Age-Related Differences in the Processing of Temporal Envelope and Spectral Cues in a Speech Segment

Matthew J. Goupell, Casey R. Gaskins, Maureen J. Shader, Erin P. Walter, Samira Anderson, and Sandra Gordon-Salant

Objectives: As people age, they experience reduced temporal processing abilities. This results in poorer ability to understand speech, particularly for degraded input signals. Cochlear implants (CIs) convey speech information via the temporal envelopes of a spectrally degraded input signal. Because there is an increasing number of older CI users, there is a need to understand how temporal processing changes with age. Therefore, the goal of this study was to quantify age-related reduction in temporal processing abilities when attempting to discriminate words based on temporal envelope information from spectrally degraded signals.

Design: Younger normal-hearing (YNH) and older normal-hearing (ONH) participants were presented a continuum of speech tokens that varied in silence duration between phonemes (0 to 60 ms in 10-ms steps), and were asked to identify whether the stimulus was perceived more as the word "dish" or "ditch." Stimuli were vocoded using tonal carriers. The number of channels (1, 2, 4, 8, 16, and unprocessed) and temporal envelope low-pass filter cutoff frequency (50 and 400 Hz) were systematically varied.

Results: For the unprocessed conditions, the YNH participants perceived the word ditch for smaller silence durations than the ONH participants, indicating that aging affects temporal processing abilities. There was no difference in performance between the unprocessed and 16-channel, 400-Hz vocoded stimuli. Decreasing the number of spectral channels caused decreased ability to distinguish dish and ditch. Decreasing the envelope cutoff frequency also caused decreased ability to distinguish dish and ditch. The overall pattern of results revealed that reductions in spectral and temporal information had a relatively larger effect on the ONH participants compared with the YNH participants.

Conclusions: Aging reduces the ability to utilize brief temporal cues in speech segments. Reducing spectral information—as occurs in a channel vocoder and in CI speech processing strategies—forces participants to use temporal envelope information; however, older participants are less capable of utilizing this information. These results suggest that providing as much spectral and temporal speech information as possible would benefit older CI users relatively more than younger CI users. In addition, the present findings help set expectations of clinical outcomes for speech understanding performance by adult CI users as a function of age.

Key words: Aging, Cochlear implants, Temporal processing.

INTRODUCTION

There is an increasing number of adults over 65 years old who use cochlear implants (CIs) to partially restore their hearing and understanding of speech. This has occurred because of the continued success of middle-aged CI users who are now aging into the older age range, and because the criteria for implantation have expanded to include older seniors. While CIs in older individuals improve speech understanding and quality

of life without undue complications (Clark et al. 2012; Lin et al. 2012; Chen et al. 2013), evidence suggests that older CI users have poorer speech understanding than younger CI users (Jin et al. 2014; Sladen & Zappler 2014, 2015).

Reduced speech understanding with advancing age has been shown to result from declines in peripheral, central auditory, and cognitive processing (CHABA 1988). Of particular interest for CI users is the well-documented age-related decline in auditory temporal processing across a wide variety of tasks (Fitzgibbons & Gordon-Salant 1995; Gordon-Salant & Fitzgibbons 2001; Schvartz et al. 2008; Gordon-Salant 2010). For example, it has been shown that aging and hearing loss reduce the ability to identify and discriminate word contrasts that are based on temporal cues (Gordon-Salant et al. 2006, 2008). Older participants required longer temporal cue durations for a number of word contrasts such as dish/ditch (silence duration), buy/pie (voice onset time), wheat/weed (vowel duration), and beat/wheat (initial consonant transition duration), particularly if the contrasts were embedded in sentences (Gordon-Salant et al. 2008).

To convey speech, CIs present users with electrical pulse trains that are modulated by temporal envelopes using a vocoder-centric approach (Loizou 2006), and it appears that temporal envelope modulation processing impacts speech understanding in CI users (Fu 2002). Because the spectral resolution of a CI is poor (Friesen et al. 2001), CI users likely rely on temporal information more than normal-hearing (NH) participants (Winn et al. 2012, 2016). Because CI users weight temporal cues more heavily than NH listeners, older CI users with relatively poor temporal processing abilities may be at risk for poor speech perception.

To remove the variability inherent to CI users that results from biological, surgical, and device factors, vocoders have been used as a method to simulate CI sound processing (Shannon et al. 1995). Numerous studies using vocoders have shown that decreasing the number of channels, typically below about 8 to 10 channels, reduces speech understanding (Shannon et al. 1995; Loizou et al. 1999; Friesen et al. 2001; Goupell et al. 2008; Stilp & Goupell 2015). The modulations conveyed by the channel envelopes can also be altered by changing the low-pass filter cutoff frequency applied to each of the temporal envelopes. Decreasing envelope cutoff (ECO), typically below about 150 Hz, also reduces speech understanding (Xu & Pfingst 2008; Stilp & Goupell 2015). Some studies have investigated the interaction between the number of channels and ECO, and found a trade-off in spectral-temporal information (Xu et al. 2005; Xu & Zheng 2007; Stone et al. 2008; Xu & Pfingst 2008; Kim et al. 2015). Such a trade-off has also been observed in actual CI users (Nie et al. 2006).

Older NH individuals are worse at understanding vowels, consonants, and sentences degraded by a vocoder than younger NH

Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland, USA.

<zdoi; 10.1097/AUD.0000000000000447>

individuals (Schvartz et al. 2008; Sheldon et al. 2008a, b). Consistent with the idea that older adults have poorer temporal processing (CHABA 1988; Gordon-Salant 2010), they are also worse at perceiving differences in envelope modulation rate (Souza & Boike 2006; Schvartz-Leyzac & Chatterjee 2012, 2015), seemingly because of decreased temporal processing abilities.

The purpose of this study was to examine the relationship between age-related declines in temporal envelope processing and single word discrimination. To test this relationship, listeners were asked to discriminate the stimulus pair dish and ditch, used in a prior investigation (Gordon-Salant et al. 2006), with the intention of clearly drawing the connection between age-related declines in temporal envelope processing to the discrimination of single words based on temporal information. The stimulus pair that showed the largest aging effects in Gordon-Salant et al. (2006) was chosen in the present study to establish that vocoding would alter the magnitude of the age-related temporal processing deficits. Younger normal-hearing (YNH) and older normal-hearing (ONH) participants completed a phoneme categorization task in which the silent interval preceding the final sibilant /ʃ/ varied from 0 to 60 ms, in 10-ms steps, as a cue to the fricative/affricate distinction. That is, the variation in the silent interval changes the listener's percept from dish (0-ms silence duration) to ditch (60-ms silence duration). ONH participants require relatively larger silence durations than YNH participants before unprocessed "ditch" is perceived, reflecting age-related temporal processing deficits (Gordon-Salant et al. 2006). The novel part of the present study is that the stimuli were also degraded with varied number of spectral channels and varied amount of temporal envelope modulations. We hypothesized that there would be no change from unprocessed to vocoded conditions with relatively high spectral and temporal information, but an age by vocoding interaction would occur for conditions with relatively low spectral information because participants would be forced to rely on temporal cues.

MATERIALS AND METHODS

Participants

Two groups of participants participated in the study: YNH participants (n = 19, 18 to 45 years, mean age and SD = 23.7±6.3) and ONH participants (n = 15, 65 to 78 years, mean age and SD = 70.4±4.1) with normal audiometric hearing, defined as <25 dB HL (ANSI, 2010) at octave frequencies from 250 to 4000 Hz. The average thresholds for both groups are shown in Figure 1. All participants were required to be native speakers of English. Cognitive testing (Montreal Cognitive Assessment [MoCA] Version 7.1 Original Version; Nasreddine et al. 2005) was used to ensure all participants had normal or near-normal cognitive function, which was defined as scores ≥22.* The YNH and ONH participants had an average MoCA score of 28.6±1.5 and 27.1±2.7, respectively.

---

* Normal cognitive functioning is defined as a MoCA score of ≥26; scores of 22–25 on the MoCA indicates that someone is at risk for mild cognitive impairment (Cecato et al. 2016). We included participants at risk for mild cognitive impairment in this study because: (1) requiring participants to have audiometrically normal hearing and a MoCA score ≥26 limits recruiting, (2) the discrimination task is relatively simple such that someone at risk for mild cognitive impairment has no problem performing it, and (3) previous studies have used such an inclusion criterion (Anderson et al. 2012). We found no correlations between performance on the perceptual task and the MoCA, further justifying our choice.

Equipment

The participants were presented stimuli that were generated on a personal computer running MATLAB version 2011B (MathWorks, Natick, MA). The stimuli were delivered by a Tucker-Davis Technologies System 3 (RP2.1, PA5, HB7; Alachua, FL) and a pair of insert earphones (ER2; Etomotyic, Elk Grove Village, IL).

Stimuli

The stimuli were the same speech tokens that were tested in a prior investigation (Gordon-Salant et al. 2006). The test stimuli were created from two natural words, dish and ditch, which were recorded in isolation by an adult American male speaker. To create a continuum of speech tokens between dish and ditch, the ditch token was modified to have the burst and frication removed, and replaced with only the frication portion of dish—this created a hybrid ditch. Then a seven-step continuum was created by removing 10-ms segments of the silence duration in the hybrid ditch. Thus, the final stimulus 1 (dish) had a silence duration of 0 ms. Stimuli 2 to 6 had silence durations of 10 to 50 ms in 10-ms increments. Stimulus 7 (ditch) had a silence duration of 60 ms. Waveforms and spectrograms of the stimulus continuum endpoints are shown in Figure 2A, B.

Once the continuum was constructed, speech segments were vocoded to degrade the speech in a manner similar to the degradation that occurs in CI speech processing. The vocoder bandpass filtered signals into 1, 2, 4, 8, or 16 contiguous channels. The corner frequencies on the band-pass filters were logarithmically spaced from 200 to 8000 Hz. The band-pass filters were 3rd-order Butterworth filters; forward-backward filtering was performed to preserve the timing of the acoustic features of the speech, resulting in a doubling of the filter order and ultimately −36 dB/oct attenuation slope on the filters. The use of forward-backward filtering likely distorted the temporal cues less than one-way filtering and thus minimized such stimulus confounds. The envelope was extracted by calculating the Hilbert envelope and low-pass forward-backward filtering using a 2nd-order Butterworth filter; the low-pass ECO was set to either 50 or 400 Hz. The motivation for using two cutoffs is that a 400-Hz ECO should show large age-related temporal processing differences because the ONH group should not be able to use the higher frequency modulations as well as the YNH group; applying the 50-Hz ECO may reduce the temporal cues to the point that the YNH group might perform similarly to the ONH group (i.e., the YNH group will appear older in their temporal processing abilities; Schvartz-Leyzac & Chatterjee 2015). The extracted envelopes modulated sinusoidal carriers that were at the geometric mean frequency of the channel. The channels were summed and the final vocoded stimulus had the same root-mean-square energy as the unprocessed stimuli. The temporal envelope of the waveform for different conditions is shown in Figure 2C. This figure shows the smoothing of the temporal envelope as the number of channels and ECO are reduced.

Procedure

MATLAB was used to control the experiment and record the participants' responses. During testing, a monitor displayed three boxes. One box read "Begin Trial" followed by the presentation of two side-by-side boxes below. The box to the left read "dish" and the box to the right read "ditch." The
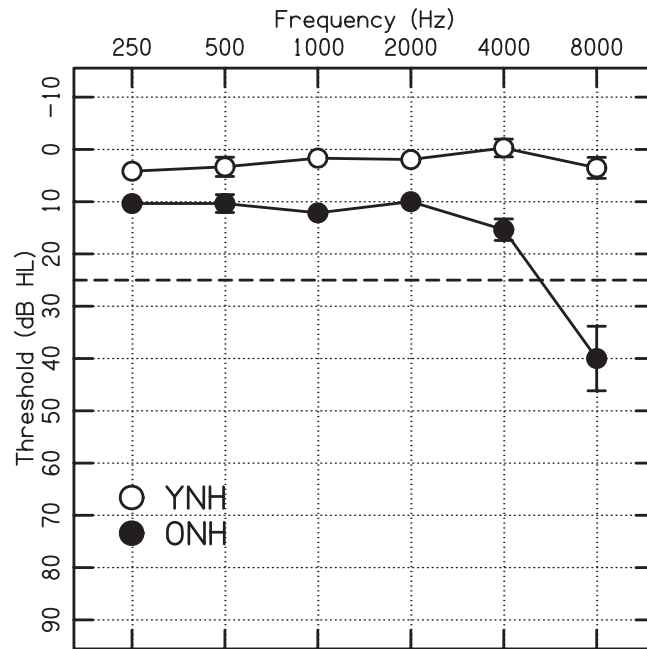
Fig. 1. Average pure-tone thresholds in dB HL (ANSI 2010). Mean thresholds are plotted as open circles for YNH participants and closed circles for ONH participants. Error bars show ±1 SE. ONH indicates older normal-hearing; YNH, younger normal-hearing.

participant initiated each trial and thus testing was self-paced. The stimuli were presented monaurally over insert earphones in a sound attenuated chamber at 65 dB-A to the better ear of the

participant. The better ear was defined as the ear with the better pure-tone hearing threshold average between 250 and 4000 Hz. Upon stimulus presentation, the participant identified whether
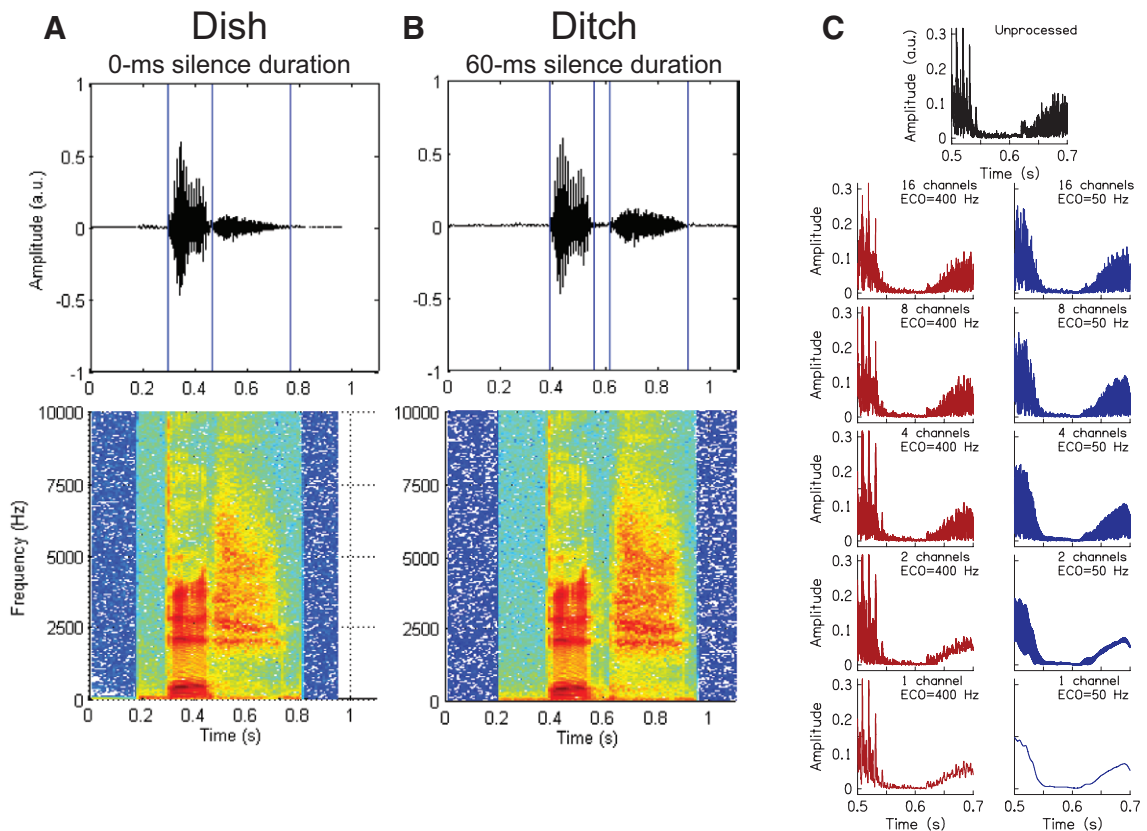

Fig. 2. Waveforms (top row) and spectrograms (bottom row) for dish (0-ms silence duration; column A) and ditch (60-ms silence duration; column B). Column C shows a 200-ms portion of the ditch stimulus, centered on the silence duration. Each panel shows the temporal envelope for the unprocessed and vocoded versions of this stimulus.

the sound was perceived as dish or ditch. If unsure, participants were told to take their best guess. Participants did not receive feedback during the experimental runs.

Before testing, each participant performed a training exercise. They were presented with the endpoint stimuli (stimulus 1—a clear dish, and stimulus 7—a clear ditch). Both of these stimuli were presented in two processing conditions: (1) unprocessed and (2) vocoded with 16 channels and ECO = 400 Hz (i.e., the vocoder condition with the most spectral and temporal information tested in this study). Participants heard 10 repetitions of these two "processing" conditions for two separate stimuli (2 × 2 × 10 = 40 trials), and received correct answer feedback after each response. The training task was conducted to verify that the participants' word recognition performance of dish and ditch was above 90% correct. A high word recognition score ensured that the participant was able to distinctly identify the difference between the endpoint speech stimuli, thus ensuring participants had a clear perception of the different consonant word pairs. If the participant has trouble identifying the endpoint stimuli of the continuum, theoretically the easiest stimuli to identify, then it can be assumed they will have more trouble during less defined boundaries of speech segments, providing unreliable data for the experiment. The training task also established stable performance for participants because of the correct answer feedback. The participant was given 15 min of training to reach the designated word recognition performance level and establish stable performance. All participants achieved near perfect discrimination during the training procedure.

After participants were familiarized with and trained on identifying the endpoint stimuli, they were presented with the entire stimulus continuum. The seven stimuli were unprocessed and vocoded by varying the number of channels (1, 2, 4, 8, or 16) and the low-pass filter ECO (50 or 400 Hz). Thus, there were 11 processing conditions. In a run, two trials of each condition were presented in random order for a total of 154 stimulus presentations (7 Silence Durations × 11 Processing Conditions × 2 Trials/Condition). Five runs were completed for a total of 770 stimulus trials. In summary, each of the seven stimuli along the continuum was presented 10 times for each type of the 11 types of processing (unprocessed and vocoded).

Analysis

For each condition, the percentage of dish responses was calculated. For a given speech contrast and condition, the resulting psychometric function was fit using a logistic regression (psignifit; Wichmann & Hill 2001). From this fit, the value at which 50% of the responses occurred was found, called the 50% crossover point; this represents the boundary marking a change in stimulus categorization. For several conditions, particularly with one or two channels, participants did not necessarily distinguish the speech contrast well and there was no crossover point. In such cases, the 50% crossover point was either inferred out to 70 ms from an extrapolation of the curve fit or set to a maximum of 70 ms. Such a decision assumes that larger silence durations are necessary to produce the perception of the word ditch.

The value of the slope was also calculated. For cases where the participant heard a distinct dish/ditch for the endpoints (i.e., they had near 100% and 0% identification for the contrast continuum endpoints), the maximum slope occurred at the 50% crossover point. For conditions where the 50% crossover did not occur within 0 to 60 ms, the maximum slope of the function between 0 and 60 ms was chosen. The psychometric function slopes computed using psignifit were at times unreasonably steep in cases where the percentage of dish responses went from near 100% to near 0% in a single step, which was a result of the numerical fitting procedure. For the present analyses, the steepest possible slope deemed appropriate was 7.5%/ms.

Inferential statistics, mixed analyses of variance (ANOVAs), were performed. In cases where Mauchly's test indicated that the assumption of sphericity had been violated, the Greenhouse-Geisser correction was used to adjust the degrees of freedom.

RESULTS

Figure 3 shows the results from the experiment. Figure 3A compares the unprocessed data from this study and those from Gordon-Salant et al. (2006). The curves for the present study are shifted to the right of the previous study, likely a result of using different presentation levels (65 versus 85 dB-A), use of different participants, or because we did not use masking noise for the YNH participants. The unprocessed data are also plotted in the other panels (Fig. 3B–K) to compare the performance for the vocoded conditions.

Focusing on just the YNH participants, Figure 3B, C, G, and H show the participants' performance for the 16- and 8-channel conditions. These vocoded conditions are quite similar to the unprocessed conditions, a notable deviation being less than 100% dish responses at 0 ms for ECO = 50 Hz (Fig. 3G, H). Figure 3D, I show the YNH participants' performance for the 4-channel condition. For this relatively small number of spectral channels, the unprocessed and ECO = 400 Hz conditions remain similar; however, the ECO = 50 Hz deviates greatly from the unprocessed condition because the curve has become much more shallow. These data are consistent with our hypothesis that temporal cues become more important when there are reduced spectral cues. Figure 3E, F, J, and K show the YNH participants' performance for the 2- and 1-channel conditions. For these conditions, slopes also became more shallow, particularly for the ECO = 50 Hz condition (Fig. 3J, K).

Now focusing on the data of the ONH participants in Figure 3, many of the age-related trends occur in several conditions. For example, ONH participants almost always had larger crossover durations and shallower slopes than the YNH participants, the largest differences occurring at 4 channels. For fewer channels, there is a lack of categorical perception, particularly for the ONH participants.

We performed a four-way mixed ANOVA using the dependent variable percentage of dish responses with factors Age (between-subjects), ECO (within-subjects), Channels (within-subjects), and Silence Duration (within-subjects). The results of the analysis are reported in Table 1 and the level of significance was chosen to be $p < 0.05$. Three main effects were significant (Age, Channels, and Silence Duration) and one was not (ECO). Of primary interest for this study were the possible significant interactions with age. The Age × Channels, Age × Silence Duration, Age × Channels × Silence Duration, and Age × Channels × ECO × Silence Duration interactions were significant. In addition, the non-age-based interactions Channels × ECO, Channels × Duration, ECO × Duration, and Channels × ECO × Duration were significant.
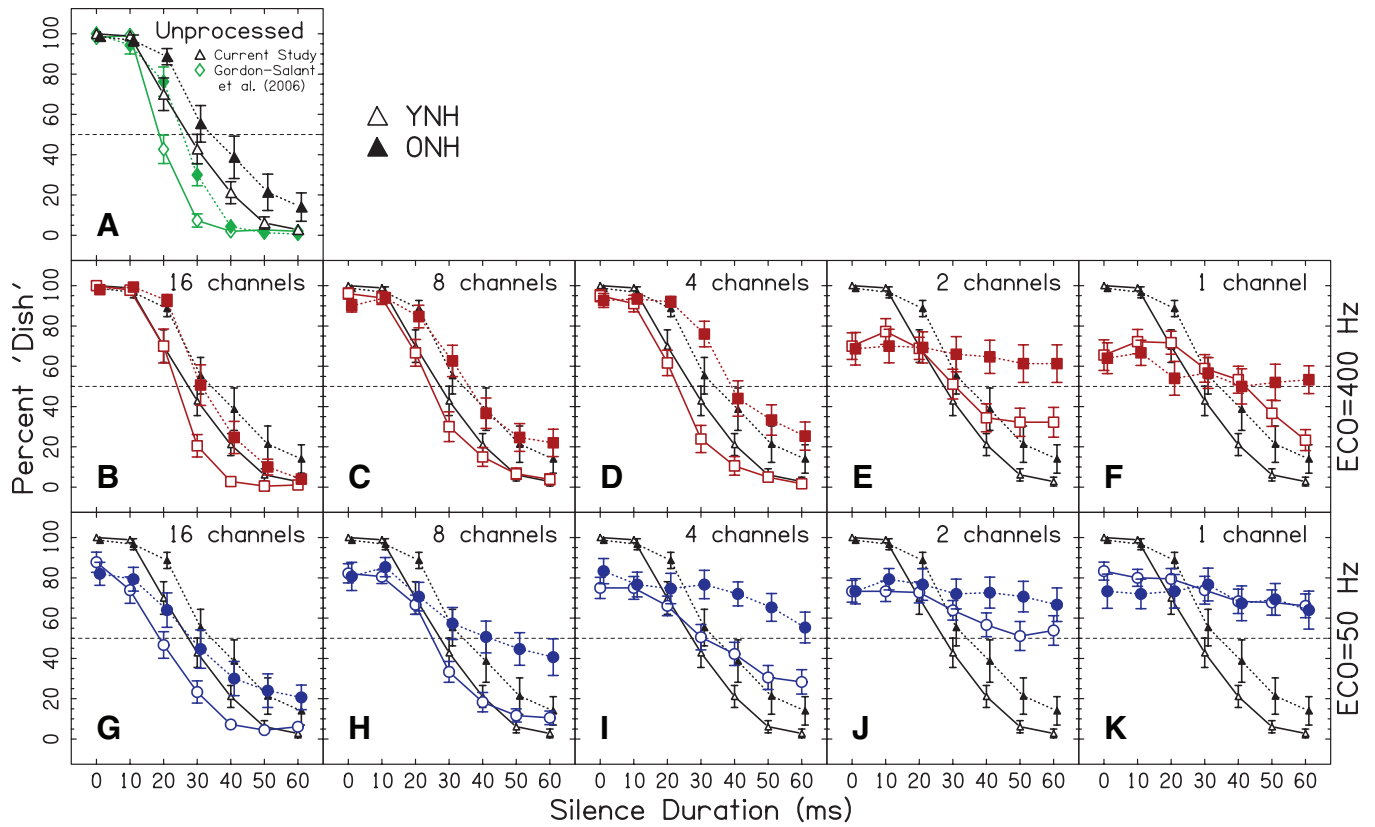
Fig. 3. Dish/ditch response functions. Data points represent the percentage of trials identified as dish as a function of silence duration, averaged over participants. Error bars show ±1 SE. Open symbols show data for YNH participants; closed symbols show data for ONH participants. A, Data for the unprocessed conditions from the present study (black upward triangles) and Gordon-Salant et al. (2006; green diamonds). B–K, Data for the vocoded conditions (larger symbols) and the unprocessed conditions (smaller symbols). Red squares show the data for the ECO = 400-Hz conditions in B–F. Blue circles show the data for the ECO = 50-Hz conditions in G–K. ONH indicates older normal-hearing; YNH, younger normal-hearing.

In an attempt to better understand the interactions, we calculated slopes and crossover points (see Analysis section) to summarize the data and more easily compare the two groups. Figure 4 shows the slopes of the performance functions in Figure 3, a comparison between the YNH and ONH groups. Figure 4C

shows the difference in slopes between the two age groups across conditions, where the YNH participants had slightly steeper slopes than the ONH participants for the unprocessed condition (0.35%/ms) and steeper slopes in all vocoding conditions except for 2 channels, ECO = 50 Hz. In other words, using ECO

TABLE 1. Results of the four-way mixed ANOVA

Factors	df		F	p	η_p^2
	Between	Within			
Age	1.00	32.00	18.32	**<0.001**	0.36
Channels	2.21	70.58	15.66	**<0.0001**	0.33
ECO	1.00	32.00	2.99	0.093	0.09
Silence Duration	2.01	64.30	196.27	**<0.0001**	0.86
Age × Channels	2.21	70.58	4.02	**0.019**	0.11
Age × ECO	1.00	32.00	0.11	0.747	0.00
Age × Silence Duration	2.01	64.30	11.63	**<0.0001**	0.27
Channels × ECO	2.60	83.02	5.56	**0.003**	0.15
Channel × Silence Duration	7.75	248.01	44.21	**<0.0001**	0.58
ECO × Silence Duration	3.77	120.69	33.78	**<0.0001**	0.51
Age × Channels × ECO	2.59	83.02	0.26	0.827	0.01
Age × Channels × Silence Duration	7.75	248.01	2.90	**0.005**	0.08
Age × ECO × Silence Duration	3.77	120.69	1.54	0.198	0.05
Channels × ECO × Silence Duration	11.50	368.12	4.00	**<0.0001**	0.11
Age × Channels × ECO × Silence Duration	11.50	368.12	3.04	**<0.001**	0.09

ANOVA, analyses of variance; df, degrees of freedom; ECO, envelope cutoff.
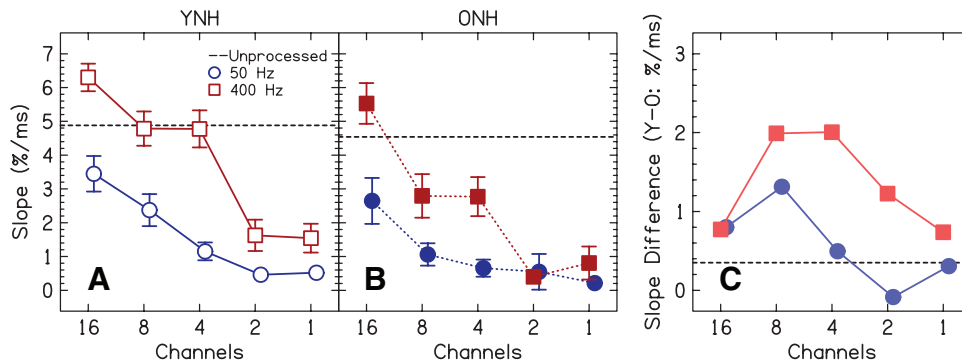Bold p values are significant at the 0.05 level.

Fig. 4. Slope of dish/ditch response functions for the YNH group (A; open symbols) and the ONH group (B; closed symbols) as a function of the number of channels. The slopes for the unprocessed data are shown by the horizontal dashed line in each panel. The vocoded conditions are also plotted for the two ECOs (50 Hz = blue circles; 400 Hz = red squares). Error bars show ±1 SE. C, Average difference in slope between the YNH and ONH listeners (positive numbers show larger age-related temporal processing deficits). Figure C uses light red and light blue symbols to signify that it is the difference between the YNH and ONH groups. ECO indicates envelope cutoff; ONH, older normal-hearing; YNH, younger normal-hearing.

= 50 Hz makes the temporal processing of the YNH participants more like the ONH participants. For the vocoded conditions, the data were analyzed with a three-way mixed ANOVA with factors Age (between-subjects), Channels (within-subjects), and ECO (within-subjects). YNH participants had significantly steeper slopes than ONH participants [$F(1,32) = 6.65$, $p = 0.015$, $\eta_p^2 = 0.17$], confirming the existence of age-related temporal processing differences. Slopes became shallower with a decreasing number of channels [$F(2.6,83.4) = 53.2$, $p < 0.0001$, $\eta_p^2 = 0.62$], indicating that it was more difficult to discriminate dish and ditch with less spectral information. Slopes became shallower with a decreasing ECO [$F(1,32) = 87.4$, $p < 0.0001$, $\eta_p^2 = 0.73$], indicating that it was more difficult to discriminate dish and ditch with less temporal information. Most importantly, there was a significant Channels × ECO interaction [$F(4,128) = 12.0$, $p < 0.0001$, $\eta_p^2 = 0.27$] where the difference in slopes between ECO = 50 and 400 Hz became smaller with decreasing number of channels. The Age × Channels [$F(2.6,83.4) = 1.12$, $p = 0.34$, $\eta_p^2 = 0.03$], Age × ECO [$F(1,32) = 3.55$, $p = 0.069$, $\eta_p^2 = 0.10$], and Age × Channels × ECO interactions [$F(4,128) = 0.99$, $p = 0.42$, $\eta_p^2 = 0.03$] were all not significant, indicating the slopes did not significantly change across the two groups depending on the vocoding parameters. This can be seen in the relatively parallel curves in Figure 4C.

Figure 5 shows the 50% crossover point of the performance functions in Figure 3, a comparison between the YNH and ONH groups. Figure 5C shows the difference in crossover points across the two age groups, where the ONH participants needed a 9.8-ms longer silence duration than the YNH participants before ditch was perceived for the unprocessed stimuli. For the vocoded conditions, the data were analyzed with a three-way mixed ANOVA with factors Age (between-subjects), Channels (within-subjects), and ECO (within-subjects). YNH participants exhibited significantly shorter silence durations to change their percept from dish to ditch than ONH participants [$F(1,32) = 19.5$, $p < 0.001$, $\eta_p^2 = 0.38$], confirming the existence of age-related temporal processing differences. Crossovers were longer with a decreasing number of channels [$F(2.8,89.5) = 60.7$, $p < 0.0001$, $\eta_p^2 = 0.66$], indicating that longer silence durations were necessary to perceive ditch when there was less spectral information. Crossovers were longer for a decreasing ECO [$F(1,32) = 14.2$, $p = 0.001$, $\eta_p^2 = 0.31$], indicating that longer silence durations were necessary to perceive ditch when there was less temporal information. There was a significant Channels × ECO interaction [$F(3.1,100.4) = 3.1$, $p = 0.029$, $\eta_p^2 = 0.09$], where listeners needed even longer silence durations to perceive ditch for ECO = 50 Hz compared with 400 Hz, particularly for 4 and 2 channels. This again supports the hypothesis that temporal cues become more necessary in the absence of spectral cues. The interactions with age were not significant [Age × Channels: $F(2.8,89.5) = 2.2$, $p = 0.099$, $\eta_p^2 = 0.06$; Age × ECO: $F(1,32) = 13.4$, $p = 0.84$, $\eta_p^2 = 0.001$; Age × Channels × ECO: $F(3.1,100.4) = 0.83$, $p = 0.49$, $\eta_p^2 = 0.03$].
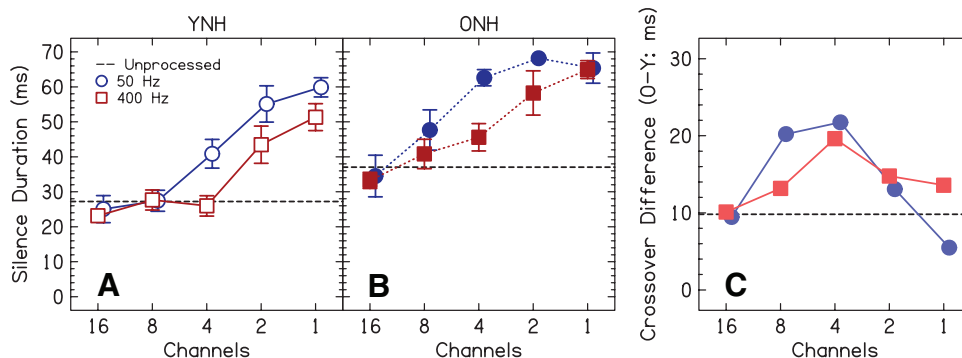

Fig. 5. Crossover of response functions for the YNH group (A; open symbols) and the ONH group (B; closed symbols) as a function of the number of channels. All conventions are the same as in Figure 4. ONH indicates older normal-hearing; YNH, younger normal-hearing.

The three-way mixed ANOVAs using slope or crossover points as the dependent variable did not show significant interactions with age, which is contrary to the significant interactions found in the four-way mixed ANOVA using percentage of dish responses. This likely occurred for several reasons. First, the slope and crossover are summary metrics and there is much less statistical power when using them. Second, these summary metrics assume well-behaved sigmoidal functions where the endpoints are highly discriminable. Our data, however, do not follow this assumption particularly as the number of channels and ECO were reduced. Third, ceiling and floor effects were much more prevalent with the summary metrics. For example, a ceiling effect on the longest crossover (set to the arbitrarily large value of 70 ms for our analysis) likely contributed to the lack of interactions with age. Finally, these summary metrics do not necessarily capture the information on the endpoints of the silence duration continuum, where some of the largest age-related differences can be seen in Figure 3. Therefore, our interpretation is that while the summary metrics shown in Figures 4 and 5 are useful for seeing some of the trends in the age-related changes in temporal processing and also allow for easier comparison with the previous study by Gordon-Salant et al. (2006), the most accurate picture of the data is demonstrated in Figure 3.

In summary, the experiment showed age-related temporal processing deficits through a main effect of age (Table 1; Figs. 3–5) and multiple significant interactions with age (Table 1; Fig. 3).

DISCUSSION

The purpose of this study was to better understand how aging, particularly age-related temporal processing deficits, might affect speech understanding in CI users. CI sound processors are vocoder-centric, meaning that temporal envelope information is primarily used to convey speech information (Shannon et al. 1995), a result of CI stimulation patterns having relatively poor spectral resolution compared with the normal ear (Friesen et al. 2001). Testing YNH and ONH participants, we presented vocoded versions of a single word contrast—dish/ditch—where the spectral (number of vocoder channels) and temporal information (low-pass filter cutoff applied to the temporal envelope or ECO) were varied. This word contrast has been previously shown to highlight age-related temporal processing deficits when presented as unaltered acoustic signals (i.e., not vocoded; Gordon-Salant et al. 2006). We hypothesized that age-related temporal processing deficits would become more apparent as the spectral information was reduced and participants were forced to utilize the temporal cues to distinguish the words.

Figure 3 shows the results of the experiment for the YNH and ONH participants, respectively. The age-related temporal processing deficits for the unprocessed conditions (Fig. 3A) were replicated from Gordon-Salant et al. (2006). The increase in silence duration needed to shift the percept from dish to ditch was 9.8 ms in the present study, which can be compared with 5.5 ms for the former study. One difference between studies is that both YNH and ONH participant response functions were shifted to longer silence durations in this study compared with the former. These changes in the response patterns between studies were likely because we tested at a lower presentation level than the other study, we tested different participants (potentially more hearing loss in our participants; Fig. 1), and

our YNH group was not presented masking noise in an attempt to equate performance between the YNH and ONH groups.

Responses for both listener groups were similar between the unprocessed conditions and conditions with a large (e.g., 16) number of channels and higher ECO (400 Hz). Deviations from the unprocessed condition grew larger as the number of channels and ECO were reduced. For example, the words became less discriminable, shown in the slopes that are plotted in Figure 4. The point at which the word ditch was perceived occurred at longer silence durations than the unprocessed condition, shown as the duration at which the response functions crossed the 50% point in Figure 5. These two metrics highlight the importance of examining the entire function: perception of stimuli with silence durations in the middle of the speech continuum revealed subtle differences in perception of the temporal gap (i.e., function slopes and crossover points), while identification of endpoint stimuli revealed large age-related differences in identification of clear examplar stimuli differing in the temporal cue.

Such information may also inform the nature of age-related temporal processing deficits. It could be that the age-related differences for the unprocessed stimuli are a result of peripheral or central processing deficits, or a mix of the two (see differences in hearing thresholds in Fig. 1). If there were no age-related central processing deficits, then no interactions with age would have been seen and one could argue a completely peripheral temporal processing deficit. The data from this experiment, however, demonstrates multiple interactions with age, which are a result of larger effects in the older participants compared with the younger as spectral and temporal information are reduced (Table 1; Fig. 3). This argues for some central component to the age-related temporal processing deficits.

Finally, these results suggest that older CI users should also be susceptible to such temporal processing deficits, which should ultimately reduce speech understanding. While previous studies have shown mixed results on whether aging and age-related temporal processing deficits affect speech understanding with a CI, there is mounting evidence that older CI listeners perform worse than their younger peers, particularly when duration of deafness is accounted for in the study (Sladen & Zappler 2015; Beyea et al. 2016). Many studies in this area use sentence recognition as a metric for CI performance. One advantage of the testing paradigm used in this study is that it targets aspects of single words that might be particularly difficult for older CI users to process. The combination of advancing age with spectrally degraded signals—which force participants to rely on temporal cues—should exacerbate the age-related temporal processing deficits that are well known to occur. The results of this study are generally consistent with Winn et al. (2012), who showed that vocoding caused NH participants to rely more on temporal cues when the spectral cues were degraded; however, they did not investigate effects of age. Winn et al. also tested CI participants and found that the NH and CI participants might be using different sets of cues to discriminate phonemes. Therefore, continuing phoneme discrimination research with both NH and CI participants, and including potential effects of aging is a promising avenue for future research because it will better inform why older CI users are performing worse than their younger peers at perceiving speech. Clearly, testing only one phoneme contrast limits the ability to generalize to other temporally based phoneme contrasts. Nonetheless, the present research highlights the added importance of technological advancement to provide

better spectral resolution for older CI users compared with younger for improved clinical outcomes. Clinicians also benefit from this information to help advise older candidates on proper expectations for speech understanding with a CI.

REFERENCES

ANSI. (2010). ANSI S3.6-2010, American National Standard Specification for Audiometers (Revision of ANSI S3.6-1996, 2004). New York, NY: American National Standards Institute.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., et al. (2012). Aging affects neural precision of speech encoding. *J Neurosci*, *32*, 14156–14164.

Beyea, J. A., McMullen, K. P., Harris, M. S., et al. (2016). Cochlear implants in adults: Effects of age and duration of deafness on speech recognition. *Otol Neurotol*, *37*, 1238–1245.

Cecato, J. F., Martinelli, J. E., Izbicki, R., et al. (2016). A subtest analysis of the Montreal cognitive assessment (MoCA): Which subtests can best discriminate between healthy controls, mild cognitive impairment and Alzheimer's disease? *Int Psychogeriatr*, *28*, 825–832.

CHABA. (1988). Speech understanding and aging. *J Acoust Soc Am*, *83*, 859–895.

Chen, D. S., Clarrett, D. M., Li, L., et al. (2013). Cochlear implantation in older adults: Long-term analysis of complications and device survival in a consecutive series. *Otol Neurotol*, *34*, 1272–1277.

Clark, J. H., Yeagle, J., Arbaje, A. I., et al. (2012). Cochlear implant rehabilitation in older adults: Literature review and proposal of a conceptual framework. *J Am Geriatr Soc*, *60*, 1936–1945.

Fitzgibbons, P. J., & Gordon-Salant, S. (1995). Age effects on duration discrimination with simple and complex stimuli. *J Acoust Soc Am*, *98*, 3140–3145.

Friesen, L. M., Shannon, R. V., Baskent, D., et al. (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. (2010). *The Aging Auditory System*. New York, NY: Springer.

Gordon-Salant, S., & Fitzgibbons, P. J. (2001). Sources of age-related recognition difficulty for time-compressed speech. *J Speech Lang Hear Res*, *44*, 709–719.

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

Goupell, M. J., Laback, B., Majdak, P., et al. (2008). Effects of upper-frequency boundary and spectral warping on speech intelligibility in electrical stimulation. *J Acoust Soc Am*, *123*, 2295–2309.

Jin, S. H., Liu, C., Sladen, D. P. (2014). The effects of aging on speech perception in noise: Comparison between normal-hearing and cochlear-implant listeners. *J Am Acad Audiol*, *25*, 656–665.

Kim, B. J., Chang, S. A., Yang, J., et al. (2015). Relative contributions of spectral and temporal cues to Korean phoneme recognition. *PLoS One*, *10*, e0131807.

Lin, F. R., Chien, W. W., Li, L., et al. (2012). Cochlear implantation in older adults. *Medicine (Baltimore)*, *91*, 229–241.

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. In A. Moller (Ed.), *Cochlear and Brainstem Implants* (pp. 109–143). Basel, Switzerland: Karger.

Loizou, P. C., Dorman, M., Tu, Z. (1999). On the number of channels needed to understand speech. *J Acoust Soc Am*, *106*(4 Pt 1), 2097–2103.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., et al. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*, *53*, 695–699.

Nie, K., Barco, A., Zeng, F. G. (2006). Spectral and temporal cues in cochlear implant speech perception. *Ear Hear*, *27*, 208–217.

Schvartz, K. C., & 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.

Schvartz, K. C., Chatterjee, M., Gordon-Salant, S. (2008). Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners. *J Acoust Soc Am*, *124*, 3972–3988.

Schvartz-Leyzac, K. C., & Chatterjee, M. (2015). Fundamental-frequency discrimination using noise-band-vocoded harmonic complexes in older listeners with normal hearing. *J Acoust Soc Am*, *138*, 1687–1695.

Shannon, R. V., Zeng, F. G., Kamath, V., et al. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Sheldon, S., Pichora-Fuller, M. K., Schneider, B. A. (2008a). Effect of age, presentation method, and learning on identification of noise-vocoded words. *J Acoust Soc Am*, *123*, 476–488.

Sheldon, S., Pichora-Fuller, M. K., Schneider, B. A. (2008b). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *J Acoust Soc Am*, *123*, 489–499.

Sladen, D. P., & Zappler, A. (2014). Older and younger adult cochlear implant users: Speech recognition in quiet and noise, quality of life, and music perception. *Am J Audiol*, *24*, 31–39.

Sladen, D. P., & Zappler, A. (2015). Older and younger adult cochlear implant users: Speech recognition in quiet and noise, quality of life, and music perception. *Am J Audiol*, *24*, 31–39.

Souza, P. E., & Boike, K. T. (2006). Combining temporal-envelope cues across channels: Effects of age and hearing loss. *J Speech Lang Hear Res*, *49*, 138–149.

Stilp, C. E., & Goupell, M. J. (2015). Spectral and temporal resolutions of information-bearing acoustic changes for understanding vocoded sentences. *J Acoust Soc Am*, *137*, 844–855.

Stone, M. A., Füllgrabe, C., Moore, B. C. (2008). Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region. *J Acoust Soc Am*, *124*, 2272–2282.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys*, *63*, 1293–1313.

Winn, M. B., Chatterjee, M., 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., Won, J. H., 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.

Xu, L., & Pfingst, B. E. (2008). Spectral and temporal cues for speech recognition: Implications for auditory prostheses. *Hear Res*, *242*, 132–140.

Xu, L., Thompson, C. S., Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J Acoust Soc Am*, *117*, 3255–3267.

Xu, L., & Zheng, Y. (2007). Spectral and temporal cues for phoneme recognition in noise. *J Acoust Soc Am*, *122*, 1758.