Stimulus context affects the phonemic categorization of temporally based word contrasts in adult cochlear-implant users

Zilong Xie,[1,a)] Samira Anderson,[2] and Matthew J. Goupell[2]

[1]*Department of Hearing and Speech, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA*

[2]*Department of Hearing and Speech Sciences, University of Maryland, 0100 Samuel J. LeFrak Hall, College Park, Maryland 20742, USA*

ABSTRACT:

Cochlear-implant (CI) users rely heavily on temporal envelope cues for speech understanding. This study examined whether their sensitivity to temporal cues in word segments is affected when the words are preceded by non-informative carrier sentences. Thirteen adult CI users performed phonemic categorization tasks that present primarily temporally based word contrasts: Buy-Pie contrast with word-initial stop of varying voice-onset time (VOT), and Dish-Ditch contrast with varying silent intervals preceding the word-final fricative. These words were presented in isolation or were preceded by carrier stimuli including a sentence, a sentence-envelope-modulated noise, or an unmodulated speech-shaped noise. While participants were able to categorize both word contrasts, stimulus context effects were observed primarily for the Buy-Pie contrast, such that participants reported more "Buy" responses for words with longer VOTs in conditions with carrier stimuli than in isolation. The two non-speech carrier stimuli yielded similar or even greater context effects than sentences. The context effects disappeared when target words were delayed from the carrier stimuli for ≥75 ms. These results suggest that stimulus contexts affect auditory temporal processing in CI users but the context effects appear to be cue-specific. The context effects may be governed by general auditory processes, not those specific to speech processing. © *2022 Acoustical Society of America*.

https://doi.org/10.1121/10.0009838

(Received 27 September 2021; revised 20 February 2022; accepted 4 March 2022; published online 25 March 2022)

[Editor: Christian Lorenzi] Pages: 2149–2158

I. INTRODUCTION

Time-varying information in speech contains cues critical to speech understanding (Rosen, 1992). For example, voice-onset time (VOT), the time elapsed between the release of the articulators and the onset of voicing, is a cue to distinguish voiced (e.g., /b/) and unvoiced (e.g., /p/) stop consonants (Lisker and Abramson, 1964). The duration of a silent interval (i.e., silence duration) is a cue to distinguish fricatives (e.g., /ʃ/) and affricates (e.g., /tʃ/) (Dorman *et al.*, 1979). While any temporal change in an acoustic signal has a concomitant spectral change (i.e., all acoustic differences in speech stimuli have both temporal and spectral changes), the relative importance of the temporal cue is highlighted by age, not hearing loss, in studies examining temporal processing deficits exhibited by older compared to younger normal-hearing adult participants (e.g., Gordon-Salant *et al.*, 2006).

Cochlear implants (CIs) severely degrade the spectral information of a signal (Goupell *et al.*, 2008; Azadpour and McKay, 2012) and largely preserve its temporal envelope information (Loizou, 2006). Many CI users can robustly recognize consonants, vowels, words, and sentences, particularly in quiet conditions (Shannon *et al.*, 1995; Friesen *et al.*, 2001). CI users appear to rely heavily on temporal cues for speech understanding (Winn *et al.*, 2012), and their ability to process temporal information contributes to speech understanding performance (Fu, 2002). For CI users to perform some speech categorization tasks, they likely have to rely even more heavily on the temporal cues than normal-hearing participants (Winn *et al.*, 2012).

Auditory temporal processing in CI users can be measured with speech categorization tasks based on manipulating a single temporal cue (e.g., Winn *et al.*, 2016; Xie *et al.*, 2019). For example, Winn *et al.* (2016) created tokens with variable VOT for the word-initial phoneme and instructed CI users to categorize the tokens into four words. The categorization responses (i.e., perceptual changes as a function of VOT) correlated with speech understanding abilities (Winn *et al.*, 2016). Based on stimuli from Gordon-Salant *et al.* (2006), Xie *et al.* (2019) used a continuum of tokens with varying silence duration before a word-final fricative/ʃ/ and instructed CI users to categorize the tokens into two words across a range of presentation levels. They found that older adult CI users exhibited reduced sensitivity to the silence duration cue than younger CI users, but only at higher presentation levels [nominally 75 and 85 dB sound pressure level (SPL)]. This indicates a level-dependent age-related temporal processing deficit in adult CI users.

Using similar categorization tasks, Gordon-Salant *et al.* (2008) evaluated temporal processing abilities across a

[a)]Electronic mail: zxie2@kumc.edu

series of temporal cues in younger to older participants with acoustic hearing. They demonstrated that placing non-informative carrier sentences (e.g., "I had not thought about the…") before words that contrast primarily in temporal cues reduce the saliency of those cues for word identification; in other words, participants may require longer cues to differentiate word contrasts in conditions with carrier sentences compared to conditions with isolated words. The affected temporal cues include VOT (e.g., Buy-Pie contrast) and silence duration cues (e.g., Dish-Ditch contrast); although not explicitly tested in that study, the stimulus context effects appear to be more prevalent for temporal cues at the word-initial position. Further, the context effects were generally exaggerated in older vs younger participants (Gordon-Salant *et al.*, 2008). To date, however, the mechanisms underlying such stimulus context effects on auditory temporal processing remain unknown. Whether such stimulus context effects extend to CI users also remains unexplored; however, since CI users rely primarily on temporal cues for speech understanding (Winn *et al.*, 2012), stimulus context effects might occur or be even larger for CI users compared to acoustic-hearing participants.

This study aimed to determine the magnitude of stimulus context effects on auditory temporal processing in CI users. A second goal of this study was to determine whether the context effects are driven by speech-specific processes or general auditory processes. To achieve that, we manipulated the type of carrier stimuli by including non-informative sentences, as well as speech-shaped noise that was modulated by the envelope of the carrier sentences and unmodulated speech-shaped noise. If speech-specific processes underlie the context effects, we hypothesize larger context effects for the speech carriers (i.e., sentences) compared to the two non-speech carriers. Alternatively, if the context effects are not specific to speech processing but result from general auditory processes (e.g., forward masking; Shannon, 1990) or task complexity, we hypothesize comparable context effects between speech and non-speech carriers. The final goal was to examine the time course of context effects by systematically delaying the target words relative to the carrier stimuli. We hypothesize reduced context effects at longer delays.

II. METHOD

A. Participants

Thirteen adult CI users (29.2 to 82.0 years, mean age = 56.0 years; eight females and five males; four left ears and nine right ears) participated in this study. All users had at least one year of CI experience (8.0 to 29.6 years) and were native speakers of American English. The duration of deafness ranged from 0 to 20.0 years. All users were implanted with Cochlear-brand Nucleus electrode arrays (Cochlear Ltd.). Unilateral CI users completed testing in the implanted ear, and bilateral CI users completed testing in the self-reported better ear. The demographics are provided in Table I. All materials and procedures were approved by the Institutional Review Board at the University of Maryland. All participants provided written informed consent and received monetary compensation for their participation.

B. Stimuli

Stimuli consisted of two continua of word contrasts that varied primarily in a single temporal cue: Buy-Pie [endpoints displayed in Fig. 1(A)] and Dish-Ditch [endpoints displayed in Fig. 1(B)]. The Buy-Pie contrast varied in the duration between the release of the articulators and the onset of voicing (i.e., VOT). The Dish-Ditch contrast varied in the duration of a silent interval preceding the final fricative /ʃ/ (i.e., silence duration). Both temporal cues were systematically manipulated from 0 (Buy or Dish) to 60 ms (Pie or Ditch) in 10-ms steps, producing a seven-step continuum. Procedures for stimulus creation have been reported in previous studies (Gordon-Salant *et al.*, 2006; Gordon-Salant *et al.*, 2008; Xie *et al.*, 2019).

An adult American male speaker produced the words "Buy," "Pie," "Dish," and "Ditch" in isolation. To generate the Buy-Pie continuum, the original word "Buy" was used as the endpoint stimulus of 0-ms VOT. A 10-ms aspiration interval was inserted between the burst release and the onset of voicing of the original word "Buy" to create the stimulus of 10-ms VOT. This aspiration portion was excised from the

TABLE I. Participant demographics.

Participant	Sex	Test ear	Age at testing (years)	Duration of deafness (years)	Duration of CI use (years)	Etiology
S1	F	Right	73.6	7	14.6	Possibly genetic
S2	M	Right	73.8	<1	15.7	Unknown
S3	M	Right	29.2	1	11.2	Genetic
S4	M	Left	57.2	2	10.8	Unknown
S5	F	Left	82.0	<1	8	Hereditary, measles
S6	F	Left	40.8	3	20.8	Unknown
S7	F	Right	66.7	1	8.7	Premature birth
S8	M	Right	80.1	1	9	Measles, antibiotics, aging
S9	F	Right	58.5	13	16.8	Meniere's disease
S10	F	Right	71.6	<1	9.1	Ototoxicity/trauma
S11	F	Right	29.5	20	9.5	Hereditary
S12	F	Right	32.2	1.3	29.6	Bacterial meningitis
S13	M	Left	33.1	1.1	9.2	Premature birth, antibiotics

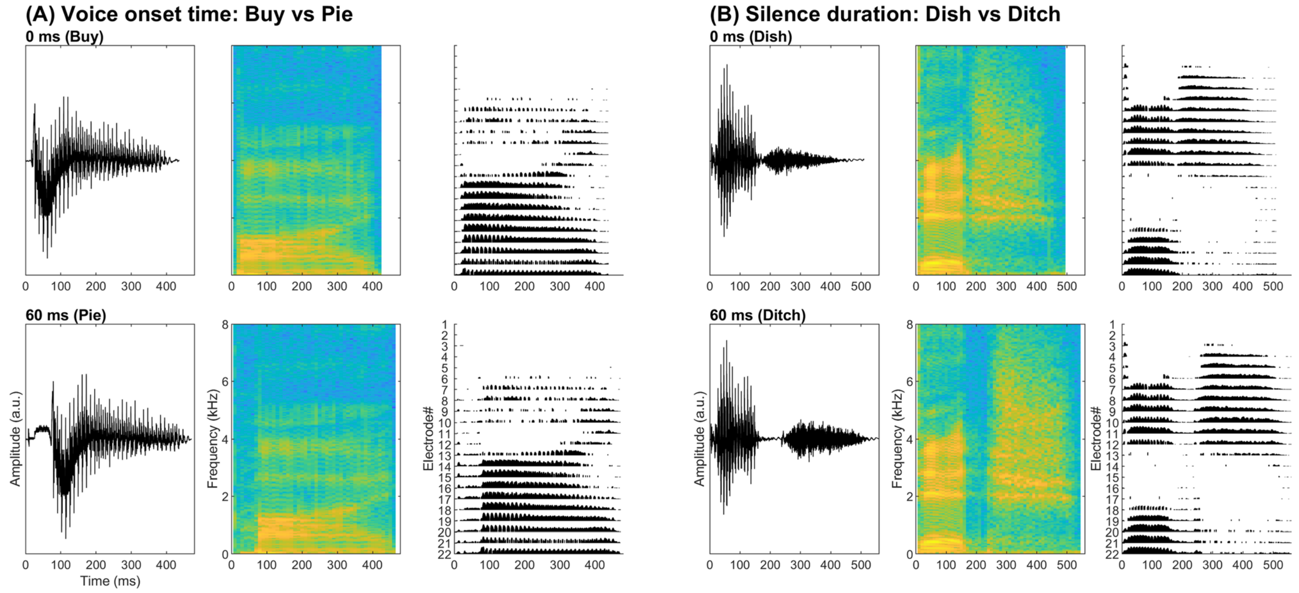

FIG. 1. (Color online) Waveforms, spectrograms, and electrodograms for the endpoint stimuli from the Buy-Pie continuum (A) and the Dish-Ditch continuum (B). The Buy-Pie continuum varies in the duration of the onset of voicing (i.e., voice-onset time). The Dish-Ditch continuum varies in the duration of a silent interval preceding the final fricative (i.e., silence duration). The electrodograms were created using an example advanced combination encoder (ACE) coding strategy from the Nucleus Matlab Toolbox.

original word "Pie." An additional 10-ms aspiration interval was inserted immediately after the burst of the 10-ms VOT stimulus to create the 20-ms VOT stimulus. This process repeated until the creation of the endpoint stimulus of 60-ms VOT ("Pie").

To generate the Dish-Ditch continuum, a hybrid word "Ditch" was created by combining the portions of initial stop, vowel, and closure duration from the original word "Ditch" with the fricative portion of the original word "Dish." This hybrid word was used as the endpoint stimulus of 60-ms silence duration ("Ditch"). A 10-ms silent interval was excised from the closure period of the hybrid word "Ditch" to create the stimulus of 50-ms silence duration. This process repeated until the creation of the end point stimulus of 0-ms silence duration (i.e., "Dish," 0-ms closure duration).

Note that the word contrasts selected for this study (e.g., Buy-Pie) may contain spectral cues (e.g., onset formant) besides the manipulated temporal cues (e.g., VOT) (Winn, 2020). However, those confounding spectral cues are likely to play a minimal role in shaping the perception of the selected word contrasts because, as outlined in the stimulus creation procedures, the same onset formant (from the original word "Buy") was used to create the seven tokens of the Buy-Pie continuum, and the only acoustic segment that varied between them was the amount of aspiration taken from the original "Pie" stimulus. As shown in Fig. 1(A), there was essentially no difference in the formant structures between the endpoints "Buy" and "Pie." Further, many CI users may have limited sensitivity to those spectral cues even if they are present (Goupell *et al.*, 2008; Azadpour and McKay, 2012; Winn and Litovsky, 2015; Winn *et al.*, 2016). Finally, we generated electrodograms (Fig. 1) for endpoints of the two word contrasts, using an example advanced combination encoder (ACE) coding strategy from the Nucleus Matlab Toolbox (Swanson and Mauch, 2006). These electrodograms suggest that the two word contrasts differ primarily on the duration of temporal cues.

C. Design

Stimuli from the two continua (Buy-Pie and Dish-Ditch) were presented in isolation (ISO-WD) or were preceded by a carrier stimulus. There were three types of carrier stimuli: a carrier sentence (SENT), a stationary speech-shaped noise modulated by the envelope of the carrier sentence (MOD-N), and an unmodulated stationary speech-shaped noise (SS-N). The carrier sentences were from Gordon-Salant *et al.* (2008) and consisted of 70 low-predictability sentences (e.g., I had not thought about the…) that do not convey semantic information to cue any of the target words (Buy, Pie, Dish, or Ditch). These carrier sentences were recorded from the same male talker who produced the target words. In Fig. 2(A), examples of the four types of stimulus context (ISO-WD, SENT, MOD-N, and SS-N) are displayed. As shown in Fig. 2(B), the three types of carrier stimuli (SENT, MOD-N, and SS-N) were matched in their long-term spectra. As shown in Fig. 2(C), the target words were also systematically delayed relative to the carrier stimuli (i.e., carrier-target delay, CTD) across 0, 75, 150, and 300 ms to examine the time course of context effects.

Stimuli from the two continua were presented in separate blocks. In each block, the four types of context (ISO-WD, and SENT, MOD-N, and SS-N at the four CTDs; 13 conditions × 7 steps = 91 trials) were mixed and

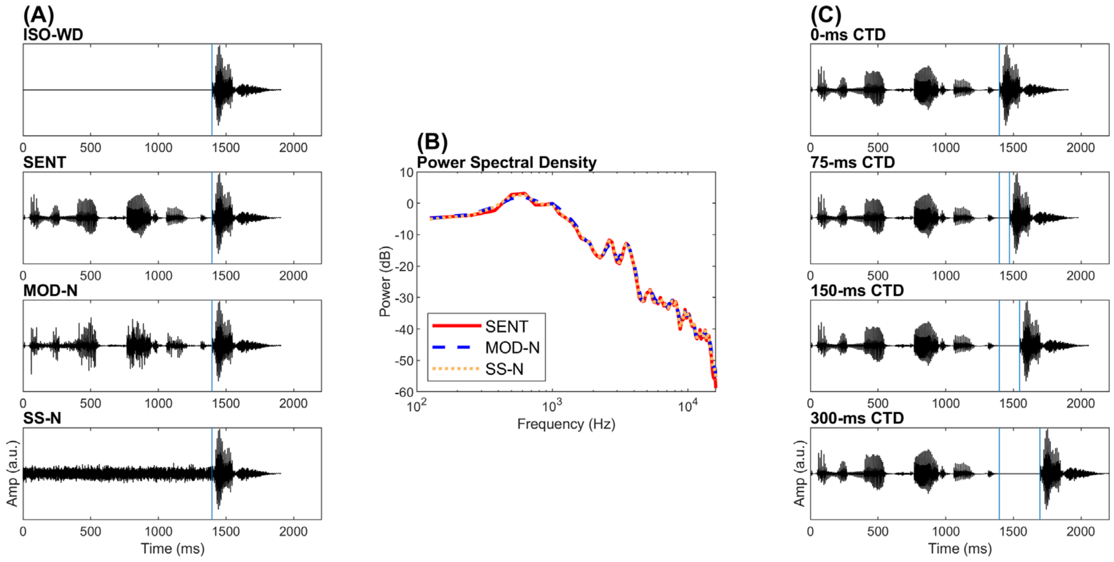

FIG. 2. (Color online) (A) Examples of the four types of context: Target words presented in isolation (ISO-WD), and preceded by a carrier sentence (SENT, e.g., "I had not thought about the…"), a speech-shaped noise modulated by the envelope of the carrier sentence (MOD-N), and an unmodulated speech-shaped noise (SS-N). The blue vertical lines indicate the offset of the carrier stimulus and the onset of the target word. (B) Examples of the spectra for the three types of carrier signals. (C) Waveforms of an example carrier sentence (SENT) followed by a target word at four carrier-target delays (CTDs): 0, 75, 150, and 300 ms. The blue vertical lines display the offset of the carrier stimulus (left line) and the onset of the target word (right line). The intervals between these two vertical lines correspond to the CTDs.

presented in random order. There were ten blocks for each continuum, resulting in ten repetitions for each stimulus. One participant had nine repetitions for the Buy-Pie contrast due to a computer error. We alternated the blocks between the two continua. Half of the participants received the Buy-Pie contrast first, and the other half received the Dish-Ditch contrast first. The stimuli were set at a level that participants reported being most comfortable.

D. Procedure

Stimuli were presented monaurally to a custom-fit research sound processor (CP910) *via* direct audio input, bypassing features like microphone directionality. We chose direct audio input for stimulus presentation over free-field loudspeakers to minimize potential confounds due to factors such as head motion in free-field listening. The processor was set to each individual's everyday clinical processor settings except that several front-end preprocessing features were de-activated, including adaptive dynamic range optimization, automatic sensitivity control, signal-to-noise-ratio noise reduction, wind noise reduction, and SCAN (automatic scene classifier system) if they were activated. This approach for stimulus presentation attempted to reduce variability introduced by clinical sound processors from individual participants.

Participants were tested individually in a sound-attenuating booth (Industrial Acoustics, Inc., Bronx, NY). The task was to identify each stimulus (i.e., the word in the ISO-WD condition or the final word in SENT, MOD-N, and SS-N conditions) as being "Buy" or "Pie" for words on the Buy-Pie continuum and as being "Dish" or "Ditch" for words on the Dish-Ditch continuum. The two continua were presented in separate blocks. Participants initiated each trial by clicking a box reading "Begin Trial" on the screen. Participants responded by clicking a box on the left or right of the screen corresponding to "Buy" or "Pie" for the Buy-Pie contrast and to "Dish" or "Ditch" for the Dish-Ditch contrast. No time limits were set for making responses. No feedback was provided.

Before testing, participants received training on the single words of the two continua separately. The training tasks were identical to the main experiment except that participants only heard the endpoint stimuli from each continuum and were provided with feedback about the correct answer at each trial. In separate blocks, the endpoint stimuli of the two continua were repeated ten times and presented in a randomized order. Participants were allowed up to 15 min to repeat the training tasks to establish stable performance. They could immediately proceed to main experiments if they achieved at least 85% accuracy for each endpoint stimulus used in the training task.

Custom scripts in MATLAB (MathWorks, Natick, MA) were created for controlling stimulus presentation and response collection. Participants completed the testing (including training and breaks) in a single session within 3 h. They were encouraged to take as many breaks as needed to minimize fatigue.

E. Statistical analysis

Separate mixed-effects logistic regression models implemented *via* lme4 (Bates *et al.*, 2014) in R version 4.0.2 (R Core Team, 2013) were used to analyze the trial-level responses of each continuum. We excluded data from three CI users for the Buy-Pie contrast and one CI user for the

Dish-Ditch contrast because they scored less than 85% accuracy on the corresponding training tasks. Hence, the final sample was 10 and 12 participants for the Buy-Pie and Dish-Ditch contrasts, respectively.

For the Buy-Pie contrast, we first examined the stimulus context effect on word categorization. The dependent variable was the response to each target word, which was coded as 1 (for "Buy" response) or 0 (for "Pie" response). Fixed factors were VOT (0 to 60 ms), context type (ISO-WD, and SENT, MOD-N, and SS-N at a 0-ms CTD), and their interaction. The VOT was centered and treated as a continuous variable. The context type (reference level = ISO-WD) was treated as a categorical variable. The random-effect structure included by-participant random slope for VOT and by-participant random intercept for context type.

Then, we examined the effects of CTD and type of carrier stimulus on word categorization. We did not combine this and the above analysis into one model that included context type (ISO-WD, SENT, MOD-N, and SS-N), CTD (0, 75, 150, and 300 ms), and VOT, because the ISO-WD context type did not include CTDs. The dependent variable was the response to each target word (i.e., "Buy" or "Pie"), which was coded as 1 (for "Buy" response) or 0 (for "Pie" response). The fixed factors were VOT (0 to 60 ms), type of carrier stimulus (SENT, MOD-N, and SS-N), CTD (0, 75, 150, and 300 ms), and their interactions. The VOT was centered and treated as a continuous variable. The carrier stimulus type (reference level = SENT) was treated as a categorical variable. The CTD was recoded [0 ms = 1 (reference level), 75 ms = 2, 150 ms = 3, 300 ms = 4] and was treated as a categorical variable. The random-effect structure included by-participant random slope for VOT, and by-participant random intercepts for carrier stimulus type and CTD.

We applied identical analyses to the Dish-Ditch contrast except that the dependent variable was "Dish" (coded as 1) or "Ditch" (coded as 0) response and the silence duration cue replaced VOT as the temporal cue of interest.

III. RESULTS

A. Buy-Pie contrast

Figure 3(A) displays the percentage of "Buy" responses as a function of VOT across the ISO-WD conditions and conditions with carrier stimuli at a 0-ms CTD. Descriptively, while participants could discriminate the words "Buy" and "Pie," they tended to perceive the words as "Buy" when the words were preceded by carrier stimuli, particularly in the SS-N condition. Figure 3(B) displays the percentage of "Buy" responses as a function of VOT for the three types of carrier stimulus at different CTDs. Descriptively, the introduction of CTDs ≥ 75 ms reduced the influence of preceding carrier stimuli on the perception of the VOT cue.

First, we examined the effect of stimulus context on the perception of temporal cues [Fig. 3(A)]. Table II shows the results (CTD = 0 ms) from the mixed-effects logistic regression model. The effect of VOT was significant ($p < 0.001$), with fewer "Buy" responses for longer VOTs in the ISO-WD condition. At the mean VOT, each type of carrier stimulus was associated with more 'Buy' responses than the ISO-WD condition (all ps < 0.05). We further examined the differences among the three types of carrier stimulus by releveling the statistical model to utilize each carrier stimulus as the baseline (reference). Results showed that at the mean VOT, the SS-N condition was associated with more "Buy" responses compared with the SENT condition [β = 1.272, standard error (SE) = 0.254, $z = 5.000$, $p < 0.001$] and the MOD-N condition ($\beta = 0.995$, SE = 0.214, $z = 4.649$, $p < 0.001$). The amount of "Buy" responses was not significantly different between the SENT and MOD-N conditions ($\beta = 0.277$, SE = 0.311, $z = 0.893$, $p = 0.372$).

There were significant two-way interactions between VOT and context type in all conditions with carrier stimuli (SENT, MOD-N, and SS-N; all ps < 0.001), demonstrating that the VOT effect was different at these conditions compared with the ISO-WD condition. These interactions suggest that while participants tended to report fewer "Buy" responses for longer VOTs, such VOT effect was reduced in conditions with carrier stimuli [Fig. 3(A)]. We further examined the differences among the three types of carrier stimulus by releveling the statistical model to utilize each carrier stimulus as the baseline (reference). Results showed that, as shown in Fig. 3(A), the VOT effect (i.e., fewer "Buy" responses for longer VOTs) was reduced in the SS-N condition compared with the SENT condition (β = 0.053, SE = 0.008, $z = 6.299$, $p < 0.001$) and the MOD-N condition (β = 0.040, SE = 0.008, $z = 4.952$, $p < 0.001$). The VOT effect was not significantly different between the SENT and MOD-N conditions (β = –0.013, SE = 0.007, z = –1.726, $p = 0.084$).

Then, we examined the effects of CTD and type of carrier stimulus on the perception of temporal cues [Fig. 3(B)]. Table III shows the results from the mixed-effects logistic regression model. The effect of VOT was significant ($p = 0.008$), with fewer "Buy" responses for longer VOTs at a 0-ms CTD in the SENT condition (reference). At the mean VOT, the SS-N condition was associated with more "Buy" responses than the SENT condition ($p < 0.001$). The effect of CTD was significant (all ps < 0.001), suggesting that the introduction of CTD (75 ms and above) was associated with fewer "Buy" responses in the SENT condition. We further examined the differences among CTDs by releveling the statistical model to utilize each CTD as the baseline (reference). The probability of "Buy" responses was not significantly different among 75-, 150-, and 300-ms CTD conditions (all ps > 0.05).

There were significant two-way interactions between VOT and carrier stimulus type in the MOD-N condition ($p = 0.02$) and the SS-N condition ($p < 0.001$). This suggests that at a 0-ms CTD, the VOT effect (i.e., fewer "Buy" responses for longer VOTs) was reduced in the MOD-N and SS-N conditions compared with the SENT condition. We further examined the difference between MOD-N and SS-N conditions by releveling the statistical model to utilize the

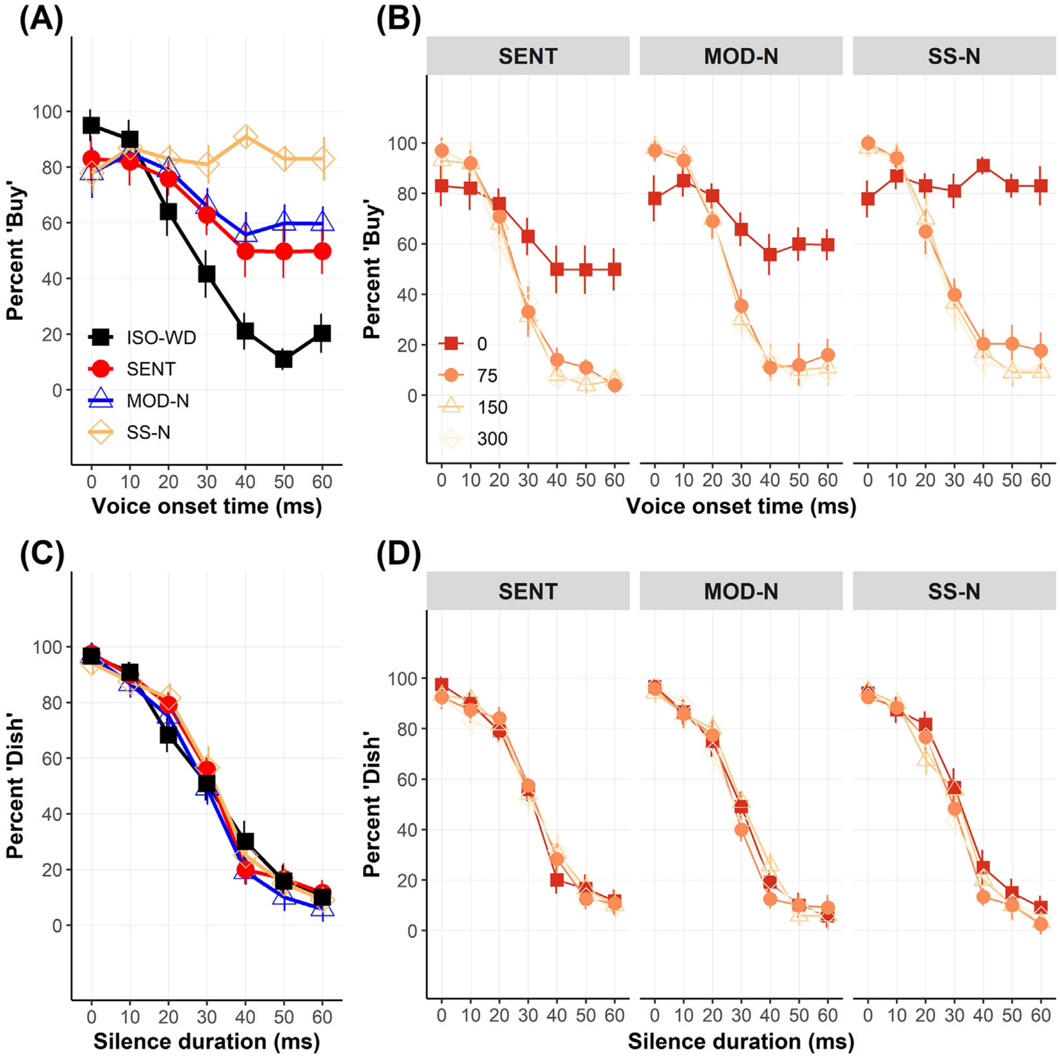

FIG. 3. (Color online) Averaged percentage of trials identified as "Buy" as a function of VOT in the Buy-Pie continuum (A and B) and as "Dish" as a function of silence duration in the Dish-Ditch continuum (panels C and D). For (A) and (C), responses were obtained at a 0-ms CTD across four contexts: Target words presented in isolation (ISO-WD; solid squares) and preceded by a carrier sentence (SENT; solid circles), a stationary speech-shaped noise modulated by the envelope of the carrier sentence (MOD-N; open triangles), and an unmodulated stationary speech-shaped noise (SS-N; open diamonds). For (B) and (D), responses were obtained across CTDs of 0, 75, 150, and 300 ms at conditions with SENT, MOD-N, and SS-N. Error bars show ±1 SE.

MOD-N condition as the baseline (reference). Results showed that the VOT effect was reduced in the SS-N condition compared with the MOD-N condition (β = 0.045, SE = 0.008, $z = 5.519$, $p < 0.001$).

There were significant two-way interactions between VOT and CTD (all $ps < 0.001$). This suggests that in the SENT condition, the VOT effect (i.e., fewer "Buy" responses for longer VOTs) was increased with the introduction of a CTD of 75 ms and above. We further examined the differences among CTDs by releveling the statistical model to utilize each CTD as the baseline (reference). Results showed that the VOT effect was not significantly different among 75-, 150-, and 300-ms CTD conditions (all $ps > 0.05$).

There were significant two-way interactions between carrier stimulus type and CTD in the SS-N condition (all $ps < 0.001$). This suggests that at the mean VOT, the probability of "Buy" responses was reduced in the SS-N condition with a CTD of 75 ms or above compared with the reference condition.

There were significant three-way interactions among VOT, carrier stimulus type, and CTD in the SS-N condition (all $ps < 0.05$). This suggests that the increase in the VOT effect from the introduction of CTDs (75 ms and above) was larger in the SS-N condition compared with the SENT condition. We further examined the difference between MOD-N and SS-N conditions by releveling the statistical model to utilize the MOD-N condition as the baseline (reference). Results showed that the increase in VOT effect from the introduction of CTDs (75 ms and above) was larger in the SS-N condition compared with the MOD-N condition (βs ranging from –0.054 to –0.032; all $ps < 0.05$).

B. Dish-Ditch contrast

Figure 3(C) displays the percentage of "Dish" responses as a function of silence duration across the ISO-WD

TABLE II. Results for the mixed-effects logistic regression model (CTD = 0 ms): "Buy" (vs "Pie") response = VOT × Context type + (1 + VOT + Context type | participant). SD, standard deviation.

Fixed effects	*β*	*SE*	z	*p*
(Intercept)	−0.214	0.31	−0.689	0.491
VOT	−0.104	0.014	−7.231	**<0.001**
Context type				
SENT – ISO-WD	1.129	0.513	2.202	**0.028**
MOD-N – ISO-WD	1.406	0.525	2.677	**0.007**
SS-N – ISO-WD	2.401	0.48	5.006	**<0.001**
VOT × Context type				
VOT × (SENT – ISO-WD)	0.054	0.009	6.141	**<0.001**
VOT × (MOD-N – ISO-WD)	0.067	0.009	7.819	**<0.001**
VOT × (SS-N – ISO-WD)	0.107	0.100	11.240	**<0.001**
Random effects	**Variance**	***SD***		
(Intercept)	0.840	0.917		
VOT	0.623	0.789		
Context type				
SENT – ISO-WD	2.376	1.541		
MOD-N – ISO-WD	2.503	1.582		
SS-N – ISO-WD	1.952	1.397		

conditions and conditions with carrier stimuli at a 0-ms CTD. Descriptively, participants were able to discriminate the words "Dish" and "Ditch." The presentation of carrier stimuli before these words has negligible influences on the perception of silence duration cue. Figure 3(D) displays the percentage of "Dish" responses as a function of silence duration across the three types of carrier stimuli at different CTDs. Descriptively, the introduction of CTDs seems to have negligible influences on the perception of silence duration cues.

First, we examined the effect of stimulus context on the perception of temporal cues [Fig. 3(C)]. The effect of silence duration was significant ($p < 0.001$), with fewer "Dish" responses for longer silence durations. All other effects were not significant (all *p*s > 0.05).

Then, we examined the effects of CTD and type of carrier stimulus on the perception of temporal cues [Fig. 3(D)]. The effect of silence duration was significant ($p < 0.001$), with fewer "Dish" responses for longer silence duration at a 0-ms CTD in the SENT condition (reference). At the mean silence duration, the MOD-N condition was associated with fewer "Dish" responses compared with the SENT condition ($p = 0.034$). Further, there were significant two-way interactions between carrier stimulus type and CTD, such that at the mean silence duration, the probability of "Dish" responses was reduced in the SS-N condition with CTDs of 75 ms ($p = 0.007$) and 150 ms ($p = 0.023$), compared with the reference condition (SENT condition with a 0-ms CTD). We further examine the difference between MOD-N and SS-N conditions by releveling the statistical model to utilize the MOD-N condition as the baseline (reference). Results showed that at the mean silence duration, the probability of "Dish" responses was reduced in the SS-N condition with a CTD of 150 ms compared with the MOD-N condition with a 0-ms CTD ($\beta = –0.466$, SE = 0.217, $z = –2.142$, $p = 0.032$). Finally, no other effects of silence duration, carrier stimulus type, CTD, or their interactions were significant (all *p*s > 0.05).

IV. DISCUSSION

The goals of this study were to examine the extent to which prior carrier stimuli affect the sensitivity to temporal cues of word segments in CI users and to determine the contribution of speech-specific vs general auditory processes to such stimulus context effects. We hypothesized that CI users' sensitivity to temporal cues would be reduced when the target words were preceded by non-informative carrier stimuli. Our results demonstrated that, compared to the isolated word stimuli, placing a carrier sentence before target words was associated with reduced sensitivity to the VOT cue (i.e., Buy-Pie contrast) but not for the silence duration cue (i.e., Dish-Ditch contrast) (Fig. 3). The reduced sensitivity to the VOT cue was reflected as shallower psychometric functions (i.e., longer crossover points and shallower slopes) for the percentage of "Buy" responses as a function of VOT [Fig. 3(A)]. The present findings are consistent with Gordon-Salant *et al.* (2008) with acoustic-hearing participants for the Buy-Pie contrast in that those participants also showed reduced sensitivity to temporal cues in conditions with carrier sentences compared to isolated words. Thus, our data suggest that stimulus context effects observed in acoustic-hearing participants extend to CI users.

Our study partially expands our understanding of auditory temporal processing from previous studies (Gordon-Salant *et al.*, 2008), because our findings shed light on the potential mechanisms underlying the stimulus context effects. In contrast with the hypothesis of speech-specific processes, we demonstrated that non-speech carriers (including modulated and unmodulated noises) yielded similar or even greater context effects than the speech carriers [Fig. 3(A)]. These data favor the argument that the context effects may be governed by general auditory processes that are not specific to speech. One potential mechanism would be forward masking (Weber and Moore, 1981; Shannon, 1990). In the context of forward masking, the carrier stimuli presented before the VOT cue (Buy-Pie contrast) can be considered as forward maskers, which affected the perception of the VOT cue and yielded context effects. Specifically, the carrier stimuli may mask the VOT cue and render long VOTs to be perceived as being shorter; thus, participants were biased to report "Buy" responses for long VOTs. Consistent with the finding that forward masking declines as delays are introduced between the forward maskers and the targets (Weber and Moore, 1981; Shannon, 1990), we found that the introduction of a 75-ms carrier-target delay almost abolished the context effects [Fig. 3(B)]. Further, we compared sound levels (root mean square energy) of the final 50 ms across the three carrier signals (SENT, MOD-N, and SS-N). There is a higher level of energy (~7 dB) for the SS-N context than SENT ($p < 0.001$) and MOD-N ($p < 0.001$). This level pattern corresponds to the findings that there was a larger context effect for SS-N

TABLE III. Results for the mixed-effects logistic regression model: "Buy" (vs "Pie") response = VOT × Carrier stimulus type × CTD + (1 + VOT + Carrier stimulus type + CTD | participant).

Fixed effects	*β*	*SE*	z	*p*
(Intercept)	0.945	0.391	2.413	0.016
VOT	−0.051	0.019	−2.652	**0.008**
Carrier stimulus type				
MOD-N – SENT	0.376	0.257	1.464	0.143
SS-N – SENT	1.472	0.289	5.101	**<0.001**
CTD				
75 ms – 0 ms	−1.527	0.431	−3.542	**<0.001**
150 ms – 0 ms	−1.865	0.503	−3.706	**<0.001**
300 ms – 0 ms	−1.821	0.518	−3.518	**<0.001**
VOT × Carrier stimulus type				
VOT × (MOD-N – SENT)	0.017	0.007	2.335	**0.02**
VOT × (SS-N – SENT)	0.063	0.008	7.429	**<0.001**
VOT × CTD				
VOT × (75 ms – 0 ms)	−0.108	0.011	10.123	**<0.001**
VOT × (150 ms – 0 ms)	−0.110	0.011	−10.198	**<0.001**
VOT × (300 ms – 0 ms)	−0.114	0.011	−10.445	**<0.001**
Carrier stimulus type × CTD				
(MOD-N – SENT) × (75 ms – 0 ms)	−0.249	0.226	−1.104	0.270
(SS-N – SENT) × (75 ms – 0 ms)	−1.073	0.240	−4.466	**<0.001**
(MOD-N – SENT) × (150 ms – 0 ms)	−0.086	0.23	−0.375	0.708
(SS-N – SENT) × (150 ms – 0 ms)	−1.128	0.248	−4.554	**<0.001**
(MOD-N – SENT) × (300 ms – 0 ms)	−0.286	0.232	−1.231	0.218
(SS-N – SENT) × (300 ms – 0 ms)	−1.359	0.248	−5.476	**<0.001**
VOT × Carrier stimulus type × CTD				
VOT × (MOD-N – SENT) × (75 ms – 0 ms)	0.002	0.014	0.132	0.895
VOT × (SS-N – SENT) × (75 ms – 0 ms)	−0.030	0.015	−2.043	**0.041**
VOT × (MOD-N – SENT) × (150 ms – 0 ms)	−0.013	0.015	−0.904	0.366
VOT × (SS-N – SENT) × (150 ms – 0 ms)	−0.068	0.016	−4.349	**<0.001**
VOT × (MOD-N – SENT) × (300 ms – 0 ms)	−0.023	0.015	−1.552	0.121
VOT × (SS-N – SENT) × (300 ms – 0 ms)	−0.066	0.015	−4.305	**<0.001**
Random effects	**Variance**	***SD***		
(Intercept)	1.419	1.191		
VOT	0.003	0.058		
Carrier stimulus type				
MOD-N – SENT	0.448	0.669		
SS-N – SENT	0.543	0.737		
CTD				
75 ms – 0 ms	1.582	1.258		
150 ms – 0 ms	2.239	1.496		
300 ms – 0 ms	2.384	1.544		

than the other two context types, providing partial support of forward masking underlying stimulus context effects. Finally, parallel to our findings, previous work suggests that the magnitude of forward masking might be smaller for speech maskers than spectrally matched steady-state (unmodulated) non-speech maskers (Grose *et al.*, 2016). Such findings are consistent with the evidence that there may be less neural adaptation to modulated sounds compared to unmodulated ones (Joris *et al.*, 2004).

A second potential mechanism is that the stimulus context added task complexity. For example, relative to the isolated word conditions, there may be increased cognitive demand to focus on the target words and ignore the task-irrelevant carrier stimuli. The amplitude modulations introduced by the carrier stimuli may interfere with the perception of the upcoming target words, as there is evidence that the amplitude modulation of one sound may mask the amplitude modulation of another sound (Dau *et al.*, 1997; Wojtczak and Viemeister, 2005). However, there are at least two potential issues with the argument of task complexity. First, it is unclear why task complexity, if it was the dominant factor, only affected the word-initial VOT cue. Second, the speech carrier stimuli contain task-irrelevant amplitude modulations and linguistic information. Based on the argument of task complexity, we might hypothesize larger context effects for the speech carrier stimuli compared to the non-speech stimuli, which was not supported by our data. Therefore, it may be that forward masking is a stronger candidate underlying the observed stimulus context effects.

However, the account of forward masking needs to accommodate at least these findings. First, for the Dish-Ditch contrast, the syllable /dɪ/ occurs before the target silence duration cue. In the context of forward masking, the syllable /dɪ/ may serve as a masker and interfere with the perception of this temporal cue. But our participants were still able to discriminate the Dish-Ditch contrast. While this finding initially appears to be at odds with the forward-masking argument, our earlier work demonstrated that increasing stimulus presentation level was associated with reduced ability to discriminate the Dish-Ditch contrast in isolation in CI users (Xie *et al.*, 2019). These earlier findings appear to be consistent with the presence of forward masking in the perception of the Dish-Ditch contrast in isolation.

Second, we did not observe stimulus context effects for the Dish-Ditch contrast from any types of carrier stimuli. The differences in the context effects between Buy-Pie and Dish-Ditch contrasts in the current study appear to be consistent with the acoustic-hearing findings from Gordon-Salant *et al.* (2008) where the context effects may be more prevalent for temporal cues at the word-initial position, though statistical analysis was not performed to compare word contrasts. From the point of forward masking, the lack of context effects for the Dish-Ditch contrast might have occurred because the forward-masking effects of the carrier stimuli were not able to spread into the silence interval of the Dish-Ditch contrast. Our data suggest that the context effects, if governed by forward masking, may be limited to a delay of ∼75 ms. This time range is shorter than the duration of the syllable /dɪ/ (∼150 ms) before the silence interval. Nevertheless, it is premature to definitively conclude that forward masking is the mechanism. Hence, future work is needed to investigate why the context effects on auditory temporal processing are cue-specific and what precise mechanisms are underlying the context effects.

There are some differences between our CI data and the acoustic-hearing data from Gordon-Salant *et al.* (2008). First, unlike our study, they reported a small but significant context effect for the Dish-Ditch contrast in acoustic-hearing participants using a sentence carrier. This discrepancy may partly be due to stimulus presentation levels. CI users in this study were tested at a most comfortable level, whereas participants in Gordon-Salant *et al.* (2008) were tested at a high level (85 dB SPL). Xie *et al.* (2019) showed that word categorization based on temporal cues depends on stimulus presentation levels. It should also be mentioned that the current dataset in our study has relatively limited power with 13 participants compared to 60 participants in Gordon-Salant *et al.* (2008). Hence, future work may directly compare context effects on auditory temporal processing between acoustic and electric hearing across various temporal cues and examine the role of stimulus level. This line of work is necessary because the underlying mechanisms for the stimulus context effects may differ across groups. For instance, compared to acoustic-hearing participants, spectral degradation of input signals through CI processers may force CI users to rely on temporal cues (Winn *et al.*, 2012), but meanwhile may reduce CI users' ability to utilize those cues (Goupell *et al.*, 2017).

Advancing age is associated with declines in auditory temporal processing of acoustic-hearing participants (Gordon-Salant *et al.*, 2008; Goupell *et al.*, 2017) and CI users (Xie *et al.*, 2019). Gordon-Salant *et al.* (2008) suggested that such age-related declines in temporal processing may be exaggerated in the presence of surrounding sentences (i.e., stimulus context) in acoustic-hearing participants. It is reasonable to infer that the presence of stimulus context may also amplify age-related temporal processing deficits among older CI users, because there is evidence to suggest that they are more susceptible to the negative impact from forward masking compared to younger CI users (Lee *et al.*, 2012; Jahn *et al.*, 2021). To test this hypothesis, future work can examine and compare the effects of aging and stimulus context on auditory temporal processing in acoustic-hearing and CI participants. This line of future work could also shed light on the mechanisms underlying the aging effects on temporal processing (Xie *et al.*, 2021).

The stimulus context effects observed in the current study might present some challenges for CI users to perceive running speech in real-life environments, wherein the perception of some temporally based words might be affected by preceding speech signals. However, this does not necessarily lead to speech understanding issues in all cases. First, other cues in real speech may compensate and facilitate individual word recognition. Second, other types of context in speech (e.g., semantic context) may override the stimulus context effects and aid in speech understanding. Future studies may examine the stimulus context effects on cue weighting and their interaction with other context types.

There are several limitations with this study that require consideration for future research. First, although it is reasonable to believe that CI users primarily used temporal cues to distinguish the word contrasts selected for this study, we could not fully exclude the possibility that some CI users used spectral cues for the categorization tasks (Winn, 2020). However, we believe the contribution from spectral profile cues should be minimal considering our stimulus creation procedures and the limited sensitivity to spectral cues in many CI users (Goupell *et al.*, 2008; Azadpour and McKay, 2012; Winn and Litovsky, 2015; Winn *et al.*, 2016). Nevertheless, the inclusion of a basic spectral resolution measure (Archer-Boyd *et al.*, 2018) would be warranted for future studies using temporally based word categorization tasks. Such studies should examine the relationship between spectral resolution and categorization performance. Second, a handful of CI users (e.g., 3 out of 13 for the Buy-Pie contrast) were excluded from analysis because they had difficulty distinguishing even the endpoint stimuli (0- and 60-ms). Relatedly, participant-related factors such as age (Xie *et al.*, 2019) and duration of deafness (Xie *et al.*, 2021) may affect temporal processing in CI users; however, they were not accounted for in the data analysis due to our relatively small sample size. Future studies should include a larger sample size to replicate our findings and investigate the extent to which those participant-related factors modulate stimulus context effects on auditory temporal processing in CI users.

In summary, this study demonstrated that stimulus context effects on auditory temporal processing occur in a group of CI users but they appear to be cue-specific. Mechanistically, the context effects appear to be governed by general auditory processes, not those specific to speech.

ACKNOWLEDGMENTS

Thank you to Alanna Schloss for her assistance with data collection. Thank you to Dr. Sandra Gordon-Salant for sharing the stimuli and providing insightful comments on the manuscript. This research was funded by the National Institute on Aging of the National Institutes of Health under Award Number R01AG051603 (M.J.G.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Archer-Boyd, A. W., Southwell, R. V., Deeks, J. M., Turner, R. E., and Carlyon, R. P. (**2018**). "Development and validation of a spectro-temporal processing test for cochlear-implant listeners," J. Acoust. Soc. Am. **144**, 2983–2997.

Azadpour, M., and McKay, C. M. (**2012**). "A psychophysical method for measuring spatial resolution in cochlear implants," J. Assoc. Res. Otolaryngol. **13**, 145–157.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2014**). "lme4: Linear mixed-effects models using Eigen and S4," R package version 1, https://github.com/lme4/lme4/ (Last viewed February 19, 2022).

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dorman, M. F., Raphael, L. J., and Liberman, A. M. (**1979**). "Some experiments on the sound of silence in phonetic perception," J. Acoust. Soc. Am. **65**, 1518–1532.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q. J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," Neuroreport **13**, 1635–1639.

Gordon-Salant, S., Yeni-Komshian, G., and Fitzgibbons, P. (**2008**). "The role of temporal cues in word identification by younger and older adults: Effects of sentence context," J. Acoust. Soc. Am. **124**, 3249–3260.

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Barrett, J. (**2006**). "Age-related differences in identification and discrimination of temporal cues in speech segments," J. Acoust. Soc. Am. **119**, 2455–2466.

Goupell, M. J., Gaskins, C. R., Shader, M. J., Walter, E. P., Anderson, S., and Gordon-Salant, S. (**2017**). "Age-related differences in the processing of temporal envelope and spectral cues in a speech segment," Ear Hear. **38**, e335–e342.

Goupell, M. J., Laback, B., Majdak, P., and Baumgartner, W.-D. (**2008**). "Current-level discrimination and spectral profile analysis in multi-channel electrical stimulation," J. Acoust. Soc. Am. **124**, 3142–3157.

Grose, J. H., Menezes, D. C., Porter, H. L., and Griz, S. (**2016**). "Masking period patterns and forward masking for speech-shaped noise: Age-related effects," Ear Hear. **37**, 48–54.

Jahn, K. N., DeVries, L., and Arenberg, J. G. (**2021**). "Recovery from forward masking in cochlear implant listeners: Effects of age and the electrode-neuron interface," J. Acoust. Soc. Am. **149**, 1633–1643.

Joris, P., Schreiner, C., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.

Lee, E. R., Friedland, D. R., and Runge, C. L. (**2012**). "Recovery from forward masking in elderly cochlear implant users," Otol Neurotol. **33**, 355–363.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Loizou, P. C. (**2006**). "Speech processing in vocoder-centric cochlear implants," Adv Otorhinolaryngol. **64**, 109–143.

R Core Team (**2013**). "R: A language and environment for statistical computing," https://www.r-project.org/ (Last viewed February 19, 2022).

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B: Biol. Sci. **336**(1278), 367–373.

Shannon, R. V. (**1990**). "Forward masking in patients with cochlear implants," J. Acoust. Soc. Am. **88**, 741–744.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Swanson, B., and Mauch, H. (**2006**). *Nucleus Matlab Toolbox 4.20 Software User Manual* (Cochlear Ltd., Lane Cove, NSW, Australia).

Weber, D. L., and Moore, B. C. (**1981**). "Forward masking by sinusoidal and noise maskers," J. Acoust. Soc. Am. **69**, 1402–1409.

Winn, M. B. (**2020**). "Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script," J. Acoust. Soc. Am. **147**, 852–866.

Winn, M. B., Chatterjee, M., and Idsardi, W. J. (**2012**). "The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing," J. Acoust. Soc. Am. **131**, 1465–1479.

Winn, M. B., and Litovsky, R. Y. (**2015**). "Using speech sounds to test functional spectral resolution in listeners with cochlear implants," J. Acoust. Soc. Am. **137**, 1430–1442.

Winn, M. B., Won, J. H., and Moon, I. J. (**2016**). "Assessment of spectral and temporal resolution in cochlear implant users using psychoacoustic discrimination and speech cue categorization," Ear Hear. **37**, e377–e390.

Wojtczak, M., and Viemeister, N. F. (**2005**). "Forward masking of amplitude modulation: Basic characteristics," J. Acoust. Soc. Am. **118**, 3198–3210.

Xie, Z., Gaskins, C. R., Shader, M. J., Gordon-Salant, S., Anderson, S., and Goupell, M. J. (**2019**). "Age-related temporal processing deficits in word segments in adult cochlear-implant users," Trends Hear. **23**, 2331216519886688.

Xie, Z., Stakhovskaya, O., Goupell, M. J., and Anderson, S. (**2021**). "Aging effects on cortical responses to tones and speech in adult cochlear-implant users," J. Assoc. Res. Otolaryngol. **22**, 719–740.