

# Word Recognition Variability With Cochlear Implants: “Perceptual Attention” Versus “Auditory Sensitivity”

Aaron C. Moberly,<sup>1</sup> Joanna H. Lowenstein,<sup>2</sup> and Susan Nittrouer<sup>2</sup>

**Objectives:** Cochlear implantation does not automatically result in robust spoken language understanding for postlingually deafened adults. Enormous outcome variability exists, related to the complexity of understanding spoken language through cochlear implants (CIs), which deliver degraded speech representations. This investigation examined variability in word recognition as explained by “perceptual attention” and “auditory sensitivity” to acoustic cues underlying speech perception.

**Design:** Thirty postlingually deafened adults with CIs and 20 age-matched controls with normal hearing (NH) were tested. Participants underwent assessment of word recognition in quiet and perceptual attention (cue-weighting strategies) based on labeling tasks for two phonemic contrasts: (1) “cop”–“cob,” based on a duration cue (easily accessible through CIs) or a dynamic spectral cue (less accessible through CIs), and (2) “sa”–“sha,” based on static or dynamic spectral cues (both potentially poorly accessible through CIs). Participants were also assessed for auditory sensitivity to the speech cues underlying those labeling decisions.

**Results:** Word recognition varied widely among CI users (20 to 96%), but it was generally poorer than for NH participants. Implant users and NH controls showed similar perceptual attention and auditory sensitivity to the duration cue, while CI users showed poorer attention and sensitivity to all spectral cues. Both attention and sensitivity to spectral cues predicted variability in word recognition.

**Conclusions:** For CI users, both perceptual attention and auditory sensitivity are important in word recognition. Efforts should be made to better represent spectral cues through implants, while also facilitating attention to these cues through auditory training.

**Key words:** Cochlear implants, Sensorineural hearing loss, Speech perception.

(*Ear & Hearing* 2016;37:14–26)

## INTRODUCTION

Cochlear implants (CIs) have dramatically improved the lives of adults who acquire sensorineural hearing loss. On the whole, postlingually deafened adults with CIs are able to recognize about 60% of spoken words and 70% of sentences presented to them in quiet in standard clinical tests (Firszt et al. 2004; Gifford et al. 2008; Holden et al. 2013). However, enormous outcome variability exists, and this variability remains largely unexplained. The primary factors that have been identified as predictors of speech perception outcomes relate to the remaining integrity of the auditory system: amount of residual hearing before implantation, prior use of hearing aids, duration of hearing loss, age at implantation, and etiology of hearing loss (Eggermont & Ponton 2003; Kelly et al. 2005; Lazard et al. 2012). However, these factors serve only as indirect predictors because it is unclear how their effects on speech

perception are imparted. And although each factor shows a significant relationship to speech perception outcomes, the sum of these factors still accounts for less than 25% of the variance in those outcomes. Moreover, these factors do not represent viable targets of intervention when seeing a patient who is considering cochlear implantation or has a poor outcome. The variance in speech perception exhibited by CI users is most often attributed to variability among patients in the fidelity of speech signal representation through CI processing and neural stimulation. Current CI speech processors operate by recovering amplitude structure (the temporal envelope—the slowly varying modulation of the signal amplitude over time) in up to 22 independent frequency channels. However, the effective number of available channels is typically limited to four to seven, mostly due to overlapping regions of neural stimulation by adjacent implant electrodes (Friesen et al. 2001). Thus, the spectral structure (the detailed frequency-specific information) delivered is significantly degraded (Wilson & Dorman 2008). Therefore, listeners with CIs lack sensitivity to many spectral cues that underlie speech perception. A reasonable assumption, then, would be that speech outcomes vary proportionately with extent of spectral degradation and that for individual patients, improving signal quality would be all that is required for improving outcomes.

Clinicians should recognize, however, that understanding speech is more complicated than simply having good “auditory sensitivity” (Lieberman et al. 1967; Remez et al. 1994). Language processing is strongly dependent on the recognition of phonemic structure, the mental organization of the fundamental units of language. In particular, the lexicon (an individual’s store of words) is organized according to phonemic structure. Thus, effective recognition of spoken words depends on the ability to recover phonemic structure from the signal. However, phonemes are not distinctly represented within the acoustic speech signal, ready to be plucked from the signal and lined up like beads on a string. In fact, there is “a marked lack of correspondence between sound and perceived phoneme” (Lieberman et al. 1967). Instead, listeners must know which components, or “acoustic cues,” of the speech signal should be attended to for the recovery of phonemic structure (Repp 1982). These cues are generated as acoustic consequences of the articulatory gestures (movements of the larynx, pharynx, palate, and tongue) used to produce speech and can consist of spectral (frequency-related), temporal (duration-related), and amplitude (intensity-related) structure. Spectral cues can be further broken down into structure described as static (stable across time) or dynamic (varying over time). Because each articulatory gesture can affect structure in each of the spectral, temporal, and amplitude domains, cues to each phonemic distinction covary across these domains. Consequently, it would seem that all cues affiliated with a distinction should be equally as effective in signaling that distinction. If that were so, the perceptual consequences of acquired hearing loss would be minimal: a temporal or amplitude cue

<sup>1</sup>Department of Otolaryngology–Head and Neck Surgery, The Ohio State University, Wexner Medical Center, Columbus, Ohio, USA; and <sup>2</sup>Speech, Language, and Hearing Sciences, University of Florida, Gainesville, Florida, USA.

preserved in the face of hearing loss and subsequent implantation could substitute for a diminished spectral cue. The current investigation examined the extent to which that may be true.

Prior to conducting this investigation with CI users, clues to its outcome could be gleaned from work with other populations. It is now known from second-language learning studies that all phonetic cues are not necessarily equal: for a given language, some cues are more informative than others. The amount of perceptual attention, or weight, given to each cue is similar for individual adult speakers of a given language, most likely because those strategies lead to the most accurate and efficient speech recognition in that language (Nittrouer 2005). Conversely, cues are not attended to if they are not informative or do not exist for a given language. A perfect example is the /r/-/l/ contrast (e.g., “rock” versus “lock”), based primarily on differences in onset frequency of the third formant (the third resonance peak in the speech signal), which is readily recognized by primary English-speaking listeners but is largely ignored by primary Japanese-speaking listeners. This phonemic contrast does not exist in the Japanese language, so native Japanese speakers have difficulty attending to the third-formant cue when learning the /r/-/l/ contrast in English as a second language. Importantly, however, this lack of perceptual attention by Japanese listeners is not explained by a lack of auditory sensitivity to the cue, as Japanese listeners are just as able as English listeners to discriminate this acoustic difference (the isolated third formants from /r/ versus /l/) in a nonspeech discrimination task (Miyawaki et al. 1975).

Postlingually deafened adults with CIs face a different but related dilemma. Most likely these individuals have developed typical strategies of perceptual attention, or “cue-weighting strategies,” for their primary language before losing their hearing. Upon implant activation, they must learn to make sense of the degraded representations of speech. It is conceivable that the availability of cues from their CIs (their auditory sensitivity) may constrain the choice of cues they can weight, so these individuals may shift their cue-weighting strategies (especially away from spectral cues) to those cues that are better represented through their CIs, namely duration and amplitude cues. Importantly, the cues that are most accessible to CI users may not be those that are most informative for efficient perception of speech in their native language.

Some evidence to support a shift in perceptual attention for early CI users was found in a study by Hedrick and Carney (1997), examining the labeling decisions for synthetic fricative-vowel syllables by four adult postlingually deafened CI users. The CI users in that study were found to weight an amplitude cue more heavily than normal-hearing (NH) listeners, while placing reduced weight on spectral cues. A more recent study assessed the use of acoustic cues for a tense/lax vowel contrast, varying in vowel-inherent spectral change and vowel duration, as well as a word-final fricative voicing contrast (with cues consisting of F1 transition, vowel duration, consonant duration, and consonant voicing) (Winn et al. 2012). In that study, CI users showed decreased use of spectral cues and greater use of durational cues than NH listeners. In neither of these studies, however, was weighting of cues examined for a relationship with speech recognition performance.

Three other studies have indirectly related perceptual attention to success in speech recognition. Kirk et al. (1992) found that CI users’ abilities to perceptually attend to formant

transitions during vowel identification predicted word recognition. Donaldson et al. (2013) examined vowel identification by adult CI users, using /dVd/ syllables that were modified to retain quasistatic spectral cues (from an 80-msec segment of the vowel center) or dynamic spectral cues (from two 20-msec segments of the vowel edges). CI users were found to have poorer vowel recognition when using either type of spectral cue alone versus in the presence of duration cues. In a study by Peng et al. (2012) investigating recognition of speech prosody, the authors examined the use of F0 contour, intensity, and duration cues in the perception of prosody using question-statement identification for CI and NH listeners. Accuracy in question-statement identification was poorer for CI than NH listeners, but CI listeners’ weighting of F0 contour predicted accuracy in question-statement identification. However, in none of these cases was it clear whether the effects on speech (or prosody) recognition were based on perceptual attention strategies or on auditory sensitivity to the relevant cues. Notably, Peng et al. (2012) acknowledged that their results did not clarify whether “(a) those CI listeners who use alternative cues are also more sensitive to those cues, or (b) all CI listeners are sufficiently sensitive to these alternative cues, but not all of them can adjust their listening strategies to attend to them” (p. 78).

To more directly address the relationships of auditory sensitivity, perceptual attention, and speech recognition performance, a recent study by Moberly et al. (2014) examined whether two distinct cues were equal in their capacity to elicit accurate speech recognition by postlingually deafened adult CI users. In that study, listeners were assessed for perceptual attention while labeling synthetic versions of the “ba”–“wa” phonemic contrast, as well as performance of word recognition in quiet. Listeners’ auditory sensitivity to spectral change was also evaluated to gauge their sensitivity to spectral structure. For the ba–wa contrast, NH adults weight a dynamic spectral cue (the rate of formant transitions—relatively slow changes in the peak resonant frequencies) more heavily than an amplitude cue (the onset rise time—the time it takes for the amplitude of the signal to reach its maximum). It was predicted that if both cues were equal in their ability to elicit accurate speech recognition (i.e., the “phonemic informativeness” of the cues), CI users would shift their attention to the amplitude cues that were well-represented by their implants, without loss of word recognition accuracy. However, this was not observed. The adults with CIs showed variable attentional (i.e., cue-weighting) strategies, and those individuals who continued to attend to the spectral cue strongly, like NH adults, showed the best word recognition. On the other hand, sensitivity to the spectral cue in that study—as measured by a discrimination task—did not independently explain variance in word recognition. Of course, that study only assessed a single phonemic contrast, limiting the generalizability of the findings.

The present study was designed to further examine, using a different set of stimuli, the contributions of perceptual attention and auditory sensitivity to word recognition for postlingually deafened adults with contemporary CIs. Three hypotheses were tested. (1) Perceptual attention to the same acoustic cues that underlie the perception of phonemic contrasts by listeners with NH would predict word recognition for a group of postlingually deafened adults with CIs. To test this hypothesis, adults with CIs and adults with NH performed perceptual labeling tasks while listening to two sets of phonemic contrasts, along with

a task of word recognition under quiet conditions. Support of this hypothesis would provide further evidence that perceptual attention to acoustic cues plays an important role in word recognition. (2) Auditory sensitivity for nonspeech correlates of the acoustic cues that underlie those phonemic contrasts would not independently predict word recognition abilities for the CI users. To test this hypothesis, auditory sensitivity was assessed using discrimination tasks and results were again analyzed as predictors of word recognition. Evidence supporting this hypothesis would further confirm that good auditory sensitivity alone is not sufficient for accurate word recognition. (3) Auditory sensitivity for nonspeech correlates would not predict perceptual attention to those related cues during the labeling tasks. Evidence supporting this hypothesis would suggest that perceptual attention is not driven solely by which cues are most accessible through implants. If the three hypotheses were supported, findings would suggest that efforts focused on improving auditory sensitivity to cues through CIs may be neither sufficient nor necessary to improve speech understanding. Rather, adjustment of attentional strategies through auditory training might assist in optimizing outcomes.

To address the first hypothesis, two sets of stimuli were chosen to examine perceptual attention during a labeling task: “cop” versus “cob” and “sa” versus “sha.” These sets were selected because each labeling decision was based on a different set of two cues. The cop–cob decision in this experiment depended on a duration cue (vocalic length) and a dynamic spectral cue (syllable-final formant transitions). The duration cue should be

readily accessible through a CI, whereas the dynamic spectral cue is likely more poorly represented through a CI (Fig. 1). In the study by Moberly et al. (2014), the ba–wa contrast was chosen to test the weighting of an amplitude cue versus a dynamic spectral cue. Whereas with the ba–wa contrast, adults with NH typically weight the cue that should be less accessible through a CI (the spectral cue), in this case, adults with NH typically weight the cue that should be more accessible to CI users (the duration cue) (Wardrip-Fruin & Peach 1984; Nittrouer 2004). Therefore, it was hypothesized that CI users would likewise weight the duration cue heavily and that the amount of perceptual attention paid to the duration cue would explain variability in word recognition, since paying attention to this cue appears to be important for adults with NH.

The sa–sha decision depended on a static spectral cue (fricative-noise spectrum) and a dynamic spectral cue (syllable-initial formant transitions) (Fig. 2). Adults with NH heavily attend to the fricative noise in the sa–sha contrast (Perkell et al. 1979; Stevens 1985; Hedrick & Ohde 1993). In this case, it was predicted that CI users would show poorer attention to the dynamic spectral cue, but it was harder to predict attention to the static spectral cue. Although this cue does involve spectral structure, it could be better represented through a CI processor than relatively low-frequency spectral structure that changes rapidly (i.e., dynamic formant transitions). Talkers tend to hold the vocal-tract configuration that creates fricative noise for a relatively longtime (hundreds of milliseconds), and the noise has a high-frequency, but broadband structure. If any spectral

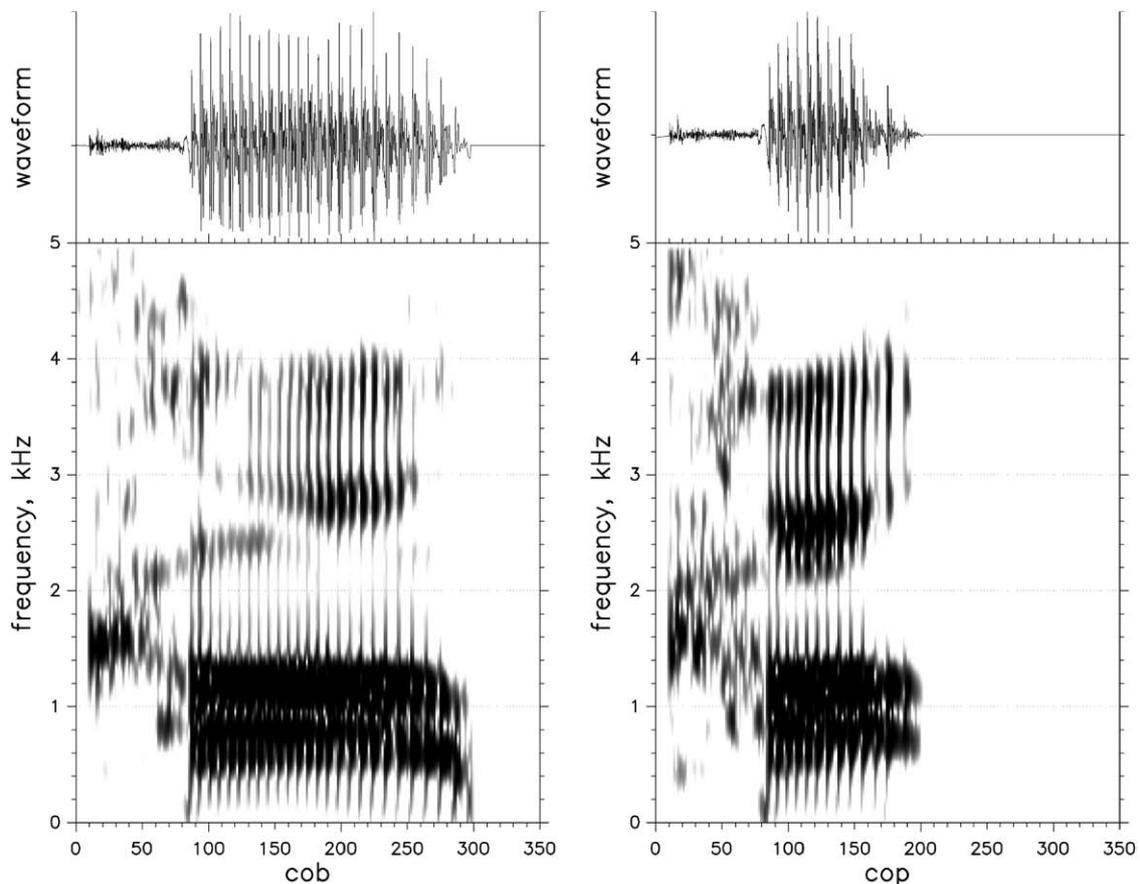


Fig. 1. Spectrograms of natural “cob” and “cop” stimuli produced by a male English speaker. Reprinted with permission from Nittrouer et al. (2014).

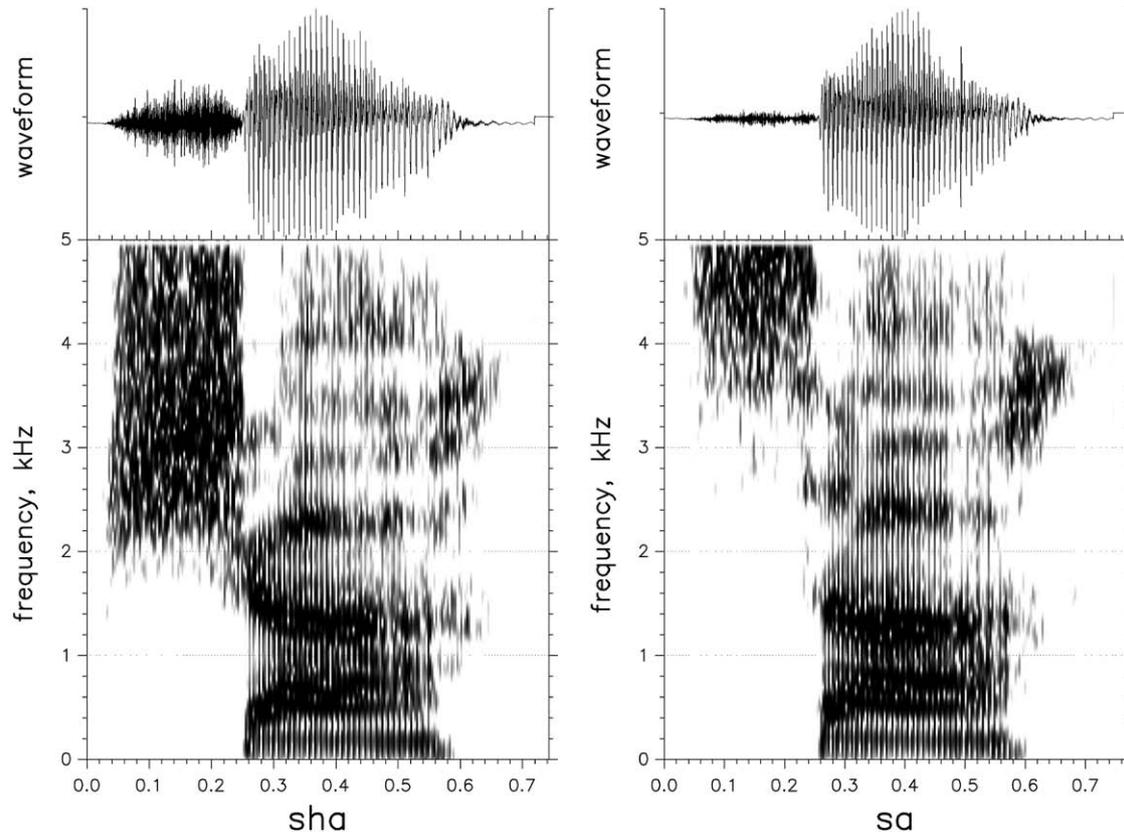


Fig. 2. Spectrograms for natural “sha” and “sa” stimuli produced by a male English speaker.

structure is accessible to CI users, this kind should be. In any event, it might be that perceptual attention to this cue would predict better word recognition for individuals with CIs because it is weighted heavily in fricative labeling by adults with NH. Use of this contrast should reveal how CI users fare in the face of a labeling decision that requires attention to two types of cues that are poorly accessible through their CIs.

In order to address the second and third hypotheses, auditory sensitivity to three acoustic cues relevant to the labeling tasks was assessed using discrimination tasks. Because listeners place stimuli heard as speech into phonemic categories, it was important that these stimuli not be recognizable as speech. Thus, nonspeech analogs of the relevant cues used in labeling were developed, using sine wave signals.

## MATERIALS AND METHODS

### Participants

Thirty adults who wore CIs and were between 52 and 88 years of age were recruited from a pool of departmental patients and were enrolled. Participants had varying etiologies of hearing loss and ages of implantation, though most experienced a progressive decline in hearing during adulthood. All participants met criteria for implantation no earlier than age 45 years, based on the appropriate sentence recognition criteria used at the time of their candidacy evaluations. All used Cochlear devices, except one who wore an Advanced Bionics device. All had CI-aided thresholds better than 35 dB HL for the frequencies 0.25 to 4 kHz, measured by audiologists within the 12 months prior to testing. All had at least 9 months of experience

using their implants. Six participants had bilateral implants, 14 used a right implant, and 10 used a left implant. A hearing aid was worn on the ear contralateral to the CI by 14 participants. All participants used an Advanced Combined Encoder speech processing strategy except for the Advanced Bionics user who used a HiRes Fidelity 120 processing strategy. Participants wore devices using their everyday mode during testing and were instructed not to change device settings throughout the experiment. Participants with CIs underwent unaided audiometry prior to testing to assess their residual hearing in each ear.

Twenty NH participants were tested as a control group. They were matched closely to age and gender of the first 20 listeners enrolled with CIs. The control participants were evaluated for NH, measured at the time of testing, defined as four-tone (0.5, 1, 2, and 4 kHz) pure-tone average (PTA) thresholds of better than 25 dB HL in the better hearing ear. Recognizing that this criterion might be difficult to achieve for some participants, it was relaxed to 30 dB HL PTA for those older than 60 years, of which only one had a better-ear PTA worse than 25 dB HL. Control participants were identified from a pool of patients with nonotologic complaints in the Department of Otolaryngology, as well as recruited using ResearchMatch, a national research recruitment database.

All participants underwent a screening test of cognitive function to determine eligibility for data inclusion in analyses. The Mini-Mental State Examination (MMSE) is a validated screening assessment of memory, attention, and ability to follow instructions, used here to rule out cognitive impairment (Folstein et al. 1975). Raw scores were converted to *T* scores, based on age and education, with a *T* score less than 29 being

suggestive of cognitive impairment (Folstein et al. 1975). A *T* score of 50 represents the mean, with a difference of 10 from the mean indicating 1 SD. Thus, a score of less than 29 is more than 2 SDs below the mean; participants with *T* scores less than 29 were excluded from data analyses.

All participants were adults whose first language was American English and who had graduated from high school, except for one NH participant who only finished 11th grade. Socioeconomic status (SES) was assessed because it may predict language abilities. The SES was quantified using a metric defined by Nittrouer and Burton (2005), which indexes the occupational status and educational level using two eight-point scales between 1 and 8. Scores of 8 were the highest occupation and educational level that could be achieved. The two scores were then multiplied, resulting in SES scores between 1 and 64. The data regarding demographics, audiologic testing, and SES are shown for the 28 CI participants whose data were included in analyses in Table 1.

### Equipment and Materials

All stimuli were presented via computer in a soundproof booth or acoustically insulated testing room. Labeling and discrimination stimuli were presented via computer over a loudspeaker positioned 1 m from the participant at zero-degree azimuth. Scoring for the auditory discrimination and speech labeling was done at the time of testing by the experimenter by directly entering responses into the computer. Along with the screening measures (audiometry and the MMSE cognitive

screen), and the measures of perceptual attention, auditory sensitivity, and word recognition, additional measures of word-reading ability (using the Wide Range Achievement Test), expressive vocabulary (using the Expressive One-Word Picture Vocabulary Test, 4th edition), and receptive vocabulary (using the Receptive One-Word Picture Vocabulary Test, 4th edition) were collected. These additional measures served as metrics of overall language proficiency and were collected as potential predictors of word recognition. For these tasks, responses were video- and audio-recorded; participants wore FM transmitters in specially designed vests that transmitted speech signals to the receivers, which provided direct input to the camera. Responses for these tasks were scored at a later time. In this way, two staff members could independently score responses to check reliability. All participants were tested wearing their usual auditory prostheses (single CI, two CIs, CI plus hearing aid, or none for the controls), which were checked at the start of testing.

### Stimuli

Two sets of stimuli (cop versus cob and sa versus sha) were selected for use in the speech labeling tasks, along with three sets of stimuli that were relevant nonspeech analogs of cues in the labeling tasks for use in the auditory discrimination tasks. All stimuli were presented at a 20-kHz sampling rate with 16-bit digitization.

**Cop-Cob Stimuli for Labeling** • Natural tokens of cop and cob spoken by a male talker were used and were derived from

**TABLE 1. Cochlear implant participant demographics**

Participant	Gender	Age (yr)	Implantation Age (yr)	SES	Side of Implant	Hearing Aid	Etiology of Hearing Loss	Better-Ear PTA (dB HL)
1	Female	62	54	24	Bilateral	No	Genetic	105
2	Female	64	62	35	Right	Yes	Genetic, progressive as adult	75
3	Male	64	61	18	Left	No	Noise, Meniere	80
4	Female	64	58	15	Right	Yes	Genetic, progressive as adult	105
5	Female	52	47	12	Left	Yes	Progressive as adult, sudden	105
6	Male	67	65	24	Right	No	Genetic, progressive as adult	84
7	Male	56	52	30	Bilateral	No	Rubella, progressive	105
8	Female	54	48	16	Right	Yes	Genetic, progressive	105
9	Male	77	67	49	Left	No	Genetic, progressive	93
10	Male	77	76	48	Right	Yes	Progressive as adult, noise, sudden	71
11	Male	88	83	30	Right	Yes	Progressive as adult	88
12	Female	66	56	12	Bilateral	No	Otosclerosis, progressive as adult	105
13	Male	52	50	30	Bilateral	No	Progressive as adult	105
16	Female	61	59	35	Right	No	Progressive as adult	105
18	Female	75	63	9	Right	No	Genetic, progressive as adult	95
19	Female	73	67	36	Left	No	Genetic, autoimmune	105
20	Male	76	74	25	Left	No	Ear infections	105
21	Male	80	58	30	Left	Yes	Meniere	69
22	Female	80	78	12	Right	Yes	Progressive as adult	65
23	Female	79	73	30	Right	No	Progressive as adult	86
24	Female	58	53	35	Bilateral	No	Progressive as adult	105
25	Male	57	56	30	Right	Yes	Autoimmune, sudden	76
26	Male	53	50	12	Bilateral	No	Noise, progressive as adult	98
29	Female	59	58	16	Right	No	Sudden hearing loss	80
30	Male	80	79	36	Right	Yes	Progressive as adult	66
31	Female	66	62	16	Left	Yes	Progressive as child and adult	84
32	Male	68	67	9	Left	Yes	Progressive as adult	73
34	Male	59	54	35	Left	Yes	Meniere, noise	81

PTA, unaided four-tone pure-tone average at 0.5, 1, 2, and 4 kHz; SES, socioeconomic status.

Nittrouer (2004). Three tokens of each word were used so that any natural but irrelevant variability in acoustic structure (e.g., intonation contour) could not cue category labeling. In each token, the release burst and any voicing during closure for the final stop (/p/ or /b/) were deleted, thus leaving only two acoustic cues for categorization: vocalic length and offset frequency of the first formant (F1). Vocalic length was manipulated either by reiterating one pitch period at a time from the most stable spectral region of the syllable (to lengthen syllables) or by deleting pitch periods from the most stable spectral region of the syllable (to shorten syllables). Seven stimuli varying in vocalic duration from roughly 82 to 265 msec were created from each token. These endpoint values were selected because they matched the mean lengths of the natural tokens. This cue (vocalic duration) was the continuously varied cue in this experiment. Going into this experiment, it was presumed to be represented well by CI processing. Final formant transitions were not modified. For the tokens used, the first formant (F1) at offset had a mean frequency of 625 Hz for cob and 801 Hz for cop. It was presumed that this cue might be difficult for CI users to recognize. In all, there were 42 cop–cob stimuli: 2 Formant transitions  $\times$  7 Lengths  $\times$  3 Tokens of each.

**Sa–Sha Stimuli for Labeling** • The other set of labeling stimuli involved a fricative-place contrast (sa–sha), originally created by Nittrouer and Miller (1997). Fourteen hybrid stimuli made up of seven natural fricative noises and two synthetic vocalic portions were used. The seven noises were created by combining natural /s/ and /ʃ/ noise. The noises were 100-msec sections, excised from natural /s/ and /ʃ/ tokens produced by a male talker. These sections of fricative noise were combined using seven /ʃ/-to-/s/ amplitude ratios, creating a continuum of noises from one that was most strongly /ʃ/-like to one that was most strongly /s/-like. This method of stimulus creation preserved details of fricative spectra, while varying the amplitude of noise in the F3 region of the following vocalic portion, a primary cue to sibilant place (Stevens 1960; Perkell et al. 1979; Stevens 1985; Hedrick & Ohde 1993; Henry & Turner 2003). This cue was the continuously varied one in this set of stimuli. Although spectral in nature, it was thought that it might be well-preserved in CI processing because it is both relatively steady state and high frequency. For the vowels in these stimuli, two synthetic sections were generated using Sensyn, a version of the Klatt synthesizer. These vocalic portions were appropriate for /a/ vowels, following the production of either /s/ or /ʃ/; that is, onset F2 and F3 varied appropriately. This cue of onset formant frequencies was the binary cue in this set of stimuli. This cue would presumably be difficult to recognize with a CI.

**Duration Stimuli for Discrimination** • Eleven nonspeech stimuli were used in the discrimination task, varying in length from 110 to 310 msec in 20-msec steps. The stimuli consisted of three steady-state sinusoids of equal amplitude of the frequencies 650, 1130, and 2600 Hz, representing typical steady-state formant values in cop or cob. These stimuli will be referred to as “dur” stimuli.

**Dynamic Spectral Stimuli for Discrimination** • Two sets of 13 dynamic spectral glide stimuli were used. These stimuli were all 150 msec long, consisting of the same steady-state sinusoids as the dur stimuli. For the first set of 13 stimuli, the last 50 msec of 12 of these stimuli had falling glides; these will be termed as the “glide-fall” stimuli. These stimuli replicated the falling

formants of the cop–cob stimuli. In each step on the continuum, the offset frequencies of the sine waves changed by 20 Hz. As an example, the lowest tone varied between 650 and 410 Hz at offset between the 1st and 13th stimulus. The other set of dynamic spectral stimuli consisted of the reverse of the glide-fall stimuli, representing rising formants, with the rising portion occurring during the first 50 msec. These stimuli will be termed as the “glide-rise” stimuli.

## General Procedures

All testing took place at the Eye and Ear Institute of the Ohio State University Wexner Medical Center. Approval was obtained from the Institutional Review Board of the Ohio State University, and informed written consent was obtained. Participants were tested during a single 2-hr session. Auditory discrimination and speech labeling measures were collected first, followed by word reading, expressive vocabulary, MMSE, receptive vocabulary, and word recognition. Auditory discrimination and labeling tasks were alternated, with order of task assigned consecutively as participants enrolled.

## Task-Specific Procedures

**Speech Labeling** • Participants were introduced to response labels and corresponding pictures for either the cop–cob or the sa–sha contrast. To visually reinforce task training and to ensure accurate recording of responses by the tester, 8 in  $\times$  8 in hand-drawn pictures were used. In the case of the cop–cob stimuli, a picture of a police officer served as the response target for cop and a picture of a corn cob served as the response target for cob. For the sa–sha stimuli, a picture of a creature from outer space served as the target for sa and a king from another country served as the target for sha.

Natural versions of cop and cob or of sa and sha were presented over the speaker, and participants were asked to label verbally what they heard, as well as to point to the corresponding picture. Training tasks were performed immediately before testing. Ten digitized natural samples were presented over the speaker to give participants practice performing the task in that format. The final practice task was that participants heard “best exemplars” of the stimulus set: that is, the endpoint stimuli along the continuum combined with the appropriate binary cues. For example, the best exemplar of sha was the stimulus with the most /ʃ/-like fricative noise and the vocalic portion that was appropriate for a preceding /ʃ/. Ten trials of these best exemplars (five of each) were presented once to participants with feedback. Next, 10 of these stimuli were presented, without feedback. Participants needed to respond to 8 of the 10 correctly to proceed to testing. They were given three chances to meet this criterion.

During testing, stimuli were presented in 10 blocks of the 14 stimuli for either cop–cob or sa–sha. Participants needed to respond correctly to 70% of the best exemplars during testing to have their data included in statistical analyses.

**Auditory Discrimination** • An AX procedure was used. In this procedure, participants judge one stimulus, which varies across trials (X), in comparison to a constant standard (A). The constant “A” stimulus serves as the category exemplar. For the dur stimuli, the “A” stimulus was the shortest member of the continuum. For both the glide-fall and glide-rise stimuli, the “A” stimulus was the one without transitions. A cardboard

response card was used to visually reinforce task training and to ensure accurate recording of responses by the tester. One half of the card showed two black squares, and the other half showed one black square and one red circle. The participant responded by pointing to the picture of the two black squares and saying “same” if the stimuli were judged as the same and by pointing to the picture of the black square and the red circle and saying “different” if the stimuli were judged as different. The experimenter controlled stimulus presentation and entered responses by keyboard. Stimuli were presented in 10 blocks of 11 (dur) or 13 (glide-fall and glide-rise) stimuli.

Before testing in each condition, participants were presented with five pairs of stimuli that were identical and five pairs of stimuli that were maximally different, in random order. They were asked to report whether the stimuli were same or different and were given feedback. Next, these same training stimuli were presented, and participants were asked to report if they were same or different, without feedback. Participants needed to respond correctly to 80% of training trials without feedback to proceed to testing. During testing, participants needed to respond correctly to at least 70% of same and maximally different stimuli to have their data included in analyses.

**Word Reading** • The Wide Range Achievement Test 4 (Wilkinson & Robertson 2006) is a standardized measure of reading skills. The participant was asked to read the words presented on a single page. Participants were scored for words correctly read aloud. All participants exhibited at least a standard score of 70 on the Wide Range Achievement Test 4 (better than 2 SDs below the mean).

**Expressive Vocabulary** • The Expressive One-Word Picture Vocabulary Test, 4th edition (Martin & Brownell 2011a) was presented next. For each item, the participant was shown a drawing and the tester asked, “What is this?” or “What is he/she doing?” depending on the picture. The participant was asked to give a single word response corresponding to what the picture represented. Testing was discontinued when the participant made six consecutive incorrect responses.

**Receptive Vocabulary** • The Receptive One-Word Picture Vocabulary Test, 4th edition (Martin & Brownell 2011b) was presented next. For each item, the participant was shown four pictures on an easel. The examiner said a word, and the participant responded by repeating the word and then stating the number (1 through 4) associated with the picture that represented that word. If the participant did not repeat the word correctly, the examiner would say the word until it was repeated correctly by the participant prior to selecting a picture. This was done to ensure that participants understood the word that was being presented. Testing was discontinued when the participant made six consecutive incorrect responses.

**Mini-Mental State Examination** • The MMSE tasks were presented in a written fashion whereby the participant read the instructions, and the tester read along out loud to ensure that the participants understood the tasks. This type of modification of the MMSE, having a version for the participants to read, has been used previously in individuals with hearing loss (De Silva et al. 2008).

**Word Recognition** • The Central Institute for the Deaf 22 Word Lists were used to obtain a measure of word recognition (Hirsh et al. 1952). All words on these lists are monosyllabic and commonly occur in the vocabularies of most speakers of

American English. The phonetic composition within each list is balanced, so the frequency of occurrence of these phonemes matches the frequency in English.

The word lists were presented via a loudspeaker at zero-degree azimuth. Each participant heard the same 50-word list. Participants were later scored on a phoneme-by-phoneme basis. All phonemes in a single word needed to be correct with no intervening phonemes in order for that word to be scored as correct. Both phoneme and whole-word scores were recorded, and percent correct whole-word recognition scores were used as dependent measures for the participants with CIs.

## Analyses

**Computation of Weighting Factors From Labeling Tasks** • To calculate measures of the perceptual weights for the duration versus spectral cues (for cop–cob) or the static spectral versus dynamic spectral cues (for sa–sha), logistic regressions were performed. The proportion of /cob/ responses given to each stimulus on each continuum (or the proportion of /sa/ responses) was used in the computation of these weighting factors. Values of the continuously set property and the binary set property served as the independent variables. The raw regression coefficient derived for each cue was used as the weighting factor. Larger weighting factors indicate that more perceptual attention was given to that cue.

**Computation of  $d'$  Values From Discrimination Tasks** • The discrimination functions of each listener were used to compute an average  $d'$  value for each condition (Holt & Carney 2005; Macmillan & Creelman 2005). The  $d'$  value was selected as the discrimination measure because it is bias-free. The  $d'$  value was calculated at each step along the continuum as the difference between the  $z$  value for the “hit” rate (proportion of different responses when A and X stimuli were different) and the  $z$  value for the “false alarm” rate (proportion of different responses when A and X stimuli were identical). A hit rate of 1.0 and a false alarm rate of 0.0 require a correction in the calculation of  $d'$  and were assigned values of 0.99 and 0.01, respectively. A value of zero for  $d'$  means participants cannot discriminate the difference between stimuli at all. A positive  $d'$  value suggests the hit rate was greater than the false alarm rate. Using the above correction values, the minimum  $d'$  value would be 0, and the maximum  $d'$  value would be 4.65. The average  $d'$  value was then calculated across all steps of the continuum, and this value was used in analyses.

**Examination for Group Differences** • A series of independent-samples  $t$  tests were performed to examine whether group mean differences existed for weighting factors and  $d'$  values between the NH and CI groups.

**Search for Predictor Variables** • A series of separate linear regression analyses were performed to uncover the nature of relationships among various demographic factors, predictor variables, and the dependent measure of word recognition.

## RESULTS

Thirty participants with CIs and 20 NH controls underwent testing. Two participants from the CI group and one from the NH group obtained a  $T$  score on the MMSE of less than 29 (suggesting cognitive impairment); thus, data from these three participants were excluded from analyses. Demographic factor means, predictor variable test scores,

and word recognition scores for the remaining 28 CI and 19 NH participants are shown in Table 2. Independent-samples *t* tests were performed first to examine whether differences existed between the CI and NH groups on demographic measures of age and SES, as well as reading ability, expressive or receptive vocabulary, and MMSE scores; no differences were found. Only word recognition, shown in Table 3, was significantly different.

For the CI group, one-way analysis of variance tests were then performed to see if side of implant (right, left, or bilateral) influenced any scores for weighting factors, *d'* values, or word recognition. No differences were found for any measures based on whether participants used a right-sided CI, a left-sided CI, or two CIs. Additionally, using independent-samples *t* tests, no differences were found on the same measures between those who wore only CIs versus a CI plus hearing aid. Therefore, data were combined across all CI participants who passed testing criteria in subsequent analyses.

### Reliability

An estimate of interscorer reliability was obtained for the tests that required audiovisual recording and later scoring of responses. Responses were reviewed by two independent scorers for 25% of all participants (13 participants). Mean percent agreement in scores for the two scorers across these participants ranged from 90 to 99% for the measures of word reading, expressive and receptive vocabulary, and word recognition. These outcomes were considered good reliability, and the scores from the staff members who initially scored all the samples were used in further analyses.

### Potential Predictors of Variance in Word Recognition: Audiologic History, Word-Reading, Vocabulary, and Cognitive Status

Before examining perceptual attention and auditory sensitivity, as well as their abilities to explain variance in word recognition, it was important to examine other potential predictors of variance in word recognition. It has previously been found that age at implantation, duration of hearing loss, and amount of residual hearing can serve as indirect predictors of some of the variance in speech recognition outcomes, so evidence was sought for the abilities of these variables to predict variance in word recognition in this group of adult CI users.

If these factors did predict variance in word recognition, the significant predictors would need to be accounted for when seeking predictive power for the test variables of interest. Separate linear regression analyses were performed with percent correct word recognition score as the dependent measure and demographic and audiologic measures as predictors, including age, age at onset of hearing loss, duration of hearing loss (computed as age minus age at onset of hearing loss), age at implantation, and better-ear residual PTA. None of these variables predicted a significant proportion of variability in word recognition.

In addition to these demographic and audiologic factors, language and cognitive factors could potentially explain variance in word recognition. Thus, separate linear regression analyses were performed with word recognition as the dependent measure. Predictor variables were word reading, expressive or receptive vocabulary (standard scores), or cognitive score (MMSE *T* score). None of these variables evaluated predicted a significant amount of variance in word recognition. Therefore, these demographic, audiologic, language, and cognitive factors were not considered during further analyses of predictors of variance in word recognition.

### Perceptual Attention (Weighting Factors)

The first question of interest was whether CI users and NH listeners demonstrated similar cue weighting for the cop–cob and sa–sha contrasts. Table 3 shows means for the CI group and the NH group for weighting factors for the two contrasts. For the cop–cob labeling task, one NH participant and five CI participants did not meet endpoint criteria, either at the level of training (criterion of 80% correct responses to the endpoint stimuli) or testing (criterion of 70% correct responses to endpoints). Inspection of weighting factors for the remaining participants and independent-samples *t* tests show that these factors were similar for participants with NH and those with CIs. Thus, the majority of CI users were able, as predicted, to make use of the duration cue for the cop–cob contrast, and they weighted this cue as heavily as NH participants. In this case, these CI users were also able to make use of the spectral cue to the same extent as the participants with NH. No significant correlation was found between attention to the duration cue and attention to the spectral cue for either the NH or CI group, suggesting that greater attention to one cue was not associated with less attention to the other.

**TABLE 2. Means and SDs of demographics, as well as language and cognitive test scores, for NH and CI groups**

	Groups		<i>t</i>	<i>p</i>
	NH	CI		
n	19	28		
Demographics, mean (SD)				
Age (yr)	63.7 (8.1)	66.7 (10.2)	1.07	0.29
SES (score)	32.8 (18.7)	25.3 (11.3)	1.71	0.09
Test scores, mean (SD)				
Reading (standard score)	104.1 (11.8)	99.6 (10.7)	1.30	0.20
Expressive vocabulary (standard score)	102.9 (19.6)	96.0 (15.3)	1.30	0.20
Receptive vocabulary (standard score)	112.9 (18.9)	106.7 (18.2)	1.07	0.29
Cognitive MMSE ( <i>T</i> score)	51.1 (8.2)	47.3 (8.5)	1.54	0.13

The *t* and *p* value columns show results of *t* test, comparing means between NH and CI groups. Degrees of freedom for all *t* tests were 45. CI, cochlear implant; MMSE, Mini-Mental State Examination; NH, normal hearing; SES, socioeconomic status.

**TABLE 3. Means and SDs of word recognition, perceptual attention weighting factors, and auditory sensitivity  $d'$  values for NH and CI groups**

	Groups		<i>t</i>	<i>p</i>
	NH	CI		
	Mean (SD)	Mean (SD)		
Test scores				
n	19	28		
Word recognition (percent correct)	97.1 (2.5)	66.5 (18.7)	7.05	<b>&lt;0.001</b>
Perceptual attention				
n	18	23		
Cop–cob spectral cue (weighting factor)	0.4 (1.0)	0.4 (0.8)	0.04	0.97
Cop–cob duration cue (weighting factor)	7.2 (3.7)	7.9 (3.4)	0.59	0.56
n	18	13		
Sa–sha static spectral cue (weighting factor)	7.9 (3.5)	5.6 (3.9)	1.77	0.087
Sa–sha dynamic spectral cue (weighting factor)	7.5 (3.0)	1.6 (1.4)	2.85	<b>0.008</b>
Auditory sensitivity				
n	19	15		
Glide fall ( $d'$ value)	3.7 (0.4)	2.5 (0.9)	4.81	<b>&lt;0.001</b>
n	19	12		
Glide rise ( $d'$ value)	3.1 (0.5)	2.6 (1.0)	1.51	0.14
n	18	27		
Dur ( $d'$ value)	2.2 (0.7)	2.4 (0.6)	1.14	0.26

The *t* and *p* value columns show results of *t* test, comparing means between NH and CI groups. Degrees of freedom for all *t* tests were *N*–2. Results are in bold where *p* < 0.05. CI, cochlear implant; NH, normal hearing.

Turning to the sa–sha labeling task, all NH participants met criteria for this task. However, 15 of 28 participants with CIs were unable to meet criteria for inclusion in data analyses. For the 13 CI users who did meet criteria on the sa–sha labeling task, weighting factors for the dynamic spectral cue were smaller than for the NH group, indicating poorer attention to this cue. There was a trend toward smaller weighting factors for the CI group as compared with the NH group for the static spectral cue, but a statistically significant difference was not found.

Because so many CI participants were unable to perform the sa–sha labeling task, *t* tests were performed to examine potential differences in demographics, audiologic factors, and test measures between the group of CI users who did and the group who did not meet criteria for sa–sha testing. Factors examined were SES, age, age at onset of hearing loss, age at implantation, duration of deafness, better-ear residual PTA, word-reading ability, expressive and receptive vocabulary, and MMSE score. No differences were found.

### Perceptual Attention and Word Recognition

The next question of interest, which addressed the first main hypothesis of the study, concerned how weighting strategies affected word recognition. Using the weighting factors derived from the cop–cob and sa–sha labeling tasks, it was asked if these weighting factors explained significant proportions of variance in word recognition scores for this group of adult CI users. Scatter plots for word recognition versus the weighting factors for the cop–cob contrast for individual CI users are shown in Figure 3A, B. Linear regression analyses were performed separately for each weighting factor, using word recognition as the dependent measure. The only weighting factor that predicted a significant amount of variance in word recognition was the weighting factor for the dynamic spectral cue in the cop–cob contrast ( $\beta = 0.54$ ,  $p = 0.007$ ). This finding suggests that the

ability of CI users to pay attention to spectral structure, even if it is not the structure they weight most heavily in the contrast, does play a role in word recognition ability.

On the other hand, the weighting factor for neither the static nor the dynamic spectral cue in the sa–sha contrast predicted word recognition (Fig. 3C, D). However, a potential problem with these regression analyses was that data needed to be excluded from the 15 CI users who did not meet criteria on the sa–sha task. With only 13 participants left in the analysis, there were simply too few to find an effect, if one existed. Therefore, two-sample *t* tests were performed comparing word recognition between the group of CI users who met criteria on the sa–sha task and the group who did not. The mean word recognition score was better for the group who met criteria (73.5% versus 60.4%), but this difference was not quite statistically significant [ $t(26) = 1.95$ ,  $p = 0.06$ ], probably due to the small sample sizes examined in this analysis. These results of weighting factors for the cop–cob and sa–sha contrasts suggest that the ability to pay attention to spectral cues is important in word recognition.

### Auditory Sensitivity to Duration and Spectral Cues

The subsequent question asked whether cue discrimination of duration (dur) and dynamic spectral cues (glide fall and glide rise) would predict variance in word recognition abilities. Twenty-seven of the 28 CI users showed sufficient discrimination of the dur cue to meet criteria for inclusion in analyses, but, overall, CI users showed greater difficulty meeting criteria for data inclusion for discrimination of the dynamic spectral cues, glide fall (13 CI users did not meet criteria) and glide rise (16 CI users did not meet criteria). All NH participants were able to complete the spectral discrimination tasks, and only one NH participant could not complete the duration discrimination task. Table 3 shows  $d'$  values from the discrimination tasks for the NH group and the CI group. For the dur cue, the groups showed

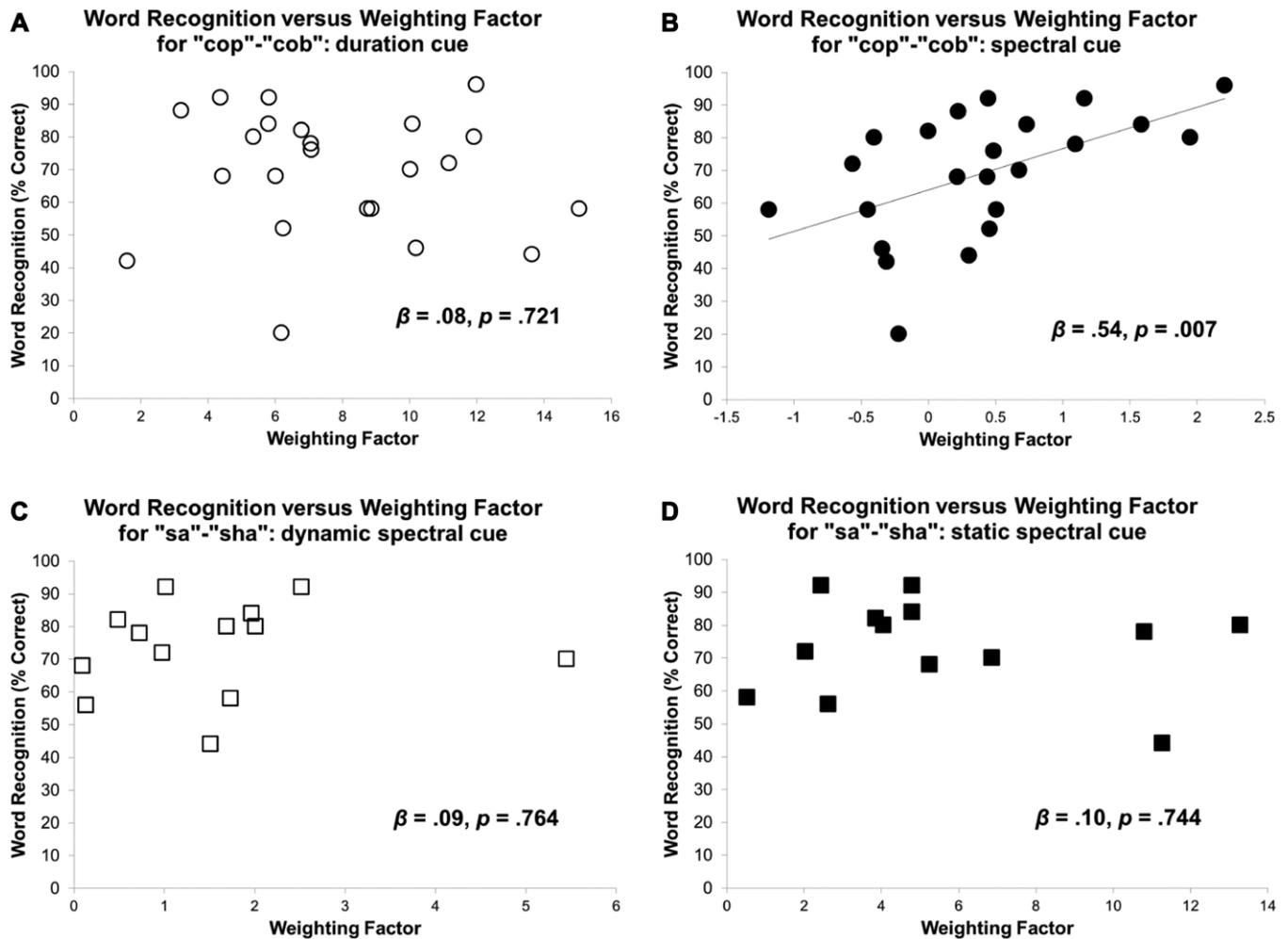


Fig. 3. Scatterplot of word recognition vs. weighting factor for individual CI users for the cop–cob contrast for the duration cue (A) and the spectral cue (B) and for the sa–sha contrast for the dynamic spectral cue (C) and the static spectral cue (D).

no significant difference in  $d'$  values and, thus, no difference in auditory sensitivity. On the other hand, even when excluding those individuals who did not meet criteria for data inclusion, the remaining 15 CI users showed poorer sensitivity than the NH group to the spectral cue glide-fall. A similar trend was found for the glide-rise cue, but only 12 CI users met criteria.

For the CI participants,  $t$  tests were again performed to examine differences in demographics, audiologic factors, and test measures between those who did or did not meet criteria for glide fall and glide rise. Factors examined were SES, age, age at onset of hearing loss, age at implantation, duration of deafness, better-ear PTA, word-reading ability, expressive and receptive vocabulary, and MMSE score. None of these differed for participants who did or did not meet criteria during the glide discrimination tasks.

### Auditory Sensitivity and Word Recognition

The second main hypothesis of the study was that auditory sensitivity to the cues underlying phonemic contrasts would not independently predict word recognition. Figure 4A, B shows scatter plots of word recognition versus  $d'$  value for the glide-fall and glide-rise cues, respectively, and Figure 4C shows word recognition versus  $d'$  value for the dur cue. In none of these

cases did the  $d'$  value predict word recognition, though there was a nearly significant relationship between sensitivity to the dur cue and word recognition. However, similar to the problem with relating sa–sha cue weighting to word recognition, the finding of a large number of CI users who were unable to meet task criteria for glide-fall and glide-rise limited the ability to examine the effect of spectral discrimination on word recognition. Therefore, mean word recognition scores were compared between the group who met criteria for glide fall (15 of 28 participants) versus the group who did not, and the group who met criteria showed significantly higher word recognition (74.1% versus 57.7%) [ $t(26) = 2.54, p = 0.017$ ]. Similarly, the group of 12 CI users who met criteria for the glide-rise task showed higher word recognition than the group of CI users who did not (80% versus 56.4%) [ $t(26) = 4.21, p < 0.001$ ]. Importantly, however, discrimination ability for the dur cue did not explain variance in word recognition. Thus, for the cues in this experiment, sensitivity to dynamic spectral cues did appear to relate to word recognition, whereas sensitivity to the duration cue did not.

### Perceptual Attention and Auditory Sensitivity

The third main hypothesis of the study was that auditory sensitivity would not independently predict perceptual attention.

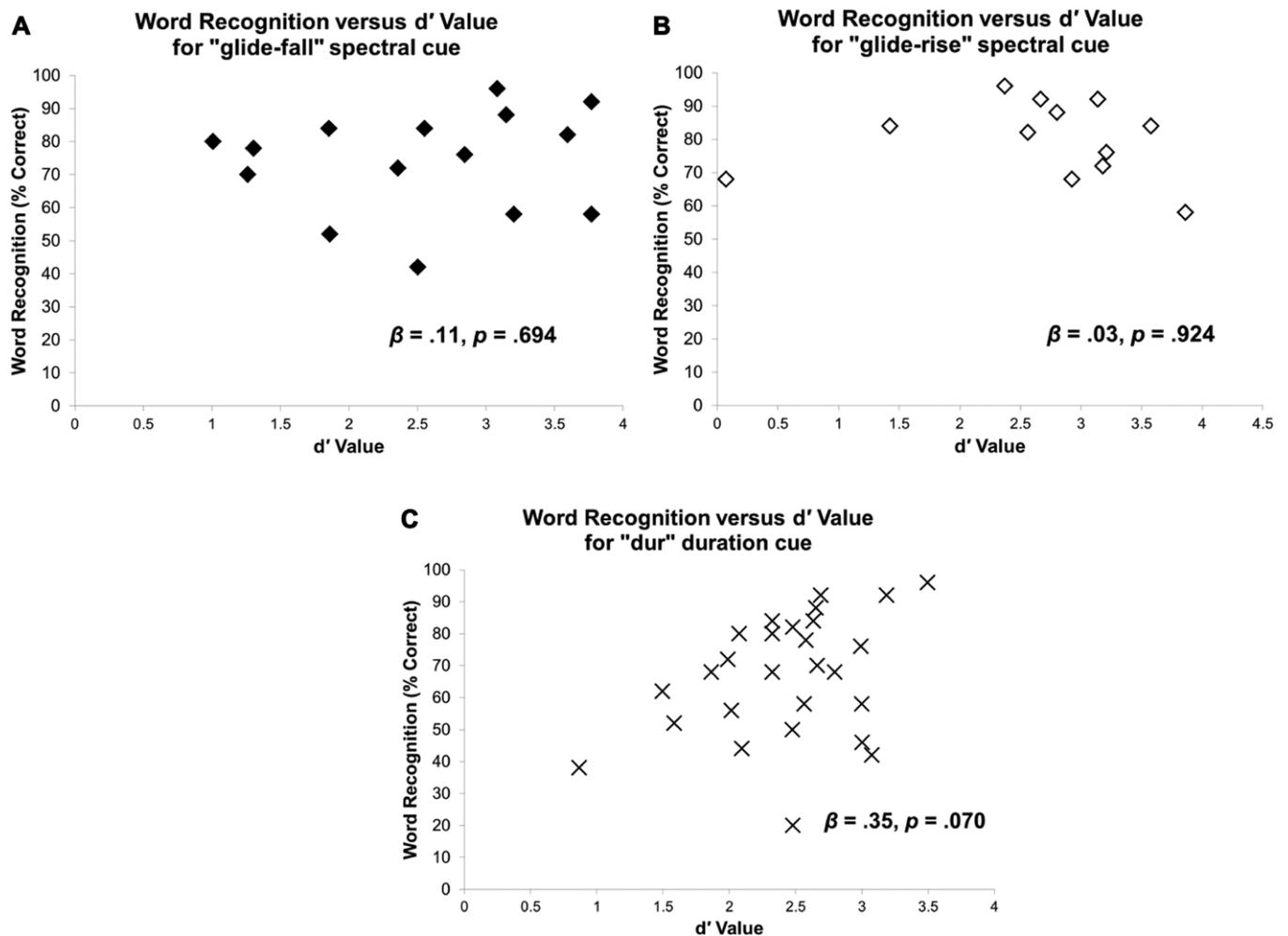


Fig. 4. Scatterplot of word recognition vs. discrimination ( $d'$  value) for individual CI users for the glide-fall cue (A), the glide-rise cue (B), and the dur cue (C).

To test this hypothesis, linear regressions were performed with weighting factors for the cop–cob contrast and the sa–sha contrast as the dependent measures and  $d'$  values for the dur, glide-fall, and glide-rise cues as predictors, for both the NH and CI groups independently. These analyses failed to find any  $d'$  values that served as significant predictors of weighting factors for either group. Thus, findings support the hypothesis that perceptual attention during labeling of a phonemic contrast does not depend on auditory sensitivity to acoustic cues.

## DISCUSSION

The experiment presented here was conducted to examine factors that might predict variability in word recognition in quiet for postlingually deafened adults using CIs. Measures were collected that are thought to underlie robust spoken language recognition. These included metrics of the amount of perceptual attention paid to acoustic cues in phonemic labeling and auditory sensitivity to those cues. Measures were collected and compared between a group of experienced CI users and a group of NH peers. Three hypotheses were tested: (1) that the use of perceptual attention strategies like those used by NH listeners while labeling phonemic contrasts would predict word recognition ability, (2) that auditory sensitivity to spectral and

duration cues would not predict word recognition, and (3) that auditory sensitivity to those same cues would not predict perceptual attention strategies. These three hypotheses were tested by measuring word recognition, along with labeling responses for the cop–cob and sa–sha phonemic contrasts, and discrimination of nonspeech spectral and duration cues.

The study by Moberly et al. (2014) examined whether two distinct cues to the ba–wa contrast, a spectral cue and an amplitude cue, were equal in their capacity to elicit accurate speech recognition. It was found that they were not equal: individuals who attended to the spectral cue strongly, like NH adults, showed the best word recognition. On the other hand, auditory sensitivity to the spectral cue in that study did not independently explain variance in word recognition. The present study extends the findings from that study by examining CI users' perception of two other phonemic contrasts. For the cop–cob contrast, adults with NH typically weight the duration cue over the dynamic spectral cue. It was expected that CI users would have relatively intact auditory sensitivity to the duration cue, which was confirmed. The first hypothesis was that stronger weighting of the duration cue, like NH adults, would predict better word recognition; however, this was not found, which suggests that the ability to weight this temporal cue, duration, was not sufficient to explain variability in word recognition. On the other hand, the ability to pay attention

to the dynamic spectral cue in the cop–cob contrast did predict approximately 25% of the variance in word recognition. For the sa–sha contrast, the phonemic decision can be made based on a dynamic spectral cue or a static spectral cue. This labeling task was found to be too difficult for many of the CI users, with only about half of them able to pass criteria for inclusion of their data in analyses. However, a trend was seen for better word recognition for those who could successfully complete the sa–sha labeling task as compared with those who could not.

The second notable finding from this study is related to auditory sensitivity for CI users to two types of cues that underlie speech perception, a duration cue and dynamic spectral cues. As expected, CI users showed intact sensitivity for the duration cue, which should be represented robustly by CIs, but overall poor sensitivity for the dynamic spectral cues, which would likely be represented poorly by CIs. With regard to the second hypothesis, that auditory sensitivity to the nonspeech cues would not predict word recognition, this hypothesis was partially supported. Sensitivity to the duration cue did not predict word recognition. The relationship between word recognition and sensitivity to the dynamic spectral cues was less clear, as several CI users could not complete those discrimination tasks. However, those participants who could complete the spectral discrimination tasks had significantly higher word recognition scores than those who could not. These findings support the idea that auditory sensitivity to dynamic spectral structure, especially spectral changes in a relatively low-frequency range, does appear to play an important, though incomplete, role in the ability to recognize speech through a CI. Sensitivity to these spectral changes is not typically assessed in either the research or clinical setting, but results suggest that measuring sensitivity to these dynamic cues may be important. The third hypothesis tested in this study examined the relationship between perceptual attention strategies and auditory sensitivity to the types of acoustic structure that underlie phonemic decisions. Interestingly, no evidence was found to support the idea that better auditory sensitivity resulted in participants directing their perceptual attention to those types of cues during phonemic labeling. This finding replicates findings for 51 children with CIs listening to the same discrimination and labeling stimuli used in this study: sensitivity to acoustic cues did not strongly predict the weight assigned to cues in those decisions (Nittrouer et al. 2014). Where both adults and children are concerned, some participants were able to attend to cues that were less salient through CIs, suggesting that factors other than simple cue saliency account for how perceptual attention is directed. Instead, individual variability in this kind of attention must play a role, and this conclusion suggests that auditory training could be valuable post implantation.

This study more broadly expands findings from previous studies of cue weighting by CI users. In agreement with the studies by Hedrick and Carney (1997) and Winn et al. (2012), CI listeners weighted amplitude or durational cues more heavily than spectral cues when making labeling decisions for phonemic categories because these cues were either more informative for making phonemic decisions or more accessible through their implants. Nonetheless, it has been found that attention to dynamic spectral structure is what predicts word recognition by CI users (Kirk et al. 1992). Finally, Donaldson et al. (2013) attributed success in vowel identification to better access to both spectral and durational cues, while accepting that cue-weighting strategies may have influenced results.

Perhaps, one of the most striking findings of this study was that approximately half of the CI users could not even complete the labeling or discrimination tasks that involved spectral cues. These results are likely (although not conclusively) representative of the abilities of adult CI users to access and attend to spectral structure more generally. This is important and should be eye-opening for clinicians because spectral structure in the speech signal supports the perception of many phonemic contrasts, and a large number of CI users simply are not sensitive to and cannot attend to that structure, which is essential for the most accurate word recognition. Moreover, no demographic, audiologic, language, or cognitive measures were found to differentiate the groups who could or could not access or attend to spectral structure. These findings suggest the need for clinical tools that could assess both sensitivity and attention to spectral structure to prognosticate outcomes and to assess patients postoperatively.

### Limitations and Future Directions

In spite of the useful results reported here, one principal problem constrained interpretation. The spectral discrimination and sa–sha labeling tasks were found to be too difficult for many of the CI users to meet criteria for data inclusion in analyses, limiting the numbers of participants whose data could be included in exploring sources of variance in word recognition. Nonetheless, the fact that this basic contrast of the English language proved so difficult for many adults with CIs should prompt further investigation into their sensitivity and attention to spectral structure in general.

## CONCLUSIONS

Variability in outcomes for patients who undergo cochlear implantation is frustrating for patients and clinicians alike. Results from both this study and the study by Moberly et al. (2014) support the idea that being able to perceptually attend to spectral structure within speech may explain some of this variability and that attention to this spectral structure is critical to accurate word recognition for adult CI users. Even if the spectral structure delivered through CIs is degraded, many implant users may still have sufficient sensitivity to spectral cues to be able to pay attention to those cues. On the other hand, the factors clinicians tend to focus on in predicting speech recognition outcomes for patients with CIs, such as duration of deafness and age at implantation, appear to play much less important and less direct roles in predicting outcomes.

The findings of this study provide two important targets for future clinical intervention. Processing and mapping strategies of current CIs should be modified to improve the delivery of spectral structure through implants. Perhaps even more importantly, adults with CIs may benefit from intensive auditory training that facilitates perceptual learning of appropriate weighting strategies for static and dynamic spectral cues, which should optimize their speech perception outcomes.

## ACKNOWLEDGMENTS

The authors acknowledge the following individuals for assistance in data collection and scoring: Jessica Apsley, Lauren Boyce, Emily Hehl, Jennifer Martin, and Demarcus Williams.

This study was supported by the Triological Society Career Development Award (to A.C.M.). In addition, this research was supported by grants R01

DC000633 and R01 DC006237 from the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health (NIH) (to S.N.). ResearchMatch, used to recruit control participants, is supported by grants UL1TR000445 and 1U54RR032646-01 from the NIH Clinical and Translational Science Award program.

The authors declare no other conflict of interest.

Address for correspondence: Aaron C. Moberly, Department of Otolaryngology, Ohio State University College of Medicine, 915 Olentangy River Road, Suite 4000, Columbus, OH 43212, USA. E-mail: aaron.moberly@osumc.edu

Received December 31, 2014; accepted June 23, 2015.

## REFERENCES

- Brownell, R. (2000a). *Expressive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.
- Brownell, R. (2000b). *Receptive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.
- De Silva, M. L., McLaughlin, M. T., Rodrigues, E. J., et al. (2008). A Mini-Mental Status Examination for the hearing impaired. *Age Ageing*, *37*, 593–595.
- Donaldson, G. S., Rogers, C. L., Cardenas, E. S., et al. (2013). Vowel identification by cochlear implant users: Contributions of static and dynamic spectral cues. *J Acoust Soc Am*, *134*, 3021–3028.
- Eggermont, J. J., & Ponton, C. W. (2003). Auditory-evoked potential studies of cortical maturation in normal hearing and implanted children: Correlations with changes in structure and speech perception. *Acta Otolaryngol*, *123*, 249–252.
- Firszt, J. B., Holden, L. K., Skinner, M. W., et al. (2004). Recognition of speech presented at soft to loud levels by adult cochlear implant recipients of three cochlear implant systems. *Ear Hear*, *25*, 375–387.
- Folstein, M. F., Folstein, S. E., McHugh, P. R. (1975). "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*, *12*, 189–198.
- 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.
- Gifford, R. H., Shallop, J. K., Peterson, A. M. (2008). Speech recognition materials and ceiling effects: Considerations for cochlear implant programs. *Audiol Neurootol*, *13*, 193–205.
- Hedrick, M. S., & Carney, A. E. (1997). Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants. *J Speech Lang Hear Res*, *40*, 1445–1557.
- Hedrick, M. S., & Ohde, R. N. (1993). Effect of relative amplitude of friction on perception of place of articulation. *J Acoust Soc Am*, *94*, 2005–2026.
- Henry, B. A., & Turner, C. W. (2003). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *J Acoust Soc Am*, *113*, 2861–2873.
- Hirsh, I. J., Davis, H., Silverman, S. R., et al. (1952). Development of materials for speech audiometry. *J Speech Hear Disord*, *17*, 321–337.
- Holden, L. K., Finley, C. C., Firszt, J. B., et al. (2013). Factors affecting open-set word recognition in adults with cochlear implants. *Ear Hear*, *34*, 342–360.
- Holt, R. F., & Carney, A. E. (2005). Multiple looks in speech sound discrimination in adults. *J Speech Lang Hear Res*, *48*, 922–943.
- Kelly, A. S., Purdy, S. C., Thorne, P. R. (2005). Electrophysiological and speech perception measures of auditory processing in experienced adult cochlear implant users. *Clin Neurophysiol*, *116*, 1235–1246.
- Kirk, K. I., Tye-Murray, N., & Hurtig, R. R. (1992). The use of static and dynamic vowel cues by multichannel cochlear implant users. *J Acoust Soc Am*, *91*, 3487–3498.
- Lazard, D. S., Vincent, C., Venail, F., et al. (2012). Pre-, per- and postoperative factors affecting performance of postlinguistically deaf adults using cochlear implants: A new conceptual model over time. *PLoS One*, *7*, e48739.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., et al. (1967). Perception of the speech code. *Psychol Rev*, *74*, 431–461.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, NJ: Erlbaum.
- Martin, N., & Brownell, R. (2011a). *Expressive One-Word Picture Vocabulary Test*. 4th ed. (EOWPVT). Novato, CA: Academic Therapy Publications.
- Martin, N., & Brownell, R. (2011b). *Receptive One-Word Picture Vocabulary Test*. 4th ed. (ROWPVT). Novato, CA: Academic Therapy Publications.
- Miyawaki, K., Strange, W., Verbrugge, R. R., et al. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Percept Psychophys*, *18*, 331–340.
- Moberly, A. C., Lowenstein, J. H., Tarr, E., et al. (2014). Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing? *J Speech Lang Hear Res*, *57*, 566–582.
- Nittrouer, S. (2004). The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *J Acoust Soc Am*, *115*, 1777–1790.
- Nittrouer, S. (2005). Age-related differences in weighting and masking of two cues to word-final stop voicing in noise. *J Acoust Soc Am*, *118*, 1072–1088.
- Nittrouer, S., & Burton, L. T. (2005). The role of early language experience in the development of speech perception and phonological processing abilities: Evidence from 5-year-olds with histories of otitis media with effusion and low socioeconomic status. *J Commun Disord*, *38*, 29–63.
- Nittrouer, S., Caldwell-Tarr, A., Moberly, A. C., et al. (2014). Perceptual weighting strategies of children with cochlear implants and normal hearing. *J Commun Disord*, *52*, 111–133.
- Nittrouer, S., & Miller, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *J Acoust Soc Am*, *101*, 2253–2266.
- Peng, S. C., Chatterjee, M., Lu, N. (2012). Acoustic cue integration in speech intonation recognition with cochlear implants. *Trends Amplif*, *16*, 67–82.
- Perkell, J. S., Boyce, S. E., Stevens, K. N. (1979). Articulatory and acoustic correlates of the [s-sh] distinction. *J Acoust Soc Am*, *65*, S24.
- Remez, R. E., Rubin, P. E., Berns, S. M., et al. (1994). On the perceptual organization of speech. *Psychol Rev*, *101*, 129–156.
- Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychol Bull*, *92*, 81–110.
- Stevens, K. N. (1985). Evidence for the role of acoustic boundaries in the perception of speech sounds. In V. A. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 243–255). New York, NY: Academic Press.
- Stevens, P. (1960). Spectra of fricative noise in human speech. *Lang Speech*, *3*, 32–49.
- Wardrip-Fruin, C., & Peach, S. (1984). Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants. *Lang Speech*, *27*, 367–379.
- Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test* (4th ed.). Lutz, FL: Psychological Assessment Resources.
- Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: Current designs and future possibilities. *J Rehabil Res Dev*, *45*, 695–730.
- 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.