Simultaneous and forward masking of vowels and stop consonants: Effects of age, hearing loss, and spectral shaping

Daniel Fogerty^{a)}

Department of Communication Sciences and Disorders, University of South Carolina, 1224 Sumter Street, Suite 300, Columbia, South Carolina 29208, USA

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

Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, 135 Rutledge Avenue, MSC 550, Charleston, South Carolina 29425-5500, USA

(Received 26 August 2016; revised 11 January 2017; accepted 14 January 2017; published online 27 February 2017)

Fluctuating noise, common in everyday environments, has the potential to mask acoustic cues important for speech recognition. This study examined the extent to which acoustic cues for perception of vowels and stop consonants differ in their susceptibility to simultaneous and forward masking. Younger normal-hearing, older normal-hearing, and older hearing-impaired adults identified initial and final consonants or vowels in noise-masked syllables that had been spectrally shaped. The amount of shaping was determined by subjects' audiometric thresholds. A second group of younger adults with normal hearing was tested with spectral shaping determined by the mean audiogram of the hearing-impaired group. Stimulus timing ensured that the final 10, 40, or 100 ms of the syllable occurred after the masker offset. Results demonstrated that participants benefited from short temporal delays between the noise and speech for vowel identification, but required longer delays for stop consonant identification. Older adults with normal and impaired hearing, with sufficient audibility, required longer delays to obtain performance equivalent to that of the younger adults. Overall, these results demonstrate that in forward masking conditions, younger listeners can successfully identify vowels during short temporal intervals (i.e., one unmasked pitch period), with longer durations required for consonants and for older adults.

© 2017 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4976082>]

[TCB]

Pages: 1133–1143

I. INTRODUCTION

In everyday listening environments speech is frequently heard in the presence of a fluctuating masker, such as a competing talker. The fluctuating masker, in combination with the fluctuations of the target speech, results in ongoing changes to the signal-to-noise ratio (SNR) for different portions of speech. That is, within brief time intervals, individual speech segments can occur at favorable or unfavorable SNRs, depending on whether the segments coincide in time with the peak (on, higher level) or valley (off, lower level) portions of the fluctuating masker. Vowels, being the most intense portions of speech, are the speech segments most likely to occur at favorable SNRs, even within the on-portion of the fluctuating masker. Furthermore, better SNRs within a dip of the noise masker (that is, between a preceding noise peak and a later occurring vowel), may also reduce the effect of forward masking (e.g., Humes *et al.*, 2010). Accordingly, vowels may be less susceptible to simultaneous and forward masking. In contrast, obstruent consonants, characterized by low intensity cues, are the speech segments most likely to occur at unfavorable SNRs (relative to the

level of simultaneous or preceding noise peaks) and be affected more by simultaneous and forward masking. Furthermore, consonant cues are frequently transient in nature compared to the longer duration vowels (Stevens, 2002; Ladefoged and Disner, 2012), and shorter stimulus durations can increase susceptibility to forward masking (Moore, 1981). In addition, consonants and vowels in sentences interact with the rate of temporal modulation present during nonsimultaneous portions of a fluctuating noise (Fogerty *et al.*, 2016). Therefore, it is likely that a combination of simultaneous and forward masking contributes to the differences in perceptual resolvability of consonant and vowel cues in the presence of fluctuating maskers.

This study was designed to determine the extent to which vowels and consonants differ in their susceptibility to simultaneous and forward masking. Such masking differences might partially explain speech recognition in fluctuating masker conditions. We hypothesized that vowel identification would be less affected by simultaneous and forward maskers than consonant identification, due to the more advantageous SNR at which vowels typically occur. A second purpose was to determine how age and hearing loss affect consonant and vowel identification in simultaneous and forward masking. This purpose is particularly important as older adults with normal and impaired hearing demonstrate decreased abilities to capitalize on improved SNRs during dips in fluctuating

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

^{b)}Also at: Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742, USA.

maskers (Festen and Plomp, 1990; Takahashi and Bacon, 1992; Dubno *et al.*, 2002; George *et al.*, 2007; Gifford *et al.*, 2007).

With regard to vowels, vowel perception in noise by younger adults with normal hearing is likely better than consonant perception in noise, due to longer and more intense vowel cues compared to the shorter, low-level cues of obstruent consonants (Blumstein and Stevens, 1979; Kewley-Port *et al.*, 1983). Access to the slow, periodic fluctuations of vowels may also benefit vowel perception in temporally complex noise (e.g., Fogerty *et al.*, 2015). However, aging appears to negatively affect the ability to use these cues for vowel perception. For older adults, the benefit obtained from momentary improvements in speech audibility decreases as forward masking thresholds increase (Dubno *et al.*, 2003). Neural encoding of vowel periodicity declines for older adults even without hearing loss (Clinard and Tremblay, 2013), which may limit the ability to use this temporal periodicity cue to identify vowels in a masker. In addition, older adults demonstrate poorer forward and backward masked thresholds for vowel identification in noise (Humes *et al.*, 2010). Thus, older adults may obtain less benefit from vowel cues, even when those cues occur at favorable SNRs. Given this evidence, we hypothesized that forward masking of vowels would result in poorer identification by older adults than by younger adults, due to age-related declines in coding slow periodic fluctuations of vowels.

Consonants (and in particular, stop consonants) are frequently characterized by brief, transient acoustic cues (Stevens, 2002; Ladefoged and Disner, 2012). For all listeners, perception of transient stop consonant cues is limited in noise maskers due to frequently poor SNRs. However, identifying brief consonant cues in the presence of a fluctuating masker may present particular difficulties for older adults due to poorer temporal processing (e.g., Pichora-Fuller, 2003; Humes *et al.*, 2010). Given that detection of brief signals within a masker is poorer for older than younger adults (Dubno *et al.*, 2002), perception of brief consonant cues may be further limited for older adults with normal and impaired hearing, even with sufficient audibility. In addition, older adults with hearing loss have shallower slopes of recovery from forward masking (Svec *et al.*, 2016). Therefore, we predicted that consonant identification would be poorer than vowel identification for all listeners, and poorer for older adults than younger adults due to poorer detection of brief acoustic cues.

Reduced audibility as a result of hearing loss also significantly affects vowel and consonant identification, particularly for high-frequency consonants (e.g., Sher and Owens, 1974). By providing appropriate amplification, speech perception by listeners with mild-to-moderate hearing loss can be equivalent to that of age-matched normal hearing listeners (Humes, 2007). In addition, under conditions that assure equal speech audibility using threshold matching noise, consonant recognition is similar for normal-hearing and hearing-impaired listeners (Dubno and Schaefer, 1992). The current study used the method of Humes (2007) to spectrally shape the speech materials (i.e., increase levels in 1/3 octave bands) to ensure audibility. However, spectral shaping does

not control for other suprathreshold factors associated with cochlear hearing loss, such as broader auditory filters, which are known to significantly affect speech perception in noise (e.g., Baer and Moore, 1993), or higher-than-normal forward masking thresholds (e.g., Glasberg *et al.*, 1987). In addition, spectral shaping introduces acoustic distortions that may be detrimental for speech perception, particularly for temporally complex speech tasks (Fogerty *et al.*, 2015). These suprathreshold factors beyond audibility may be especially relevant for perception of isolated syllables, where the compensatory role of top-down linguistic and cognitive factors may be limited (e.g., Bronkhorst *et al.*, 2002; Trevino and Allen, 2013).

In summary, the current study examined consonant and vowel identification under simultaneous and forward masking conditions. Older adults with normal and impaired hearing were tested to examine the effects of age and hearing loss in comparison to younger listeners with normal hearing. Spectral shaping was provided to minimize the effects of reduced audibility that are associated with hearing loss. Finally, a younger control group listened to spectrally shaped speech to assess potential perceptual consequences of shaping-related acoustic distortions.

II. METHODS

A. Listeners

Four groups of listeners participated in the study: younger adults with normal hearing (YNH; $N = 20$, mean age = 22 years, range = 18–28), older adults with normal hearing (ONH; $N = 20$, mean age = 68 years, range = 62–82), older adults with hearing loss (OHI; $N = 20$, mean age = 72 years, range = 63–88), and a separate group of younger adults with normal hearing who listened to speech that was spectrally shaped based on the mean audiogram of the OHI listeners (YSC; $N = 20$, mean age = 21 years, range = 18–25). Three additional OHI listeners were unable to obtain criterion scores in quiet and were not included in the analyses. 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., test ear). Audiometric thresholds for each listener group are reported in Table I.

To control for audibility across all participants, all groups (except YSC) listened to speech that was spectrally shaped based upon their individual audiograms. The YSC group listened to speech that underwent a standard spectral shaping based on the mean audiometric thresholds of the OHI group. According to the Speech Intelligibility Index (SII; ANSI, 1997), maximum speech intelligibility is obtained when the root-mean-square speech spectrum is 15 dB above the listener's hearing threshold. Spectral shaping of the stimuli was designed to restore audibility of the speech signal to this criterion through at least 3.15 kHz. For all experimental conditions, spectral shaping was applied to the combined speech + noise signal.

TABLE I. Mean audiometric thresholds (dB HL) for each of the four listener groups. Standard deviations are displayed in parentheses.

Frequency (kHz)	YNH	YSC	ONH	OHI
0.25	8 (6)	9 (6)	9 (6)	18 (9)
0.5	6 (6)	6 (6)	10 (6)	19 (11)
1	4 (7)	5 (6)	9 (7)	24 (13)
2	4 (5)	7 (7)	9 (5)	35 (12)
3	2 (5)	6 (7)	10 (7)	44 (8)
4	3 (7)	4 (8)	14 (7)	52 (6)
6	5 (7)	5 (8)	20 (7)	57 (7)
8	4 (8)	8 (8)	33 (16)	65 (10)

Figure 1 displays the mean audiometric thresholds in the test ear for each participant group along with the average speech spectrum following spectral shaping. Note that because audiometric thresholds were used for 1/3 octave band level calculations of the speech, two of the normal hearing groups (YNH and ONH) listened to speech that received little or no spectral shaping in the frequency range up to 3 kHz. However, due to somewhat elevated thresholds above 3 kHz for the ONH group, some spectral shaping was applied to the speech in this region. This spectral shaping helped to ensure adequate speech audibility for the ONH participants for higher frequency components (i.e., consonants). Following spectral shaping, the average overall speech presentation level was 71.5 dB sound pressure level (SPL) for YNH (SD = 2.0 dB) and ONH (SD = 2.2 dB) and 77.5 dB SPL for OHI (SD = 6.0 dB) and YSC (SD = 0.0 dB, all YSC listeners received identical spectral shaping).

B. Stimuli

1. Speech

Eighteen syllables spoken by a single male talker were selected from the UCLA version of the Nonsense Syllable Test (NST; see [Dubno and Schaefer, 1992](#)). Three voiced stop consonants /b, d, g/ were paired with each of three vowels /i, a, u/ in consonant-vowel (CV) and vowel-consonant

FIG. 1. Mean long-term speech spectrum levels (lines without symbols) and mean hearing thresholds (lines with symbols) for the four listener groups. Spectral shaping was based on each individual's audiometric threshold for YNH, ONH, and OHI listener groups. All YSC listeners received the same spectral shaping based on the mean audiogram of the OHI listeners.

(VC) syllables. The original productions of these syllables were characterized by somewhat prolonged phonemes compared to what would be expected in fluent speech for sentence materials. Syllables varied in length from 468 to 636 ms. However, for sentence materials, average vowel and consonant durations are closer to 97 and 62 ms, respectively ([Fogerty and Kewley-Port, 2009](#)), resulting in average CV or VC syllables of approximately 159 ms. One of the purposes of the current study was to investigate temporal processes that may operate for natural speech productions, such as sentences. Because stimulus durations can influence forward masking ([Kidd and Feth, 1982](#); [Zwicker, 1964](#)), syllable materials were reduced in duration (procedures detailed below) to more closely model the timescales and temporal interactions that are likely to operate during typical speech conditions with forward and/or simultaneous maskers. Pitch cues that can provide supplemental acoustic-phonetic cues (e.g., [Winn et al., 2013](#)) to syllable identification were also removed by flattening the fundamental frequency to 100 Hz. This step helped to ensure that performance was based on temporal interactions between the speech and noise.

Temporal and pitch processing was implemented using STRAIGHT ([Kawahara et al., 1999](#)), a speech analysis and synthesis program. Given the long duration of the syllables, linear time compression to 150 ms would result in removal of brief transient stop cues associated with the consonant burst. To preserve these acoustic-phonetic cues in the speech signal, nonlinear time compression was implemented that modeled the nonlinear effects of naturally producing more rapid speech (e.g., [Janse, 2004](#)). For natural, fluent speech, time constraints in articulatory movements result in greater preservation of consonant duration relative to vowel duration at faster speaking rates ([Gay, 1978](#); [Max and Caruso, 1997](#)). Syllable intervals were defined by vowel periodicity and nonlinear time compression reduced syllable durations to 150 ms (± 2 ms). Syllables contained between 7–9 pitch pulses (i.e., number of voicing periods). The combined duration of the voiced closure and noise burst was 71.2 ms (SD = 9.8 ms); the duration of the vocalic portion of the syllable was 78.3 ms (SD = 9.5 ms). A paired t-test demonstrated no significant difference in the duration of these two portions of the signal ($p > 0.05$). Example waveforms for the final stimuli in quiet are displayed in the right panels of Fig. 2.

2. Masker

A masking noise was created that matched the long-term spectrum of a concatenation of all of the processed (but not shaped) speech files (i.e., all 18 syllables). Given the homogeneity of the speech stimuli, mostly characterized by the spectrum levels of three vowels, the resulting masking noise was vowel-like, with broad resonant peaks, but the spectrum did not represent one of the tested vowels more than another. The noise spectrum ensured sufficient energetic masking of the vowel formants and is displayed in Fig. 3 in comparison to the syllable spectrum for the three vowels spoken in a /bV/ syllable. The masker also preserved some random fluctuations in periodicity and amplitude, as can be

FIG. 2. Example of the stimulus conditions tested. Vowel and consonant identification was tested for CV and VC syllables in quiet (far right) and in noise with three stimulus offset delays. Offset delays were defined as the time between the end of the noise and the end of the syllable.

observed from the bottom panel in Fig. 3 which displays the amplitude envelope derived by halfwave rectification and lowpass filtering using a sixth-order Butterworth filter with a 256-Hz cutoff. Thus, the masker can be thought of as having some fluctuating properties (not unlike Gaussian noise, see Stone *et al.*, 2011). This processing resulted in a spectrally and temporally complex masker. On each trial, a random

200-ms segment of this noise was selected from a 24 s noise file. A 2-ms raised cosine on/off ramp was then applied to the broadband noise masker.

3. Combined speech and noise presentation

Consonant and vowel identification was tested in quiet and in noise with the relative timing of the speech and noise determined by three stimulus offset delays: 10, 40, and 100 ms. The offset delay defined the time between the end of the noise and the end of the syllable. As syllables were 150 ms in duration, identification of initial phonemes was primarily determined by simultaneous masking, whereas final phoneme identification involved a combination of simultaneous and forward masking, which was defined by the offset delay. Stimuli were summed at these three offset delays using an SNR of -5 dB relative to the average syllable level. SNRs were equalized for all testing based on the average syllable level in order to model everyday listening contexts that preserve intrinsic differences in the level of vowels and consonants. Figure 2 displays these stimulus combinations for an example CV and VC syllable. From the figure, it can be seen that 10 and 40 ms delays resulted in partial simultaneous masking of the final vowel or consonant segment for CV and VC syllables, respectively. In contrast, for the 100 ms delay, only the initial segment was presented with simultaneous masking. Thus, final vowel or consonant identification with the 100 ms delay is dominated by the potential effects of forward masking.

C. Procedures

Experimental testing was completed in a sound attenuating booth, either at the University of South Carolina (for YNH and YSC groups) or at the Medical University of

FIG. 3. The top panel displays spectra for the masking noise (thick black) and for the three vowels spoken in a /bV/ context. The bottom panel displays the amplitude envelope low-pass filtered at 256 Hz for a random 200-ms selection of the noise.

South Carolina (for ONH and OHI groups). Both laboratories were similarly equipped and used identical stimulus presentation and spectral shaping software. The participants listened to monaural 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). Testing was conducted in two phases for syllable presentation in quiet and noise, respectively. Syllable testing was blocked by CV or VC syllable and by consonant or vowel identification. The order of the blocks was counterbalanced across participants. Each block consisted of a three-alternative forced-choice task, with buttons labeled by the three vowels or three consonants. In order to facilitate participant attention to the target segment, an orthographic cue to the syllabic structure (“CV” or “VC”) was displayed on the screen during testing.

For quiet testing, an initial training block was completed prior to the experimental block in order to familiarize participants to the task and processed syllables. During this training block, the correct consonant or vowel response was visually highlighted (i.e., the correct response button turned yellow) during stimulus presentation. Participants were required to listen to the stimulus and select the correct (i.e., highlighted) response. During experimental testing, no feedback was provided in order to minimize perceptual learning during experimental testing and provide stable performance between test blocks. If a participant did not obtain a phoneme identification score greater than 60%, the training and testing blocks were completed a second time. If the second score was greater than 60%, it was recorded as the participant’s performance in quiet and served as a baseline measure of phoneme identification. Three OHI listeners could not obtain this performance level and were not included in the analysis with the remaining 20 OHI listeners.

Testing in noise consisted of identifying consonants or vowels with various durations of temporal overlap with the masker. Consonant and vowel identification blocks were again counterbalanced, but the noise overlap was tested in the following order: 10, 40, 100 ms, proceeding from the most temporal overlap to the least. Thus, testing proceeded from the most to least difficult temporal delay, which minimized potential learning effects between test blocks. Prior to each experimental block, participants completed a familiarization block without feedback. All stimuli were presented three times within an experimental block and all experimental conditions were cycled through twice. If a given condition had a difference in performance between the two cycles that was greater than 20 percentage points, a third run of that condition was completed. The average performance on the two closest experimental runs was used as the participant’s score for each condition.

III. RESULTS

The results for the different experimental conditions were analyzed to determine the effect of phoneme type (vowel versus consonant) and syllable type (CV versus VC) for the four listener groups. Percent correct scores for all conditions were transformed to rationalized arcsine units

(RAU) to stabilize the error variance prior to analysis (Studebaker, 1985). Bonferroni correction was used for all tests involving multiple comparisons (i.e., t-tests discussed below).

A. Phoneme identification in quiet

Baseline scores in quiet were examined first. Mean initial and final vowel identification for all groups was at ceiling performance (>99%). Mean initial consonant identification was between 77% and 93% across groups, whereas mean final consonant identification ranged from 66% and 79%. These scores, converted to RAU, are displayed as the isolated symbols in Fig. 4. A mixed model analysis of variance (ANOVA) demonstrated main effects of the syllable type [$F(1, 76) = 52.0, p < 0.001$], phoneme type [$F(1, 76) = 7.3, p < 0.01$] and listener group [$F(3, 76) = 6.5, p < 0.001$]. Significant interactions were also observed between phoneme type and group [$F(3, 76) = 7.3, p < 0.001$] and between syllable type and phoneme type [$F(1, 76) = 39.4, p < 0.004$]. Independent samples t-tests demonstrated that YNH listeners performed significantly better than the other three groups for consonant identification in CV syllables ($p < 0.001$). YNH and YSC listeners also performed better than ONH listeners for vowel identification in CV syllables ($p < 0.001$). No other group differences were statistically significant. Across all groups, vowel identification was better than consonant identification regardless of the syllable type ($p < 0.001$), and consonant identification was better with the consonant in the initial position than in the final position ($p < 0.001$).

B. Phoneme identification in noise

1. Analysis of initial phonemes

Identification of initial phonemes primarily indicates effects of simultaneous masking. However, up to ~30 ms of the end of the initial phoneme was unmasked at an offset delay of 100 ms. Therefore, the 100-ms delay could preserve information in the formant transitions between the two phonemes. As can be viewed in Fig. 4, identification of vowels was better than consonants for all listener groups across stimulus offset delays, although this difference in initial phoneme identification did not reach significance for YNH and OHI groups at the 10-ms delay. Overall, these results parallel the better identification performance for vowels in quiet.

Similar to the analysis of identification in quiet, a 2 (phoneme) $\times$ 3 (offset delay) $\times$ 4 (group) mixed model ANOVA was conducted (Table II). *Post hoc* contrasts demonstrated that the most notable difference between these conditions is the significantly better performance for both younger groups compared to the older groups for consonant and vowel identification at the 100-ms offset delay ($p < 0.05$). This group difference indicates that younger listeners were able to capitalize on a release from masking for the formant transitions between the syllable segments and potentially the final portion of the initial segment. In contrast, older adults were only able to use these cues to a minimal degree in simultaneous masking to improve

FIG. 4. Identification scores for consonants in the initial (CV) and final (VC) positions and vowels in the initial (VC) and final (CV) positions in quiet and noise for three offset delays. Results are displayed in separate panels for the four listener groups. Chance performance is indicated by the horizontal dotted line. Error bars indicate standard error of the mean.

performance from the 40-ms condition. Indeed, whereas initial vowel and consonant identification by younger listeners significantly improved from 40- to 100-ms delays [vowel: $t(39) = 6.8$, $p < 0.001$, $d = 1.1$; consonant: $t(39) = 7.0$, $p < 0.001$, $d = 1.1$], older adults made only minimal gains at the longer delay for vowel and consonant identification [vowel: $t(39) = 3.1$, $p < 0.01$, $d = 0.5$; consonant: $t(39) = 2.2$, $p = 0.04$, $d = 0.3$].

2. Analysis of final phonemes

The stimulus offset delays resulted in syllable final phoneme identification occurring with partial to no overlap with the noise. Thus, these conditions reflect either a combination of simultaneous and forward masking (i.e., 10-ms and 40-ms delays) or only forward masking (i.e., 100-ms delay). Examination of Fig. 4 demonstrates that the identification of final vowels in CV syllables is above chance for all groups when only 10-ms (i.e., one pitch period) at the end of the

vowel is unmasked. This finding is in agreement with earlier reports of forward masking of vowels for younger and older adults (Humes *et al.*, 2010). However, older adults performed significantly poorer than younger adults in this condition [older: 57 ± 16 rau; younger: 69 ± 15 rau, $t(78) = 3.5$, $p < 0.001$]. Consistent with this observation, the earlier report by Humes *et al.* (2010) found that older adults required an additional 20 ms delay to obtain performance similar to the younger adults. Vowel identification was markedly improved with 40-ms delays for all groups. However, additional benefit was not seen with a 100-ms delay. This finding suggests that the partial simultaneous masking in the 40-ms delay had little effect on performance. Instead, performance at these two delays may have been dominated by the effect of forward masking on the initial transition into the vowel. Evidence for this hypothesis can be observed by improved performance for quiet presentations relative to the 100-ms delay condition for ONH, OHI, and YSC groups ($p < 0.01$). (All groups had significantly poorer final vowel identification at a 40-ms delay compared to quiet, $p < 0.01$.)

Final consonants were mostly characterized by a weak voiced closure and a final transient noise burst. In the 10-ms delay condition, the final burst was mostly available in quiet, so scores were primarily determined by forward masking. However, only the younger participants were able to use this final burst to identify the final consonant when tested at 40 ms. Forward masking may have reduced perception of the burst following a 40-ms delay for older adults. This observation may be related to slower recovery from forward masking, as observed with fluctuating maskers for older adults with hearing loss (Svec *et al.*, 2016). Identification of the final consonant by older adults was observed with the 100-ms delay, indicating that some information from the formant

TABLE II. F values and effect sizes (η^2) from the ANOVA results for initial and final phoneme position conditions. ($p < 0.01$, n.s. = not significant.)

Effect	df	Initial position		Final position	
		F	η^2	F	η^2
Phoneme (P)	1, 76	108.4	0.59	685.3	0.90
Offset (O)	2, 152	85.1	0.53	421.9	0.85
Group (G)	3, 76	12.5	0.33	17.3	0.41
PxG	3, 76	n.s.		n.s.	
OxG	6, 152	8.5	0.25	11.1	0.31
PxO	2, 152	8.2	0.10	40.3	0.35
PxOxG	6, 152	n.s.		3.6	0.12

transitions from the vowel into the consonant may be essential for older adults.

Identification of final vowels with a 10-ms delay was higher for YSC subjects than for all other groups, including YNH ($p < 0.001$). Younger listeners performed better than both older groups for 40-ms and 100-ms delays for consonant and vowel identification ($p < 0.01$), except that YSC performance for vowel identification at 40 ms was not significantly different from ONH performance after correction for multiple comparisons. Performance for the OHI group was comparable to that of the ONH group, except for poorer performance for final consonant identification at 100 ms ($p < 0.001$). The YSC group also had poorer performance than the YNH group for this 100-ms condition, although this comparison did not reach significance after correction for multiple comparisons. In this case, acoustic distortions stemming from spectral shaping may have limited expected performance gains from increased audibility for the OHI group. However, the reduced performance for OHI as compared to YSC and ONH for final consonants with 100-ms offset delay suggests effects of both age and hearing loss on recovery from forward masking.

3. Comparison between conditions

One notable observation is that OHI listeners are the only group that did not match their quiet performance for final consonants with the 100-ms offset delay ($p < 0.001$). However, performance was poorer than in quiet for all groups for higher intensity final vowels in the 100-ms condition ($p < 0.01$), which reflected only forward masking, although differences for the YNH group were not significant ($p > 0.05$). Even for OHI listeners, effect sizes were larger for forward masking of the vowel (i.e., difference between quiet and 100 ms) than for the consonant ($d = 1.2$ versus $d = 0.9$, respectively). Thus, preserved properties of the initial vowel may facilitate identification of the final consonant when presented in quiet. Access to these vowel transitional cues in the 100-ms delay condition may be sufficient to obtain consonant identification scores comparable to quiet (except for OHI listeners).

Figure 5 replots the identification data from Fig. 4 according to phoneme position across the consonant and vowel conditions. Comparison across these four panels indicates differences between consonant and vowel identification in noise. All groups demonstrate some ability to use just 10 ms (i.e., one unmasked pitch pulse) at the end of the vowel for vowel identification. This results in comparable performance for initial vowels with a 100-ms delay that provided some release of simultaneous masking at the end of the vowel. In contrast, all groups required a 40-ms delay of the noise in the final position to match their performance for initial consonants with 100-ms delay (although this later comparison was limited to younger adults who had above chance performance).

These results may highlight greater forward masking for final consonant cues. In addition, transitional cues from the final vowel aid in the identification of initial consonants (see also [Delattre et al., 1955](#)). Thus, while listeners obtain

FIG. 5. Identification scores are displayed in four panels for vowels and consonants in initial and final positions. Within each panel performance is compared among the four participant groups. Chance performance is indicated by the horizontal dotted line. Error bars indicate the standard error of the mean.

benefit from a single pitch pulse for vowel identification, consonant identification in noise is most robust when additional information from the neighboring vowel is available or when additional time for recovery from forward masking is provided.

4. Correlational analysis

A correlational analysis was conducted to determine the association between the different measures of phoneme identification in noise for the younger (YNH and YSC) and older (ONH and OHI) listeners. Phoneme identification scores were collapsed across delays to provide global measures of phoneme identification in the presence of noise. Correlation coefficients in Table III demonstrate significant associations for the younger listeners (below the diagonal) between final vowel and final consonant identification and between either of the two final segments and initial vowel identification. However, identification of initial consonants in noise was independent of the three other measures. The significant associations may indicate a common perceptual ability in recovery from forward masking for the younger adults. One

TABLE III. Correlation coefficients between the accuracy in different noise conditions collapsed across onset delays. Coefficients below the diagonal are for younger adults and above the diagonal (in italics) are for older adults.

	V_initial	V_final	C_initial	C_final
V_initial		<i>0.33^a</i>	<i>0.15</i>	<i>0.23</i>
V_final	0.50 ^b		<i>0.31</i>	<i>0.21</i>
C_initial	−0.09	−0.02		<i>0.43^b</i>
C_final	0.39 ^a	0.48 ^b	−0.11	

^a $p < 0.05$.

^b $p < 0.01$.

reason the initial consonant condition might not have clustered with the remaining measures is that younger adults required about 40-ms delay to obtain significant masking release for consonants (as seen with the final consonants across delays) but only required 10-ms delay for vowels. In contrast, correlations for older adults (in italics in Table III) demonstrated associations between initial and final position for both vowel and consonant identification and no associations between consonant and vowel identification. This result may indicate different effects of aging for vowel and consonant perception. Finally, it is notable that correlations between composite phoneme identification scores and pure-tone average (PTA) thresholds (1, 2, and 4 kHz) were not significant for the older adults ($p > 0.05$). The overall similar performance between OHI and ONH groups suggests that spectral shaping was generally sufficient to compensate for reduced audibility due to elevated thresholds, with two notable exceptions. PTA was significantly negatively correlated with (1) initial consonant recognition in quiet ($r = -0.33$, $p < 0.05$) and (2) final consonant recognition at the 100-ms delay ($r = -0.50$, $p < 0.001$). These results suggest that spectral shaping provided to OHI listeners with higher PTAs may not have been sufficient and that reduced audibility may have resulted in poorer consonant identification by some OHI subjects in these conditions. However, in these two conditions, YSC listeners also performed poorer than their age-matched counterparts. Thus, given that higher PTAs require more spectral shaping, these significant correlations raise the concern that spectral shaping distorts some acoustic cues important for consonant recognition under certain conditions.

IV. DISCUSSION

The main pattern of results demonstrates that vowel and consonant identification in simultaneous and forward maskers is poorer for older adults than younger adults. Analysis of the different conditions suggests the involvement of periodicity and formant change, two speech cues known to facilitate the processing of consonants and vowels. Performance was similar for ONH and OHI groups when audibility was ensured by providing appropriate spectral shaping. One exception was for identification of final consonants, which was also reduced for the YSC group as compared to the YNH group and highlights potential acoustic or perceptual distortions related to spectral shaping rather than elevated thresholds. Overall, these results demonstrate the importance of speech cues related to periodicity and formant change and the acoustic effects of spectral shaping on vowel and consonant identification by the four subject groups.

A. Temporal envelope periodicity processing

Poorer vowel identification during forward masking may be related to poorer periodicity coding by older than younger adults. For example, older listeners demonstrate poorer periodicity coding of resolved harmonics in a fundamental frequency discrimination task than younger adults (Vongpaisal and Pichora-Fuller, 2007). Reduced coding of the temporal envelope periodicity for modulation rates above

100 Hz (i.e., the periodicity of a speaker's pitch) has been observed for older adults using behavioral (Schvartz and Chatterjee, 2012) and neural measures (Walton *et al.*, 2002; Purcell *et al.*, 2004; Leigh-Paffenroth and Fowler, 2006; Grose *et al.*, 2009). These results suggest that a reduction in periodicity processing combined with slower recovery from forward masking may limit vowel identification in noise for older adults with normal and impaired hearing.

In addition to perceptual aspects of periodicity processing, we also noted acoustic distortions of periodicity following spectral shaping. Naturally produced vowels contain some degree of aperiodicity due to turbulent airflow at the glottis (Hillenbrand, 1987), which can be quantified using the harmonic-to-noise ratio (Yumoto *et al.*, 1984; Awan and Frenkel, 1994). Noise components in the high frequency spectrum can be emphasized during spectral shaping, which results in an amplified noise component. For example, examination of the /dV/ syllables indicated a decrease in the harmonic-to-noise ratio of 5 dB during the vowel portion of the syllable, which is a characteristic of reduced periodicity. Nevertheless, YSC and OHI listeners performed similarly to their age-matched groups for vowel identification. However, for consonant identification, reduced acoustic periodicity may combine with reduced periodicity processing by some older listeners or for some spectral shaping parameters. These combined factors could have had an effect on processing consonant cues and resulted in poorer consonant identification for OHI than for YSC groups.

B. Formant change

We also observed that older adults received less benefit compared to younger adults from transitional cues at the end of the vowel for identifying final consonants. This result is consistent with prior work showing that older listeners are less able than younger adults to use dynamic formant change to aid in consonant identification in silent center CVC syllables (Fox *et al.*, 1992). Thus, poorer processing of final consonants might also be due to the combined effects of forward masking of the consonant burst and poorer processing of dynamic vowel cues used for consonant identification. Incidentally, the reduced processing of dynamic formant change is also associated with poorer vowel processing (Fox *et al.*, 1992), which may relate to poorer initial vowel identification for older adults with the 100-ms delay. In this condition, the final portion of the initial vowel is in quiet, where transitional dynamics into the final consonant would be present. Poorer processing of these dynamic cues to vowel identification may help to explain this effect, potentially in combination with reduced periodicity processing.

C. Spectral shaping

The purpose of testing the YSC group was to determine the perceptual effect in normal control subjects of spectrally shaping the speech stimulus, which raises its overall level in addition to altering its spectral shape. Thus, as noted earlier, the YSC group only provided an acoustic control and did not model changes in suprathreshold auditory function by the impaired ear.

Overall, both YSC and OHI groups (groups with matched spectral shaping) performed similarly to their age-matched counterparts. This general trend is in agreement with several studies that have documented that audibility is the primary peripheral factor contributing to poorer speech understanding by older adults (reviewed by [Humes, 2013](#)). This observation also suggests that suprathreshold auditory deficits not controlled by spectral shaping, such as higher forward masking thresholds ([Dubno and Schaefer, 1992](#)) and poorer frequency resolution ([Baer and Moore, 1993](#)), had a minimal effect on performance for the older listeners with hearing loss for the majority of the tested conditions.

However, there were select cases that indicate a negative effect of spectral shaping. This result was most notable for the identification of final consonants with a 100-ms delay by YSC and OHI listeners. YSC listeners also performed poorer than YNH listeners for initial consonants in quiet. Spectral shaping may have altered the consonant burst spectra and disrupted important spectral shape cues used for discriminating voiced stops ([Blumstein and Stevens, 1979](#)). This hypothesis may be supported by the finding that older adults with hearing loss, under similar spectrally shaped conditions, have poorer spectral shape discrimination compared to younger subjects ([Shrivastav et al., 2006](#)). Furthermore, in previous work spectral shape discrimination was moderately correlated with nonsense syllable identification ([Shrivastav et al., 2006](#)). Similar performance declines with spectral shaping for YSC listeners have also been observed when investigating consonant cues in sentences ([Fogerty et al., 2015](#)).

It is also possible that reduced identification with spectrally shaped materials is partly due to the novelty of the shaped stimulus. While some initial training was provided to the listeners first in quiet, this exposure was brief, and performance for consonant identification did not saturate at a maximum performance level. The Ease of Language Understanding model, which incorporates episodic models of memory, predicts that the acoustic mismatch between the stored mental representation of the syllable and the spectrally shaped stimulus could have resulted in poorer processing ([Rönnberg et al., 2013](#)). Future studies will need to assess the benefit of additional exposure and training with shaped stimuli. The presence of acoustic and perceptual distortions may contribute to increased listening effort among individuals with hearing loss (e.g., [Kuchinsky et al., 2013](#)). Furthermore, even when speech recognition has been maximized with appropriate amplification, small acoustic distortions can impact the speed of lexical decisions ([Fogerty et al., 2014](#)) and impair recall of items that were initially recognized successfully (e.g., [Rabbitt, 1968](#); [Pichora-Fuller et al., 1995](#); [Surprenant, 2007](#)). Combined, these findings demonstrate the importance of quantifying acoustic and perceptual distortions imposed by spectral shaping to fully understand how signal processing may limit the benefit derived by improved audibility. Importantly, speech recognition measures should index processing speed, memory, and listening effort—abilities that may be sensitive to the perceptual consequences of listening to amplified and spectrally altered speech. This recommendation is particularly relevant

for high-frequency, low-intensity consonants cues, which are most affected by spectral shaping.

V. SUMMARY AND CONCLUSIONS

The results of this study revealed differential effects of simultaneous and forward masking for perception of consonants and vowels and effects of age and hearing loss. As expected, vowels are more resistant to simultaneous and forward masking effects than consonants. This result likely occurs due to the greater intensity and temporal periodicity of vowels, which improve their salience during masking. In addition, forward masking of consonants extends over longer durations following the noise offset compared to masking for vowels (between 10 and 40 ms). Compared to younger adults, the results from older adults with normal and impaired hearing are consistent with a slower recovery from forward masking as observed for fluctuating forward maskers ([Svec et al., 2016](#)). The similarity in performance for older adults with normal and impaired hearing suggests that the broader auditory filters typically associated with cochlear hearing loss do not fully account for reduced vowel identification. Instead, age-related declines for processing vowels in noise may occur due to poorer coding of vowel temporal periodicity. In addition, older adults with hearing loss are particularly vulnerable to forward masking of relatively brief consonants, even when audibility of these consonant features is ensured by spectral shaping.

Future studies will need to confirm these effects for speech recognition in the presence of fluctuating noise that occurs during real-world listening. Results of the current experiment suggest that consonant identification may require longer periods when the fluctuating masker level is low, whereas vowel cues remain robust during shorter fluctuations. Furthermore, older adults may require greater durations of preserved vowel cues. Poorer identification of brief consonants coupled with poor coding of temporal periodicity in vowels may limit the communication abilities of older adults in fluctuating noise. For sentence intelligibility in fluctuating maskers, performance may relate most directly to the perception of segments during temporal intervals at favorable SNRs, i.e., “glimpses.” The effective duration of these favorable temporal intervals for successful phoneme identification can be quite short for younger listeners (i.e., one pitch period), whereas longer durations may be required for processing consonant cues and for older adults who have greater effects of forward masking and reduced periodicity coding.

ACKNOWLEDGMENTS

This work was supported, in part, by an Advancing Academic-Research Career Award through the American Speech-Language-Hearing Association (D.F.) and by National Institutes of Health/NIDCD Grants Nos. R01 DC015465 (D.F.) and R01 DC000184 (J.R.D.). This work was also supported by the South Carolina Clinical and Translational Research (SCTR) Institute (NIH Grant No. UL1 RR029882). This investigation was partially conducted in a facility constructed with support from Research Facilities Improvement Program Grant No. C06 RR14516 from the NIH National Center for Research Resources.

- ANSI (1997). ANSI S3.5-1997, *Methods for the Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York).
- Awan, S. N., and Frenkel, M. L. (1994). "Improvements in estimating the harmonics-to-noise ratio of the voice," *J. Voice* **8**, 255–262.
- Baer, T., and Moore, B. C. (1993). "Effects of spectral smearing on the intelligibility of sentences in noise," *J. Acoust. Soc. Am.* **94**, 1229–1241.
- Blumstein, S. E., and Stevens, K. N. (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," *J. Acoust. Soc. Am.* **66**, 1001–1017.
- Bronkhorst, A. W., Brand, T., and Wagener, K. (2002). "Evaluation of context effects in sentence recognition," *J. Acoust. Soc. Am.* **111**, 2874–2886.
- Clinard, C. G., and Tremblay, K. L. (2013). "Aging degrades the neural encoding of simple and complex sounds in the human brainstem," *J. Am. Acad. Audiol.* **24**, 590–599.
- Delattre, P. C., Liberman, A. M., and Cooper, F. S. (1955). "Acoustic loci and transitional cues for consonants," *J. Acoust. Soc. Am.* **27**, 769–773.
- Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (2002). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," *J. Acoust. Soc. Am.* **111**, 2897–2907.
- 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., and Schaefer, A. B. (1992). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," *J. Acoust. Soc. Am.* **91**, 2110–2121.
- 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.
- Fogerty, D., Ahlstrom, J. B., Bologna, W. J., and Dubno, J. R. (2015). "Sentence intelligibility during segmental interruption and masking by speech-modulated noise: Effects of age and hearing loss," *J. Acoust. Soc. Am.* **137**, 3487–3501.
- Fogerty, D., Ahlstrom, J. B., Bologna, W. J., and Dubno, J. R. (2016). "Glimpsing speech in the presence of nonsimultaneous amplitude modulations from a competing talker: Effect of modulation rate, age, and hearing loss," *J. Speech Lang. Hear. Res.* **59**, 1198–1207.
- 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., Montgomery, A. A., and Crass, K. A. (2014). "Effect of initial-consonant intensity on the speed of lexical decisions," *Atten. Percep. Psychophys.* **76**, 852–863.
- Fox, R. A., Wall, L. G., and Gokcen, J. (1992). "Age-related differences in processing dynamic information to identify vowel quality," *J. Speech Lang. Hear. Res.* **35**, 892–902.
- Gay, T. (1978). "Effect of speaking rate on vowel formant movements," *J. Acoust. Soc. Am.* **63**, 223–230.
- George, E. L., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (2007). "Auditory and nonauditory factors affecting speech reception in noise by older listeners," *J. Acoust. Soc. Am.* **121**, 2362–2375.
- Gifford, R. H., Bacon, S. P., and Williams, E. J. (2007). "An examination of speech recognition in a modulated background and of forward masking in younger and older listeners," *J. Speech Lang. Hear. Res.* **50**, 857–864.
- Glasberg, B. R., Moore, B. C., and Bacon, S. P. (1987). "Gap detection and masking in hearing-impaired and normal-hearing subjects," *J. Acoust. Soc. Am.* **81**, 1546–1556.
- Grose, J. H., Mamo, S. K., and Hall J. W., III (2009). "Age effects in temporal envelope processing: Speech unmasking and auditory steady state responses," *Ear Hear.* **30**, 568–575.
- Hillenbrand, J. (1987). "A methodological study of perturbation and additive noise in synthetically generated voice signals," *J. Speech Lang. Hear. Res.* **30**, 448–461.
- 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., Kewley-Port, D., Fogerty, D., and Kinney, D. (2010). "Measures of hearing threshold and temporal processing across the adult lifespan," *Hear. Res.* **264**, 30–40.
- Janse, E. (2004). "Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech," *Speech Commun.* **42**, 155–173.
- Kawahara, H., Masuda-Katsuse, I., and De Cheveigne, A. (1999). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," *Speech Commun.* **27**, 187–207.
- Kewley-Port, D., Pisoni, D. B., and Studdert-Kennedy, M. (1983). "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants," *J. Acoust. Soc. Am.* **73**, 1779–1793.
- Kidd, G., Jr., and Feth, L. L. (1982). "Effects of masker duration in pure-tone forward masking," *J. Acoust. Soc. Am.* **72**, 1384–1386.
- Kuchinsky, S. E., Ahlstrom, J. B., Vaden, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., and Eckert, M. A. (2013). "Pupil size varies with word listening and response selection difficulty in older adults with hearing loss," *Psychophysiol.* **50**, 23–34.
- Ladefoged, P., and Disner, S. F. (2012). *Vowels and Consonants*, 3rd ed. (Wiley, New York).
- Leigh-Paffenroth, E. D., and Fowler, C. G. (2006). "Amplitude-modulated auditory steady-state responses in younger and older listeners," *J. Am. Acad. Audiol.* **17**, 582–597.
- Max, L., and Caruso, A. J. (1997). "Acoustic measures of temporal intervals across speaking rates: Variability of syllable- and phrase-level relative timing," *J. Speech Hear. Res.* **40**, 1097–1110.
- Moore, B. C. (1981). "Interactions of masker bandwidth with signal duration and delay in forward masking," *J. Acoust. Soc. Am.* **70**, 62–68.
- Pichora-Fuller, M. K. (2003). "Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension," *Int. J. Audiol.* **42**, 59–67.
- 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.
- Purcell, D. W., John, S. M., Schneider, B. A., and Picton, T. W. (2004). "Human temporal auditory acuity as assessed by envelope following responses," *J. Acoust. Soc. Am.* **116**, 3581–3593.
- Rabbitt P. (1968). "Channel capacity, intelligibility and immediate memory," *Q. J. Psychol.* **20**, 241–248.
- Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., and Rudner, M. (2013). "The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances," *Front. Sys. Neurosci.* **7**, 31.
- Schvartz, K. C., and Chatterjee, M. (2012). "Gender identification in younger and older adults: Use of spectral and temporal cues in noise-vocoded speech," *Ear Hear.* **33**, 411–420.
- Sher, A. E., and Owens, E. (1974). "Consonant confusions associated with hearing loss above 2000 Hz," *J. Speech Hear. Res.* **17**, 669–681.
- Shrivastav, M. N., Humes, L. E., and Kewley-Port, D. (2006). "Individual differences in auditory discrimination of spectral shape and speech-identification performance among elderly listeners," *J. Acoust. Soc. Am.* **119**, 1131–1142.
- Stevens, K. N. (2002). "Toward a model for lexical access based on acoustic landmarks and distinctive features," *J. Acoust. Soc. Am.* **111**, 1872–1891.
- Stone, M. A., Füllgrabe, C., Mackinnon, R. C., and Moore, B. C. J. (2011). "The importance for speech intelligibility of random fluctuations in 'steady' background noise," *J. Acoust. Soc. Am.* **130**, 2874–2881.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Surprenant A. M. (2007). "Effects of noise on identification and serial recall of nonsense syllables in older and younger adults," *Aging Neuropsychol. Cogn.* **14**, 126–143.
- Svec, A., Dubno, J. R., and Nelson, P. B. (2016). "Inherent envelope fluctuations in forward maskers: Effects of masker-probe delay for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **139**, 1195–1203.
- 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.
- Trevino, A., and Allen, J. B. (2013). "Individual variability of hearing-impaired consonant perception," in *Seminars in Hearing* (Thieme Medical, Stuttgart, Germany), Vol. 34, No. 2, pp. 74–85.
- Vongpaisal, T., and Pichora-Fuller, M. K. (2007). "Effect of age on F0 difference limen and concurrent vowel identification," *J. Speech Lang. Hear. Res.* **50**, 1139–1156.
- Walton, J. P., Simon, H., and Frisina, R. D. (2002). "Age-related alterations in the neural coding of envelope periodicities," *J. Neurophysiol.*

- 88, 565–578, available at <http://jn.physiology.org/content/jn/88/2/565.full.pdf>.
- Winn, M. B., Chatterjee, M., and Idsardi, W. J. (2013). “Roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering,” *J. Speech Lang. Hear. Res.* **56**, 1097–1107.
- Yumoto, E., Sasaki, Y., and Okamura, H. (1984). “Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness,” *J. Speech Lang. Hear. Res.* **27**, 2–6.
- Zwicker, E. (1964). “‘Negative afterimage’ in hearing,” *J. Acoust. Soc. Am.* **36**, 2413–2415.