Author Manuscript

The Effect of Residual Acoustic Hearing and Adaptation to Uncertainty on Speech Perception in Cochlear Implant Users: Evidence from Eye-Tracking

Bob McMurray[1], **Ashley Farris-Trimble**[2], **Michael Seedorff**[3], and **Hannah Rigler**[4]

[1]Departments of Psychological and Brain Sciences, Communication Sciences and Disorders, and Linguistics, University of Iowa, Iowa City, Iowa, USA

[2]Department of Linguistics, Simon Fraser University, Burnaby, British Columbia, Canada

[3]Department of Biostatistics, University of Iowa, Iowa City, Iowa, USA

[4]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, Iowa, USA.

Abstract

Objectives—While outcomes with cochlear implants (CIs) are generally good, performance can be fragile. The authors examined two factors that are crucial for good CI performance. First, while there is a clear benefit for adding residual acoustic hearing to CI stimulation (typically in low frequencies), it is unclear whether this contributes directly to phonetic categorization. Thus, the authors examined perception of voicing (which uses low-frequency acoustic cues) and fricative place of articulation (s/ʃ, which does not) in CI users with and without residual acoustic hearing. Second, in speech categorization experiments, CI users typically show shallower identification functions. These are typically interpreted as deriving from noisy encoding of the signal. However, psycholinguistic work suggests shallow slopes may also be a useful way to adapt to uncertainty. The authors thus employed an eye-tracking paradigm to examine this in CI users.

Design—Participants were 30 CI users (with a variety of configurations) and 22 age-matched normal hearing (NH) controls. Participants heard tokens from six b/p and six s/ʃ continua (eight steps) spanning real words (e.g., *beach/peach*, *sip/ship*). Participants selected the picture corresponding to the word they heard from a screen containing four items (a b-, p-, s- and ʃ-initial item). Eye movements to each object were monitored as a measure of how strongly they were considering each interpretation in the moments leading up to their final percept.

Results—Mouse-click results (analogous to phoneme identification) for voicing showed a shallower slope for CI users than NH listeners, but no differences between CI users with and without residual acoustic hearing. For fricatives, CI users also showed a shallower slope, but unexpectedly, acoustic + electric listeners showed an even shallower slope. Eye movements showed a gradient response to fine-grained acoustic differences for all listeners. Even considering only trials in which a participant clicked "b" (for example), and accounting for variation in the

---

category boundary, participants made more looks to the competitor ("p") as the voice onset time neared the boundary. CI users showed a similar pattern, but looked to the competitor more than NH listeners, and this was not different at different continuum steps.

Conclusion—Residual acoustic hearing did not improve voicing categorization suggesting it may not help identify these phonetic cues. The fact that acoustic + electric users showed poorer performance on fricatives was unexpected as they usually show a benefit in standardized perception measures, and as sibilants contain little energy in the low-frequency (acoustic) range. The authors hypothesize that these listeners may over-weight acoustic input, and have problems when this is not available (in fricatives). Thus, the benefit (or cost) of acoustic hearing for phonetic categorization may be complex. Eye movements suggest that in both CI and NH listeners, phoneme categorization is not a process of mapping continuous cues to discrete categories. Rather listeners preserve gradiency as a way to deal with uncertainty. CI listeners appear to adapt to their implant (in part) by amplifying competitor activation to preserve their flexibility in the face of potential misperceptions.

Keywords

Adaptation; Cochlear implants; Phoneme categorization; Residual hearing; Speech perception

INTRODUCTION

Accurate speech perception is a complex problem. Listeners must integrate dozens of brief and transient acoustic cues, deal with variability across talkers, contexts, and speaking rates, and map the auditory input onto tens of thousands of words in the lexicon. All of this must be done in a half a second or less for a given word. For most listeners, this appears simple and effortless. However, for listeners with hearing impairment, it can be enormously difficult.

Cochlear implants (CIs) are an increasingly common approach for remediating severe to profound hearing loss (Niparko 2009). These devices supplant normal hair cell activity to electrically excite the auditory nerve. CIs yield impressive speech perception accuracy for many listeners. This accuracy is fragile, however, breaking down when listeners are faced with background noise (Fu et al. 1998; Friesen et al. 2001; Stickney et al. 2004) or open-set tasks (Helms et al. 1997; Balkany et al. 2007). Understanding the cause of this fragility—and how CI users adapt to it—is a crucial issue.

This study examines two factors that are important for successful speech perception in CI users. First, it investigates the information available to listeners, specifically, the combination of electric hearing (the CI) and residual acoustic hearing (if present). Second, CI users may adapt to their degraded input, and we use an eye-tracking paradigm based on McMurray et al. (2002; Clayards et al. 2008) to pinpoint mechanisms by which this adaptation may occur. We focus on one aspect of speech perception: the identification of phonetic categories across continuous acoustic variation. We start by reviewing work on electric/acoustic hearing, and on adaptation to CIs. We then discuss phonetic categorization in normal hearing (NH) listeners and CI users to make predictions about the role of acoustic

hearing and adaptation. Finally, we present an eye-tracking experiment that addresses both questions.

The Information Available to Listeners: Electric and Acoustic Hearing

CIs degrade the information in the speech signal. Thousands of frequencies that would be discriminable by NH listeners are compressed to a much smaller number of electrodes (typically 22, although this varies), and many CI users do not appear to benefit from all of their electrodes (Fishman et al. 1997; Friesen et al. 2001; Mehr et al. 2001). Lower frequencies are often not transmitted by CI (depending on the device and processing strategy), and for the remaining frequencies the continuous natural acoustic signal is transformed to a pulse train. Thus, an important factor-limiting CI users' speech perception accuracy is signal quality (Dorman & Loizou 1997; Fishman et al. 1997; Friesen et al. 2001).

A number of CI configurations augment the electric signal with acoustic hearing. One method is to amplify whatever acoustic hearing is available in the contralateral ear to the implant (typically in the low frequencies). This *bimodal* configuration offers a combination of acoustic and electric hearing in separate ears. More recently, *hybrid CIs* (Gantz & Turner 2004; Gantz et al. 2006; 2009; and see Woodson et al. 2010 for a review) preserve residual hearing in the implanted ear. This residual acoustic hearing is often in the low frequency range and usually amplified by a hearing aid.

Adding acoustic hearing to a CI (acoustic + electric hearing) improves speech perception in bimodal (Tyler et al. 2002; Gifford et al. 2007; Dorman et al. 2008, 2014; Zhang et al. 2010a, b) and hybrid (Gantz & Turner 2004; Turner et al. 2004; Gantz et al. 2006; Dorman et al. 2008; Dunn et al. 2010; Gifford et al. 2013) configurations. Listeners with better acoustic hearing show better speech perception (Gifford et al. 2013; Dorman et al. 2014).

While the evidence for the acoustic + electric benefit is clear, the mechanism is less so. Because the talker's pitch is primarily conveyed by low frequencies, acoustic hearing could help pull an attended speaker out of background noise (Turner et al. 2004). Similarly, binaural information may help localize a signal from background noise (Dunn et al. 2010; Gifford et al. 2013). This is not the whole story, however. Zhang et al. (2010a) show an acoustic + electric benefit with words in quiet, even when acoustic input was low-pass filtered to 125 Hz. They suggest that low-frequency acoustic information offers better cues to voicing and manner of articulation, as well as talker-voice characteristics that can help parse the higher frequencies carried by the implant. However, this remains speculative. In all of these cases, the acoustic + electric benefit was based on standardized speech perception tests like the CNC word lists or AZ-Bio sentences. There have not been any investigations on how acoustic hearing may impact recognition of specific phonetic information.

Adaptation

While signal quality is a crucial issue, the auditory/speech perception system may also adapt to degraded input: speech perception accuracy improves over months or years after

implantation in adult CI users (Tyler et al. 1997; Hamzavi et al. 2003; Oh et al. 2003). At least two mechanisms could account for these adaptations.

First, listeners must adapt to a new set of acoustic-phonetic cues. For example, while NH listeners commonly use formant frequencies and pitch to identify vowels, neither is available to CI users who use a coarser distribution of activity across electrodes. CI users may also reweight their cue-combination metrics to use temporal cues like duration or amplitude envelope over spectral cues which are poorly carried by the implant). This has been documented in a number of studies (Peng et al. 2012; Winn et al 2012; Moberly et al. 2014), which suggest CI users may, in some cases, adopt a weighting by reliability scheme (Toscano & McMurray 2010), adapting to use whatever cues are most reliable.

A second possibility is that CI users adopt high level strategies for dealing with uncertainty. When hearing spoken words, NH listeners activate multiple candidates that briefly match the input (McClelland & Elman 1986; Marslen-Wilson 1987). For example, when hearing *candle,* during the onset (*can-*) they activate potential matching words (e.g., *canning, candy, canticle*). This is not purely dictated by the bottom-up input: activation for competitors persists after they are ruled out by the input (Luce & Cluff 1998; Dahan & Gaskell 2007), and competition is modulated by lexical frequency and by processes like inhibition (Luce & Pisoni 1998; Dahan et al. 2001). Thus, this competition is a cognitively rich process that could be modulated to help listeners cope with uncertainty.

Farris-Trimble et al. (2014) offered hints that CI users may adapt these competition dynamics to cope with uncertainty about the input. CI users heard isolated words (e.g., *sandal*) and chose the corresponding picture from a set that included onset (*sandwich*) and offset (*candle*) competitors. Fixations to these pictures were monitored as a measure of how strongly each competitor was considered over time (the visual world paradigm or VWP, Tanenhaus et al. 1995). CI listeners showed late fixations to lexical competitors, even though they were correctly clicking on the target object. The authors suggest that CI users may maintain more activation for competitors (than NH listeners) in case they need to revise an interpretation. For example, if they mishear /b/ in *beach* as /p/, they may erroneously activate *peach*; when they get information indicating *beach* later, it may be difficult to reactivate it. However, if they never fully commit to *peach* (keeping *beach* partially active) reactivation would be easier. This seems like a useful way to deal with uncertainty; however, it is not clear how this may interact with CI listeners' encoding of acoustic cues. Our second goal is to probe such adaptations more precisely.

Phonetic Categorization

The current investigation examined both the effect of acoustic + electric hearing and adaptation in the context of phonetic categorization. This domain allowed us to isolate phonetic cues for which acoustic hearing may be useful, and to test several types of adaptation. For these purposes, we define phonetic categorization as the process of mapping some range of acoustic cue values onto a single category (e.g., a feature, phoneme, or word).

Phonetic categorization is often studied by manipulating one or more acoustic cues to create a continuum from one speech sound to another. Listeners label these tokens, generating a

sigmoidal identification function (Liberman et al. 1957; Repp 1982). CI users typically show much shallower identification slopes than NH listeners. This has been observed for a range of distinctions, including place of articulation (b/d, p/t: Dorman et al. 1988; Hed-rick & Carney 1997; Desai et al. 2008), stop voicing (b/p: Iverson 2003; but see Dorman et al. 1991), manner of articulation (b/w: Moberly et al. 2014; Nittrouer, personal communication), liquids (r/l, w/j: Munson & Nelson 2005), consonant clusters (s/st: Munson & Nelson 2005), fricative place of articulation (s/ʃ: Hedrick & Carney 1997; Lane et al. 2007), vowels (Lane et al. 2007; Winn et al. 2012), and word-final fricative voicing (s/z: Winn et al. 2012). This shallower identification function is not observed for all CI users (Dorman et al. 1991; Iverson 2003), or for all speech contrasts and cues (Dorman et al. 1988; Hedrick & Carney 1997; Munson & Nelson 2005; Winn et al. 2012), and these functions can sharpen to near normal levels with experience (Lane et al. 2007). Nonetheless, this shallower slope is robust across studies, and identification slopes predict speech perception accuracy (Lane et al. 2007; Moberly et al. 2014).

The typical interpretation of such curves is that a discrete (categorical) function is the goal of speech perception, and any deviation from that (a shallow function) derives from noise in cue encoding (Dorman et al. 1988; Iverson 2003; Munson & Nelson 2005)[*]. The assumption is that NH listeners encode cues accurately and compare them to a discrete boundary to obtain sharp categories. If cue encoding is noisy, a VOT of 10 msec could be erroneously encoded as 5 or 20. When the actual cue value is far from the boundary, this noise plays a minimal role—if a true VOT of 0 msec is occasionally heard as −10 or 10 it will still be a /b/. However, the same encoding noise applied to a VOT of 15 msec will occasionally "flip" the identification decision—a true VOT of 15 msec that is occasionally misheard as 25 msec will now be categorized as a /p/, resulting in a shallower slope.

This account is almost certainly a part of the explanation for CI users shallower identification slopes, and we can use it to test the degree to which acoustic hearing contributes to phonetic categorization. By comparing contrasts like voicing (that take advantage of low frequency information) to contrasts like fricative place of articulation (s/ʃ) that do not, we can test the hypothesis of Zhang et al. (2010a) that part of the acoustic + electric benefit is that acoustic information can directly participate in *phonetic* processing. This hypothesis predicts steeper slopes for voicing judgments in acoustic + electric listeners, but no difference for fricatives. Thus, this study examines voicing and fricative place in both electric only and acoustic + electric configuration using a phonetic classification paradigm.

In recent years, however, the psycholinguistic community has pointed out that phonetic categorization may not be well characterized by a discrete boundary model. If so, then noisy encoding may not entirely explain differences exhibited by CI users. Phonetic categories have a graded prototype structure (Miller 1997), and listeners appear to retain, not discard, fine-grained within-category differences, even as their overt categorization may appear

---

[*] A number of authors (Peng et al. 2012; Winn et al. 2012; Moberly et al. 2014) also suggest that a shallower slope along one acoustic cue dimension can derive simply from CI users giving that dimension less weight, in favor of other dimensions. However, the usual explanation for such down-weighted dimensions is that they are down-weighted because they are poorly encoded (and therefore less reliable).

discrete (Andruski et al. 1994; McMurray et al. 2008, 2002). This implies that shallower slopes may be the desired goal as they preserve more of this structure.

Supporting this, McMurray et al. (2002) presented listeners with VOT continua spanning two words (e.g., *beach/peach, bear/pear*) in a VWP task using screens containing pictures of both the /b/- and /p/- items as well as two additional filler items (e.g., *lamp* and *ship*). Fixations to each object were analyzed only for trials in which the participant responded appropriately relative to their own boundary (e.g., for trials on the /b/ side of their boundary, they clicked on the /b/-item). Even for trials that were all categorized identically, NH listeners fixated the competitor (/p/ for a /b/ stimulus) more when the VOT approached the category boundary than when it was more unambiguous. This analysis assumed the discrete psychophysical model, but found that, over and above that (after changes in identification were accounted for) listeners preserved differences in VOTs in patterns of lexical competition (Fig. 1A; see also McMurray et al. 2008, 2014).

This finding is difficult to explain in a simple psychophysical (discrete) model of categorization. If the category can be computed accurately, there is no need to retain within-category detail. In fact, several studies suggest that such retention of detail may be an adaptive way to deal with uncertainty. McMurray et al. (2009) presented NH listeners with long, overlapping words like *barricade* and *parakeet.* In such words, if the initial VOT were not accurately perceived (e.g., a VOT of 10 msec, perceived as /p/), the listener may commit to the wrong word (*parakeet*) for several hundred msec until the disambiguating material (-*cade*) arrives. This study showed that recovery from such errors was faster when the VOT was closer to the boundary; within category differences in VOT affect the degree of initial commitment to each word, and these partial commitments affect recovery after the disambiguating material. As listeners are more likely to make such errors near the boundary than further from it, then this gradient pattern becomes adaptive, helping listeners keep their options open at places where they are more likely to be uncertain.

Clayards et al. (2008) examined whether this gradient approach to perception could be adapted in a training study. They tested two groups of listeners with an eye-tracking paradigm similar to McMurray et al. (2002). One group heard VOTs from two tightly clustered categories—on each trial the VOT was likely to be close to the prototypical values. The second group received highly variable VOTs simulating a less reliable cue. Clayards et al. found that both the slope of the identification function and the within-category sensitivity in the fixations were sensitive to this variability. When listeners were less certain (because the VOTs were more variable), they showed shallower slopes and more sensitivity to within-category detail.

Although these studies examined NH individuals, they have clear implications for CI users. They suggest that one response to the uncertainty faced by CI users would be to alter the dynamics of categorization and/or lexical access to maintain activation for competitors in case listeners need to recover from a misinterpretation. Even if the ultimate decision is discrete, CI users could alter the dynamics of the process to "hedge their bets." Furthermore, Clayards et al. suggest that there may be value in tuning this heightened competition to

specific locations in the acoustic-cue space, "hedging" most at points where the listener is most likely to be wrong.

There have been few assessments of CI users' speech perception from this perspective of graded categories. One exception is Lane et al. (2007), who examined s/ʃ and u/i continua with labeling and discrimination functions, as well as a phoneme goodness rating to assess the graded structure of categories. CI users showed less within-category discrimination and more graded and overlapping phoneme goodness functions than NH listeners. Thus, CI users' categories may be less discrete and more overlapping than NH listeners'. However, it is unclear whether the overlap results from listeners identifying a sound accurately (but keeping competitors available) or from noisy in the encoding of acoustic cues. With only one measure on each trial, it was not possible (for example) to determine if a token that was rated a poor /s/ because it was perceived as an /s/ but it was not a good one, or if it was a poor /s/ because it was perceived as an /ʃ/ on that specific trial.

The McMurray et al. (2002) paradigm deals with this by conditionalizing the analysis of the eye movements on the ultimate response. Given that the participant indicated that the stimulus was an /s/, what are their fixations like? Thus, the second goal of this study was to use this paradigm to determine if CI users adapt the gradiency of their speech categories over and above the contributions of noisy cue encoding; and if they do so, to determine the precise nature of the adaptations. There are several ways in which CI listeners might adapt their categorization processes. Figure 1A shows schematized results from McMurray et al. (2002). The *x* axis shows the stimulus (e.g., VOT) relative to the category boundary and the *y* axis is the proportion of fixations to the competitor. As the stimulus approaches the boundary, listeners make increasing fixations to the competitor. Figure 1B shows predictions that might be derived from Clayards et al. (2008): CI users modulate their sensitivity to within-category detail such that at regions near the boundaries, and they maintain heightened competitor activation (to hedge against their heightened uncertainty), but this falls off at more prototypical values.

It is also possible that CI users cannot encode cues accurately enough to take advantage of this strategy. This might give rise to one of two patterns. First, Figure 1C suggests CI users may heighten activation for competitors overall, even as they cannot perceive within-category detail as accurately as NH listeners, and therefore show a weaker effect of the continuous cue value. This would be predicted from Lane et al. (2007) showing poorer within-category discrimination and less well-organized categories. Second, Figure 1D presents a model in which CI users have sufficient within-category sensitivity to preserve the graded structure of the category, but adapt to their implant by increasing activation to close lexical competitors without regard to location along the continuum. These strategies could differ for different phonetic cues, or for different types of CI users (e.g., hybrid, electric only, etc.).

Logic

We tested these hypotheses using a variant of the McMurray et al. (2002) paradigm. Participants heard tokens from one of several VOT and s/ʃ continua (eight steps) spanning two words. They identified the referent of each auditory stimulus from a screen containing

b-, p-, s-, and ʃ-initial items. Participants' mouse clicking was used analogously to phoneme identification. This allowed us to examine the effect of acoustic + electric hearing on phonetic categorization by comparing the identification slope between CI users with and without acoustic hearing and NH listeners. We also examined fixations to assess whether there was any restructuring of the category over and above the role of noise in the input (Fig. 1). Here, by controlling for participants' overt responses and analyzing fixations relative to participants' own boundaries, we minimize the chance that any increased fixations near the boundary were due to a noisy encoding that caused the stimulus to be miscategorized. Any differences in the fixations over and above the ultimate response can then be seen as an adaptation in categorization. This is conservative; shallower slopes in the identification functions are likely due both to noise in cue encoding, and altered categorization dynamics. This analysis assumes that variance is all due to noisy encoding, to look for unique contributions of adapted categorization dynamics.

We used fricative and voicing continua for three reasons. First, fricative place of articulation gives us a close match to Lane et al. (2007), while VOT continua were used in several prior CI studies (Dorman et al. 1991; Iverson 2003). Thus, we were confident we would see differences between CI users and NH listeners. Second, the most extensive work on within-category sensitivity in NH listeners has been done with voicing (Andruski et al. 1994; McMurray et al. 2002, 2008, 2014). Third, and most importantly, one of our goals was to ask whether acoustic hearing offers a direct benefit for phonetic categorization. This design offered one continuum (voicing) in which low-frequency information was likely to be useful (c.f., Zhang et al. 2010a), and one continuum (fricative place) where it would not (voiceless fricatives have almost no energy below 1500 Hz).

MATERIALS AND METHODS

Participants

CI users were recruited through the Department of Otolaryngology at the University of Iowa Hospitals and Clinics. NH participants were recruited through advertisements in the community. Participants included 30 CI users (17 female) and 22 NH controls (17 female). The average age of the CI participants was 53.5 years (range: 20 to 69) and the average age of the NH group was 54.1 years (range: 45 to 66). These did not significantly differ ($t(49) < 1$). All CI users were postlingually deafened.

CI users represented a variety of device configurations (Table 1). Ten were in an electric-only configuration, of whom five used a unilateral implant with no hearing aid and five used bilateral implants. Twenty subjects were acoustic + electric CI users; eight used a CI with a hearing aid contralateral to the implant (a bimodal configuration) and 12 used a hybrid CI with a hearing aids both ipsilateral and contralateral to the implant. Bimodal, bilateral, and unilateral CI users used a wide range of devices reported in Supplemental Digital Content S1 (http://links.lww.com/EANDH/A198). Hybrid CI users used the Nucleus EAS system (10-mm electrode with eight electrodes; $N = 7$); the Nucleus L24 (16 mm/18 functional electrodes; $N = 3$); or the Nucleus S12 (10 mm/10 electrodes; $N = 2$). Hybrid CI users all wore an additional hearing aid on the contralateral ear. All CI users had at least a year of experience using their CI and most had several. Audiograms conducted with the CI showed

very low thresholds with an average PTA of 25.5 dB (SD = 6.2), and no participants with thresholds higher than 35 dB. A complete description of the CI users is provided in Online Supplemental Digital Content S1 (http://links.lww.com/EANDH/A198). We also ran a small battery of assessments of language and nonverbal abilities to use as moderators. These are described in Supplemental Digital Content S1 and the moderating analyses in Supplemental Digital Content S4 (http://links.lww.com/EANDH/A198).

NH participants reported NH at the time of testing and were native monolingual English speakers. We conducted a hearing screening on these participants. Seventeen of them passed by the ASHA (1990) guidelines at all four frequencies for both ears. An additional five did not pass at one or more frequencies (no participant failed at more than two across both ears). Follow-up work with these participants revealed their thresholds to be below 40 dB at these frequencies. As these thresholds were below the level requiring clinical intervention, these participants were retained for analysis. All CI users and NH listeners reported normal or corrected-to-normal vision in at least one eye.

Design

We used six minimal pairs for voicing and six for fricative place. Pairs were selected to be easily picturable. For each minimal pair, an eight-step continuum was constructed by manipulating acoustic cues to either VOT or fricative place of articulation. Each participant heard each member of the 12 continua six times for a total of 576 trials. On each trial, the screen contained pictures of both members of one voicing and one fricative pair. This allowed the fricatives to serve as the unrelated visual objects for the voicing trials (and vice versa). To avoid highlighting the unique relationship between members of minimal pair, the particular assignment of voicing and fricative pairs was fixed throughout the experiment (but randomly selected across subjects), as in McMurray et al. (2002).

Stimuli

Auditory stimuli were constructed from recordings of a male native English speaker with a standard Midwest dialect. Words were recorded in a sound-attenuated room with a Kay CSL 4300B A/B board at a sampling rate of 44.1 kHz. Each word was recorded in a carrier phrase ("He said ____") several times and then extracted from the phrase for further modifications.

VOT continua were constructed by progressive cross-splicing (McMurray et al. 2008). One token of each endpoint was selected that best matched on pitch, duration, and formant frequencies. Next, a specified duration of material was cut from the onset of the voiced token (e.g., *beach*) and replaced with a corresponding segment from the voiceless token (*peach*). This was done at approximately 8 msec increments, taking care to only cut at zero crossings of the signal. This led to an eight-step continuum ranging from 0 to 56 msec of VOT.

Fricatives were created using a procedure outlined in Galle (2014; Fig. 2). First, the frication portions of the /s/ and /ʃ/ tokens were excised. These were equated on length by excising (at 0 crossings) material from the center of the longer of the two tokens. Next, we extracted the

long-term average spectra from each fricative and computed their spectral means (Fig. 2A). We then aligned the spectra to have the same spectral mean (Fig. 2B) and averaged them in proportions ranging from 0% /s/ to 100% /s/ in eight steps (Fig. 2C). This created a continuum with the same spectral mean but with a spectral shape varying from /s/ to /ʃ/. Next, the locations of these spectra were shifted in frequency space in eight steps from the spectral mean of an /s/ to that of an /ʃ/ (Fig. 2D). After that, white noise was filtered through these spectra. Finally, we extract the amplitude envelope from the original /s/ and the /ʃ/ tokens, averaged them, and imposed the average onto the filtered noise to create a new fricative. This was spliced on to the vocoid from a neutral production (for *self/shelf*, it was spliced on to the *-elf* from *help*). After both continua were constructed, 100 msec of silence was added to the beginning and end of each word. The final stimuli were normalized to the same RMS amplitude.

Visual stimuli were developed using a standard laboratory procedure that ensures clear, representative images of our words (Farris-Trimble et al. 2014). We first selected images from a commercial clipart database. Several images were selected for each word and lab members viewed these images and select the most prototypical image for each word in a focus group. The selected image was then edited to ensure visual continuity among the picture stimuli and to minimize visual distractions. All images were approved by one of three members of the laboratory with extensive experience in the VWP.

Procedure

After giving informed consent, participants were seated in front of a 17″ computer monitor at 1280×1024 resolution and a standard keyboard and mouse. Auditory stimuli were presented over Bose loud speakers amplified by a Sony STR-DE197 amplifier/receiver at a volume level most comfortable to the listener. Volume was initially set to 70 dB, and participants could adjust the volume to a comfortable listening level during a brief training procedure. A padded chin rest was placed at the end of the testing table (29″ from the monitor) and experimenters adjusted its height to a comfortable position. Subjects were allowed to move around during breaks (every 32 trials). Researchers then calibrated the eye tracker. After calibration, both written and verbal instructions were given.

Before testing, every subject performed an eight-trial training procedure to familiarize them with the task and to allow them to adjust the volume. The words and pictures used in training were different from those used in testing. After this training, participants underwent a second phase of training to familiarize them with the visual stimuli used during the experiment. In this phase, subjects advanced through the images of the experimental accompanied by the written word and a single audio presentation (an unmanipulated recording).

On test trials, participants saw four pictures accompanied by a red dot in the center of the screen. When the dot turned blue (after 500 msec) they clicked on it to play an auditory stimulus (randomly selected). After hearing the word, subjects clicked on the referent with the mouse. Subjects were instructed to take their time and perform the task as naturally as possible.

Eye Movement Recording and Analysis

Eye movements were recorded with an SR Research Eye-link 1000 desktop mounted eye tracker. The standard nine-point calibration was used. To account for the natural drift of the eyes during the experiment and to maintain a good calibration, drift corrections were run every 32 trials. If the participant failed a drift correction, the eye tracker was recalibrated. Pupil and corneal reflections were sampled at 250 Hz to determine point-of-gaze. Eye movements were automatically classified into saccades, fixations, and blinks by the Eyelink control software using the default parameters. For analysis, these movement types were combined into "looks" which began at the onset of the saccade and ended at the end of the subsequent fixation (McMurray et al. 2002, 2008). When determining which object a look was directed to, image boundaries on the screen were extended by 100 pixels to account for noise in the eye track. This did not result in any overlap among the areas of interest.

RESULTS

We report two primary analyses. First, we examined the mouse-click responses which are analogous to 2AFC phoneme identification. Next, we examined eye movements relative to each subject's category boundary as a measure of sensitivity to the acoustic distinctions within a category. In both cases, the primary experimental variable was the relevant phonetic cue (VOT, s/ʃ step), and the primary individual difference was listener type. As our goal was to examine electric + acoustic stimulation, we investigated three listener groups: NH listeners, electric-only CI users (both unilateral and bilateral) with no residual acoustic hearing (CI_E), and acoustic + electric CI users (both bimodal and hybrid CIs), who had some acoustic hearing in either the ipsi- or contralateral ear (CI_{AE}). We explored differences within CI groups in Supplement S3 (http://links.lww.com/EANDH/A198).

Identification Performance

Stop Voicing—• Our analyses of the b/p continua eliminated any trial in which participants selected a response that was not relevant for that continuum (an /s/- or /ʃ/- initial item). This eliminated four trials from NH listeners (M = 0.18 trials/ participant of a total of 288 trials) and 13 from the CI users (M = 0.41 trials/participant). Figure 3A shows the proportion of /p/ responses as a function of VOT step and device group. Overall differences were quite small; the only apparent difference is one in slope between NH listeners and CI users as a whole. However, visualized this way, it is possible that such slope differences actually derive from an averaging artifact; for example, if CI participants have equally sharp categorization functions (as NH listeners) but more variation in the location of their boundaries, the averaged data will appear to have a shallower slope. Thus, Figure 3B plots the same data with each step of the continuum recoded relative to each participant's own boundary (adjusted for variable boundaries among continua), using the procedure described in the next section. The boundary is at an rStep of 0; −1 is one step toward the voiced end of the continuum, and +1 represents one step toward the voiceless end. Here, a clearer difference in slope can be observed between NH and CI users.

These data were analyzed with a binomial mixed effects model using the LME4 package (ver 1.1–7) of R (ver 3.1.0). This model included b/p step as a fixed effect (centered), and

two contrast codes to capture hearing group. The first captured CI users (+0.5, both groups) versus NH listeners (−0.5) and was centered. The second captured the CI_{AE} versus CI_E distinction (+0.5/−0.5, NH coded as 0, overall mean of 0). Potential random effects included participant and continuum. Before examining the fixed effects, we conducted a series of models exploring different random effects structures to determine the most complex (conservative) random effects structure that best fit the data. These examined random intercepts for subject and continua, and random slopes of step on both terms and random slopes of hearing group on continuum. The maximal model (random slopes of step on subject and continuum, and of hearing group on continuum) offered a significantly better fit than simpler models (χ^2(18) = 213.70; $p < 0.0001$).

As expected, this model showed a significant main effect of step ($B = 2.61$, SE = 0.19, $Z = 13.8$; $p < 0.0001$). The CI versus NH contrast was also significant ($B = −2.79$, SE = 0.71, $Z = −3.9$; $p < 0.0001$), but the CI_E versus CI_{AE} was not ($B = 0.32$, SE = 0.75, $Z = 0.43$; $p = 0.67$). This pattern of main effects suggests that CI users had a right-shifted boundary (more /b/ responses), but that the two CI groups did not differ. Most importantly, there was a significant step × CI versus NH interaction ($B = −1.48$, SE = 0.46, $Z = −3.2$; $p = 0.0014$), indicating that the boundary was steeper for NH listeners than both groups of CI users (Fig. 3B). There was no step × CI_E versus CI_{AE} interaction ($B = 0.41$, SE = 0.35, $Z = 1.18$; $p = 0.24$). Thus, the largest predictor of differences (in both slope and boundary) was whether or not the listener wore a CI, not which type they used. Unexpectedly, the availability of low-frequency acoustic information did not appear to alter voicing categorization.

Fricative Place of Articulation—• The fricative analysis excluded the small number of trials in which participants clicked /b/ or /p/ items (NH: 2 trials, $M = 0.091$ trials/participant; CI: 13, $M = 0.41$ trials/participant of 288 total trials). Figure 4 shows the identification performance for each hearing group both as a function of absolute step and relativized to participants' boundaries, respectively. Here, marked differences between listeners can be seen. Unexpectedly, the electric-only CI users appear to track NH performance fairly closely (after accounting for a difference in category boundary), but the acoustic–electric CI users (who typically show better accuracy) show much shallower categorization slopes.

This was evaluated with a binomial mixed effects model. Here, our descriptive analysis suggested that the acoustic–electric group seemed to stand out so we used a different set of contrast codes to examine listener: one contrast captured the acoustic + electric CI users versus the average of the NH listeners and electric-only CI users, and the other captured the difference between electric-only CI users and NH listeners. Again the maximal random effects structure offered a significantly better fit than simpler models (χ^2(18) = 693.6; $p < 0.0001$).

This model revealed a significant main effect of step ($B = 1.83$, SE = 0.16, $Z = 11.7$; $p < 0.0001$). There were no main effects of either CI contrast (CI_{AE} versus other: $B = −0.39$, SE = 0.76, $Z = −0.51$, $p = 0.61$; NH versus CI_E: $B = 0.77$, SE = 0.87, $Z = 0.88$, $p = 0.38$) suggesting similar boundaries for each. Both step × CI group interaction terms were significant (step × CI_{AE} versus other: $B = −1.36$, SE = 0.22, $Z = −6.1$, $p < 0.0001$; step × NH versus CI_E: $B = −1.22$, SE = 0.28, $Z = −4.3$, $p < 0.0001$). This indicates that the AE group

had significantly shallower slopes than the mean of the other two groups, and that the electric-only CI users had a shallower slope than the NH listeners (although as Fig. 4B suggests, this effect was numerically smaller than the acoustic + electric effect). A follow-up model comparing CI_E and CI_{AE} users showed a significant interaction of this contrast with step ($B = -0.57$, SE = 0.23, $Z = -2.5$; $p = 0.013$), confirming that the acoustic + electric group had an even shallower slope than the electric-only group.

Summary—• Identification results for the *voicing* continua suggested that both groups of CI users had significantly shallower slopes to their identification functions. Unexpectedly, the acoustic + electric users did not differ from electric-only users, suggesting that residual acoustic information may play a minimal role in voicing categorization. Effects of CI use, however, were small, suggesting that CI users, while impaired, are fairly good at detecting voicing.

Results for fricatives, however, were somewhat unexpected. As with voicing, both CI groups (together) showed significantly shallower slopes. However, the acoustic + electric CI users were much shallower than the other CI users. While this effect appeared in both types of acoustic + electric configuration (although stronger for bimodals; see Supplement S2; http://links.lww.com/EANDH/A198), we were worried that this might reflect the influence of one or more outlier subjects. Thus, we examined individual subject identification curves.

As Figure 5C suggests, this did not appear to be the case—individual users with both bimodal and hybrid configurations exhibited shallower and more variable slopes, while the NH and electric-only CI users showed much more uniformly steep slopes (Figs. 5A, B, respectively). Within the hybrid group, we also examined the specific devices (not shown), and found that users of two of the devices (the Nucleus EAS and the Nucleus Hybrid S12 Implants; both from Cochlear Americas) showed similarly shallow slopes. Both of these implants use 10-mm electrodes. While the slightly longer Nucleus Hybrid L24 (a 16-mm implant with more electrodes) showed a profile more similar to electric-only users, no firm conclusions could be drawn as there were only three subjects, particularly given that the bimodal users had electrodes of similar lengths (but still displayed shallow identification slopes). Thus, it does not appear that the poor fricative categorization derives from any specific device (bimodal configurations or specific hybrid implants) or from a small number of CI users. Rather this finding applies more broadly to a variety of situations in which listeners must integrate acoustic and electric hearing. This finding is somewhat surprising because in most standardized speech perception measures residual acoustic hearing is helpful.

Eye Movement Analysis

Our analysis of the eye movements focused on the degree to which CI and NH listeners are sensitive to differences in VOT or frication spectrum within a category. Given a set of trials that were unambiguously heard as /b/ (for example), to what degree are listeners still sensitive to changes in VOT? This allows us to examine the residual sensitivity to fine-grained changes in either VOT or frication spectra. Under a discrete boundary model, a shallower identification slope is entirely due to noise in encoding the cues, noise that

occasionally results in a cue being encoded as on the other side of the boundary and "flipping" the identification. By conditionalizing the analysis of VOT/frication on the ultimate response, these trials with a flipped response are excluded. This allows us to observe if CI users are making any more fundamental changes to how cues are mapped to categories (or words), over and above any changes in how the cue values are encoded.

To accomplish this, we used a technique similar to several previous studies (McMurray et al. 2002, 2008, 2014). First, we eliminated variability between participants and/or items in the location of the category boundary by recoding continuum step (either VOT or frication spectra) in terms of distance from the boundary. Second, we eliminated all of the trials in which the participant chose the competing response (given the boundary). For example, if the relative VOT was −1 (a /b/ one step from that participant's boundary), we eliminated any trial in which the response was /p/. Finally, we examined the effect of relative step (VOT or frication spectra) on this subset of the trials.

To compute category boundaries, we fit a four-parameter logistic function to each participant's identification data (Eq. 1). The four-parameter logistic (with variable maximum and minimum asymptotes, *max* and *min*) was used rather than the more standard two-parameter version (which asymptotes at 1 and 0) to account for the fact some of our listeners did not reach perfect asymptotes.

$$P\left(p \quad \text{response}\right) = \frac{max - min}{1 + exp\left(4 \cdot \frac{slope}{max - min} \cdot \left(\text{boundary} - \text{step}\right)\right)} + min \quad (1)$$

We expected different boundaries by both subject and continua. In prior studies (McMurray et al. 2008, 2014), this was handled by computing separate fits for each continuum for each subject. This was not possible here because this study had roughly half the trials as these studies. This was due to the fact that while prior studies examined typical college students who can complete over 1000 trials, we anticipated that our population (which was older, and some of whom had hearing impairments) would not be able to complete this many. To estimate boundaries with this smaller dataset, we instead fit logistic functions to each subject (averaging across continua) and to each continua (averaging across subjects). We then computed the boundary for a given subject × continua by adding the deviation of that continuum's boundary from the grand mean to that subject's boundary. Logistic functions were fit with a constrained gradient descent method that minimized the least squares error, while constraining function parameters to ensure that the function could not exceed 0 or 1 and that the boundary was located somewhere within the stimulus space.

Descriptives—• Figure 6 shows the representative timecourse of fixations for a number of subconditions. This figure is not meant to evaluate group differences, but rather give a rough picture of how each group responded to the task. Figure 6A, for example, shows the NH listeners after having heard a /b/ with a VOT of 0 msec. By about 500 msec, NH listeners are already fixating the target (/b/) more than the competitor (/p/) or unrelated (/s, ʃ/) items. There are few looks to the competitor (relative to the baseline unrelated items) as the VOT was unambiguous and listeners quickly converged on the correct word. Figure 6B shows

that CI users exhibit a slower rise of fixations to the target (replicating Farris-Trimble et al. 2014), and maintain looks to the competitor much longer. The fricatives (Figs. 6C, D) show more looks to the competitors overall but similar group differences.

Of course, our primary analysis concerned the way that competitor fixations are modulated by within-category changes in the stimuli. Figure 7 shows representative data from the /ʃ/ side of the fricative continua. It plots fixations to the competitor objects over time (e.g., /s/ when the stimulus was on the /ʃ/ side of the subjects' boundary) as a function of distance from the boundary (rStep, as described above; see Supplement S3 for full color figures of all of the data; http://links.lww.com/EANDH/A198). NH listeners (Fig. 7A) showed a gradient pattern with more fixations to the competitor for rSteps near the boundary (±1). CI listeners (Fig. 7B) showed a similar gradient pattern, but competitor fixations persisted for longer (more consideration of the competitor). This suggests that CI users are responding qualitatively similarly to NH listeners, but there are some quantitative differences.

To more precisely identify these group differences, we averaged across time (but within group and rStep) to compute the amount of looking to the competitors in general, the area under these curves. For the voicing continua, this was computed over a 300 to 2300 msec window. This window was selected because it takes 200 msec to plan and launch an eye movement and there was 100 msec of silence at the onset of the sound files (thus it would take 300 msec for signal driven eye movements to emerge). The 2000 msec duration was selected for consistency with prior studies (McMurray et al. 2002, 2008, 2014). Fricatives used a slightly later window of 600 to 2600 msec. This was selected because ongoing work in our lab suggests that listeners appear to wait several hundred msec to make signal-driven eye movements for fricatives (Galle 2014).

The results of this process are shown in Figure 8. Consistent with prior work, subjects showed a gradient effect of distance from the boundary—as the speech sound approached the boundary listeners of all types tended to look more to the competitor object and CI users tended to look to competitors even more overall, perhaps resembling the hypothetical Figure 1D.

Statistical Approach—• To address these issues statistically, these area-under-curve estimates were examined as the dependent variable in a linear mixed effects model. Separate models were run for each side of each continuum, yielding four models (for /b/, /p/, /s/, /ʃ/). All models included rStep as a continuous predictor (centered) as well as two contrast codes for CI group (CI versus NH and CI_E versus CI_{AE}) as in the prior analyses. Because we were worried that the effect of rStep may be nonlinear, we explored additional polynomial terms for rStep, finding that the addition of a quadratic effect of rStep (and its interactions with CI status) accounted for additional variance for the /b/ ($\chi^2(6) = 25.2$; $p = 0.0003$) and /s/ ($\chi^2(6) = 53.1$; $p < 0.0001$) analyses. As a result, in just these analyses, we included both linear and quadratic effects of rStep. In preliminary analyses, we also found that looks to the unrelated objects varied significantly between CI users and NH listeners. To ensure that any group differences were specifically targeted to the competitor and did not reflect more general uncertainty we included looks to the unrelated object as a covariate. Consequently, the final model had rStep, CI status (two variables) and their interactions, plus a main effect of

unrelated looking. We included random slopes of rStep on participant and random intercepts for continuum as this model offered a significantly or marginally significantly better fit to the data than models with random intercepts alone (/b/: $\chi^2(2) = 7.89$, $p = 0.019$; /p/: $\chi^2(2) = 18.09$, $p < 0.0001$; /ʃ/: $\chi^2(2) = 10.84$, $p = 0.0044$; /s/: $\chi^2(2) = 4.78$, $p = 0.092$). All p values for fixed effects were estimated using the Satterwaithe approximation for the d.f., implemented in the lmerTest package (ver 2.0.11) in R. Critically, if CI users show heightened (Fig. 1B) or reduced (Fig. 1C) sensitivity to within-category detail as an adaptation to their input we should observe an interaction of rStep with CI status. In contrast, if CI users simply heighten activation for competitors without respect to continuum step (Fig. 1D), we should see main effects but no interaction.

Voicing—• Our first analysis examined the VOT continuum (Fig. 8A). Here, due to variability in the boundary among participants and continua, there were only a handful of listeners with rSteps less than –3.5[†] (most listeners' boundaries were too close to the voiced side). Thus, rSteps less than this were excluded. The results of this analysis are shown in Table 2. As expected, we found a main effect of rStep on both sides of the boundary (/b/: $p < 0.0001$; /p/: $p < 0.0001$), and a quadratic effect on the /b/ side ($p = 0.035$). As rStep approached 0 (the boundary), participants made more fixations to the competitor. There was also a large difference between CI and NH listeners (/b/: $p = 0.0001$; /p/: $p < 0.0001$), but no difference between electric-only and acoustic + electric configurations (/b/: $p = 0.49$, /p/: $p = 0.29$). As Figure 7A shows, CI users look at competitors more, regardless of VOT (and over and above what would be predicted by their unrelated looks). On the voiced side, the rStep × CI_E versus CI_{AE} interaction was significant (/b/: $p = 0.01$) along with the quadratic rStep × CI_E versus CI_{AE} interaction (/b/: $p = 0.02$). All other interactions were nonsignificant. For /b/s, acoustic + electric CI users appeared to have a slightly greater effect of rStep: they tracked within-category differences in VOT better than the other listeners.

Fricatives—• The analysis of the fricatives showed a similar pattern with respect to rStep, but larger differences between CI groups. The results of this analysis are shown in Table 3 and Figure 8B. As before, we found a significant effect of rStep (/ʃ/: $p < 0.0001$; /s/: $p < 0.0001$), and the quadratic effect was significant for /s/ ($p < 0.0001$). Also as before, CI listeners showed greater competitor fixations than NH listeners (/ʃ/: $p = 0.0003$; /s/: $p = 0.0002$). Unlike the voicing analysis, there was also a significant difference between acoustic + electric and electric-only CI users for /ʃ/ and a marginally significant difference for /s/ (/ʃ/: $p = 0.016$; /s/: $p = 0.082$). Here, mirroring their shallower identification slopes, CI_{AE} users showed even more fixations to the competitor than CI_E users. Neither CI contrast interacted with rStep.

GENERAL DISCUSSION

The results of this investigation were clear, although somewhat unexpected. With respect to the identification measures, we found consistent evidence for shallower slopes in CI users for both voicing and fricative continua. The availability of acoustic hearing did not moderate

---

[†]Because individuals' category boundaries were not limited to discrete integer values, most rSteps were continuous valued. For example, a participant with a boundary at step 3.6 would have rSteps of –1.6, –.6, +.4, +1.4, +2.4, etc.

this for the voicing continua. Contrary to our hypotheses (and predictions by Zhang et al. 2010a), this suggests that at least for this phonetic cue, there is little benefit to residual acoustic hearing. One might argue that this is simply a ceiling effect—CI users are already as good as they can get at voicing identification. However, CI users as a whole did show a shallower slope, so there was clearly room for acoustic hearing to reverse this effect. Despite this null effect of acoustic + electric stimulation, we are hesitant to argue that there would be no acoustic benefit for voicing perception under more challenging listening conditions, or if other phonetic cues in this low-frequency range (e.g., the first formant) had been manipulated. Nonetheless, these results argue that any direct effect of acoustic hearing on voicing perception may be small or variable across listeners and listening conditions.

While there was little effect of acoustic hearing on voicing identification, there was a strong effect for fricative identification. Counterintuitively, acoustic + electric CI users showed much shallower identification slopes. This was observed in both bimodal and hybrid users. To the best of our knowledge, this is the first study to find poorer performance in acoustic + electric CI users than electric-only configurations. Perhaps even more counterintuitively, this decrement was limited to a speech cue with almost no information in the low frequencies that CI_{AE} listeners have access to acoustically. Fricatives (and voiceless sibilants in general) have almost no information below 1500 Hz, and an inspection of the acoustic-only audiograms of for the CI_{AE} listeners suggested that they have little acoustic hearing above 800 Hz.

Thus, CI_{AE} users' marked and unique performance decrement with fricatives is something of a mystery. One possibility is that the acoustic + electric CI users simply had poorer auditory detectability at the high frequencies needed for fricative identification. However, this was not the case—PTAs for frequencies above 2000 Hz were actually slightly better in *acoustic + electric* CI users ($M = 22.3$ dB, SD = 5.5) than *electric-only* CI users ($M = 27.7$ dB, SD = 6.5 dB; $t(25) = 2.3$; $p = 0.032$) It is tempting to chalk this up to the shorter hybrid implants which may have fewer electrodes in the high-frequency electric ranges (which are essential for discriminating /s/ from /ʃ/). However, this is unlikely the cause since the decrement was also observed (and was slightly stronger; see Supplement S3; http://links.lww.com/EANDH/A198) in bimodal users who have a full-sized CI. Rather, we speculatively offer two overlapping explanations.

First, CI_{AE} users may have come to overweight acoustic information for speech perception. This works quite effectively for many sounds, but is uniquely bad for voiceless fricatives where there is no acoustic information in the frequency ranges they can hear. That is, these performance decrements may indicate a cue-integration strategy that is useful in many situations but fails in this one circumstance. In that sense, the difference between identification slopes for fricatives and other speech sounds (e.g., the voicing continua) may potentially serve as a useful clinical marker for how much listeners are relying on acoustic rather than electric hearing. This may help identify CI users that are not getting the most out of their implants.

A second and related account is that fricatives are also unique from other speech sounds in that they feature a long period of time with no periodic voicing information. Consequently,

during the frication, the acoustic input does not unambiguously sound like speech, as opposed to environmental noise (c.f., Galle 2014, for evidence for this in NH listeners). For electric-only CI users, this would not cause a problem—since they do not perceive periodic voicing as such, fricatives are no different than any other speech sounds. However, for acoustic + electric users, again the lack of periodic voicing may be particularly problematic and their perceptual systems may not immediately recognize the frication as speech. Here, the confluence of factors—input that briefly does not sound like speech plus the poor spectral resolution of the CI—may lead to poorer phoneme perception.

Under either account, however, it is clear that we can no longer see the benefit of residual acoustic hearing as merely an additive factor; rather the availability of low-frequency acoustic hearing has effects even on speech sounds that do not contain any information in those frequencies. Clearly, this indicates a long-term problem of learning to weigh and integrate acoustic and electric information to maximize the information gleaned from each modality.

With respect to the fixations, CI users, like NH listeners, show a gradient effect of rStep on fixations to the competitor. This cannot be an averaging artifact of variation in category boundaries across people/continua, and our analyses accounted for differences in the final response. Even though all of these tokens were ultimately perceived as a /b/ (for example), listeners still fixated the /p/ more for 10 msec of VOT than for 0 msec. This finding raises challenges to a psychophysical account of speech perception in which categorization is based on accurate encoding of cues and a discrete boundary. Rather, people do not discard fine-grained differences that lie within a category; they preserve such differences and activate categories in a graded prototype-like structure (Andruski et al. 1994; Miller 1997; McMurray et al. 2002). Our results extend such findings beyond NH individuals to listeners who use CIs, and are also the first to show that this applies to fricatives.

Such gradiency may be a useful way of dealing with uncertainty by maintaining competing alternatives when these alternatives are most likely to be needed (Clayards et al. 2008; McMurray et al. 2009). Moreover, the fact that these effects are recorded at the level of lexical items (fixations to pictures of the referents) suggests that these fine-grained differences are not discarded as part of an autonomous, prelexical speech categorization system, but rather what we think of as speech perception and word recognition unfold simultaneously. Thus, our findings suggest a rather different framing of speech perception that is clearly relevant to CI users. This echoes the call by Nittrouer and Pennington (2010) to move research on communication disorders toward more sophisticated psycholinguistic models, although this study (and the work on which it is built) suggests a view closer to interactive activation (McClelland & Elman 1986), exemplar (Goldinger 1998) and predictive coding models (McMurray & Jongman 2011) than the motor theoretic approach they advocate.

Beyond this sensitivity to fine-grained detail, we also found differences as a function of CI status. Our analyses started with the assumption that CI users' shallower identification slope is entirely due to noise in encoding speech cues like VOT, by controlling for listeners' boundaries and predicating our analysis on the overt response. They then asked if CI

listeners differ in their response to continuous stimulus difference over and above this encoding noise. As such differences are independent of overt accuracy, they likely represent differences in speech or lexical processes that may be an adaptation to the degraded input. For both VOT and frication spectra, the evidence clearly favored a model in which CI listeners raise competitor activation overall, but this increase insensitive to the degree of ambiguity in the cue value (Fig. 1D). CI listeners showed heightened fixations to the competitor but there was little evidence that this differed as a function of VOT or frication spectra. Moreover, as we controlled for looking to unrelated objects, this was not just a matter of general uncertainty (although as we have previously argued, this notion is difficult to distinguish from increased lexical activation for competitors: McMurray et al. 2014).

This increase in looks to the competitors also varied as a function of acoustic hearing in a way that mirrored the identification data. For the VOT continua, there were no differences between acoustic–electric and electric-only configurations. For the fricative continua, we found greater competitor fixations for CI_E users and even greater fixations for CI_{AE} users. This close correspondence between the online eye-movement behavior and the end-state identification results suggests that CI users are adapting their level of commitments fairly precisely to the situations in which they are likely to be wrong.

Nonetheless, the argument that CI users are raising competitor activation overall without altering phonetic representations is based on a null effect (of the interaction of rStep and hearing group). While null effects must always be interpreted with caution, there are reasons to trust this effect. First, we have a large number of CI users and more participants than Clayards et al. (2008), who did find an interaction. Second, we used extremely powerful statistical analyses, accounting for individual subject slopes and basic looking patterns. Third, a similar null effect has also been reported in other populations facing communicative difficulties (children with SLI: McMurray et al. 2014). Finally, and most importantly, the overall degree of fixations to the competitor was highly sensitive to CI status as a whole, and in the fricatives, to the type of CI. Thus, it is clear that CI effects can be observed in this population, they just do not appear to modulate sensitivity to fine-grained detail.

Accepting these findings, this heightened competitor consideration could be adaptive. It is quite likely that CI listeners often mishear or miscategorize phonemes. If they were to fully commit to a single word on the basis of this, when later (semantic or discourse) information arrives, they may have difficulty recovering and activating the correct item. In contrast, by keeping likely competitors around, they may be able to reactivate them more efficiently. Our study suggests that CI users engage in such a strategy, but that for them this strategy does not appear to be differentially modulated by how ambiguous the phonetic cues are—CI listeners simply activate nearby words more while preserving a similar sensitivity to fine-grained detail. The adaptive nature of this heightened activation is underscored by Figures 6 and 7 which examine the timecourse of this effect. Importantly, they suggest that the increased fixations shown by CI users is not driven by the peak looking over time, but rather that CI users tend to persist in looking for longer durations. This would seem to fit well with a model of preserving flexibility in case decisions must be revisited.

CONCLUSION

So why would not CI users modulate this heightened activation as a function of fine-grained detail? Clearly such a strategy could be advantageous. Maintaining competitor activation potentially slows recognition, and thus, it may be more optimal to do so at the places where a misperception is most likely (e.g., near the boundary). Hence, there are good reasons to maintain competitors more near these regions of the continua (Clayards et al. 2008). It is possible that CI users do not do this because they cannot hear such fine-grained differences. However, that seems unlikely since they did show strong gradient effects of rStep. A second possibility is that this adaptation takes place primarily at the lexical level. Our earlier study on adolescents with SLI suggests their deficits in comprehension may be lexical rather than perceptual (McMurray et al. 2010, 2014), and derives from differences in how words are activated, compete, and decay. Crucially, they show a similar pattern of results to the CI users here. In this case, it may be easier to hedge one's bets by modulating global activation/ competition dynamics than by remapping phonetic categories (particularly given CI users' challenges in speech perception). This is underscored by the apparent fine tuning of these activation dynamics by the acoustic + electric users who appear to show heightened effects for fricatives (where they struggle).

More broadly, however, these results suggest a complex view of speech perception. Fine-grained detail is preserved at the level of lexical processes, and these processes can be modulated by CI users to keep their options open during speech perception. We also demonstrate that the effect of residual acoustic hearing is not simply additive—CI_{AE} users do more poorly on a speech contrast with no low-frequency information. This supports a view of speech perception which is not a purely bottom-up process of mapping speech cues to discrete categories in a bottom-up process. Rather, cues are integrated in a much more gradient and nonlinear way, as speech perception and word recognition unfold simultaneously.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

Andruski JE, Blumstein SE, Burton M. The effect of subphonetic differences on lexical access. Cognition. 1994; 52:163–187. [PubMed: 7956004]

Balkany T, Hodges A, Menapace C, et al. Nucleus Freedom North American clinical trial. Otolaryngol Head Neck Surg. 2007; 136:757–762. [PubMed: 17478211]

Clayards M, Tanenhaus MK, Aslin RN, et al. Perception of speech reflects optimal use of probabilistic speech cues. Cognition. 2008; 108:804–809. [PubMed: 18582855]

Dahan D, Gareth Gaskell M. The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. J Mem Lang. 2007; 57:483–501. [PubMed: 18071581]

Dahan D, Magnuson JS, Tanenhaus MK, et al. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Lang Cognitive Proc. 2001; 16:507–534.

Desai S, Stickney G, Zeng FG. Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. J Acoust Soc Am. 2008; 123:428–440. [PubMed: 18177171]

Dorman MF, Loizou PC. Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants. Am J Otol. 1997; 18(6 Suppl):S113–S114. [PubMed: 9391623]

Dorman MF, Hannley MT, McCandless GA, et al. Auditory/phonetic categorization with the Symbion multichannel cochlear implant. J Acoust Soc Am. 1988; 84:501–510. [PubMed: 3170943]

Dorman MF, Dankowski K, McCandless G, et al. Vowel and consonant recognition with the aid of a multichannel cochlear implant. Q J Exp Psychol A. 1991; 43:585–601. [PubMed: 1775658]

Dorman MF, Gifford R, Lewis K, et al. Word recognition following implantation of conventional and 10-mm hybrid electrodes. Audiol Neurootol. 2008; 14:181–189. [PubMed: 19005252]

Dorman MF, Loizou P, Wang S, et al. Bimodal cochlear implants: The role of acoustic signal level in determining speech perception benefit. Audiol Neurootol. 2014; 19:234–238. [PubMed: 24992987]

Dunn CC, Perreau A, Gantz B, et al. Benefits of localization and speech perception with multiple noise sources in listeners with a short-electrode cochlear implant. J Am Acad Audiol. 2010; 21:44–51. [PubMed: 20085199]

Farris-Trimble A, McMurray B, Cigrand N, et al. The process of spoken word recognition in the face of signal degradation. J Exp Psychol Hum Percept Perform. 2014; 40:308–327. [PubMed: 24041330]

Fishman KE, Shannon RV, Slattery WH. Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. J Speech Lang Hear Res. 1997; 40:1201–1215. [PubMed: 9328890]

Friesen LM, Shannon RV, Baskent D, et al. Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. J Acoust Soc Am. 2001; 110:1150–1163. [PubMed: 11519582]

Fu QJ, Shannon RV, Wang X. Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. J Acoust Soc Am. 1998; 104:3586–3596. [PubMed: 9857517]

Galle, M. PhD dissertation. University of Iowa; 2014. Integration of asynchronous cues in fricative perception in real-time and developmental-time: Evidence for sublexical memory/ integration systems..

Gantz BJ, Turner C. Combining acoustic and electrical speech processing: Iowa/nucleus hybrid implant. Acta Otolaryngol. 2004; 124:344–347. [PubMed: 15224850]

Gantz BJ, Turner C, Gfeller KE. Acoustic plus electric speech processing: Preliminary results of a multicenter clinical trial of the Iowa/nucleus hybrid implant. Audiol Neurootol. 2006; 11(Suppl 1): 63–68. [PubMed: 17063013]

Gantz BJ, Hansen MR, Turner CW, et al. Hybrid 10 clinical trial: preliminary results. Audiol Neurootol. 2009; 14(Suppl 1):32–38. [PubMed: 19390173]

Gifford RH, Dorman MF, McKarns SA, et al. Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing. J Speech Lang Hear Res. 2007; 50:835–843. [PubMed: 17675589]

Gifford RH, Dorman MF, Skarzynski H, et al. Cochlear implantation with hearing preservation yields significant benefit for speech recognition in complex listening environments. Ear Hear. 2013; 34:413–425. [PubMed: 23446225]

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychol Rev. 1998; 105:251–279. [PubMed: 9577239]

Hamzavi J, Baumgartner WD, Pok SM, et al. Variables affecting speech perception in postlingually deaf adults following cochlear implantation. Acta Otolaryngol. 2003; 123:493–498. [PubMed: 12797584]

Hedrick MS, Carney AE. Effect of relative amplitude and formant transitions on perception of place of articulation by adult listeners with cochlear implants. J Speech Lang Hear Res. 1997; 40:1445–1457. [PubMed: 9430763]

Helms J, Müller J, Schön F, et al. Evaluation performance with the COMBI 40 cochlear implant in adults: A multicentric clinical study. J Otorhinolaryngol Head Neck Surg. 1997; 59:23–35.

Iverson P. Evaluating the function of phonetic perceptual phenomena within speech recognition: An examination of the perception of /d/-/t/ by adult cochlear implant users. J Acoust Soc Am. 2003; 113:1056–1064. [PubMed: 12597198]

Lane H, Denny M, Guenther FH, et al. On the structure of phoneme categories in listeners with cochlear implants. J Speech Lang Hear Res. 2007; 50:2–14. [PubMed: 17344544]

Liberman AM, Harris KS, Hoffman HS, et al. The discrimination of speech sounds within and across phoneme boundaries. J Exp Psychol. 1957; 54:358–368. [PubMed: 13481283]

Luce PA, Cluff MS. Delayed commitment in spoken word recognition: Evidence from cross-modal priming. Percept Psychophys. 1998; 60:484–490. [PubMed: 9599997]

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear Hear. 1998; 19:1–36. [PubMed: 9504270]

Marslen-Wilson WD. Functional parallelism in spoken word-recognition. Cognition. 1987; 25:71–102. [PubMed: 3581730]

McClelland JL, Elman JL. The TRACE model of speech perception. Cogn Psychol. 1986; 18:1–86. [PubMed: 3753912]

McMurray B, Jongman A. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. Psychol Rev. 2011; 118:219–246. [PubMed: 21417542]

McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. Cognition. 2002; 86:B33–B42. [PubMed: 12435537]

McMurray B, Aslin RN, Tanenhaus MK, et al. Gradient sensitivity to within-category variation in words and syllables. J Exp Psychol Hum Percept Perform. 2008; 34:1609–1631. [PubMed: 19045996]

McMurray B, Tanenhaus MK, Aslin RN. Within-category VOT affects recovery from "lexical" garden paths: Evidence against phoneme-level inhibition. J Mem Lang. 2009; 60:65–91. [PubMed: 20046217]

McMurray B, Samelson VM, Lee SH, et al. Individual differences in online spoken word recognition: Implications for SLI. Cogn Psychol. 2010; 60:1–39. [PubMed: 19836014]

McMurray B, Munson C, Tomblin JB. Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. J Speech Lang Hear Res. 2014; 57:1344–1362. [PubMed: 24687026]

Mehr MA, Turner CW, Parkinson A. Channel weights for speech recognition in cochlear implant users. J Acoust Soc Am. 2001; 109:359–366. [PubMed: 11206164]

Miller JL. Internal structure of phonetic categories. Lang Cognit Proc. 1997; 12:865–869.

Moberly AC, Lowenstein JH, Tarr E, et al. Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing? J Speech Lang Hear Res. 2014; 57:566–582. [PubMed: 24686722]

Munson B, Nelson PB. Phonetic identification in quiet and in noise by listeners with cochlear implants. J Acoust Soc Am. 2005; 118:2607–2617. [PubMed: 16266181]

Niparko, J. Cochlear Implants: Principles and Practices. 2nd ed.. Lippincott Williams & Wilkins; Philadelphia, PA: 2009.

Nittrouer S, Pennington B. New approaches to the study of childhood language disorders. Curr Dir Psychol Sci. 2010; 19:308–313. [PubMed: 22707856]

Oh SH, Kim CS, Kang EJ, et al. Speech perception after cochlear implantation over a 4-year time period. Acta Otolaryngol. 2003; 123:148–153. [PubMed: 12701730]

Peng SC, Chatterjee M, Lu N. Acoustic cue integration in speech intonation recognition with cochlear implants. Trends Amplif. 2012; 16:67–82. [PubMed: 22790392]

Repp BH. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychol Bull. 1982; 92:81–110. [PubMed: 7134330]

Stickney GS, Zeng FG, Litovsky R, et al. Cochlear implant speech recognition with speech maskers. J Acoust Soc Am. 2004; 116:1081–1091. [PubMed: 15376674]

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, et al. Integration of visual and linguistic information in spoken language comprehension. Science. 1995; 268:1632–1634. [PubMed: 7777863]

Toscano JC, McMurray B. Cue integration with categories: A statistical approach to cue weighting and combination in speech perception. Cognit Sci. 2010; 34:436–464.

Turner CW, Gantz BJ, Vidal C, et al. Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. J Acoust Soc Am. 2004; 115:1729–1735. [PubMed: 15101651]

Tyler RS, Parkinson AJ, Woodworth GG, et al. Performance over time of adult patients using the Ineraid or nucleus cochlear implant. J Acoust Soc Am. 1997; 102:508–522. [PubMed: 9228814]

Tyler RS, Parkinson AJ, Wilson BS, et al. Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization. Ear Hear. 2002; 23:98–105. [PubMed: 11951854]

Winn MB, Chatterjee M, Idsardi WJ. The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. J Acoust Soc Am. 2012; 131:1465–1479. [PubMed: 22352517]

Woodson EA, Reiss LA, Turner CW, et al. The hybrid cochlear implant: A review. Adv Otorhinolaryngol. 2010; 67:125–134. [PubMed: 19955729]

Zhang T, Dorman MF, Spahr AJ. Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. Ear Hear. 2010a; 31:63–69. [PubMed: 20050394]

Zhang T, Spahr AJ, Dorman MF. Frequency overlap between electric and acoustic stimulation and speech-perception benefit in patients with combined electric and acoustic stimulation. Ear Hear. 2010b; 31:195–201. [PubMed: 19915474]
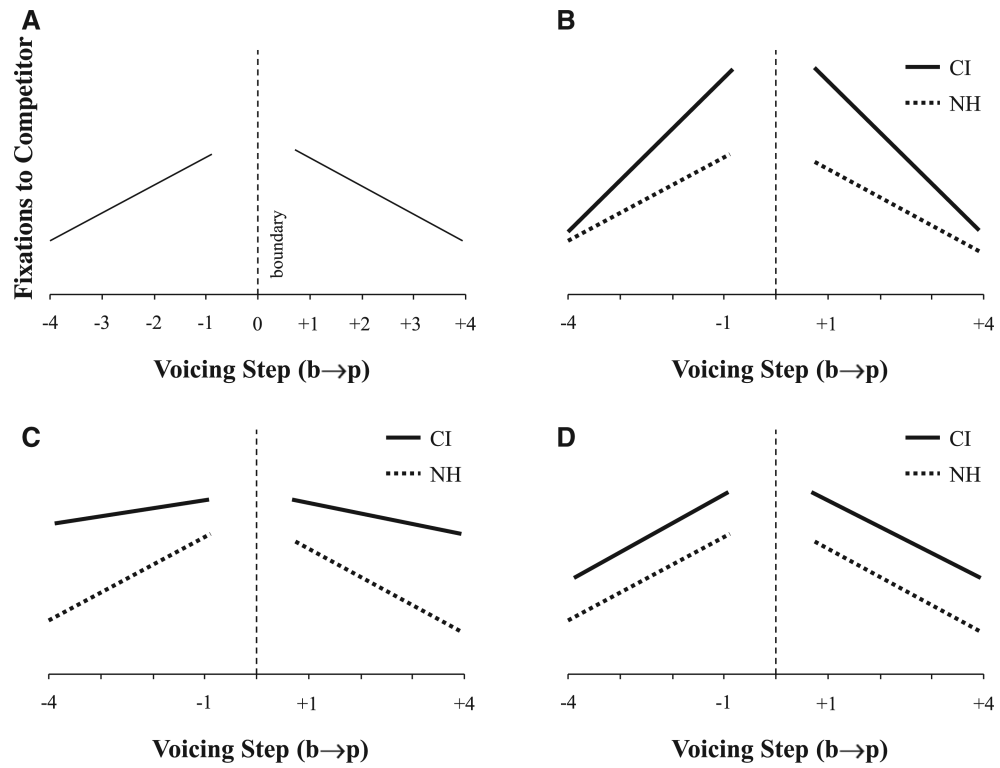
Fig. 1.
Hypothesized looks to the competitor as a function of distance from the category boundary. A, Schematic of prior results with NH individuals showing a gradient effect of distance. B, Predictions if CI users show heightened sensitivity to fine-grained detail. C, Predictions if CI users show less sensitivity to fine-grained detail, but heightened lexical activation for competitors. D, Predictions if CI users increase activation for lexical competitors but do not modulate sensitivity to fine-grained detail. CI indicates cochlear implant; NH, normal hearing.
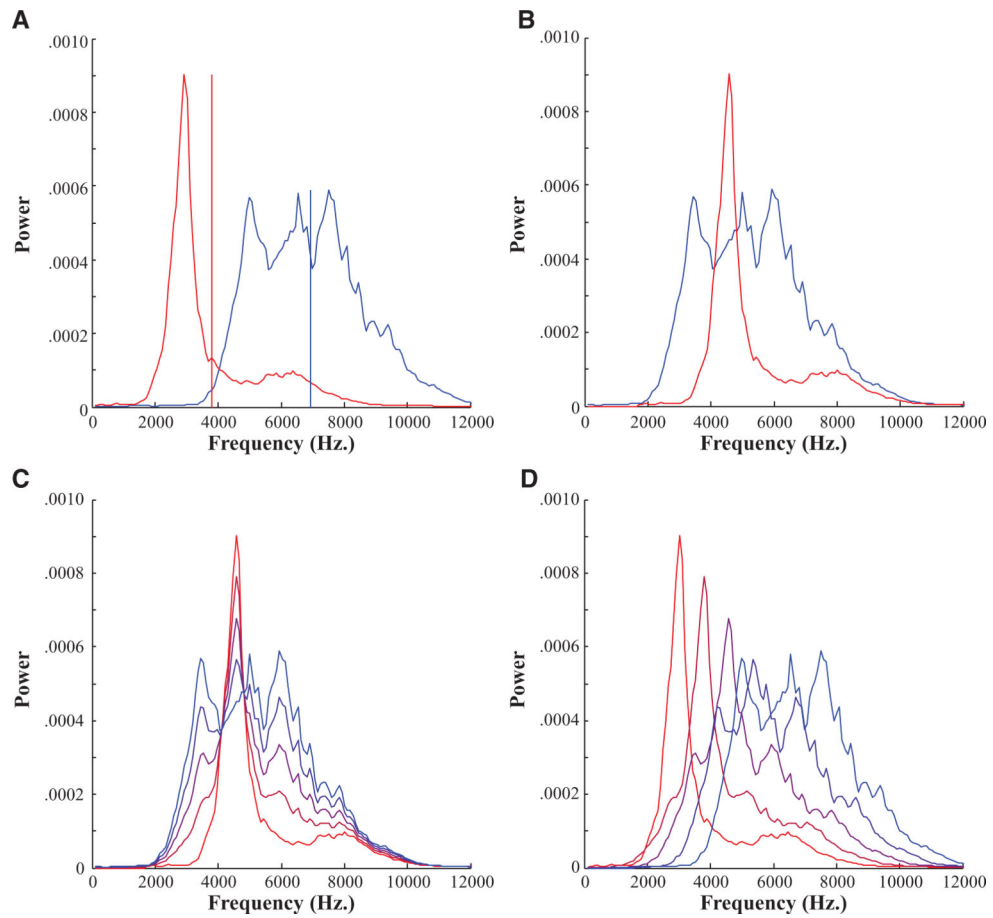
Fig. 2.
Schematic of construction of fricatives. A, First Long Term Average Speech Spectrum (LTASS) were extracted for the /s/ and /ʃ/ tokens and the spectral means (vertical lines) were identified. B, Next spectra were aligned to have the same means. C, Spectra were averaged in eight steps. D, Finally, spectra were realigned to have spectral means varying in eight steps.
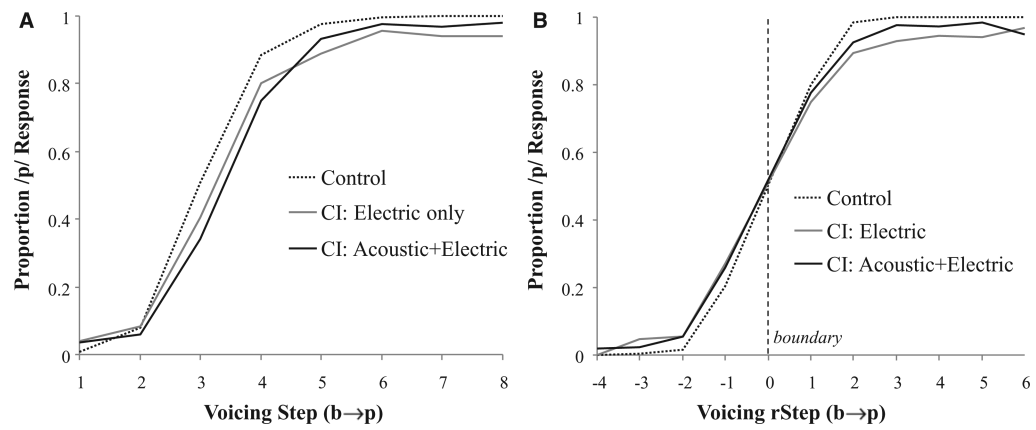
Fig. 3.
Mouse click (identification) of the voicing continuum. A, As a function of absolute voicing step; (B) with step computed relative to each subject × continuum category boundary (rStep).
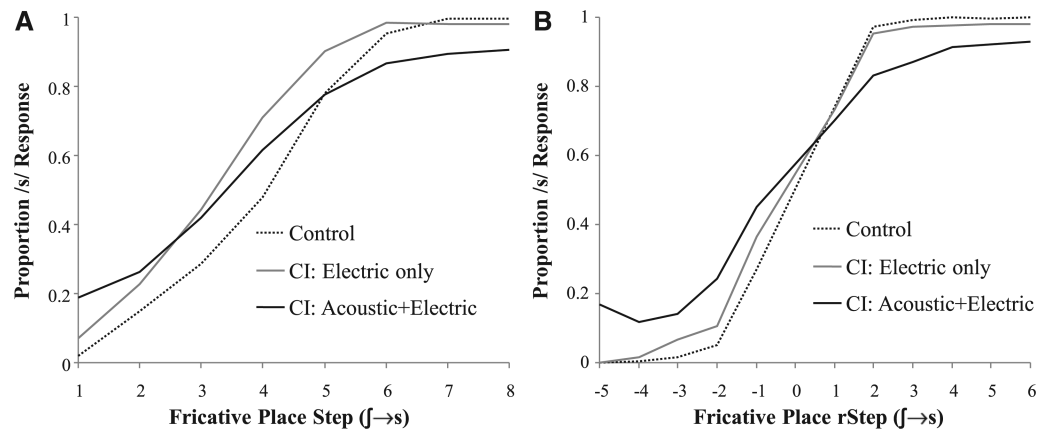
Fig. 4.
Mouse click (identification) of the fricative continuum. A, As a function of absolute step;
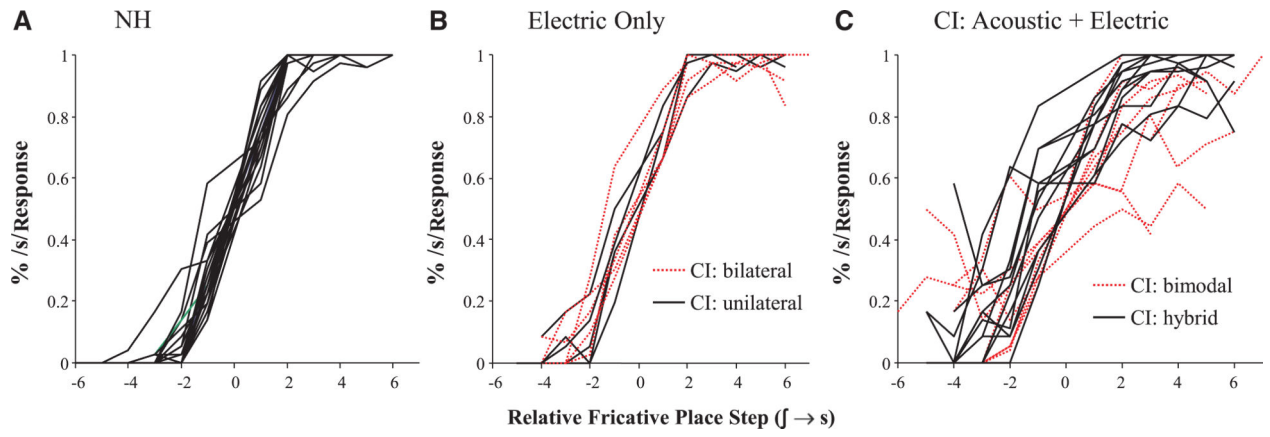(B) with step computed relative to each subject × continuum category boundary.

Fig. 5.
Individual identification data for each of the three groups of listeners. Step is computed relative to each subjects/item category boundary.
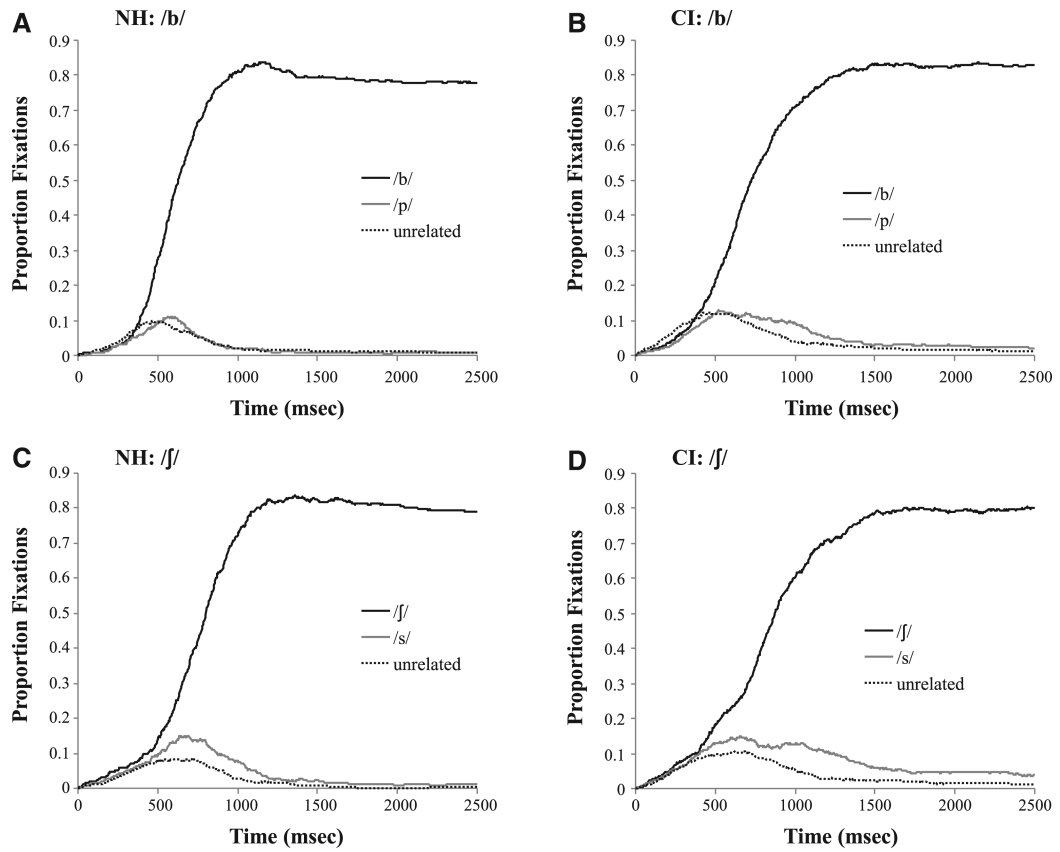
Fig. 6.
Timecourse of fixating the target, competitor, and unrelated objects. A, For NH listeners after hearing a good /b/ (step 1). B, For all CI users after hearing a good /b/. C, For NH listeners after hearing a good /ʃ/ (step 1). D, For all CI users after hearing a good /ʃ/. CI indicates cochlear implant; NH, normal hearing.
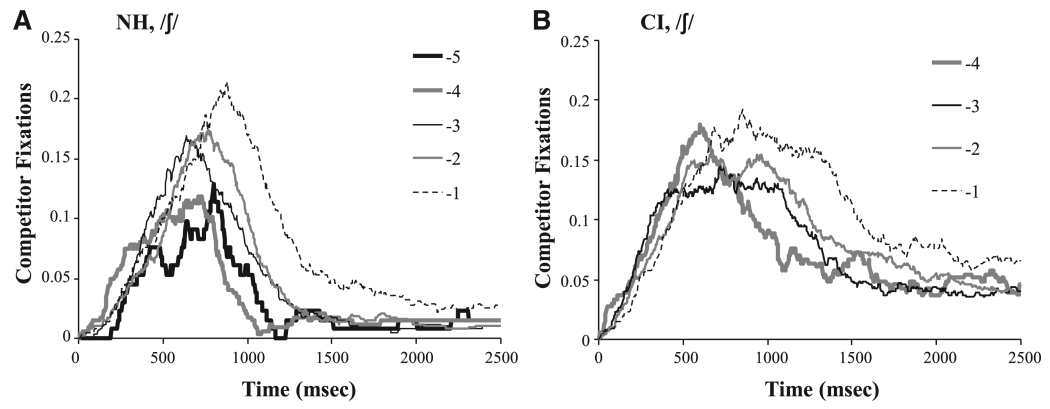
Fig. 7.
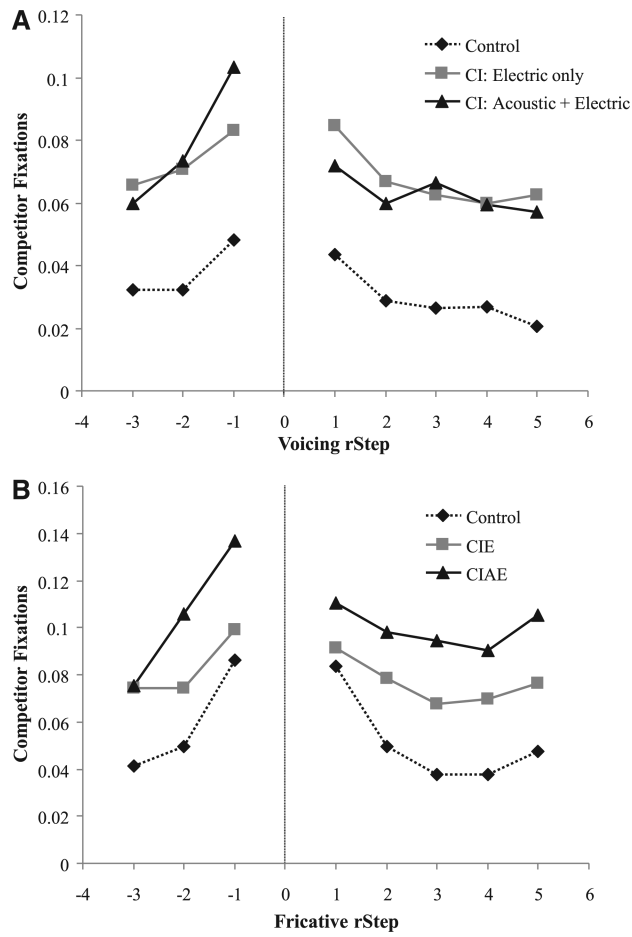Looks to the competitor as a function of time and relative step for the /ʃ/ side of the fricative continua.

Fig. 8.
Looks to the competitor (area under the curve) as a function of rStep and listener group for the voicing continua (A) and the fricative continua (B).

TABLE 1

Summary of participants

Group	N	Average Years of Device Use
Normal hearing	22	
Electric only		
Unilateral	5	14.2
Bilateral	5	6.8
Acoustic + electric		
Bimodal	8	3.4
Hybrid	12	5.8

TABLE 2

Mixed effects models examining looks to the competitor object for the voiced (b/p) continua

Effect	*B*	SE	*T*	df	*P*
Voiced (/b/)					
rStep	0.014	0.003	4.8	77.6	<0.0001*
rStep^2	0.007	0.003	2.2	43.5	0.035*
CI vs. NH	0.031	0.008	4.1	50.5	0.0001*
CI$_E$ vs. CI$_{AE}$	−0.007	0.010	−0.7	47.6	0.49
Unrelated looks (covariate)	0.312	0.063	4.9	725.9	<0.0001*
rStep × CI vs. NH	0.009	0.006	1.5	83.0	0.15
rStep × CI$_E$ vs. CI$_{AE}$	0.019	0.008	2.5	65.3	0.014*
rStep^2 × CI vs. NH	0.000	0.007	0.1	47.9	0.95
rStep^2 × CI$_E$ vs. CI$_{AE}$	0.022	0.009	2.5	43.6	0.015*
Voiceless (/p/)					
rStep	−0.004	0.001	−4.3	44.9	<0.0001*
CI vs. NH	0.030	0.006	5.2	51.8	<0.0001*
CI$_E$ vs. CI$_{AE}$	−0.008	0.008	−1.1	50.1	0.29
Unrelated looks (covariate)	0.207	0.042	4.9	1596	<0.0001*
rStep × CI vs. NH	0.000	0.002	0.0	43.7	0.99
rStep × CI$_E$ vs. CI$_{AE}$	0.002	0.003	0.9	46.8	0.36

CI indicates cochlear implant; NH, normal hearing.

*
$P < .05$.

TABLE 3

Results of two linear mixed models examining competitor looks for the fricative continua

Effect	*B*	SE	*T*	df	*P*
Postalveolar					
rStep	0.019	0.003	6.2	95.9	<0.0001[*]
CI vs. NH	0.031	0.008	3.8	50.3	0.0003[*]
CI_E vs. CI_{AE}	0.027	0.011	2.5	52.3	0.016[*]
Unrelated looks (covariate)	0.448	0.082	5.5	902.4	<0.0001[*]
rStep × CI vs. NH	0.001	0.006	0.2	87.7	0.84
rStep × CI_E vs. CI_{AE}	0.010	0.008	1.2	105.9	0.24
Alveolar					
rStep	−0.007	0.002	−4.5	58.0	<0.0001[*]
$rStep^2$	0.005	0.001	5.1	68.3	<0.0001[*]
CI vs. NH	0.040	0.010	4.1	53.1	0.0002[*]
CI_E vs. CI_{AE}	0.023	0.013	1.8	51.5	0.082[†]
Unrelated looks (covariate)	0.158	0.064	2.4	1468	0.014[*]
rStep × CI vs. NH	0.006	0.003	1.8	58.6	0.073[†]
rStep × CI_E vs. CI_{AE}	0.004	0.004	0.9	52.8	0.38
$rStep^2$ × CI vs. NH	−0.002	0.002	−0.9	69.9	0.36
$rStep^2$ × CI_E vs. CI_{AE}	0.000	0.003	−0.1	58.9	0.92

CI indicates cochlear implant; NH, normal hearing.

[*] P < .05

[†] P < .01.