Research Article

Glimpsing Speech in the Presence of Nonsimultaneous Amplitude Modulations From a Competing Talker: Effect of Modulation Rate, Age, and Hearing Loss

Daniel Fogerty,^a Jayne B. Ahlstrom,^b William J. Bologna,^{b,c} and Judy R. Dubno^b

Purpose: This study investigated how listeners process acoustic cues preserved during sentences interrupted by nonsimultaneous noise that was amplitude modulated by a competing talker.

Method: Younger adults with normal hearing and older adults with normal or impaired hearing listened to sentences with consonants or vowels replaced with noise amplitude modulated by a competing talker. Sentences were spectrally shaped according to individual audiograms or to the mean audiogram from the listeners with hearing impairment for a younger spectrally shaped control group. The modulation spectrum of the noise was low-pass filtered at different modulation cutoff frequencies. The effect of noise level was also examined.

Results: Performance declined when nonsimultaneous masker modulation included faster rates and was maximized when masker modulation matched the preserved primary speech modulation. Vowels resulted in better performance compared with consonants at slower modulation cutoff rates, likely due to suprasegmental features. Poorer overall performance was observed with increased age or hearing loss, and for listeners who received spectrally shaped speech.

Conclusions: Nonsimultaneous amplitude modulations from a competing talker significantly interacted with the preserved speech segment, and additional listener factors were observed for age and hearing loss. Importantly, listeners may obtain benefit from nonsimultaneous competing modulations when they match the preserved modulations of the sentence.

The primary conversational message of speech, or target speech, is often heard simultaneously with one or more competing talkers. To maximize communication in such environments, listeners must successfully separate these two competing signals. Because amplitude modulation provides primary speech information (e.g., Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), the ability to process amplitude modulations of the target speech in the presence of modulations associated with the competing talkers may be crucial for speech recognition in these complex environments. Previous work has demonstrated that amplitude modulations of the competing speech often interfere with the ability to process modulation cues of the target speech, referred to as modulation masking (e.g.,

Gustafsson & Arlinger, 1994; Kwon & Turner, 2001). The effects of simultaneous masking of sentences by amplitude-modulated signals have been assessed in several studies (e.g., Festen & Plomp, 1990; Jin & Nelson, 2010; Summers & Molis, 2004). However, nonsimultaneous masking also occurs under these conditions, with intense portions of the masker resulting in temporal masking of preserved portions of the speech signal (e.g., Bacon, Opie, & Montoya, 1998; Dubno, Horwitz, & Ahlstrom, 2003; Gifford, Bacon, & Williams, 2007). The purpose of the current study was to investigate how amplitude modulation during those intense portions of competing noise interfere with listeners' ability to process the relatively preserved portions of speech that occur during the less intense portions of the masker.

Toward this end, the current investigation used a noise-replacement paradigm to selectively replace portions of the speech with noise. This method avoided factors related to simultaneous masking during intense portions of the masking noise and modeled how listeners might use speech that is preserved between noise-masked intervals. Furthermore, interactions of the interrupted noise with the acoustic properties of the preserved speech intervals were

^aUniversity of South Carolina, Columbia

^bMedical University of South Carolina, Charleston

^cUniversity of Maryland, College Park

Correspondence to: Daniel Fogerty; fogerty@sc.edu

Editor: Nancy Tye-Murray

Associate Editor: Mitchell Sommers

Received July 22, 2015

Revision received November 18, 2015

Accepted January 4, 2016

DOI: 10.1044/2016_JSLHR-H-15-0259

Disclosure: The authors have declared that no competing interests existed at the time of publication.

investigated by limiting the preserved speech to vowels and consonants. These speech segments are characterized by different acoustic-phonetic cues (e.g., Stevens, 2002) and by different rates of amplitude modulation (Crouzet & Ainsworth, 2001) by virtue of their different distributions of spectral energy. The effects of age and hearing loss were also examined by comparing results for younger adults with normal hearing (YNH), older adults with normal hearing (ONH), and older adults with impaired hearing (OHI). All listeners received spectral shaping that amplified the speech spectrum in one-third-octave bands to ensure that the speech was at least 15 dB above audiometric thresholds through at least 4 kHz, similar to hearing-aid amplification that would be provided by the desired sensation level method (Seewald, Ramji, Sinclair, Moodie, & Jamieson, 1993; see also Humes, 2007). To investigate the perceptual consequences of spectral shaping and the associated elevation in sound-pressure level, a “spectral control” group of younger listeners with normal hearing (YSC) was tested using the average spectral shaping provided to the OHI listeners. Although the speech signals for YSC and OHI listeners are similar, spectral shaping alone does not model impaired suprathreshold auditory processing associated with sensorineural hearing impairments such as broadened tuning and spectral smearing (Baer & Moore, 1994) or reduced temporal processing of glimpses (George, Festen, & Houtgast, 2006).

We hypothesized an interaction between masker modulation rate and the preserved speech segment, with a greater effect of masker modulation at higher rates for consonants compared with vowels due to the faster amplitude modulations inherent in consonant segments. Furthermore, we expected that older listeners would perform as well as younger listeners in the vowel conditions, in accordance with earlier reports of preserved sentence recognition of vowel-only sentences by older adults with normal or impaired hearing (Fogerty, Kewley-Port, & Humes, 2012).

Effect of Masker Modulations

When listening to speech in the presence of an amplitude-modulated masker, such as a competing talker, portions of the speech signal that are coincident with less intense portions of a modulated masker are relatively preserved at favorable signal-to-noise ratios (SNRs), whereas remaining portions of the speech signal may be significantly masked. Moreover, speech recognition is believed to be strongly related to the portions of speech (i.e., “glimpses”) that are preserved at these favorable SNRs (e.g., Cooke, 2003; Moore, 2008). Indeed, results of many past studies (e.g., Bashford & Warren, 1987; Dirks, Wilson, & Bower, 1969; Miller & Licklider, 1950) and more recent investigations (e.g., Benard & Başkent, 2015; Nagaraj & Knapp, 2015; Shafiro, Sheft, Risley, & Gygi, 2015; Wilson & Irish, 2015) demonstrate listeners’ ability to process speech glimpses and maintain good speech recognition in the presence of modulated maskers. Most of these studies used interrupted speech, where portions of the speech signal were

selectively replaced with silence or noise. However, recognition of interrupted speech and recognition of speech in the presence of continuous fluctuating maskers are strongly correlated (Jin & Nelson, 2010), which provides support for the ecological validity of this experimental method.

Speech intelligibility in competing amplitude-modulated conditions is, in part, determined by modulation masking (Gustafsson & Arlinger, 1994). Indeed, modulation masking of speech is observed even in the random amplitude fluctuations of Gaussian noise typically assumed to be a purely energetic masker (Stone, Füllgrabe, & Moore, 2012; Stone & Moore, 2014). Successful speech recognition may rely on the ability to process temporal fine structure cues during these brief glimpses of speech at favorable SNRs (e.g., Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006), which is correlated with speech reception thresholds in modulated noise for OHI listeners (Hopkins & Moore, 2011). Although the ability to benefit from the improved SNR during less intense portions of the modulated masker is a significant predictor of speech recognition (e.g., Lorenzi et al., 2006; Moore, 2008; Nelson, Jin, Carney, & Nelson, 2003), the higher-level peaks of the masker may exert nonsimultaneous masking that may limit the glimpsing or dip-listening benefit and reduce speech recognition. Early studies proposed a role for forward and backward masking during the glimpsed portions of speech between masker bursts (Dirks et al., 1969). Consistent with this hypothesis, associations between forward masking and recognition of interrupted or modulated-noise-masked sentences have been observed (Bacon et al., 1998; Dubno et al., 2003; Gifford et al., 2007). In addition, temporal (i.e., nonsimultaneous) modulation masking is known to occur for nonspeech stimuli (Wojtczak & Viemeister, 2005). These observations suggest that dip listening is influenced by the ability to process “glimpsed” acoustic information, such as the temporal fine structure, as well as the ability to resolve speech information in the presence of temporal masking caused by the high-intensity portions of the masker.

Separating the concurrent effects of simultaneous and nonsimultaneous masking of speech in modulated noise is difficult because both of these factors potentially co-occur and influence speech recognition. As discussed earlier, noise replacement during interrupted speech objectively defines the speech glimpses and isolates the effects of nonsimultaneous masking by the noise intervals from simultaneous masking effects. To the extent that speech recognition can be predicted on the basis of the available speech glimpses (Cooke, 2003), this method models how listeners might glimpse speech in the presence of nonsimultaneous masking from relatively intense portions of a concurrent modulated noise.

This evidence leads to the hypothesis that modulations from a competing talker may interfere with detection of modulations within the target speech, resulting in poorer speech recognition. The greatest interaction is expected to occur when modulation rates of the speech and noise are matched.

Temporal Nature of Speech Segments

Slow, varying speech modulation rates, corresponding to syllables, are correlated across frequency bands, whereas faster modulation rates, corresponding to phonemes, vary more independently in the high-frequency bands (Crouzet & Ainsworth, 2001). Different amplitude modulation rates code different information about the segmental (i.e., consonants and vowels) and suprasegmental aspects of speech (Rosen, 1992). As such, different modulation rates of a competing talker may interact with (e.g., modulation mask) different corresponding features in the target talker's production.

A secondary goal of this study was to investigate how nonsimultaneous modulation masking interacts with the type of speech segment that is preserved. Limiting preserved speech glimpses to specific segments, such as consonants and vowels, provides a means to determine the extent to which the competing noise interacts with perception of these speech segments. Many consonants are characterized by rapid, transient features that occur on fast timescales, whereas vowels are characterized by slow varying acoustic features at slower timescales (Stevens, 2002). In addition, the variation of vowel amplitude across the sentence provides important suprasegmental information related to the stress and rhythm of the sentence (Port, 2003), which also facilitates sentence recognition (e.g., Fogerty, 2013, 2014). Because modulation masking occurs within modulation rate-selective channels (e.g., Dau, Kollmeier, & Kohlrausch, 1997), it is likely that different rates of amplitude modulation, contained within the competing amplitude-modulated signal of a competing talker, selectively interact with the different temporal properties of consonants, vowels, and temporal suprasegmental features that are concurrently realized during these speech segments.

In this article, speech glimpses are defined according to segmental boundaries, similar to other methods that have selectively examined the contribution of specific speech segments (e.g., Cole, Yan, Mak, Fenty, & Bailey, 1996; Strange, Jenkins, & Johnson, 1983; Wilson, Arcos, & Jones, 1984). Because these segments were tested within sentences, this method is not particularly sensitive to isolating interactions with specific acoustic-phonetic properties of consonants and vowels but, rather, provides a general view for interactions of the noise modulation with different segmental and suprasegmental properties of sentences that are associated with different temporal rates (Rosen, 1992). The current study was designed to investigate the interaction between modulation rates of the competing speech and preserved speech segments.

Effect of Age, Hearing Loss, and Spectral Shaping

Older adults may be as able as younger adults to process temporal envelope information in quiet conditions when differences in audibility are minimized (Takahashi & Bacon, 1992), but they may experience greater difficulties in the presence of amplitude-modulated maskers (e.g., Dubno et al., 2003; George, Zekveld, & Kramer, 2007;

Gustafsson & Arlinger, 1994). For many individuals with hearing loss, hearing aids provide amplification in the higher frequencies to ensure audibility of the full speech spectrum (Humes, 2007). However, such spectral shaping of the speech can reduce temporal modulation cues, which may affect speech recognition in temporally complex listening conditions (Fogerty, Ahlstrom, Bologna, & Dubno, 2015). To further address this question, the current investigation included a second group of younger listeners with normal hearing who listened to speech that was spectrally shaped according to the average audiogram of the listeners with impaired hearing and that was presented at the same sound-pressure level. Including this control group listening to higher-level, spectrally shaped speech provides an empirical method for separating the contributions of reduced audibility, suprathreshold perceptual effects of hearing impairment, and effects of spectral shaping on modulation cues.

Method

Listeners

Forty-one adults participated in the current study, divided among four groups: YNH ($N = 10$; mean age = 22 years, range = 18–25), ONH ($N = 10$, mean age = 70 years, range = 63–83), OHI ($N = 10$; mean age = 73 years, range = 65–85), and YSC (younger spectral control group with normal hearing, matched to the mean audiogram of the OHI listeners; $N = 11$, mean age = 21 years, range = 18–24). All listeners with normal hearing had pure-tone thresholds ≤ 20 -dB HL at octave frequencies from 0.25 to 4.0 kHz. Inclusion criteria required listeners with hearing impairment 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).

For subjects in all groups, speech spectra were spectrally shaped as determined by their individual audiograms; speech spectra for the YSC group were set according to the mean audiometric thresholds of the OHI group. Figure 1A displays the mean 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 the speech spectra for the YNH and ONH groups were closely matched, at least in the frequency range up to 4 kHz.

Speech

Stimuli consisted of sentences selected from the TIMIT database (Garofolo et al., 1993) that were recorded at a sampling frequency of 16000 Hz and later divided into lists of 18 sentences (Gilbert, Tamati, & Pisoni, 2013). The 13 sentence lists of 18 sentences each (76 keywords) used in this study were balanced for keyword frequency, familiarity, and talker gender. No sentences were repeated in any condition. The TIMIT database provides time markings for segmental boundaries that were confirmed by expert phoneticians. TIMIT boundaries were adjusted within 1 ms

Figure 1. (A) Audiometric thresholds (symbols) for the four listener groups. The long-term average speech spectrum after spectral shaping is displayed for the four groups. Spectral shaping for the YSC listeners was determined by the average audiogram of the OHI listeners. (B) Envelope modulation spectrum; arrows indicate the modulation frequency corresponding to maximum performance on vowel (red) and consonant (blue) conditions, respectively, for each of the listener groups. (C) The stimulus waveform for an example sentence with the vowels preserved (black) and the consonants replaced with a speech-modulated noise (gray). (D) Keyword recognition accuracy is plotted according to the different low-pass filter modulation cutoff frequencies tested. Square symbols display the continuous speech condition and circles indicate results for the segmentally interrupted speech materials; open = 16-dB signal-to-noise ratio (SNR), filled = 0-dB SNR. Error bars display the standard error of the mean. RAU = rationalized arcsine units. SPL = sound pressure level.

to the nearest local minima (i.e., zero-crossing) for processing. Consistent with previous classifications using a segment-replacement paradigm (Kewley-Port, Burkle, & Lee, 2007), vowel segments included monophthongs, diphthongs, rhotics, and intervocalic glottal stops; consonant segments included obstruents, nasals, and semivowels.

Modulated Noise

A speech-shaped noise was created that matched the duration and power spectrum of two concatenated sentences spoken by the target talker that were not used for target stimuli. The temporal envelope of the concatenated speech sample was extracted by half-wave rectification followed by low-pass filtering using a fourth-order Butterworth filter with a cutoff modulation frequency of 4, 8, 16, or 32 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 it contained the average spectral and temporal characteristics of the target talker of each trial. The modulation spectrum for these sentences is displayed

in Figure 1B.¹ For the unmodulated 0-Hz condition, a speech-shaped noise matching the concatenated sentence was used without subsequent amplitude modulation.

Stimuli and Design

The target speech sentences and competing noise were combined to create 13 different experimental conditions. First, we tested a baseline condition in which speech was presented simultaneously with a 0-dB-SNR continuous modulated noise using a low-pass filter modulation cutoff of 16 Hz. This condition tested the combined effects of simultaneous and nonsimultaneous masking that occurs

¹This modulation spectrum was derived from a concatenation of three PRESTO lists with silent delays removed between each sentence. The Hilbert transform was used to extract the envelope, which was then low-pass filtered at 32 Hz. Concatenating sentences can introduce artifacts in the modulation spectrum related to the sentence duration. For this reason, only modulation rates above 2 Hz are displayed. This spectrum is well above the modulation rates determined by sentence duration; the shortest duration was 1.36 s, corresponding to a maximum sentence rate of 0.74 Hz.

when listening to speech in the presence of amplitude modulations from a competing talker.

The remaining experimental conditions segmentally interrupted the speech signal to preserve either vowel or consonant segments and replaced the alternate segment type with the modulated noise. Note that in these latter conditions, the modulated masker was a nonsimultaneous replacement signal (see Figure 1C). The first set of segmental conditions presented speech in unmodulated “steady-state noise” defined here as 0 Hz. The level of the noise was presented at 0-dB SNR (i.e., matching level) and 16-dB SNR (i.e., lower level) relative to the sentence level prior to interruption. This resulted in four experimental conditions (two noise levels $\times$ two segment types: consonant or vowel). The second set of segmental conditions presented speech at only the 0-dB-SNR speech-modulated noise level using 4-, 8-, 16-, and 32-Hz modulation cutoff frequencies. This resulted in eight additional experimental conditions (four modulation rates $\times$ two segment types).

Procedures

Testing of the younger participants was completed at the University of South Carolina, and testing of the older participants was completed at the Medical University of South Carolina. Testing at both locations was completed inside a sound-attenuating booth using similar testing equipment and identical stimulus presentation software. The participants listened to stimuli presented at a sampling rate of 48828 Hz via a Sennheiser HDA 200 headphone over the test ear (usually the right ear) routed first through a TDT System III digital-to-analog processor (RP2 or RX6) and headphone buffer (HB7 or HB5). Systems were calibrated to present the speech (prior to signal processing and spectral shaping) at 70-dB SPL. All sentence materials were presented to listeners in blocks randomized according to the different masker modulation rates. Sentences for each modulation rate were randomized, with consonant and vowel conditions intermixed during presentation. Demonstration trials were presented without feedback to provide participants some familiarization with the signal processing and testing procedure. Listeners were required to repeat each sentence aloud, were encouraged to guess, and were scored live by the experimenters according to keywords. Responses were also audio recorded for offline analysis if required. 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 transformed to rationalized arcsine units (RAU) to stabilize the error variance (Studebaker, 1985).

Spectral Shaping and Calibration

A calibration noise was first created to match the long-term average spectrum of the original sentence files. Output of the Sennheiser HDA 200 headphone for this calibration noise was adjusted to be presented at 70-dB SPL. Individual stimulus files were then spectrally shaped on the

basis of the listener’s audiometric thresholds. Stimulus files were band-pass filtered into 20 one-third-octave frequency bands using a 128th-order finite impulse response (FIR) filter within the range of 100–8000 Hz. The output level of each band was adjusted to match an audibility target level that was either equivalent to the original speech level in that band or adjusted to be no greater than 20 dB above the individual’s hearing threshold within that band, whichever resulted in the greater 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 Figure 1A). This procedure is similar to the desired sensation level algorithm for fitting hearing aids. Speech and noise were mixed and then spectrally shaped after signal processing.

Results and Discussion

Continuous Speech and Noise

Results for the baseline condition of uninterrupted speech in speech-modulated noise are plotted as the filled square symbols in Figure 1D and were analyzed using a one-way analysis of variance (ANOVA). Results demonstrated a main effect of listener group, $F(3, 40) = 9.9$, $p < .001$. Independent samples’ t -tests demonstrated significant group differences between YNH and YSC, $t(19) = 4.5$, $p < .001$, $d = 2.0$, and between ONH and OHI, $t(18) = 2.8$, $p < .05$, $d = 1.3$, with higher scores for YNH and ONH groups. No significant differences were observed between the two groups with normal hearing (YNH and ONH) or the two groups with the most significant spectral shaping (YSC and OHI).

Segmental Interruption with a Speech-Modulated Masker

Segmental interruption conditions were examined in two sets of analyses. The first set examined the effect of noise level in the unmodulated 0-Hz condition (see open and filled symbols at 0 Hz in Figure 1D). The second set examined the effect of the noise modulation cutoff frequency (see symbols connected by the line in Figure 1D).

Effect of Noise Level

The unmodulated 0-Hz (i.e., “steady-state”) masker was presented at 16-dB and 0-dB SNR (i.e., average speech root mean square); results are displayed as the open and filled symbols at 0 Hz, respectively, in Figure 1D. A 2 (SNR) $\times$ 2 (segment replaced) $\times$ 4 (group) mixed model ANOVA demonstrated a significant main effect of segment, $F(1, 37) = 440.3$, $p < .001$; SNR, $F(1, 37) = 10.5$, $p < .01$; and group, $F(1, 37) = 541.9$, $p < .001$. All interactions were also significant ($p < .05$) and were explored further through specific planned contrasts. All groups demonstrated significantly better performance when vowels were preserved compared with consonants for both SNRs ($p < .01$), although this comparison did not reach significance for OHI listeners at

0-dB SNR ($p > .05$). Interestingly, the effect of SNR was mainly isolated to conditions wherein the consonants were preserved and the vowels were replaced by noise. Performance was better when the vowels were replaced by the 0-dB-SNR noise for all groups ($p < .05$). Note that this noise level was still below the average level of the replaced vowel (vowels on average are about 5 dB above the average sentence level). In contrast, the noise level did not significantly affect performance in conditions that preserved the vowel and replaced the consonant for YNH, YSC, and ONH listeners ($p > .05$) with the exception of OHI listeners. The OHI group performed better for vowel conditions with the lower noise level (16-dB SNR), $t(9) = 4.2$, $p < .01$, $d = 1.2$.

Paired comparisons were also examined between groups. An effect of age (YNH vs. ONH) was only observed for consonants at 16-dB SNR, $t(18) = 4.6$, $p < .001$, $d = 2.0$, with YNH listeners performing better by 15 RAU. This may indicate that performance by older listeners is significantly affected when the temporal envelope of the speech signal is highly degraded (i.e., high-level vowels are replaced by very low-level noise, an average 21-dB difference in segment level). An effect of hearing loss was also observed; OHI performed significantly poorer than ONH listeners across the four conditions ($p < .01$). To determine if this difference was related to spectral shaping for the OHI group, scores for YNH, YSC, and OHI groups were compared. First, YSC listeners performed poorer than YNH by 9.4–11.6 RAU, but scores for the two groups were not significantly different after Bonferroni correction for multiple comparisons (adjusted $p > .05$). Second, YSC listeners performed significantly better than OHI listeners in all conditions ($p < .01$), except when vowels were preserved at 16-dB SNR. Note that this is the same condition that demonstrated an improvement in sentence recognition, only for OHI listeners, with the lower consonant replacement noise. Thus, the poorer results for OHI subjects were not accounted for by potential acoustic distortions imposed by spectral shaping.

Effect of Masker Modulation Rate

The effect of the masker modulation cutoff frequency was examined using a 2 (segment) $\times$ 5 (modulation cutoff rate: 0–32 Hz) $\times$ 4 (group) mixed model ANOVA. Main effects were observed for segment, $F(1, 37) = 211.8$, $p < .001$; modulation rate, $F(4, 148) = 30.5$, $p < .001$; and group, $F(1, 37) = 507.3$, $p < .001$. Significant interactions were also observed with modulation rate ($p < .001$), which were further explored by subsequent contrasts.

As observed in Figure 1D, across all groups and modulation cutoffs, listeners performed significantly better for the vowel conditions at the slower modulation rates, with performance for the two segment types converging at higher masker modulation cutoff frequencies. The greatest convergence was seen for the younger listeners, starting around 8 Hz. Interestingly, whereas OHI listeners demonstrated some convergence at 16 Hz, better performance for vowel conditions was still observed at 32 Hz. This is a

clear deviation from the pattern observed with the other listener groups. This is noteworthy when compared with the adults with normal hearing (YNH and ONH), who demonstrated significant declines in performance for vowels from 16 Hz to 32 Hz ($p < .01$). An apparent trend in the data across modulation rates is a significant decrease in performance for both segmental conditions when the fastest masker modulation rates were present (i.e., 32-Hz cutoff).

An examination of the trends in performance across modulation rates within groups demonstrates that not all masker modulation rates resulted in the same performance. Performance tended to reach a local maximum at one of the modulation rates. An exception to this inverted U-shaped trend was observed for older listeners, who generally obtained maximum performance at 0 Hz, potentially indicating greater difficulty processing speech with more temporally complex nonsimultaneous maskers. Using a logarithmic axis, third-order polynomials were fit to the group data to quantify the modulation cutoff rate that yielded maximum performance, that is, the low-pass filter cutoff modulation rate, after 0 Hz, that maximized performance. These optimal modulation rates correspond to primary modulation rates observed in the temporal envelope modulation spectrum (Figure 1C), suggesting that maintaining primary speech modulations through the noise-replacement intervals may facilitate speech continuity. In general, older listeners had maximum performance at higher cutoff frequencies, although these still corresponded to peak rates in the modulation spectrum.

Absolute performance at each modulation cutoff rate was also compared between groups. After Bonferroni correction for multiple comparisons, results indicate similar performance among the two younger groups and between the two groups with normal hearing. The exception was significantly better performance for YNH compared with ONH for consonants at 8 Hz ($p < .05$), which was close to the peak of the YNH function. In contrast, OHI listeners performed significantly poorer than ONH and YSC listeners at all cutoff rates ($p < .05$), with the exception of the 8-Hz rate for the vowel condition, where OHI listeners approached YSC performance (although still 14 RAU poorer), which was close to the peak of the OHI function. Overall, these results indicate a significant effect of nonsimultaneous masker modulation rate and of hearing loss that was not accounted for by spectral shaping.

Examining Group Effects

To further explore the contribution of hearing loss, age, and spectral shaping, correlational analyses were run separately for the four listener groups. High, positive correlations (generally above 0.6) were observed between the majority of the vowel and consonant conditions for each listener group. Therefore, composite vowel and consonant scores were derived by averaging performance across the five cutoff rates (0–32 Hz). Independent samples' *t*-tests of these average scores confirmed similar trends observed with the raw data, except that the trend for poorer performance by YSC compared with YNH listeners reached

significance with the average data: vowel: $t(19) = 2.7, p < .05, d = 1.2$; consonant: $t(19) = 2.5, p < .05, d = 1.1$. These average data were then entered into a correlational analysis with age, low- and high-frequency pure-tone averages (PTA_{low}: 0.25, 0.5, and 1 kHz; PTA_{high}: 1, 2, and 4 kHz), and performance on the continuous speech in modulated noise condition. Overall, the average consonant and vowel scores were highly positively correlated ($r = .68$ to $.90$ across groups) and therefore cannot be viewed as independent factors. No other correlations were significant for the two younger groups. For the ONH group, age was significantly negatively correlated with the three speech measures ($r = -.69$ to $-.90$), with no significant correlation with either PTA. In contrast, PTAs were significantly negatively correlated with the average vowel score for OHI listeners (PTA_{low}: $r = -.89$, PTA_{high}: $r = -.75, p < .05$), yet age was not significantly correlated.

Discussion

This study examined the effect of nonsimultaneous speech-modulated noise on recognition of interrupted speech. The results demonstrated that the noise modulation rate significantly reduced listeners' ability to benefit from speech glimpses, with differential effects for the preserved segment type and experimental group. Overall, ONH listeners performed more poorly for consonants than YNH listeners. OHI listeners performed poorest, which could not be attributed to reduced audibility. Acoustic distortions of the speech signal due to spectral shaping were also examined by comparing results from YNH, YSC, and OHI listeners. A trend toward poorer overall performance for YSC compared with YNH listeners was observed, but these differences did not entirely account for poorer scores for OHI listeners. Overall, the results suggest that a listener's ability to use speech glimpses is determined by the acoustic cues of the preserved speech segment, as well as by nonsimultaneous masking by the modulated masker.

Effects of Nonsimultaneous Maskers on Speech Glimpses

The primary purpose of this study was to determine the effect of amplitude modulations of a nonsimultaneous masker on perception of preserved speech glimpses. Prior work has suggested that forward masking is related to the perceptual processing of speech glimpses (Dubno et al., 2003; Gifford et al., 2007) and that the effect of masker modulation is also observed in conditions of forward masking, at least for simple nonspeech stimuli (Wojtczak & Viemeister, 2005). The current study extended this work by demonstrating modulation masking of speech glimpses in favorable SNRs that occurs between higher-intensity portions of a modulating masker. Nonsimultaneous masking contributes to sentence recognition when listening to speech in a modulated noise background.

In the current study, speech recognition was best when masker modulation rates matched the intrinsic modulation

rates of the preserved speech. Nevertheless, interactions with subject age suggest that age-related differences in suprathreshold temporal processing abilities are also contributing factors. These results differ from those observed for simultaneous modulated maskers, wherein modulation masking was largest at rates equal to those of the target (e.g., Gustafsson & Arlinger, 1994; Kwon & Turner, 2001). Instead, the current study demonstrates that modulations of a competing signal that match the modulation rate of the preserved speech are sufficient for facilitating the continuity of the preserved speech signal. This occurred even though differences in the phase of the modulation were likely between the replacement noise and preserved speech signals and that the replacement noise did not preserve any acoustic-phonetic cues to the target speech. This finding could potentially be incorporated into speech enhancement algorithms, providing a way to benefit from speech signals when only partial information is preserved. If the missing information cannot be recovered from the noisy signal, it could potentially be modified to match temporal properties of the preserved speech to facilitate speech continuity, and possibly intelligibility.

The perceived continuity of interrupted speech may be increased through processes of phonemic restoration, where interrupted speech is perceived as continuous through a more intense competing noise (e.g., Bashford, Warren, & Brown, 1996). In the current experiment, the use of different noise levels in the unmodulated 0-Hz condition provided a means to test this hypothesis. On the basis of phonemic restoration, we would predict minimal changes in performance for the consonant conditions, because even the highest noise level tested was not sufficient to mask the noise-replaced vowel segments. Conversely, improved performance would be predicted with increased noise levels for the vowel conditions, because these conditions involved noise replacement of relatively low-level consonants. In contradiction to phonemic restoration predictions, speech recognition improved at higher noise levels only when the consonants were preserved. In addition, higher noise levels resulted in reduced performance for OHI listeners when the vowels were preserved. Thus, phonemic restoration does not appear to account for performance in these conditions.

An alternative interpretation is that the continuity of the sentence was facilitated by a failure to segregate the noise modulations from the speech, particularly when the modulation rate of the noise matched the primary modulation rate of the preserved speech. This is consistent with findings of improved performance when interrupted noise segments preserve the amplitude modulations of the sentence (Bashford et al., 1996; Shinn-Cunningham & Wang, 2008), particularly slow-rate suprasegmental modulations provided by the vowels (Fogerty, 2014).

Hearing Loss and Spectral Shaping

The effect of hearing loss was notable across all comparisons in this study in the poorer performance for OHI listeners, even though spectral shaping was applied

to ensure audibility through at least 4 kHz. Furthermore, PTA (low and high frequencies) was significantly correlated with vowel scores. This is noteworthy given the lower frequency and higher-intensity composition of vowels compared with consonants. At least three factors can explain these results.

First, hearing loss in the higher frequencies for OHI listeners required spectral shaping of the speech to meet a single target sensation level across all participants (i.e., 20-dB SL). In contrast, OHI listeners had better thresholds in the low- to mid-frequency region and did not require as much spectral shaping in this region (see Figure 1A). As a result, speech sensation levels in lower frequencies varied more across participants than speech sensation levels in higher frequencies, because lower-frequency speech sensation levels were determined by their individual thresholds rather than by spectral shaping of the speech. Indeed, mean sensation levels for OHI listeners for lower frequencies ($M = 25$ -dB SL; $SD = 5.3$ dB) and higher frequencies ($M = 21$ -dB SL; $SD = 1.8$ dB) differed significantly, $t(9) = 3.3$, $p < .01$, $d = 1.1$; sensation level standard deviations were also larger in the lower than higher frequencies. Moreover, speech sensation levels for lower and higher frequencies were each correlated with vowel scores ($r = .76$ – $.79$, $p < .05$), but not consonant scores ($p > .05$). Whereas sensation levels were sufficiently audible, lower sensation levels for some OHI listeners may have contributed to poorer vowel scores.

Second, higher PTAs result not only in reduced speech audibility (i.e., reduced sensation levels), but they also may reflect greater severity of the hearing impairment, which is associated with poorer suprathreshold auditory function, such as broader auditory filters, increased spectral smearing, and declines in temporal fine structure sensitivity. These processing changes may adversely affect the use of vowel spectral cues and, potentially, consonant place cues. Consonants also contain temporal acoustic–phonetic cues (Rosen, 1992; Stevens, 2002) and, therefore, may have been less affected by reduced auditory processing in the spectral domain.

Third, associations with PTA may relate to effects of spectral shaping, given that individuals with higher PTAs listened to speech with more substantial spectral manipulations. Results from YSC listeners provided additional evidence for this explanation, specifically, their lower average vowel and consonant scores, and scores for continuous speech that were statistically equivalent to those for OHI listeners (but poorer than YNH and ONH listeners). Spectral shaping, while effective at increasing speech audibility for OHI listeners and minimizing audibility differences across groups (Humes, 2013), might limit the benefit in temporally complex conditions due to altered acoustic signals. For example, spectral shaping reduces the amplitude-modulation depth (i.e., consonant-vowel intensity ratio) of the broadband speech signal by amplifying high-frequency, low-intensity consonants (see also Fogerty et al., 2015). In addition, listeners' unfamiliarity with the spectrally shaped stimulus may have resulted in a perceptual mismatch with stored mental representations of speech (Rönnberg et al., 2013), leading to poorer scores. If so, additional familiarization and auditory training may be necessary with spectrally

shaped speech (or speech with a new hearing aid). However, many studies have indicated that spectral shaping is sufficient to account for audibility after minimal exposure, with OHI listeners performing similarly to ONH listeners (reviewed by Humes, 2013).

Along with perceptual factors, cognitive processing is likely involved in sentence processing when listeners are tracking and filling in partial speech information. The ways in which perceptual and cognitive factors support sentence recognition for different listener groups will ultimately need to be defined. However, the perceptual and cognitive requirements of the tasks used in the current study would not be expected to change by altering the cut-off frequency of modulated noise, which would minimize contributions of cognitive factors in these results.

The results of this study suggest that reduced audibility does not entirely account for the declines in speech recognition that were observed for older adults in these temporally complex conditions involving partial speech information. Additional studies are needed to more completely define how spectral shaping provided by hearing aids might limit benefits of improved speech audibility, possibly due to acoustic distortions of the speech stimulus. In temporally complex conditions, other aspects of perceptual declines associated with age and cochlear pathology may be factors that are not currently addressable by hearing-aid amplification.

In summary, our results demonstrate that the rate of nonsimultaneous noise modulation affects performance. That is, even if speech is well preserved at optimal SNRs during the dips of the masker, the ability to resolve speech information is determined, in part, by the temporal properties of the surrounding higher-intensity portions of noise. The results of this study indicate that maximal speech recognition occurs when the masking noise preserves the glimpsed speech modulation rates. Therefore, preservation of speech modulation rates may facilitate the continuity of interrupted speech.

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 (NIH)/National Institute on Deafness and Other Communication Disorders Grants R03 DC012506 (awarded to Daniel Fogerty) and R01 DC000184 (awarded to Judy R. Dubno). This work was also supported by the South Carolina Clinical and Translational Research (SCTR) Institute (NIH Grant UL1 RR0299882). This investigation was partially conducted in a facility constructed with support from Research Facilities Improvement Program Grant C06 RR14516 from the NIH National Center for Research Resources.

References

- Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech, Language, and Hearing Research*, 41, 549–563.

- Baer, T., & Moore, B. C. (1994). Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *The Journal of the Acoustical Society of America*, 95, 2277–2280.
- Bashford, J. A., & Warren, R. M. (1987). Multiple phonemic restorations follow the rules for auditory induction. *Perception & Psychophysics*, 42, 114–121.
- Bashford, J. A., Warren, R. M., & Brown, C. A. (1996). Use of speech-modulated noise adds strong “bottom-up” cues for phonemic restoration. *Perception & Psychophysics*, 58, 342–350.
- Benard, M. R., & Başkent, D. (2015). The effect of visual cues on top-down restoration of temporally interrupted speech, with and without further degradations. *Hearing Research*, 328(10), 24–33.
- Cole, R., Yan, Y., Mak, B., Fanty, M., & Bailey, T. (1996). The contribution of consonants versus vowels to word recognition in fluent speech. In *Proceedings of the ICASSP'96* (pp. 853–856). Atlanta, GA: ICASSP.
- Cooke, M. (2003). Glimpsing speech. *Journal of Phonetics*, 31, 579–584.
- Crouzet, O., & Ainsworth, W. A. (2001). On the various influences of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation. In *Workshop on Consistent and Reliable Acoustic Cues for Sound Analysis*. Aalborg, Denmark.
- Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, 102, 2892–2905.
- Dirks, D. D., Wilson, R. H., & Bower, D. R. (1969). Effect of pulsed masking on selected speech materials. *The Journal of the Acoustical Society of America*, 46, 898–906.
- Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2003). Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing. *The Journal of the Acoustical Society of America*, 113, 2084–2094.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88, 1725–1736.
- Fogerty, D. (2013). Acoustic predictors of intelligibility for segmentally interrupted speech: Temporal envelope, voicing, and duration. *Journal of Speech, Language, and Hearing Research*, 56, 1402–1408.
- Fogerty, D. (2014). Importance of envelope modulations during consonants and vowels in segmentally interrupted sentences. *The Journal of the Acoustical Society of America*, 135, 1568–1576.
- Fogerty, D., Ahlstrom, J. B., Bologna, W. J., & Dubno, J. R. (2015). Sentence intelligibility during segmental interruption and masking by speech-modulated noise: Effects of age and hearing loss. *The Journal of the Acoustical Society of America*, 137, 3487–3501.
- Fogerty, D., Kewley-Port, D., & 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. *The Journal of the Acoustical Society of America*, 132, 1667–1678.
- Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., & Dahlgren, N. L. (1993). *The DARPA TIMIT acoustic-phonetic continuous speech corpus*. Philadelphia, PA: Linguistic Data Consortium.
- George, E. L., Festen, J. M., & Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 120, 2295–2311.
- George, E. L., Zekveld, A. A., & Kramer, S. E. (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *The Journal of the Acoustical Society of America*, 121, 2362–2375.
- Gifford, R. H., Bacon, S. P., & Williams, E. J. (2007). An examination of speech recognition in a modulated background and of forward masking in younger and older listeners. *Journal of Speech and Hearing Research*, 50, 857–864.
- Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013). Development, reliability, and validity of PRESTO: A new high-variability sentence recognition test. *Journal of the American Academy of Audiology*, 24(1), 26–36.
- Gustafsson, H. Å., & Arlinger, S. D. (1994). Masking of speech by amplitude-modulated noise. *The Journal of the Acoustical Society of America*, 95, 518–529.
- Hopkins, K., & Moore, B. C. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *The Journal of the Acoustical Society of America*, 130, 334–349.
- Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology*, 18(7), 609–623.
- Humes, L. E. (2013). Understanding the speech-understanding problems of older adults. *American Journal of Audiology*, 22, 303–305.
- Jin, S. H., & Nelson, P. B. (2010). Interrupted speech perception: The effects of hearing sensitivity and frequency resolution. *The Journal of the Acoustical Society of America*, 128, 881–889.
- Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 122, 2365–2375.
- Kwon, B. J., & Turner, C. W. (2001). Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference? *The Journal of the Acoustical Society of America*, 110, 1130–1140.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, 103, 18866–18869.
- Miller, G. A., & Licklider, J. C. R. (1950). The intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, 22, 167–173.
- Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, 9, 399–406.
- Nagaraj, N. K., & Knapp, A. N. (2015). No evidence of relation between working memory and perception of interrupted speech in young adults. *The Journal of the Acoustical Society of America*, 138, EL145–EL150.
- Nelson, P. B., Jin, S. H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 113, 961–968.
- Port, R. F. (2003). Meter and speech. *Journal of Phonetics*, 31, 599–611.
- Rönneberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., . . . Rudner, M. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7, 31.
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 336, 367–373.

- Seewald, R. C., Ramji, K. V., Sinclair, S. T., Moodie, K. S., & 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*. London, Ontario, Canada: The University of Western Ontario.
- Shafiro, V., Sheft, S., Risley, R., & Gygi, B. (2015). Effects of age and hearing loss on the intelligibility of interrupted speech. *The Journal of the Acoustical Society of America*, 137, 745–756.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304.
- Shinn-Cunningham, B. G., & Wang, D. (2008). Influences of auditory object formation on phonemic restoration. *The Journal of the Acoustical Society of America*, 123, 295–301.
- Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, 111, 1872–1891.
- Stone, M. A., Füllgrabe, C., & Moore, B. C. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America*, 132, 317–326.
- Stone, M. A., & Moore, B. C. (2014). On the near non-existence of “pure” energetic masking release for speech. *The Journal of the Acoustical Society of America*, 135, 1967–1977.
- Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America*, 74, 695–705.
- Studebaker, G. A. (1985). A “rationalized” arcsine transform. *Journal of Speech and Hearing Research*, 28, 455–462.
- Summers, V., & Molis, M. R. (2004). Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level. *Journal of Speech, Language, and Hearing Research*, 47, 245–256.
- Takahashi, G. A., & Bacon, S. P. (1992). Modulation detection, modulation masking, and speech understanding in noise in the elderly. *Journal of Speech and Hearing Research*, 35, 1410–1421.
- Wilson, R. H., Arcos, J. T., & Jones, H. C. (1984). Word recognition with segmented-alternated CVC words: A preliminary report on listeners with normal hearing. *Journal of Speech, Language, and Hearing Research*, 27, 378–386.
- Wilson, R. H., & Irish, S. E. (2015). Recognition performance on words interrupted (10 ips, 50% duty cycle) with two interruption patterns referenced to word onset: Young listeners with normal hearing for pure tones and older listeners with sensorineural hearing loss. *International Journal of Audiology*, 54(12), 933–941.
- Wojtczak, M., & Viemeister, N. F. (2005). Forward masking of amplitude modulation: Basic characteristics. *The Journal of the Acoustical Society of America*, 118, 3198–3210.