Sentence intelligibility during segmental interruption and masking by speech-modulated noise: Effects of age and hearing loss

Daniel Fogerty^{a)}

Department of Communication Sciences and Disorders, University of South Carolina, 1224 Sumter Street, Columbia, South Carolina 29208, USA

Jayne B. Ahlstrom, William J. Bologna,^{b)} and Judy R. Dubno

Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, 135 Rutledge Avenue, Charleston, South Carolina 29425, USA

(Received 5 December 2014; revised 23 March 2015; accepted 11 May 2015)

This study investigated how single-talker modulated noise impacts consonant and vowel cues to sentence intelligibility. Younger normal-hearing, older normal-hearing, and older hearing-impaired listeners completed speech recognition tests. All listeners received spectrally shaped speech matched to their individual audiometric thresholds to ensure sufficient audibility with the exception of a second younger listener group who received spectral shaping that matched the mean audiogram of the hearing-impaired listeners. Results demonstrated minimal declines in intelligibility for older listeners with normal hearing and more evident declines for older hearing-impaired listeners, possibly related to impaired temporal processing. A correlational analysis suggests a common underlying ability to process information during vowels that is predictive of speech-in-modulated noise abilities. Whereas, the ability to use consonant cues appears specific to the particular characteristics of the noise and interruption. Performance declines for older listeners were mostly confined to consonant conditions. Spectral shaping accounted for the primary contributions of audibility. However, comparison with the young spectral controls who received identical spectral shaping suggests that this procedure may reduce wideband temporal modulation cues due to frequency-specific amplification that affected high-frequency consonants more than low-frequency vowels. These spectral changes may impact speech intelligibility in certain modulation masking conditions.

© 2015 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4921603>]

[MAH]

Pages: 3487–3501

I. INTRODUCTION

Speech is often heard in the presence of environmental noise that fluctuates in amplitude over time. One of the most common fluctuating noise sources that can pose a significant challenge to speech intelligibility is a competing speech signal from another talker. There are several reasons for the difficulty posed to listeners in speech-on-speech listening conditions, one of which involves modulation masking (e.g., Gustafsson and Arlinger, 1994; Kwon and Turner, 2001). In these conditions, the amplitude modulation properties of the masker interfere with a listener's ability to perceive the natural amplitude fluctuations of the target talker (Gnasia *et al.*, 2008). These slow-rate amplitude changes in speech, termed the temporal envelope, provide highly salient information essential for speech intelligibility (Shannon *et al.*, 1998). Indeed, fluctuating and interrupted noise sources interfere with a listener's ability to process the temporal envelope of speech (e.g., Nelson and Jin, 2004; Gilbert *et al.*, 2007; Gnasia *et al.*, 2008), requiring listeners to adjust their perceptual processing to other preserved speech cues (Fogerty,

2011). Adding complexity to the processing requirements of these conditions is that speech information carried by the temporal envelope is not likely to be evenly distributed across the entire speech sample. Rather, some time intervals of the envelope are characterized by rapid periods of acoustic change (i.e., entropy) that are important for conveying speech information (Chen and Loizou, 2012; Stilp, 2014). This is because the temporal envelope is produced by specific articulatory events and generally results in rapid change at consonant-vowel boundaries (Chen and Loizou, 2012). The examination of specific speech events significantly aids in our understanding of how this temporal signal is distributed across the sentence.

One basic and fundamental division of speech sounds is represented by consonants and vowels (Stevens, 2002; Ladefoged and Disner, 2012). While these sound categories are overlapping in time (e.g., Cooper *et al.*, 1952; Strange *et al.*, 1983) and features (e.g., Stevens, 2002), this general classification of speech sounds has highlighted important differences in how acoustic properties present during these segments contribute to intelligibility. Evidence from statistical learning and interruption paradigms suggests that consonants provide important information for lexical access, while suprasegmental properties conveyed during vowels are important for providing the temporal and grammatical structure

^{a)}Electronic mail: fogerty@sc.edu

^{b)}Also at: Department of Hearing and Speech Sciences, University of Maryland, College Park, MD.

of the sentence (e.g., Nespor *et al.*, 2003; Toro *et al.*, 2008; Fogerty and Humes, 2010; Fogerty *et al.*, 2012; Fogerty, 2014). However, it is important to note that these acoustic properties vary continuously and are not discretely bounded by segmental markings supplied by phoneticians. Therefore, the current study is an investigation of the relative information that occurs during vocalic or consonantal intervals of speech and not related to an exclusive role of vowels or consonants per se. When we discuss vowels or consonants in the present context, we refer to acoustic properties that are predominantly present during speech segments of that category.

The present study specifically investigates how listeners are able to use the cues conveyed during consonants and vowels in the presence of amplitude modulations from a competing talker. Furthermore, hearing impairment and amplification may have specific consequences for processing these cues, which represents a second focus of this study.

A. Contribution of temporal modulations during vowels

The temporal envelope conveys slow-rate information about the rhythm and stress of speech (Rosen, 1992), which is temporally aligned to the vowels in the sentence (Port, 2003). There is increasing evidence that vowels contain temporal envelope cues that are crucial for sentence intelligibility (Fogerty and Humes, 2012a; Fogerty, 2013, 2014); while in isolated word contexts, information in the temporal envelope appears to be more evenly distributed among the segments (Fogerty and Humes, 2010, 2012a). The envelope clearly provides information about vowel and consonant phoneme identification (e.g., Shannon *et al.*, 1998); however, vowels take on added importance in sentence contexts (Kewley-Port *et al.*, 2007) likely due to additional suprasegmental cues present in sentences (Fogerty and Humes, 2012a), such as speech rhythm. These slower rate amplitude modulations may serve to provide disambiguating acoustic information about syntactic structure (Sharpe *et al.*, 2014). In contrast, the transient acoustic properties of most consonants (Stevens, 2002) produce relatively faster local fluctuations in the speech amplitude that are intrinsic to individual phonemes and are therefore less likely to provide additional contextual cues about the global sentence construction. This represents an acoustic difference between consonants, which provide faster fluctuations at short time scales that are important for phoneme identification, and vowels, which provide slow fluctuations at longer timescales that span the entire sentence and are important for sentence intelligibility.

The importance of vowels, which provide a large advantage over consonants in sentence contexts, is now well established for English (e.g., Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009; Fogerty *et al.*, 2012) and Mandarin (Chen *et al.*, 2013; Fogerty and Chen, 2014). This contribution has been investigated using a segment-replacement paradigm wherein all of the vowels (or all of the consonants) in the sentence are replaced with steady-state noise. In these conditions, listeners must rely on cues present in the preserved segments to understand the sentence. Introducing amplitude modulations from the temporal

envelope of the missing segments improves intelligibility, but only when these temporal envelope modulations come from missing vowels, not consonants (Fogerty, 2014), demonstrating the importance of vowel amplitude modulations for intelligibility. This also suggests that listeners process and (mis)assign the masker modulations to the target sentence (Shinn-Cunningham and Wang, 2008) so that, in this experimental case, the masker modulations are informative about the target sentence. However, in more natural contexts, interfering noise sources, such as from competing talkers, provide conflicting information in the temporal envelope.

B. Effect of masking or replacement by speech-modulated noise

The current study investigated how listeners are able to use cues provided by consonants and vowels when a competing amplitude modulated noise is present. In addition, this study extends these methods to examine cases of speech masked simultaneously by a competing modulated signal, as compared to the steady-state noise replacement paradigm used in previous studies (e.g., Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009). These additional conditions provide a means to ascertain the ability to generalize results from an interruption paradigm to more natural speech-on-speech masking conditions.

Four speech-masker combinations were used in the present study to investigate the relative contribution of vowels and consonants during speech-modulated noise, when (1) speech and noise were continuous (CONT), (2) continuous speech was presented with an interrupted speech-modulated noise (CONT_{speech}), (3) interrupted speech was presented with a continuous speech-modulated noise (INT_{speech}), and (4) interrupted speech was alternated with interrupted noise (INT_{speech&noise}). Segmental conditions for speech-masker combinations two through four investigated the relative roles of vowels and consonants. These combinations specifically investigated how competing modulations interfere with consonant and vowel cues during simultaneous masking as well as during segmental replacement.

The use of simultaneous modulated masking noise provides a method to assess how competing modulations interfere with concurrent cues provided by vowel or consonant segments. While simultaneous modulated masking noise impairs speech intelligibility (e.g., Gustafsson and Arlinger, 1994; Kwon and Turner, 2001), it is not clear how this may interfere with acoustic events that occur during consonant or vowel segments, particularly given the different spectral and temporal qualities of these segmental categories. In addition, the use of modulated replacement noise provides a method to investigate the effect of non-simultaneous modulations on the ability of listeners to use the acoustic information in preserved consonant or vowel segments of speech. While previous investigations have demonstrated intelligibility gains when replacement noise is modulated by the missing temporal envelope (Fogerty, 2014; Shinn-Cunningham and Wang, 2008), it is not clear if introducing competing modulations will have any effect different from the random amplitude fluctuations of a “steady-state” noise, which can also act as a

modulation masker (Stone *et al.*, 2012). Importantly, these conditions detail how temporal amplitude modulations of a competing talker influence a listener's ability to use acoustic information present during consonants or vowels for sentence intelligibility. An additional correlational analysis investigated the association of listener performance across these different stimulus conditions. The analysis of associations among the conditions may highlight whether speech-in-noise performance is a result of general speech perception abilities or is specific to the stimulus characteristics of an individual condition.

C. Effect of age and hearing loss

Given the high importance of temporal envelope cues for intelligibility, and the multiple timescales of this temporal information that are present during sentences (e.g., Peelle and Davis, 2012; Poeppel, 2014), it is also important to understand how different listener groups are able to process this signal. Evidence suggests that older adults and adults with hearing loss are able to adequately process temporal envelope information (Peters and Hall, 1994; Boettcher *et al.*, 2001) if adequate audibility is ensured (Takahashi and Bacon, 1992), but may have more difficulty in the presence of modulated maskers (e.g., Gustafsson and Arlinger, 1994; Dubno *et al.*, 2003; George *et al.*, 2007). Important for this study, older adults with normal and impaired hearing also appear able to process vowel cues in sentences as well as younger adults, whereas they have more difficulty processing sentence-level consonant cues and both segment types in isolated words (Fogerty *et al.*, 2012). Thus, sentence-level vowel cues represent an area of preserved processing for older listeners. However, as older adults have more difficulty in complex temporal listening conditions (e.g., Gordon-Salant and Fitzgibbons, 1999), we may expect to find more difficulties in the context of a competing signal. The current study explores age and hearing loss related factors to the processing of consonant and vowel cues for a more comprehensive set of modulated maskers.

D. Effect of amplification via spectral shaping

In addition to age and hearing loss, this study also investigates the effect of spectrally shaping the speech stimuli to ensure audibility. Spectral shaping is a common experimental control to ensure that speech materials are presented above the hearing thresholds of an individual listener and is sometimes designed to model the frequency-specific amplification of hearing aids (Humes, 2007). The effect of this spectral shaping (increasing the sensation level) was explored in the current study by including a group of younger normal hearing adults who listened to speech that was spectrally shaped according to the mean audiogram of the older hearing-impaired listeners. While generally successful at ensuring audibility (Humes, 2013), spectral shaping can alter how listeners perceptually weight the temporal envelope in different spectral regions (Fogerty and Humes, 2012b). Spectral shaping amplifies the high frequencies more than the low frequencies for the typical older adult with high frequency hearing loss. On average, given that

consonants contain more energy in the high frequencies than vowels, this spectral shaping increases the average level of the consonants more than the vowels. This results in a reduction in the consonant-vowel intensity ratio, which could change the nature of amplitude cues at the consonant-vowel boundary that capture essential speech information (Chen and Loizou, 2012). While some modification of the consonant-vowel ratio can facilitate intelligibility of syllables (e.g., Gordon-Salant, 1986), such processing can also have significant consequences for the speed of lexical processing (Fogerty *et al.*, 2014), which may be particularly important when listening to longer sentence-length materials or for older adults. Thus, spectral shaping could have important implications for temporal envelope processing, particularly when a competing amplitude modulated signal is present.

E. Summary

Overall, previous research has highlighted the importance of temporal amplitude modulations. Furthermore, vowels appear to convey slow amplitude modulations that span the entire sentence and provide important contributions to sentence intelligibility that complement phoneme-level cues. Amplitude modulated signals from a competing talker may specifically interfere with a listener's ability to process the envelope cues of the target sentence, particularly for older adults who have more difficulty in temporally complex listening environments. The spectral shaping that is modeled after hearing aid algorithms to ensure sufficient audibility of the speech signal also interacts with the amplitude modulations of the speech signal and may have significant implications for processing envelope cues that are available for listeners (Fogerty and Humes, 2012b). This is in addition to other changes in amplitude modulation that might result from hearing aid processing, such as amplitude compression (e.g., Jenstad and Souza, 2007; Sabin *et al.*, 2013). The current study was designed to explore these issues by investigating the contributions of consonants and vowels in the presence of a competing amplitude modulated signal that either replaced or was presented simultaneously with the target speech. In addition, age, hearing loss, and spectral shaping were investigated to determine their effects on sentence intelligibility under these conditions.

II. METHODS

A. Listeners

Eighty-three listeners participated in the current study, divided among four groups: young normal-hearing adults (YNH; $N=20$; mean age = 22 yr, range = 18–28), older normal-hearing adults (ONH; $N=23$, mean age = 68 yr, range = 62–82), older hearing-impaired adults (OHI; $N=20$; mean age = 72 yr, range = 63–88), and young normal-hearing spectral controls matched to the mean audiogram of the OHI listeners (YSC; $N=20$, mean age = 21 yr, range = 18–25). All normal hearing listeners had pure-tone thresholds ≤ 20 dB hearing level (HL) at octave frequencies from 0.25 to 4.0 kHz. Inclusion criteria required hearing-impaired listeners to have audiometric thresholds < 55 dB

HL at 0.25 kHz, and <60 dB HL at 0.5, 1, 2, 3, and 4 kHz in at least one ear (i.e., the test ear). Eight of the OHI listeners had previous experience using hearing aids (Median duration of use: 15.0 months; range: 4.6–67.3 months). To control for audibility across all listeners, all groups (except YSC) received spectral shaping based upon their individual audiogram. The YSC group received a standard spectral shaping based on the mean audiometric thresholds of the OHI group. Figure 1 displays the audiometric thresholds in the test ear for each participant group along with the average speech spectrum for the corpus following spectral shaping for each listener group. Note that while the audiometric thresholds were used for 1/3 octave band level calculations, the two normal hearing groups (YNH and ONH), received little or no spectral shaping in the frequency range up to 4 kHz. Therefore, these groups can be considered to have functionally equivalent acoustic conditions in this frequency range. However, due to somewhat elevated thresholds above 3.0 kHz for the ONH group, some spectral shaping was applied in this region. This helped to ensure adequate audibility for the ONH for even high frequency components (i.e., consonants). Following this procedure, the average overall speech presentation level was 70 dB sound pressure level (SPL) for YNH, 73 dB SPL for ONH, and 83 dB SPL for OHI and YSC groups. Spectral shaping of the stimuli was designed to restore audibility of the speech signal through at least 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). For all experimental conditions, spectral shaping was applied to the combined speech+noise signal. Comparisons between YNH and ONH groups allowed for the examination of age differences, comparisons between ONH and OHI allowed for the examination of the effects of cochlear pathology when audibility has been restored, and comparisons between

YNH and YSC allows for the investigation of how spectral shaping might influence speech intelligibility. Finally, comparisons between YSC and OHI groups examined the extent to which accounting for reduced audibility can explain performance in the OHI group.

B. Stimuli

Stimuli consisted of sentences selected from the TIMIT database (Garofolo *et al.*, 1993). These sentences have been previously divided into lists of 18 sentences each for use in a test, PRESTO, which was designed to assess speech perception with high-variability linguistic and indexical properties (Gilbert *et al.*, 2013). The seven sentence lists of 18 sentences (76 keywords) each used in this study were balanced for keyword frequency, familiarity, and talker gender. No sentences were repeated in any condition. TIMIT sentences have been included in several studies using segmentally interrupted speech to selectively preserve vowel segments (e.g., Fogerty and Kewley-Port, 2009) as the database provides time markings for segmental boundaries that were confirmed by expert phoneticians. TIMIT phonetic symbols were classified into vowel and consonant categories. There were 20 vowels (iy, ih, eh, ey, ae, aa, aw, ay, ah, ao, oy, ow, uh, uw, ux, er, ax, ix, axr, ax-h) and 32 consonants (b, d, g, p, t, k, dx, q, jh, ch, s, sh, z, zh, f, th, v, dh, hh, hv, l, r, y, w, el, eng, nx, m, n, ng, em, en). Each sentence within a list was spoken by a novel talker.

C. Signal processing and design

Two of the sentences recorded in the TIMIT database were spoken by every talker. For each trial, these two sentences were selected from the target talker of that specific trial and concatenated into one longer sound file. A speech-shaped noise was then created that matched the duration and power spectrum of this concatenated sound file to be used as the competing signal on the individual experimental trial. The temporal envelope of the concatenated speech sample was then extracted by half-wave rectification followed by low-pass filtering using a fourth order Butterworth filter with a cutoff modulation frequency of 16 Hz. This wideband temporal envelope was then used to amplitude modulate the speech-shaped noise that was constructed for the given talker. The final speech-modulated noise was unintelligible but contained the average spectral and temporal characteristics of the target talker of each trial.

The target speech sentence and competing noise were then combined to create seven different experimental conditions distributed across four different speech-masker combinations. Six of these conditions involved segmentally interrupting either the speech or noise using the segmental boundaries of the target sentence defined by the TIMIT database and adjusted within 1-ms to the nearest local minima (i.e., zero-crossing). The stimulus combinations for these experimental conditions are displayed in Fig. 2 and defined in the following text.

- (1) The first speech-masker combination examined sentence intelligibility for speech presented with continuous

FIG. 1. Quiet thresholds for the four subject groups (lines with data points). Lines without data points are mean one-third-octave band levels for the spectrally shaped speech for the younger normal hearing group (black dashed line) and the older normal hearing group (black solid line). Spectral shaping was identical for the younger shaped control and the older hearing impaired groups (solid grey line).

speech-modulated noise. To create these stimuli, a random portion of the speech-modulated noise that matched the target talker was extracted and added to the target sentence at 0 dB SNR relative to the average sentence level (70 dB SPL). For this combination, speech and noise were continuous (CONT).

- (2) The second speech-masker combination involved two conditions that investigated continuous speech in the presence of an interrupted speech-modulated noise (CONT_{speech}). In these two conditions, the competing noise was interrupted to preserve either the vowels or the consonants of the target sentence. Therefore, in the vowel condition, the vowels of the target sentence were presented in quiet, while the consonants of the target sentence were presented simultaneously with the speech-modulated noise that was added at -5 dB SNR relative to the average sentence level, which matched the average vowel level (Fogerty *et al.*, 2012). Likewise, in the consonant condition, consonants were presented in quiet, with the vowels presented at -5 dB SNR (re: sentence level). It is important to note that this processing resulted in a competing noise that was modulated by two signals. The noise was turned “on” or “off” according to the segments of the target sentence, but the “on” intervals of the noise also retained the envelope modulation derived from the competing sentence.
- (3) The third speech-masker combination investigated the intelligibility of interrupted sentences in the presence of a

continuous speech-modulated noise (INT_{speech}). Again, two conditions were tested that preserved either the vowels or the consonants of the target sentence but removed the other segment type. The interrupted sentence that preserved the target segment type but replaced the alternative segment type with silence was then added to a random selection of the speech-modulated noise that matched the duration and talker characteristics of the target sentence. The noise was presented at $+5$ dB SNR (re: sentence level prior to interruption), which matched the average level of the consonants (Fogerty *et al.*, 2012). Therefore, for consonant and vowel presentations, continuity of the noise masker, but not the target speech, was preserved.

- (4) The final speech-masker combination alternated interrupted portions of speech and noise (INT_{speech&noise}). Two experimental conditions that preserved either the consonants or vowels of the target sentence were again created. For the consonant condition, the consonants of the sentence were presented in quiet and the vowels were replaced with the interrupted speech-modulated noise, while the reverse was true for the vowel condition. The noise was added at 0 dB SNR (re: sentence level prior to interruption).

D. Procedures

Experimental testing for YNH and YSC groups was completed at the University of South Carolina, and testing for ONH and OHI groups was completed at the Medical University of South Carolina. Both laboratories were similarly equipped and used identical stimulus presentation software. All testing was completed in a sound attenuating booth. The participants listened to stimuli presented at a sampling rate of 48 828 Hz via a Sennheiser HDA 200 headphone routed first through a TDT System III digital-to-analog processor (RP2 or RX6) and headphone buffer (HB7 or HB5). The level of the speech (prior to replacement and spectral shaping) was calibrated to 70 dB SPL. Stimuli were presented monaurally to the test ear. All sentence materials were presented to listeners in blocks according to the four speech-masker combinations. Sentences for each speech-masker combination were randomized, with consonant and vowel conditions intermixed during presentation, and the order of blocks across participants was also randomized on an individual basis. Participants first completed demonstration trials to familiarize them with the stimulus processing and the speech-masker combination prior to completing the experimental conditions. Each of the seven experimental conditions consisted of 18 sentences (76 key words).

Listeners were required to repeat each sentence and were encouraged to guess. Sentence presentation was self-paced by pressing a button on the custom interface. Listener responses were scored online as well as audio recorded for offline analysis by trained raters who scored the keywords correctly repeated. Keywords were selected for analysis here as PRESTO lists were designed to balance lexical properties of the keywords. A strict scoring procedure was used that required listeners to repeat each word exactly correct (e.g., no missing or extra suffixes). Percent correct scores were

FIG. 2. Waveforms of an example sentence, “His artistic accomplishments guaranteed him entry into any social gathering,” processed according to the four different speech-masker combinations. The speech stimulus is displayed in black, and the speech-modulated noise masker is overlaid in grey. During testing, these two signals were concurrently presented to a single test ear. The top condition displays continuous speech in the presence of a continuous speech-modulated noise. Examples of the remaining three combinations involving some type of segmental replacement are provided using the preserved vowel stimulus as an example.

transformed to rationalized arcsine units (RAU) to stabilize the error variance (Studebaker, 1985).

E. Spectral shaping and calibration

A calibration noise was first created that matched the long-term average speech spectrum for a concatenation of all of the original sentence files prior to interruption or noise processing. One-third octave band levels for this calibration noise were measured at the output of the earphone to determine the maximum linear output in each band (overall level: 108 dB SPL). Based on this measure, the attenuation factor to obtain a target speech level output of 70 dB SPL was calculated (i.e., -38 dB).

The stimulus files were filtered into one-third octave band frequencies using a 128th order finite impulse response (FIR) filter using a Hamming window. Attenuation was applied individually to each of 20 one-third octave bands from 100 to 8000 Hz to match a target speech presentation level of 70 dB SPL (by attenuating by 38 dB) or to match a target audibility level that did not exceed 20 dB above the individual's hearing threshold within that band, whichever resulted in the smallest attenuation (i.e., highest presentation level). This spectral shaping ensured that one-third octave band speech levels exceeded audiometric thresholds by 15 dB up to at least 4000 Hz (see Fig. 1 for final presentation levels). All speech materials were spectrally shaped after signal processing and the addition of noise. This ensured that the SNR was maintained across all listeners (i.e., both speech and noise were spectrally shaped).

III. RESULTS AND DISCUSSION

A. Continuous speech and noise

Results for the first speech-masker combination, CONT, are displayed in Fig. 3. These results were analyzed using a one-way analysis of variance with group as the between-subjects variable. The analysis demonstrated a significant effect of group, $F(3, 82) = 5.8$, $p = 0.001$. Planned comparisons were completed using independent samples t -tests to examine differences between the groups. Bonferroni corrections were used here, and for all subsequent analyses, to control for multiple comparisons. YNH listeners performed better ($M = 86.8$ RAU, $SD = 8.4$) compared to YSC listeners ($M = 75.1$, $SD = 14.5$), $t(38) = 3.1$, $p < 0.01$. The same was true for ONH ($M = 84.2$, $SD = 13.8$) compared to OHI ($M = 69.4$, $SD = 21.3$), $t(41) = 2.8$, $p < 0.01$. However, no significant differences were obtained between the two normal hearing groups, YNH and ONH, $p > 0.05$, and between the two groups that received the most spectral shaping, OHI and YSC, $p > 0.05$.

These results indicate that for the high variability sentences used in this study, younger and older adults with normal hearing perform similarly when uninterrupted speech is presented concurrently with a continuous speech-modulated masker. This is in contrast to previous investigations that have found age-related declines for speech in the presence of continuous modulated noise (e.g., George *et al.*, 2006). Similar performance of YNH and ONH groups in this study could be due to the presence of some minimal high frequency spectral

FIG. 3. Sentence intelligibility scores (in rau) for continuous speech and speech-modulated noise sources (CONT) for four subject groups. Error bars indicate the standard error of the mean.

shaping in the ONH group, which typically has not been provided in previous studies. Elevated thresholds in the high frequencies are believed to account for some of the variance in older adult performance (e.g., George *et al.*, 2006).

Previous investigations have also demonstrated that when OHI listeners receive spectral shaping, they usually are able to perform as well as their age-matched peers with normal hearing (e.g., Amos and Humes, 2007; Humes and Coughlin, 2009; Fogerty *et al.*, 2012). Instead the current study aligns with the few studies that have failed to demonstrate this effect (e.g., Humes *et al.*, 2007; Lee and Humes, 2012). However, the addition of the YSC group allows us to interpret this finding more clearly. It appears that the poorer performance of the OHI listeners is not simply a result of cochlear pathology that continues to manifest suprathreshold impairments after audibility is experimentally controlled. Rather, the impaired performance may be specific to the spectral shaping procedure as young normal hearing listeners without cochlear pathology, tested under identical spectral conditions (YSC), performed at the same level as the OHI group and lower than the YNH group. Therefore, at least for this condition, spectral shaping may play a role in reduced sentence intelligibility. One possible reason for this adverse effect is that spectral shaping amplifies high frequencies more so than low frequencies; this reduces the consonant-vowel ratio. This may be particularly true for YSC listeners. However, the involvement of this factor is less clear for OHI listeners who have larger-than-normal differences in consonant and vowel levels due to reduced audibility in the higher frequencies. Consonant-vowel differences are restored to values closer to those for normal-hearing listeners with spectral shaping. A second difference with this procedure is an increase in overall stimulus level from 70 dB SPL for YNH listeners to an average of 83 dB SPL for the OHI and YSC listeners. Thus, spectral shaping increased the overall level for both groups and increased the sensation level more so for the YSC than the OHI group. Performance of young normal-hearing listeners may begin to decline at very high presentation levels (Studebaker *et al.*, 1999; Dubno *et al.*,

2005a), particularly for high-frequency speech bands that are emphasized by the spectral shaping (Dubno *et al.*, 2005b). However, performance of OHI listeners can be stable or improve with increases in speech presentation levels up to 87.5 dB SPL due to improved audibility (Summers and Cord, 2007). Therefore, the current stimulus levels appear to be within an acceptable range to avoid most performance decrements with high presentation levels. The remaining experimental tests examined the effect of age, cochlear pathology, and spectral shaping on the ability to use acoustic information present during consonants and vowels under various noise conditions.

B. Use of consonant and vowel cues in continuous speech during the presence of an interrupted speech-modulated masker

The second combination (CONT_{speech}) presented uninterrupted speech continuously with a speech-modulated noise that was selectively presented during either the vowels or the consonants of the target sentence (see Fig. 4). Results for this combination were analyzed using a mixed model analysis of variance (ANOVA) with the segment preserved (consonant or vowel) as the repeated measure and listener group as a between-subjects factor. Results demonstrated a main effect of group, $F(3, 79) = 16.7$, $p < 0.001$, and of segment, $F(1, 79) = 42.5$, $p < 0.001$, with a significant interaction, $F(3, 79) = 9.7$, $p < 0.001$.

Planned-comparison *t*-tests that examined performance differences between the four groups demonstrated no significant differences for the two normal-hearing groups (YNH and ONH) when either the consonants or vowels were presented in quiet (Bonferroni-adjusted $p > 0.05$). For the younger listeners, spectral shaping significantly reduced performance relative to the YNH listeners for the vowel condition, $t(38) = 3.7$, $p = 0.001$, but not for the consonant condition, $t(38) = 0.6$, $p > 0.05$. In contrast to the continuous speech and noise condition, cochlear pathology appeared to play a role that was not explained by reduced audibility.

FIG. 4. Sentence intelligibility scores (in rau) for continuous speech presented during an interrupted speech-modulated noise that preserved either consonant or vowel segments (CONT_{speech}) for four subject groups. Error bars indicate the standard error of the mean.

First, OHI again performed significantly poorer than the ONH group for consonants, $t(41) = 3.4$, $p = 0.001$, and vowels, $t(41) = 4.8$, $p < 0.001$. Second, OHI listeners also performed poorer than YSC listeners under similar spectral conditions for consonants, $t(38) = 4.0$, $p < 0.001$, and vowels, $t(38) = 3.7$, $p = 0.001$.

Performance within groups was also examined using paired-samples *t*-tests to investigate differences in sentence intelligibility when vowels or consonants are presented in quiet. No significant difference between vowel and consonant conditions was observed for either of the two normal hearing groups (YNH and ONH, $p > 0.05$). However, spectral shaping may result in reduced performance when the vowels are preserved and the consonants are masked [YSC: $t(19) = 6.4$, $p < 0.001$; OHI: $t(19) = 5.4$, $p < 0.001$]. Therefore amplification of high frequencies may not always result in improved performance as expected when those high frequency cues (i.e., consonants) are selectively masked by a speech-modulated masker.

The observation of similar performance for consonant and vowel conditions for normal-hearing listeners is notable for several reasons. First, high performance in the vowel condition was observed with relatively low-level consonants presented at unfavorable SNRs (an average of -10 dB). While some acoustic information present during the consonants was likely available, resulting in meaningful consonant confusions (e.g., Phatak *et al.*, 2008), these results are also consistent with the high intelligibility of vowel-only sentences (e.g., Kewley-Port *et al.*, 2007). Second, the similarly high performance in the consonant condition, where the local SNR during vowel segments was estimated at 0 dB SNR, may indicate the relative preservation of cues during the vowel and during transitional information at the consonant-vowel boundary that significantly contributes to intelligibility (e.g., Chen and Loizou, 2012; Lee and Kewley-Port, 2009).

Previous studies have implemented testing similar to the consonant condition here using IEEE sentences in 20-talker babble for normal hearing listeners (Li and Loizou, 2008) and older listeners with normal and impaired hearing (Saripella *et al.*, 2011). Across both studies, authors found that preservation of obstruent consonants significantly contributes to intelligibility, providing gains of 20–40 percentage points compared to overall scores for speech in continuous noise at -5 dB SNR. Much of the contribution of obstruent consonants in those studies was isolated to the benefit of low-frequency information and voicing feature transmission in addition to possible contributions of vowel transitions. Therefore, similar contributions from voicing may have played a role in the current study, particularly with the additional inclusion here of sonorant consonants.

C. Use of consonant and vowel cues in interrupted speech during the presence of a continuous speech-modulated masker

This third combination (INT_{speech}) investigated interrupted speech that preserved either the consonants or the vowels and presented them concurrently with a continuous speech-modulated noise masker. Therefore, these highly

degraded conditions presented an interrupted speech signal, with the preserved portions masked by noise. Listener performance on this combination is displayed in Fig. 5. Results were again analyzed using a mixed model ANOVA with the segment preserved (consonant or vowel) as the repeated measure and listener group as a between-subjects factor. Results demonstrated a main effect of group, $F(3, 79) = 9.4$, $p < 0.001$, and of segment, $F(1, 79) = 680.5$, $p < 0.001$, with a significant interaction, $F(3, 79) = 3.2$, $p < 0.05$.

Performance between the groups was again compared. Results demonstrated an effect of age for the consonant conditions with older adults performing significantly poorer than their younger counterparts [YNH vs ONH: $t(41) = 2.7$, $p < 0.01$; YSC vs OHI: $t(38) = 6.3$, $p < 0.001$]. However, no effect of age was observed when the vowels were preserved (Bonferroni-adjusted $p > 0.05$). The presence of suprathreshold deficits due to cochlear pathology appears to exacerbate difficulties processing the preserved consonants in the presence of a continuous speech-modulated masker as OHI listeners also performed significantly poorer than ONH listeners only for the consonant condition, $t(41) = 4.7$, $p < 0.001$. Particularly notable across all of these comparisons is the preserved ability to use vowel cues. The only comparison that resulted in poorer vowel performance was between YNH and YSC groups, demonstrating a possible impact of spectral shaping on using available vowel cues, $t(38) = 3.0$, $p < 0.01$. This is consistent with earlier results when the speech was presented continuously, but noise was used to selectively mask the consonants and preserve the vowels.

Within-group tests also examined the relative difference between consonants and vowels for the four different listener groups. All 83 listeners performed better when vowels were preserved and presented in continuous speech-modulated noise compared to when the consonants were preserved ($p < 0.001$). This result is likely due to the naturally lower intensity level of consonants, which resulted in a greater degree of energetic masking. The size of this difference between consonant and vowel performance was large and

consistent across groups (Cohen's $d = 1.6$ – 1.8). Groups showed about a 31–34 percentage-point improvement for the vowel condition with somewhat reduced difference scores for the YSC listeners (24 percentage points). This appears to be related to poorer performance on the vowel condition by the YSC group.

D. Use of consonant and vowel cues in interrupted speech alternated with an interrupted speech-modulated masker

This final experimental combination removed the contribution of energetic masking by alternating speech and noise sources ($INT_{\text{speech\&noise}}$). In these conditions, either vowel or consonant segments were preserved with the other segment type replaced with the speech-modulated noise. Group performance is displayed in Fig. 6. Following the statistical analysis procedures of the earlier speech-masker combinations, results of the ANOVA demonstrated a main effect of group, $F(3, 79) = 21.6$, $p < 0.001$, and of segment, $F(1, 79) = 20.2$, $p < 0.001$, with a significant interaction, $F(3, 79) = 5.8$, $p = 0.001$.

Comparisons between groups demonstrated poorer older adult performance for consonants [YNH vs ONH: $t(41) = 6.8$, $p < 0.001$, Cohen's $d = 1.4$; YSC vs OHI: $t(38) = 6.6$, $p < 0.001$, $d = 1.4$] and vowels [YNH vs ONH: $t(41) = 2.4$, $p < 0.05$, $d = 0.7$; YSC vs OHI: $t(38) = 3.1$, $p < 0.01$, $d = 0.9$]. While significant age effects were observed for these modulated masking conditions, smaller (i.e., medium) effect sizes were observed for vowel conditions compared to the large effect sizes for consonant conditions. This trend is consistent with greater age-related declines for consonant as compared to vowel conditions (Fogerty *et al.*, 2012). OHI listeners also performed more poorly than ONH listeners [consonant: $t(41) = 4.0$, $p < 0.001$, $d = 1.0$; vowel: $t(41) = 2.9$, $p < 0.01$, $d = 0.8$]. In contrast with the previous two speech-masker combinations that demonstrated a negative effect of spectral shaping on the use of vowel cues, YSC listeners performed more poorly

FIG. 5. Sentence intelligibility scores (in rau) for speech that was interrupted to preserve either consonant or vowel segments and presented concurrently with a continuous speech-modulated noise (INT_{speech}) for four subject groups. Error bars indicate the standard error of the mean.

FIG. 6. Sentence intelligibility scores (in rau) for alternating speech and speech-modulated noise sources that preserved either the consonant or vowel segments of the target sentence ($INT_{\text{speech\&noise}}$) for four subject groups. Error bars indicate the standard error of the mean.

than YNH listeners for the consonant condition, $t(38) = 3.5$, $p = 0.001$, $d = 1.0$, but not the vowel condition ($p > 0.05$).

Comparisons within-groups demonstrated no difference for vowel versus consonant performance for the young listeners ($p > 0.05$). However, both older listener groups had better performance for vowel compared to consonant conditions [ONH: $t(22) = 5.6$, $p < 0.001$; OHI: $t(19) = 3.2$, $p < 0.01$].

Taken together, these results suggest that while older adults appear to have poorer performance overall, their ability is best preserved for the vowel conditions. Suprathreshold factors may impact older listeners' ability to process vowel cues. However, altering the spectrum of the speech signal also appears to have an adverse effect on processing preserved consonant cues in this alternating speech and noise condition for younger (YSC) listeners. Furthermore, while spectral shaping likely provided some benefit for older (OHI) listeners due to improved audibility (Humes, 2007), this benefit was not enough to account for all performance differences between ONH and OHI listeners. This may indicate other factors associated with cochlear pathology that limited OHI performance or factors associated with the spectral shaping that may have limited improvements from what would be expected from increased audibility alone. However, it is important to note that even non-individualized high-frequency amplification provides significant advantages to speech recognition for hearing-impaired listeners (Horwitz *et al.*, 2008).

This speech-masker combination also presents an interesting comparison with results of previous investigations that have replaced vowels or consonants with noise (e.g., Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009). These previous studies used a low-level (e.g., 16 dB SNR) speech-shaped noise for replacement and obtained better performance for sentences that preserved only the vowels compared sentences that preserved only the consonants by a factor of 2. This speech-masker combination in the current experiment modeled this replacement paradigm. However, instead of using an unmodulated "steady-state" noise, the current study modulated the replacement noise with the temporal envelope of a competing talker; the level of the noise was presented at 0 dB SNR. Performance for YNH listeners was markedly different in the current study compared to that in previous investigations. Performance was at 65%–74% for vowels and 31%–34% for consonants in these previous studies compared to the 37% and 39% performance levels for vowels and consonants, respectively, in the current study.

The dramatic reduction in performance for the vowel condition is possibly due to the introduction of contradictory temporal modulations of the replacement noise as listeners appear to use temporal amplitude modulations of replacement noise for sentence recognition (Shinn-Cunningham and Wang, 2008), but only if these modulations come from vowels (Fogerty, 2014; at least for the modulation rates investigated here). This may reflect a form of modulation interference where listeners have greater difficulty using vowel temporal modulations when interleaved with competing modulations of a different talker.

E. Correlations among the measures

This study used several different noisy listening conditions to investigate how listeners are able to use consonant and vowel cues for sentence intelligibility. Of interest in this study is whether performance on these conditions reflects general processing abilities or whether performance is specific to the stimulus characteristics of the individual condition. This has large implications for the generalizability of these results to everyday listening conditions. Of particular interest was how performance on the individual segmental conditions relates to performance for continuous speech presented concurrently with a continuous speech-modulated noise (CONT). This most closely models natural listening conditions of speech-on-speech masking. To assess these associations, correlations were first investigated among the seven experimental conditions, along with age and high-frequency pure-tone averages at 1, 2, and 4 kHz (PTA). Due to significant group differences, all correlations were run separately for the four different listener groups. Age was not significantly correlated with any of the measures. In contrast, PTA was significantly correlated with five of the seven measures for the OHI group ($r = -0.47$ to -0.64 , $p < 0.05$), excluding CONT and the vowel INT_{speech} conditions. PTA was not significantly correlated with age ($r = -0.13$, $p > 0.05$). Due to the association of PTA with the experimental measures for this group, PTA was used as a covariate, and partial correlations were conducted for this group. No other group had significant correlations with PTA except the YNH group for one condition (vowel INT_{speech&noise}: $r = -0.48$, $p = 0.03$). However, PTA had a minimal effect on overall performance for this group and was therefore not entered into the statistical model for the YNH group.

Results from the correlational analysis are displayed in Table I. As can be observed, the YNH and OHI groups demonstrated correlations among several of the vowel conditions and correlations of the vowel conditions with the continuous speech in noise condition (CONT). Few significant correlations were observed with the consonant conditions. In contrast, YSC and ONH groups demonstrated significant correlations across all of the measures. However, these correlations appear strongest for the vowel and CONT conditions and weaker for the consonant conditions. Note that significant correlations were obtained for the consonant condition when the speech was continuous. However, this condition also preserved the vowels of the sentence, albeit noise masked, and can therefore be thought of as a similar speech-in-noise condition like the CONT condition.

Due to the intercorrelations between the vowel and consonant conditions, it may be that some shared variance between these measures led to the elevated correlations. Therefore, correlations were also calculated from composite vowel and consonant measures that pooled performance estimates for the highly related conditions. Toward this end, average vowel and average consonant scores were calculated over the three different speech-masker combinations. Two sets of partial correlations were then calculated to determine the association between the average vowel scores and the CONT condition, controlling for shared variance on the consonant condition; and between average consonant scores and the CONT conditions, controlling for performance on the

TABLE I. Correlations among the seven experimental measures are displayed for the four listener groups. Italic text above the diagonal indicates the two listener groups that received spectral shaping (YSC, OHI). Non-italic text below the diagonal indicates the two listener groups with normal-hearing who received minimal or no spectral shaping (YNH, ONH). * $p < 0.05$, ** $p < 0.01$.

		Vowel				Consonant		
		CONT	CONT _{speech}	INT _{speech&noise}	INT _{speech}	CONT _{speech}	INT _{speech&noise}	INT _{speech}
(a) Younger listener correlations								
Vowel	CONT		0.844**	0.610**	0.747**	0.681**	0.558*	0.539*
	CONT _{speech}	0.456*		0.644**	0.752**	0.788**	0.648**	0.686**
	INT _{speech&noise}	0.409	0.361		0.745**	0.559*	0.711**	0.495*
	INT _{speech}	0.574**	0.502*	0.576**		0.553*	0.518*	0.551*
Consonant	CONT _{speech}	0.185	0.170	0.497*	0.213		0.636**	0.642**
	INT _{speech&noise}	0.387	0.217	0.490*	0.063	0.230		0.778**
	INT _{speech}	0.146	0.135	0.263	0.445*	0.180	0.217	
(b) Older listener correlations								
Vowel	CONT		0.903**	0.776**	0.878**	0.775**	0.374	0.104
	CONT _{speech}	0.888**		0.808**	0.951**	0.811**	0.355	0.141
	INT _{speech&noise}	0.874**	0.829**		0.794**	0.736**	0.442	−0.084
	INT _{speech}	0.812**	0.835**	0.813**		0.736**	0.184	0.076
Consonant	CONT _{speech}	0.768**	0.786**	0.797**	0.659**		0.318	0.110
	INT _{speech&noise}	0.715**	0.615**	0.727**	0.686**	0.675**		0.061
	INT _{speech}	0.458*	0.483*	0.596**	0.457*	0.524*	0.489*	

vowel conditions (see Table II). Note that partial correlations also included PTA for the OHI listeners, but this resulted in little additional explanation of the variance. Results demonstrated that when the shared variance on the consonant conditions was controlled for, average performance on the vowel conditions was strongly associated with performance on CONT for all four listener groups ($r = 0.52$ – 0.78 , $p < 0.05$). However, when the shared variance on the vowel conditions was controlled, partial correlations of average consonant performance and CONT performance were not significantly correlated for any of the four experimental groups ($r = 0.10$ – 0.26 , $p > 0.05$). Therefore, there may be some shared variance between vowel and consonant conditions, potentially indicating a general ability to process speech with only partial information. However, the ability to process speech information contained during vowel segments appears to be particularly associated with performance when listening to uninterrupted speech in the presence of a speech-modulated masker.

TABLE II. Partial correlations of the mean consonant (C_{mean}) and mean vowel (V_{mean}) scores with the CONT condition that presented continuous speech in speech-modulated noise. Correlations were performed by controlling for the other mean score (i.e., C_{mean} correlations controlled for V_{mean} scores). Partial correlations also controlled for high-frequency PTA for the OHI group (in parentheses). Significant correlations demonstrate an association between average segment scores and the abilities to process speech in modulated noise. * $p < 0.05$, ** $p < 0.01$.

	C_{mean}	V_{mean}
YNH	0.10	0.52*
YSC	0.16	0.61**
ONH	0.13	0.77**
OHI	0.15 (0.26)	0.77 (0.78)**

Given the high variability of these speech materials (i.e., unpredictable sentences spoken by unfamiliar talkers with varying dialects), variability in listener performance within a given condition was expected. Therefore the systematic variance to be explained for performance in the CONT condition was determined by calculating the Pearson correlation in listener performance between even and odd trials (YNH: $r = 0.54$, $p < 0.05$; YSC: $r = 0.68$, $p < 0.01$; ONH: $r = 0.69$, $p < 0.01$; OHI: $r = 0.91$, 0.90 controlling for PTA, $p < 0.01$). It is notable that the majority of the systematic variance on this naturalistic listening condition (i.e., listening to speech during a continuous speech-modulated masker) is explained by mean performance on the vowel conditions.

Figure 7 displays scatterplots for the partial correlations of vowel and consonant mean performance with performance on the CONT condition. Across all groups, performance on the vowel conditions accounted for a significant proportion of the variance on this continuous speech-in-noise condition. Therefore, understanding how listeners successfully process vowel information during sentences is central to understanding how listeners process sentences in common listening environments, such as when speech is masked by a single competing talker.

IV. GENERAL DISCUSSION

This study investigated the speech recognition abilities of younger and older adults in the presence of interruption and/or fluctuating noise backgrounds. All listeners were provided spectral shaping to ensure audibility through at least 4 kHz. Given the normal hearing thresholds of YNH and ONH listeners, this resulted in little to no spectral shaping (some minimal spectral shaping was provided above 4 kHz for the ONH group due to elevated audiometric thresholds at

FIG. 7. Results for the partial correlational analysis that explains the independent contribution of mean vowel or mean consonant scores to performance for continuous speech, which was presented in continuous speech-modulated noise (CONT). For this figure, residual scores were calculated independently for all four groups and then collapsed for linear regression across the four groups. Nearly 50% of the variance on the continuous speech condition is accounted for by vowel scores, controlling for consonant performance; little variance is explained by consonant performance, controlling for vowel performance.

those frequencies). In contrast, OHI and YSC listeners received spectral shaping (increased sensation levels) and were tested on acoustically equivalent speech materials. The comparison of these groups allowed for the investigation of age, hearing loss, and spectral shaping on sentence intelligibility scores.

A. Effect of age

Comparison of the YNH and ONH groups demonstrated a detrimental effect of age for two of the seven experimental conditions, even when accounting for minor elevations in audiometric thresholds in the higher frequencies for ONH listeners. The two conditions that demonstrated poorer sentence recognition for older listeners (INT_{speech} and $INT_{\text{speech\&noise}}$) involved processing interrupted speech based on preserved consonant segments. All other conditions presented speech with at least some partially preserved vowel segments. The results suggest that older listeners may have increased difficulty in processing the rapid, transient cues of consonants when speech is interrupted. These

findings are consistent with Fogerty *et al.* (2012) who found poorer performance for older listeners with interrupted sentences that preserved consonants, but no performance declines for sentences that preserved vowels. This finding is also consistent with George *et al.* (2006), who found that age was significantly associated with reduced performance in modulated backgrounds. The other factor related to performance in the study by George *et al.* (2006) was temporal resolution, particularly in the higher stimulus frequencies. This would be particularly associated with the transient high-frequency content of consonants that contain faster amplitude modulation rates than vowels (Drullman *et al.*, 1994). Vowels, which provide low-frequency, slow-varying temporal modulations, contain temporal rates that unfold across the sentence at syllabic and phrase levels. These amplitude modulation rates may be particularly important for processing contextual information in sentences (Fogerty *et al.*, 2012; Dille and Pitt, 2010) such as that provided by the grammatical structure of the sentence (Sharpe *et al.*, 2014). The finding that older adults performed as well as younger adults when vowels were preserved, which may preserve slow-rate amplitude modulated cues important for processing contextual information in sentences, is consistent with several other studies demonstrating equivalent or better use of supporting linguistic context by older listeners relative to young listeners (e.g., Wingfield *et al.*, 1991; Wingfield *et al.*, 1994; Pichora-Fuller *et al.*, 1995; Humes *et al.*, 2007).

B. Effect of cochlear pathology

OHI listeners who received spectral shaping that ensured audibility consistently performed more poorly than ONH listeners. These results highlight the possibility that suprathreshold deficits may underlie some of the difficulty OHI listeners face in modulated-noise backgrounds. It is consistent with the finding that individuals with mild hearing loss have more difficulty than normal hearing listeners in modulated noise backgrounds (Takahashi and Bacon, 1992; Dubno and Ahlstrom, 1997; Lorenzi *et al.*, 2006). Festen and Plomp (1990) discussed three determinants of poorer performance of hearing impaired listeners in modulated background noises: audibility, temporal resolution, and comodulation masking release. Audibility, arguably the primary determinant of performance declines by OHI listeners (Humes, 2007), was explicitly controlled for by the spectral shaping procedure used in the current study. While this procedure generally controls for audibility (Humes, 2013), there remains the possibility of elevated thresholds in the high frequencies (above 4 kHz) that still contributed to performance. However, whereas audiometric thresholds (PTA) were correlated with performance, this measure did not adequately explain the systematic variance in OHI performance across the experimental conditions (i.e., performance was highly correlated even after controlling for PTA). Therefore, factors more related to poorer temporal resolution and/or masking release are likely to be involved (see also Gustafsson and Arlinger, 1994), particularly as this study employed two types of amplitude modulated interference: the speech-modulated fluctuating masker and the interrupted speech

signal. It is possible that the temporally complex nature of these conditions, with multiple interruption and modulation rates, poses particular difficulty for older hearing impaired listeners. This would be consistent with findings suggesting poorer performance in temporally complex listening conditions (e.g., [Gordon-Salant and Fitzgibbons, 1999](#)).

C. Effect of spectral shaping

It is somewhat surprising that providing spectral shaping, i.e., increasing the sensation level, for younger normal hearing listeners resulted in a decrement in performance (YSC compared to YNH). YSC listeners performed more poorly than YNH listeners when the speech and masker were continuous, for consonant conditions when the speech and masker alternated in presentation, and for vowel conditions when the consonants of continuous sentences were selectively masked by an interrupted masker. Therefore, it is possible that the negative consequences of spectral shaping may have also contributed to poorer performance of OHI relative to ONH listeners. This group effect among the older listeners was not observed by [Fogerty et al. \(2012\)](#) for segmental interruption conditions that used a steady-state masker.

Ensuring audibility of these materials undoubtedly improved performance for the OHI listeners overall and controlled for the negative consequences of reduced audibility ([Humes, 2007](#)). This can be observed by the minimal predictive power of PTA in explaining OHI performance. While the poorer performance in the OHI and YSC groups could be a result of individual differences in performance ([Humes, 2013](#)), it is important to understand potential factors that may limit benefit in select conditions. These are reviewed in Secs. [IV C 1–IV C 3](#).

1. Effect of spectral shaping on the consonant-vowel intensity ratio

The spectral shaping procedure implemented here increased the sensation levels of high-frequency bands more than low-frequency bands, which would reduce the consonant-vowel intensity ratio (as consonant cues mainly comprise the high frequency bands). This can be clearly seen from [Fig. 1](#) where the spectral levels for the normal-hearing groups gradually decline in amplitude for the high frequencies, while an increase in amplitude in the high-frequencies above 2 kHz (i.e., reduced consonant-vowel ratio) is observed after spectral shaping for the hearing-impaired group. Not only is the consonant-vowel ratio potentially important for speech recognition ([Gordon-Salant, 1986](#); [Montgomery and Edge, 1988](#)), altering the ratio also results in a reduction in the modulation depth of the temporal envelope. This can have significant negative consequences in noisy listening conditions ([Sammath et al., 1999](#); [Fogerty et al., 2014](#)) and is particularly important for the identification of some consonants ([Hickson and Byrne, 1997](#)).

To illustrate the potential for the temporal effect of spectral shaping, a consonant-vowel syllable (i.e., “fee”) was processed according to the spectral shaping provided to the YSC group based upon the mean audiometric thresholds for the OHI listeners in this study. [Figure 8](#) displays the running

rms level of 16-ms non-overlapping windows relative to the mean rms of the stimulus. The solid and dashed lines indicate the relative rms changes across the syllable for naturally spoken and spectrally shaped versions, respectively. Noting the separation of the solid and dashed lines, spectral shaping effectively reduced the consonant-vowel ratio by 5.6 dB, largely through increasing the level of the consonant relative to a constant vowel level. Even minor consonant-vowel ratio reductions of 3 dB have been shown to influence the speed of lexical decisions ([Fogerty et al., 2014](#)). For this stimulus, it also had the effect of providing a more gradual onset to the stimulus. Both of these features could have important implications for onset detection, which is believed to be important for processing speech (e.g., [Fitzgibbons et al., 2007](#)).

2. Effect of spectral shaping on the temporal envelope modulation

Amplitude changes of the temporal envelope have been shown repeatedly to be essential for sentence intelligibility (e.g., [Dilley and Pitt, 2010](#); [Shannon et al., 1995](#); [Shannon et al., 1998](#); [Fogerty, 2014](#); [Stilp, 2014](#)). However, reduction of the envelope modulation depth may result in particularly challenging conditions when in the presence of other amplitude modulated maskers, such as in the current study. Indeed, the interrupted conditions that resulted in poorer performance for the groups with spectral shaping (OHI and YSC) involve the conditions where spectral shaping (i.e., consonant amplification) would result in providing the least change in amplitude modulation across the sentence. For example, consider when speech was continuous, but the masker was interrupted to be present only during the consonants (i.e., vowel condition) ([Fig. 4](#)). Spectral shaping would increase the level of the consonant. This increased level, combined with the additional noise present only during the consonant intervals, would reduce amplitude modulation across the sentence. No deleterious effect of spectral shaping for OHI listeners was observed in [Fogerty et al. \(2012\)](#). This

FIG. 8. Example effect of spectral shaping on the relative amplitudes of speech for the syllable “fee.” The solid line depicts the amplitude contour for the naturally produced syllable, and the dashed line depicts the amplified stimulus.

may have been due to the use of a low noise level relative to the speech level, which would have a minimal impact on the overall temporal envelope. Thus, it is conceivable that spectral shaping, while implemented in the frequency domain, may have significant implications for the temporal structure of speech due to the alternating consonant and vowel segments that largely define the temporal envelope of the sentence. Indeed, amplitude changes at the consonant-vowel boundary are significant predictors of intelligibility (Chen and Loizou, 2012).

3. Effect of spectral shaping on presentation level

The rationale for the use of spectral shaping for the OHI group is to increase sensation levels at frequencies with elevated audiometric thresholds. Providing this form of frequency-selective amplification is predicted to minimize performance declines associated with decreased audibility (Humes, 2007), but it also increases overall presentation level. There is a long history of research documenting impaired speech intelligibility for normal-hearing and hearing-impaired listeners at high presentation levels (e.g., French and Steinberg, 1947; Hawkins and Stevens, 1950; Pollack and Pickett, 1958; Studebaker *et al.*, 1999; Dubno *et al.*, 2005a; Summers and Cord, 2007). Such reduced performance has also been associated with high-frequency emphasis in hearing aids (e.g., Ching *et al.*, 1998; Hogan and Turner, 1998; Molis and Summers, 2003). Thus, it is possible that the presentation levels of the high frequency “consonant” bands contributed to reduced performance for listeners with the most spectral shaping, and particularly for YSC listeners who already had adequate sensation levels in these frequency bands. However, the 83 dB SPL presentation level for OHI and YSC listeners in the present study is below the range that such presentation level effects are typically seen for OHI listeners (Summers and Cord, 2007).

D. Relative contributions of consonants and vowels

Investigation of correlations among the seven experimental conditions demonstrated a strong consistency across listener groups for separable perceptual abilities associated with processing sentences with preserved consonants or vowels. Stronger correlations were found for conditions that preserved vowel segments to some degree (see Table I). Given these segment-related associations, the mean consonant and vowel scores were used to determine how performance with speech information preserved under those degraded conditions related to general speech intelligibility during speech-modulated maskers. Results clearly demonstrated a stronger association of listeners’ performance on vowel conditions with overall speech intelligibility in these natural speech-on-speech-modulated noise conditions. These results suggest that access to vowel cues is important. A listener’s ability to process those preserved cues is a significant predictor for sentence intelligibility in complex listening environments, accounting for nearly half of the variance in performance across all listener groups.

Access to phonetic information of consonants is important for intelligibility and possibly could be a limiting factor

when audibility is not ensured. This is particularly true as OHI listeners have the most difficulty with the consonant conditions. In contrast, ONH listeners had preserved performance, relative to YNH listeners, for vowel conditions in this study (and OHI listeners in steady-state noise conditions observed by Fogerty *et al.*, 2012). Thus, these vowel cues appear to be essential for intelligibility in fluctuating backgrounds, possibly in part due to their overall higher intensity. Fortunately, older adults have preserved processing for these cues, which these listeners rely on to a greater extent, as observed by the high correlation coefficients for vowel conditions in these groups.

V. SUMMARY AND CONCLUSIONS

The results from this study demonstrate that when adequate audibility is ensured, such as through spectral shaping, older listeners are able to adequately understand sentences presented in temporally complex backgrounds (i.e., during interrupted and/or modulated noise masking). Performance for older adults is most preserved when high level portions of the speech signal are preserved (such as vowels). However, performance for older adults is reduced more compared to younger adults when they must rely on high-frequency consonant cues. Older adults with hearing loss had consistently poorer performance across most conditions compared to their age-matched normal-hearing counterparts. Some of this decline may be related to suprathreshold auditory abilities, such as poorer temporal resolution and masking release. However, the effect of these reduced abilities may be compounded due to interactions with spectral shaping used to increase sensation levels. Analysis of the results for younger listeners who received spectral shaping equivalent to that of the older hearing impaired group also demonstrated poorer performance for some select conditions. These conditions highlight the potential for spectral shaping to reduce the modulation depth of the temporal envelope. This spectral shaping reduces amplitude cues at the transitions between consonants and vowels, which is a strong predictor of intelligibility (Chen and Loizou, 2012). Decreasing these modulation cues will likely make it even more difficult for hearing impaired listeners, who may have reduced modulation masking release and temporal resolution, to process the informative temporal envelope of the speech signal, particularly in the context of a competing modulated masker.

Overall, results demonstrate the importance of access to vowel cues, particularly the amplitude modulation imposed by these high level intervals of speech. Performance with these cues was strongly associated with speech understanding in more ecological conditions of continuous speech in the presence of a competing speech-modulated masker. Importantly, older adults have relatively preserved abilities in processing this type of segmental information. In contrast, a reduction in the use of consonant cues may limit performance for older listeners, particularly those with hearing loss, even when audibility may be assured with frequency-specific amplification. However, processing of consonant cues appears specific to particular requirements of the speech and masker combination and was much less associated with

performance in the speech-in-modulated-noise condition, which better reflects real-world listening.

ACKNOWLEDGMENTS

This work was supported, in part, by an Advancing Academic-Research Career Award through the American Speech-Language-Hearing Association and by National Institutes of Health/NIDCD Grant Nos. R03 DC012506 (D.F.) and R01 DC000184 (J.R.D.).

- Amos, N. E., and Humes, L. E. (2007). "Contribution of high frequencies to speech recognition in quiet and noise among listeners with varying degrees of high-frequency sensorineural hearing loss," *J. Speech Lang. Hear. Res.* **50**, 819–834.
- ANSI (1997). S3.5, *American National Standard Methods for the Calculation of the Speech Intelligibility Index* (Acoustical Society of America, New York).
- Boettcher, F. A., Poth, E. A., Mills, J. H., and Dubno, J. R. (2001). "The amplitude-modulation following response in young and aged human subjects," *Hear. Res.* **153**, 32–42.
- Chen, F., and Loizou, P. C. (2012). "Contributions of cochlea-scaled entropy and consonant-vowel boundaries to prediction of speech intelligibility in noise," *J. Acoust. Soc. Am.* **131**, 4104–4113.
- Chen, F., Wong, L. L., and Wong, E. Y. (2013). "Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility," *J. Acoust. Soc. Am.* **134**, EL178–EL184.
- Ching, T., Dillon, H., and Byrne, D. (1998). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," *J. Acoust. Soc. Am.* **103**, 1128–1140.
- 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.
- Dilley, L. C., and Pitt, M. A. (2010). "Altering context speech rate can cause words to appear or disappear," *Psychol. Sci.* **21**, 1664–1670.
- Drullman, R., Festen, J. M., and Plomp, R. (1994). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670–2680.
- Dubno, J. R., and Ahlstrom, J. B. (1997). "Additivity of multiple maskers of speech," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 253–272.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2003). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing," *J. Acoust. Soc. Am.* **113**, 2084–2094.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2005a). "Word recognition in noise at higher-than-normal levels: Decreases in scores and increases in masking," *J. Acoust. Soc. Am.* **118**, 914–922.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2005b). "Recognition of filtered words in noise at higher-than-normal levels: Decreases in scores with and without increases in masking," *J. Acoust. Soc. Am.* **118**, 923–933.
- Festen, J. M., and Plomp, R. (1990). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.* **88**, 1725–1736.
- Fitzgibbons, P. J., Gordon-Salant, S., and Barrett, J. (2007). "Age-related differences in discrimination of an interval separating onsets of successive tone bursts as a function of interval duration," *J. Acoust. Soc. Am.* **122**, 458–466.
- Fogerty, D. (2011). "Perceptual weighting of the envelope and fine structure across frequency for sentence intelligibility: Effect of interruption at the syllabic-rate and periodic-rate of speech," *J. Acoust. Soc. Am.* **130**, 489–500.
- Fogerty, D. (2013). "Acoustic predictors of intelligibility for segmentally interrupted speech: Temporal envelope, voicing, and duration," *J. Speech Lang. Hear. Res.* **56**, 1402–1408.
- Fogerty, D. (2014). "Importance of envelope modulations during consonants and vowels in segmentally interrupted sentences," *J. Acoust. Soc. Am.* **135**, 1568–1576.
- Fogerty, D., and Chen, F. (2014). "Vowel spectral contributions to English and Mandarin sentence intelligibility," in *Proceedings of the 15th Annual Conference of the International Speech Communication Association (InterSpeech)*, Singapore, pp. 499–503.
- 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. (2012a). "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 Humes, L. E. (2012b). "A correlational method to concurrently measure envelope and temporal fine structure weights: Effects of age, cochlear pathology, and spectral shaping," *J. Acoust. Soc. Am.* **132**, 1679–1689.
- Fogerty, D., and Kewley-Port, D. (2009). "Perceptual contributions of the consonant-vowel boundary to sentence intelligibility," *J. Acoust. Soc. Am.* **126**, 847–857.
- Fogerty, D., Kewley-Port, D., and Humes, L. E. (2012). "The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss," *J. Acoust. Soc. Am.* **132**, 1667–1678.
- Fogerty, D., Montgomery, A. A., and Crass, K. A. (2014). "Effect of initial-consonant intensity on the speed of lexical decisions," *Atten. Percept. Psychophys.* **76**, 852–863.
- French, N., and Steinberg, J. (1947). "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* **19**, 90–119.
- Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., and Dahlgren, N. L. (1993). *The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus* (Linguistic Data Consortium, Philadelphia).
- George, E. L., Festen, J. M., and Houtgast, T. (2006). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **120**, 2295–2311.
- George, E. L., Zekveld, A. A., and Kramer, S. E. (2007). "Auditory and non-auditory factors affecting speech reception in noise by older listeners," *J. Acoust. Soc. Am.* **121**, 2362–2375.
- Gilbert, G., Bergeras, I., Voillery, D., and Lorenzi, C. (2007). "Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues," *J. Acoust. Soc. Am.* **122**, 1336–1339.
- Gilbert, J. L., Tamati, T. N., and Pisoni, D. B. (2013). "Development, reliability, and validity of PRESTO: A new high-variability sentence recognition test," *J. Am. Acad. Audiol.* **24**, 26–36.
- Gnansia, D., Jourdes, V., and Lorenzi, C. (2008). "Effect of masker modulation depth on speech masking release," *Hear. Res.* **239**, 60–68.
- 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.
- Gordon-Salant, S., and Fitzgibbons, P. J. (1999). "Profile of auditory temporal processing in older listeners," *J. Speech Lang. Hear. Res.* **42**, 300–311.
- Gustafsson, H. Å., and Arlinger, S. D. (1994). "Masking of speech by amplitude-modulated noise," *J. Acoust. Soc. Am.* **95**, 518–529.
- Hawkins, J., and Stevens, S. (1950). "The masking of pure tones and of speech by white noise," *J. Acoust. Soc. Am.* **22**, 6–13.
- Hickson, L., and Byrne, D. (1997). "Consonant perception in quiet: Effect of increasing consonant to vowel ratio with compression amplification," *J. Am. Acad. Audiol.* **8**, 222–232.
- Hogan, C., and Turner, C. (1998). "High-frequency audibility: Benefits for hearing-impaired listeners," *J. Acoust. Soc. Am.* **104**, 432–441.
- Horwitz, A. R., Ahlstrom, J. B., and Dubno, J. R. (2008). "Factors affecting the benefits of high-frequency amplification," *J. Speech Lang. Hear. Res.* **51**, 798–813.
- 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**, 590–603.
- Humes, L. E. (2013). "Understanding the speech-understanding problems of older adults," *Am. J. Audiol.* **22**, 303–305.
- Humes, L. E., Burk, M., Coughlin, M. P., Busey, T. A., and Strauser, L. E. (2007). "Auditory speech-recognition and visual text-recognition measures in younger and older adults: Similarities and differences between modalities and the effects of presentation rate," *J. Speech Lang. Hear. Res.* **50**, 283–303.
- Humes, L. E., and Coughlin, M. P. (2009). "Aided speech-identification performance in single-talker competition by elderly hearing-impaired listeners," *Scand. J. Psych.* **50**, 485–494.
- 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, T. Z., and Lee, J. H. (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.

- Kwon, B. J., and Turner, C. W. (2001). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," *J. Acoust. Soc. Am.* **110**, 1130–1140.
- Ladefoged, P., and Disner, S. F. (2012). *Vowels and Consonants*, 3rd ed. (Wiley and Sons, New York).
- Lee, J. H., and Humes, L. E. (2012). "Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background," *J. Acoust. Soc. Am.* **132**, 1700–1717.
- Lee, J. H., and Kewley-Port, D. (2009). "Intelligibility of interrupted sentences at subsegmental levels in young normal-hearing and elderly hearing-impaired listeners," *J. Acoust. Soc. Am.* **125**, 1153–1163.
- Li, N., and Loizou, P. (2008). "The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise," *J. Acoust. Soc. Am.* **124**, 3947–3958.
- Lorenzi, C., Husson, M., Ardoint, M., and Debrulle, X. (2006). "Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception," *Int. J. Audiol.* **45**, 487–495.
- Molis, M. R., and Summers, V. (2003). "Effects of high presentation levels on recognition of low- and high-frequency speech," *ARLO* **4**, 124–128.
- Montgomery, A. A., and Edge, R. A. (1988). "Evaluation of two speech enhancement techniques to improve intelligibility for hearing-impaired adults," *J. Speech Hear. Res.* **31**, 386–393.
- Nelson, P. B., and Jin, S.-H. (2004). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **115**, 2286–2294.
- Nespor, M., Peña, M., and Mehler, J. (2003). "On the different roles of vowels and consonants in speech processing and language acquisition," *Ling. Linguaggio* **2**, 201–227.
- Peelle, J. E., and Davis, M. H. (2012). "Neural oscillations carry speech rhythm through to comprehension," *Front. Psych.* **3**, 320.
- Peters, R. W., and Hall, J. W. (1994). "Comodulation masking release for elderly listeners with relatively normal audiograms," *J. Acoust. Soc. Am.* **96**, 2674–2682.
- Phatak, S. A., Lovitt, A., and Allen, J. B. (2008). "Consonant confusions in white noise," *J. Acoust. Soc. Am.* **124**, 1220–1233.
- 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.
- Poeppel, D. (2014). "The neuroanatomic and neurophysiological infrastructure for speech and language," *Curr. Opin. Neurobiol.* **28**, 142–149.
- Pollack, I., and Pickett, J. (1958). "Masking of speech by noise at high sound levels," *J. Acoust. Soc. Am.* **30**, 127–130.
- Port, R. F. (2003). "Meter and speech," *J. Phon.* **31**, 599–611.
- Rosen, S. (1992). "Temporal information in speech: Acoustic, auditory, and linguistic aspects," *Philos. Trans. R. Soc. London, Ser. B* **336**, 367–373.
- Sabin, A. T., Gallun, F. J., and Souza, P. E. (2013). "Acoustical correlates of performance on a dynamic range compression discrimination task," *J. Acoust. Soc. Am.* **134**, 2136–2147.
- Saripella, R., Loizou, P., Thibodeau, L., and Alford, J. (2011). "The effects of selective consonant amplification on sentence recognition in noise by hearing-impaired listeners," *J. Acoust. Soc. Am.* **130**, 3028–3037.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Shannon, R. V., Zeng, F. G., and Wygonski, J. (1998). "Speech recognition with altered spectral distribution of envelope cues," *J. Acoust. Soc. Am.* **104**, 2467–2476.
- Sharpe, V. P., Den Ouden, D. B., and Fogerty, D. (2014). "The role of acoustics in processing syntactically ambiguous sentences," *Proc. Meet. Acoust.* **21**, 060006.
- Shinn-Cunningham, B. G., and Wang, D. (2008). "Influences of auditory object formation on phonemic restoration," *J. Acoust. Soc. Am.* **123**, 295–301.
- 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. E. (2014). "Information-bearing acoustic change outperforms duration in predicting intelligibility of full-spectrum and noise-vocoded sentences," *J. Acoust. Soc. Am.* **135**, 1518–1529.
- Stone, M. A., Füllgrabe, C., and Moore, B. C. (2012). "Notionally steady background noise acts primarily as a modulation masker of speech," *J. Acoust. Soc. Am.* **132**, 317–326.
- 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.
- Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (1999). "Monosyllabic word recognition at higher-than-normal speech and noise levels," *J. Acoust. Soc. Am.* **105**, 2431–2444.
- Summers, V., and Cord, M. T. (2007). "Intelligibility of speech in noise at high presentation levels: Effects of hearing loss and frequency region," *J. Acoust. Soc. Am.* **122**, 1130–1137.
- Takahashi, G. A., and Bacon, S. P. (1992). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," *J. Speech Hear. Res.* **35**, 1410–1421.
- 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.
- Wingfield, A., Aberdeen, J. S., and Stine, E. A. (1991). "Word onset gating and linguistic context in spoken word recognition by young and elderly adults," *J. Gerontol.* **46**, P127–P129.
- Wingfield, A., Alexander, 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.