "Word onset gating and linguistic context in spoken word recognition by young and elderly adults," *J. Gerontol.* **46**, P127–P129.
- Wingfield, A., Alexander, A. H., and Cavigelli, S. (1994). "Does memory constrain utilization of top-down information in spoken word recognition? Evidence from normal aging," *Lang. Speech* **37**, 221–235.
- Wingfield, A., Lombardi, L., and Sokol, S. (1984). "Prosodic features and the intelligibility of accelerated speech: Syntactic versus periodic segmentation," *J. Speech Hear. Res.* **27**, 128–134.
- Zue, V. W., and Seneff, S. (1988). "Transcription and alignment of the TIMIT database," *Proceedings of the Second Meeting on Advanced Man-Machine Interface through Spoken Language*, pp. 11.1–11.10.
- Zurek, P. M., and Delhorne, L. A. (1987). "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," *J. Acoust. Soc. Am.* **85**, 1548–1559.