JSLHR

Research Article

Speech Rate Normalization and Phonemic Boundary Perception in Cochlear-Implant Users

Brittany N. Jaekel,[a] Rochelle S. Newman,[a] and Matthew J. Goupell[a]

Purpose: Normal-hearing (NH) listeners rate normalize, temporarily remapping phonemic category boundaries to account for a talker's speech rate. It is unknown if adults who use auditory prostheses called cochlear implants (CI) can rate normalize, as CIs transmit degraded speech signals to the auditory nerve. Ineffective adjustment to rate information could explain some of the variability in this population's speech perception outcomes.
Method: Phonemes with manipulated voice-onset-time (VOT) durations were embedded in sentences with different speech rates. Twenty-three CI and 29 NH participants performed a phoneme identification task. NH participants heard the same unprocessed stimuli as the CI participants or stimuli degraded by a sine vocoder, simulating aspects of CI processing.
Results: CI participants showed larger rate normalization effects (6.6 ms) than the NH participants (3.7 ms) and had shallower (less reliable) category boundary slopes. NH participants showed similarly shallow slopes when presented acoustically degraded vocoded signals, but an equal or smaller rate effect in response to reductions in available spectral and temporal information.
Conclusion: CI participants can rate normalize, despite their degraded speech input, and show a larger rate effect compared to NH participants. CI participants may particularly rely on rate normalization to better maintain perceptual constancy of the speech signal.

No two talkers produce the exact same speech signals: Differences in speaking rate, age, dialect, and other factors can create acoustic variability in speech (Jacewicz, Fox, O'Neill, & Salmons, 2009). In terms of speech rate, the listener can potentially resolve some of this acoustic variability using *rate normalization* (Miller, 1981; Newman & Sawusch, 2009). Rate normalization is the process by which the perception of speech sounds with similar acoustical properties is altered on the basis of sentence context and the talker's rate of speech.

Speech rate can affect the perception of short contrastive speech sounds called *phonemes*. The durations of certain phonemes are particularly affected by speech rate (Crystal & House, 1982; Miller & Grosjean, 1981). As some phonemic contrasts are signaled by their respective durations (Klatt, 1976; Summerfield, 1981), the parameters by which those phonemes are identified could change under various speech rate contexts. For example, the word-initial velar stop consonants /g/ and /k/ are largely contrasted on the basis of durational differences in *voice-onset time* (VOT), with /k/ having a longer VOT duration than /g/ (Summerfield, 1981). VOT is defined as the length of time between a stop consonant's noisy release burst and the beginning of the periodic pulses associated with vocal fold vibration. A potential perceptual difficulty arises with changes in speech rate because a longer VOT duration could occur either because the talker intended to produce a /k/ rather than a /g/, or because the talker was simply speaking more slowly. Thus, the listener needs to interpret the underlying cause of the longer VOT duration. If a listener perceived phonemic categories in a fixed way (i.e., did not rate normalize), word misperceptions could occur across speaking rates. In reality, normal-hearing (NH) listeners *adapt* to speech rate, remapping certain phonemic boundaries in response to new speech rates (Miller & Volaitis, 1989; Newman & Sawusch, 2009). Rate normalization is thus a crucial tool for speech perception (but see Nakai & Scobbie, 2016 for an alternative viewpoint). Furthermore, it may be an obligatory process (Sawusch & Newman, 2000) and has been observed in infants as young as 2 to 4 months old (Miller, 1981; Eimas & Miller, 1980).

[a]Department of Hearing and Speech Sciences, University of Maryland, College Park
Correspondence to Brittany N. Jaekel: jaekel@umd.edu
Editor: Nancy Tye-Murray
Associate Editor: Richard Dowell
Received December 11, 2015
Revision received May 4, 2016
Accepted October 14, 2016
https://doi.org/10.1044/2016_JSLHR-H-15-0427

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

The extent to which signal degradation affects rate normalization is unknown. A degraded speech signal may contain fewer acoustic phonetic properties important for speech, and thus phoneme boundaries may be less likely to be informed or altered by incoming rate information. One group that experiences degraded speech signals are severe-to-profoundly hearing-impaired individuals who have been fitted with *cochlear implants* (CI), auditory prostheses that can partially restore speech understanding. The CI relays auditory information by electrically stimulating the auditory nerve. However, because of technological constraints, CIs typically provide spectrally and temporally degraded signals, and thus an informationally sparse representation of speech to the listener. The speech information is spectrally presented on 12 to 22 different electrodes or channels, which is far worse spectral resolution than what occurs in typical hearing. Furthermore, CI users generally have functional access to only about eight channels at any one time (e.g., Friesen, Shannon, Başkent, & Wang, 2001). The fast acoustical changes in the signal, called the *temporal fine structure*, are omitted in most CI speech processing strategies, and thus only the slowly varying changes in the temporal envelope remain (Loizou, 2006). Despite the spectral and temporal degradation, speech can be highly intelligible through a CI in quiet listening conditions, though speech perception outcomes across CI users are quite variable (e.g., Liu, Del Rio, Bradlow, & Zeng, 2004; Nie, Barco, & Zeng, 2006). Tests of speech intelligibility in this population typically involve identifying sentences or single words out of either open or closed sets, and are measured in percent words correct (Ching, van Wanrooy, & Dillon, 2007). When speech is not identified correctly, these tests unfortunately do not usually help us understand why the error occurred.

How CI users perceive specific phonemes, and which acoustic phonetic cues might be available for identifying phonemes, can be measured using more fine-grained methods such as confusion matrices (Blamey & Clark, 1990; Dorman, Dankowski, McCandless, & Smith, 1989; Munson, Donaldson, Allen, Collison, & Nelson, 2003), cue-weighting paradigms (Moberly et al., 2014; Winn, Chatterjee, & Idsardi, 2012; Winn & Litovsky, 2015), and phoneme continua with a single manipulated cue (Iverson, 2003). Furthermore, phoneme perception has also been studied using NH groups listening to degraded stimuli, in an attempt to control certain signal properties (e.g., spectral resolution) or with synthetic stimuli, to tightly control available acoustic cues to the listener (Nittrouer & Lowenstein, 2014; Souza & Rosen, 2009; Van Tasell, Soli, Kirby, & Widin, 1987; Winn, Chatterjee, & Idsardi, 2013; Winn & Litovsky, 2015).

For good consonant perception, access to spectral information appears to be key (Zeng & Galvin, 1999). However, specifically for identifying consonant voicing, it appears that few "channels" of spectral information are needed (Dorman, Loizou, & Rainey, 1997). Rosen (1989) implicated the temporal cues of periodicity and temporal fine structure as most important for conveying information about consonant voicing and first formant transitions, and suggested that reductions in spectral resolution should lead CI users to generally increase their use of available temporal information, such as the speech envelope. The contrast of interest in the present study is /g/ versus /k/, which is based mostly on the voiced–voiceless distinction, or consonant voicing. The presence of voicing is likely well represented by the CI, at least in identification tests of phonemes produced in an isolated syllable. This is supported by evidence from confusion matrices showing that CI users rarely identify /g/ as a /k/ or vice versa (Munson et al., 2003; Tyler & Moore, 1992; Välimaa, Määttä, Löppönen, & Sorri, 2002).

Whether the potentially well-represented voicing information by the CI results in phoneme perception similar to NH listeners is less clear, as is how perception of this voicing information might be affected by speaking rate. CI users showed variability in phoneme boundary locations when identifying the voiced /dɑ/ and voiceless /tɑ/ across a synthetic continuum with only VOT duration purposely manipulated (Iverson, 2003), and overall usually required longer VOT durations to change their perception from /dɑ/ to /tɑ/ than NH listeners. VOT duration is considered a temporal cue, to which CI users likely have access (Rosen, 1989). However, all CI users did not appear to identify phonemes in the same way, indicating that different users may have access to different or additional cues in the speech signal. For example, Iverson (2003) believed some CI participants in his study may have been sensitive to certain changes in the first formant frequency, which would add additional evidence for the listener identifying a voiced versus voiceless consonant.

Which acoustic-phonetic cues are most helpful for CI users (and for NH listeners presented *vocoded* or degraded speech meant to model aspects of CI processing) has been more thoroughly studied with cue-weighting paradigms. Though NH listeners and CI users may perform similarly on recognizing certain phonemes, the cues CI users are utilizing to identify phonemes may be very different. Moberly et al. (2014) found much variability in the cues used by CI users to identify /bɑ/ versus /wɑ/, syllables that contain both amplitude and spectral contrasts. Some CI users weighted formant rise time heavily, which mirrored behavior in NH listeners, and others weighted amplitude envelope information more heavily (a temporal cue predicted to be well represented in CI processing). Other CI listeners used neither cue. The reduced spectral information transmission of CI processors was thought to be the cause for CI users' weighting of "coarse" spectral cues over cues requiring fine spectral resolution (Winn & Litovsky, 2015). Likewise, when NH listeners heard speech with reduced spectral resolution similar to a CI, their utilization of the fine spectral cue was reduced. There are indications from these tasks that CI users with strong word recognition may have more access to spectral information, and with better spectral information, are able to utilize spectral cues to better identify phonemes. Moberly et al. (2014) observed that CI users who heavily weighted spectral cues over amplitude cues had the highest word recognition scores, as did CI users who strongly weighted the fine spectral cue (Winn & Litovsky, 2015).

If CI users are relying on different properties of the incoming speech signal for phoneme identification, we may

observe differences in the rate normalization effect on their phoneme boundaries compared to NH listeners. Sources of speech rate information include general rhythmic information of the preceding phrase, on the basis of syllabic durations and stress patterns (Kidd, 1989), and duration of the phonemic segment(s) immediately adjacent to the target sound—both preceding and subsequent (K. P. Green, Stevens, & Kuhl, 1994; Lotto, Kluender, & Green, 1996; Miller & Liberman, 1979; Sawusch & Newman, 2000). Durations of consonants and vowels, as well as syllable durations and stress information, are expected to be primarily temporal in nature, and represented by the CI processor in the form of temporal envelopes (Rosen, 1989). Thus, to clarify, though we believe CI users will perceive and apply the speech rate information provided in the contextual sentence, how that speech rate information is applied to phoneme boundaries might differ from NH listeners. If speech rate information is not applied to phoneme boundaries in a typical way, this could help explain CI users' particular difficulty with understanding rapidly spoken speech (Iwasaki, Ocho, Nagura, & Hoshino, 2002).

A final concern is the shallow slopes of the phonemic boundaries reported for CI users and NH listeners presented vocoded speech (Iverson, 2003; Winn et al., 2012). Shallow slopes indicate a more continuous than categorical phonemic boundary (Munson & Nelson, 2005), and may imply that phonemic boundaries are less clear or reliable (Gordon-Salant, Yeni-Komshian, & Fitzgibbons, 2008). How the presence of shallow slopes affects the ability to adjust phonemic category mappings in response to speech rate is of interest.

To summarize, it is unknown how CI users' perceptions of certain phonemes are affected by speaking rate. It could be hypothesized that CI participants may show atypically smaller rate normalization compared to NH participants. If a VOT category boundary's slope is shallow, then changes in the boundary's location due to rate might be very subtle and less perceptible to the listener. However, an alternative hypothesis is that CI participants may show atypically large rate normalization compared to NH participants. With reduced spectral resolution, and typically a larger reliance on temporal cues, the effects of speech rate (which largely affect temporal cues) may be exaggerated. It has been shown previously that speech signal degradation may *increase* the importance of utilizing a duration cue such as VOT. Miller and Wayland (1993), studying NH listeners, showed how a transition duration cue was particularly affected by speech rate, and became increasingly important when part of a degraded signal. They studied how perception of a naturalistically produced "bɑ/wɑ" contrast was affected by speaking rate in the presence of speech-shaped noise, which was meant to degrade the quality of the speech signal. When the contrast was presented in quiet, a negligible rate effect was observed. Miller and Wayland argued that this indicated that listeners were using rate-independent properties to complete the task, such as the location of the formant frequency at vowel onset. The rate effect emerged when the contrast was presented in noise, and the rate-dependent property—transition duration—became the prominent cue used by listeners. Thus, when speech stimuli became degraded, NH listeners discarded certain cues for distinguishing phonemic contrasts and used a rate-dependent property as a cue instead. We could similarly expect CI participants to utilize and rely exclusively on a rate-dependent property such as VOT duration and show a stronger rate effect with a degraded signal, especially because spectral resolution would be poorer, and thus perhaps fewer redundant acoustic phonetic cues are available in the signal. This would indicate that rate normalization is not only possible in this group but is relied upon to a greater extent than in NH listeners during speech perception.

Experiment 1: Rate Normalization in CI Participants

The first experiment examined the extent to which CI participants rate normalized. The target phonemes chosen for this experiment were /g/ and /k/, which are primarily differentiated by VOT duration, with /g/ having a shorter VOT duration than /k/. Using these speech sounds as endpoints of a continuum of VOT durations, we have the ability to measure the location, reliability, and relationship of CI participants' phonemic boundaries to speech rate changes.

Method

Participants

Twenty-three people with CIs participated in this study. The mean age of CI participants was 55.5 years (*SD* = 16.8 years), with a range of 21 to 81 years. Additional demographic information such as age at testing, duration of deafness estimated by patient report, years of CI experience, and CI brand is presented in Table 1. Spoken word recognition scores, as measured with Institute of Electrical and Electronics Engineers (IEEE) sentences (IEEE Subcommittee on Subjective Measurements, 1969), were obtained for each participant and are also presented in Table 1.

Stimuli

The speech stimuli were a subset of those used in Newman and Sawusch (2009). A male Midwestern American English speaker naturalistically produced the sentence "I heard him say the word /gaɪp/" at fast, normal, and slow rates. The nonword /gaɪp/ was excised from the sentences, resulting in three precursor sentences of the following lengths and rates: 753 ms or 8.0 syllables/s (fast), 971 ms or 6.2 syllables/s (medium), and 1220 ms or 4.9 syllables/s (slow). Although the entire sentence duration changed with speaking rate, prior work suggests that a driving factor in speaking rate perception is the duration of the immediately preceding phonemes (i.e., those occurring in "word"; K. P. Green et al., 1994; Lotto et al., 1996; Miller & Liberman, 1979; Sawusch & Newman, 2000). The durations of these preceding phonemes are presented in Table 2.

The speaker next produced the nonwords /gaɪp/ and /kaɪp/ (henceforth, *gipe* and *kipe*) in isolation, which were both 200 ms in duration. Using nonwords as stimuli reduced

Table 1. Demographics of cochlear-implant (CI) participants.

Code	Sex	Age at testing (yrs)	Age at onset (yrs)	Duration of deafness (yrs)	No. of implants	Age at (first) implant	CI use duration (yrs)	CI brand	Activated electrodes / Total electrodes[a]	IEEE Scores (%)
CBQ	M	21	0	4	Unilateral	4	17	Advanced Bionics	16 of 16	63
CAR	M	23	3	11	Unilateral	14	9	Cochlear	22 of 22	83
CAT	M	27	10	11	Unilateral	21	6	Cochlear	21 of 22	77.5
CBP	F	35	11	9	Unilateral	20	15	Cochlear	20 of 22	95
CBO	M	41	2.5	36.5	Bilateral	39	2	Advanced Bionics	16 of 16	13
CBN	M	50	0	39	Unilateral	39	11	Cochlear	22 of 22	40
CBJ	F	52	17	33	Unilateral	50	2	Med El	10 of 12	91.5
CAX	M	53	49	0	Bilateral	49	4	Cochlear	22 of 22	80
CBA	F	54	0	52	Bilateral	52	2	Cochlear	22 of 22	38.5
CBK	F	56	22	28	Unilateral	50	6	Cochlear	22 of 22	40
CBI	M	56	48	6.5	Unilateral	54.5	1.5	Med El	10 of 12	78.5
CBF	M	57	5	47	Unilateral	52	5	Cochlear	22 of 22	45
CAY	F	57	31	17	Bilateral	48	9	Cochlear	21 of 22	61.5
CBG	F	61	4	53	Unilateral	57	4	Cochlear	22 of 22	32.5
CBR	F	62	0	56	Bilateral	56	6	Cochlear	22 of 22	72.5
CBV	F	62	18	40	Unilateral	58	4	Cochlear	22 of 22	66.5
CAJ	F	63	0	47	Unilateral	47	16	Cochlear	22 of 22	0
CAK	M	69	57	1	Unilateral	58	11	Cochlear	21 of 22	32.5
CAO	F	70	6	59	Unilateral	65	5	Cochlear	22 of 22	31.5
CAM	F	70	40	23	Unilateral	63	7	Cochlear	21 of 22	62.5
CBC	F	77	69	1	Bilateral	70	7	Cochlear	22 of 22	21
CBD	M	79	74	0	Unilateral	74	5	Cochlear	21 of 22	52
CBB	M	81	77	1	Unilateral	78	3	Cochlear	22 of 22	69.5
Mean		55.5	23.6	25.2		48.6	6.8			56.7
SD		16.8	26.3	21.3		18.8	4.5			23.4

Note. IEEE = Institute of Electrical and Electronics Engineers.

[a]In preferred ear only.

Table 2. Duration of phonemes (ms) immediately preceding the target contrast.

Unprocessed speech	/wɜd/	Preceding /w/	Preceding /ɜ/	Preceding /d/	Target /g/ or /k/
Slow	315	125	90	100	14–72
Medium	250	90	75	85	14–72
Fast	165	80	35	50	14–72

the chance of *lexical access effects*—the perception of the stimuli being biased by one's level of familiarity with the words (Ganong, 1980). The transitional probabilities of the first pairs of phonemes—that is, the probability that the diphthong /aɪ/ will follow /g/ or /k/—were similar for the two nonwords, 0.0005 and 0.0006, respectively (Storkel & Hoover, 2010). The number of lexical neighbors for *gipe* and *kipe* (i.e., the number of words that exist when one phoneme is changed, added to, or deleted from the input) was 14 and 16, respectively (Vaden, Halpin, & Hickok, 2009). Various phoneme recognition studies indicate that adult CI users do not often confuse /g/ and /k/ for one another (Munson et al., 2003; Tyler & Moore, 1992; Välimaa et al., 2002), and that /g/ and /k/ are more likely to be confused with phonemes that are also voiced or voiceless stops (respectively), but which differ in place of articulation. This means that CI participants were not expected to have particular difficulty with identifying well-articulated pronunciations of *gipe* and *kipe*.

The two nonwords were manipulated to create a velar stop consonant series extending from the veridical *gipe* to the veridical *kipe*, with more ambiguous stimuli existing in the continuum between them. The *gipe* and *kipe* endpoints presented a durational contrast in terms of VOT, with *gipe* having a 14-ms VOT and *kipe* having a 72-ms VOT. After the veridical *gipe*, the next word in the series maintained a 14-ms VOT, but the release burst of *gipe* was deleted and replaced with a 14-ms release burst from *kipe*. The third word in the series additionally deleted the first vocal pulses of *gipe*, and replaced these vocal pulses with release burst plus aspiration from *kipe*, effectively creating a longer VOT (25 ms), while still maintaining overall syllable duration. This process of replacing additional vocal pulses with release plus aspiration from *kipe* was repeated for the rest of the series, creating VOTs of 36, 45, 54, 63, and 72 ms. In total, eight nonwords comprised the consonant series. Though VOT is believed to be the primary cue to consonant voicing (Francis, Kaganovich, & Driscoll-Huber, 2008; Klatt, 1976; Nagao & de Jong, 2007; Summerfield, 1981), a brief description of a potential secondary cue—the first formant frequencies of the target stimuli—is warranted (Hillenbrand, 1984; Lisker, 1975). The first formant frequency of the vowel onset (through 30 ms post-onset, analyzed with a Gaussian window) was 560 Hz in *gipe* and increased to 670 Hz in *kipe*. The first formant frequency was stable (~560 Hz) across the first three steps of the continuum, increased gradually from Steps 3 through 6, and was again stable (~670 Hz) across Steps 6 through 8. Waveforms and spectrograms of the *gipe* and *kipe* stimuli are presented in Figure 1.

The eight nonwords were appended individually to the end of the three precursor sentences, resulting in 24 different stimuli. The 24 items (3 Precursor sentence rates × 8 Target items with varied VOT) in random order composed a single block, and 24 blocks were presented. Thus, CI participants heard a total of 576 trials.

CI participants had the input delivered through the direct audio input of their devices via personal audio cables and listened to stimuli unilaterally. Stimuli were presented to participants over an external soundcard (Roland/Edirol, UA-25 EX, Los Angeles, CA) and amplifier (Crown Audio, Elkhart, IN), and had a sampling rate of 44.1 kHz prior to digital-to-analog conversion. Bilateral CI participants were directly connected via their self-reported better ear only. CI participants were asked to adjust volume to a comfortable listening level.

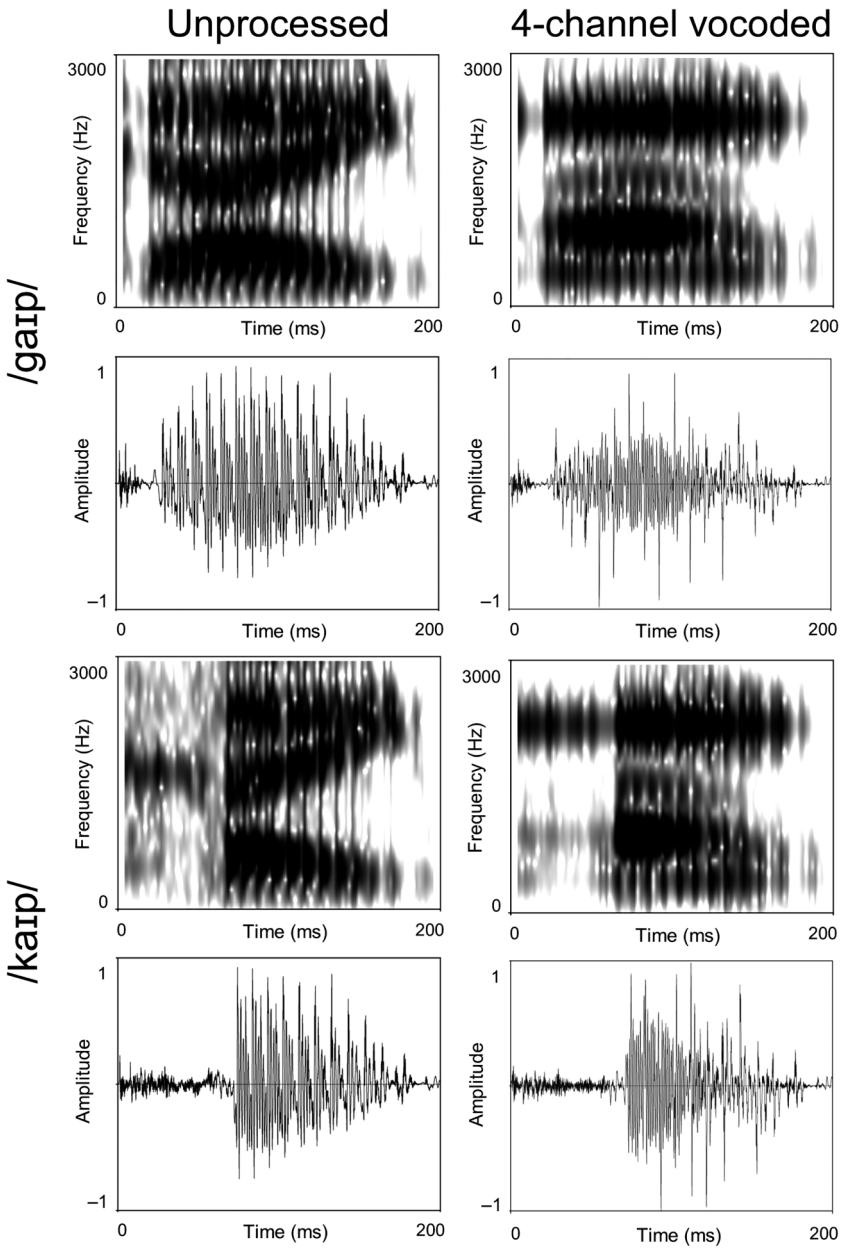
Procedure

Participants were seated in a double-walled sound-attenuating booth (Industrial Acoustics, Inc., Bronx, NY), and responded using a computer and mouse. The task was two-alternative forced-choice and was controlled by a computer via MATLAB (The Mathworks, Inc., Natick, MA). The testing screen was presented on a monitor and participants clicked the "Begin" button to play each sentence. Participants then identified whether the speaker said *gipe* or *kipe*. These two response options were presented as separate buttons on either side of the screen. Testing was self-paced, and participants were allowed to take breaks as needed.

Two techniques were used for analyzing the data: Generalized linear mixed-effects models with logit links revealed how participant factors such as chronological age and durations of deafness affected *gipe* and *kipe* responses, and repeated measures mixed analyses of variance (ANOVAs) examined how factors affected specific averaged results of interest. ANOVAS have been used in previous studies on rate normalization (e.g., Newman & Sawusch, 2009), and their inclusion here makes comparisons with prior work easier. This latter analysis required the calculation of the 50% cross-over point of the best fitting psychometric function tracing percent *kipes* heard across the VOT continuum—that is, the VOT category boundary—and the slope of this psychometric function (Wichmann & Hill, 2001), for each listening condition for each participant. VOT category boundaries and slopes were calculated using *psignifit*[1] in MATLAB, in which the data were fit by a four-parameter cumulative Gaussian

[1]The psychometric function slopes computed using *psignifit* were at times unreasonably steep, particularly in cases where the percent *kipes* went from near 0% to near 100% in a single step, which was a result of the numerical fitting procedure. For the present analyses, the steepest possible slope deemed appropriate was 0.11, or an 11% increase in percent *kipes* heard per millisecond, which was the maximum possible change in performance between any two steps in the *gipe*/*kipe* continuum (a 100% increase in percent *kipes* heard could only occur across an approximately 9-ms change in VOT duration, or one step).

Figure 1. The top two rows show the waveforms and spectrograms for the original unprocessed *gipe* (left) and for the four-channel sine vocoded *gipe* (right). The bottom two rows show the waveforms and spectrograms for the unprocessed (left) and four-channel sine vocoded (right) *kipe* stimuli.

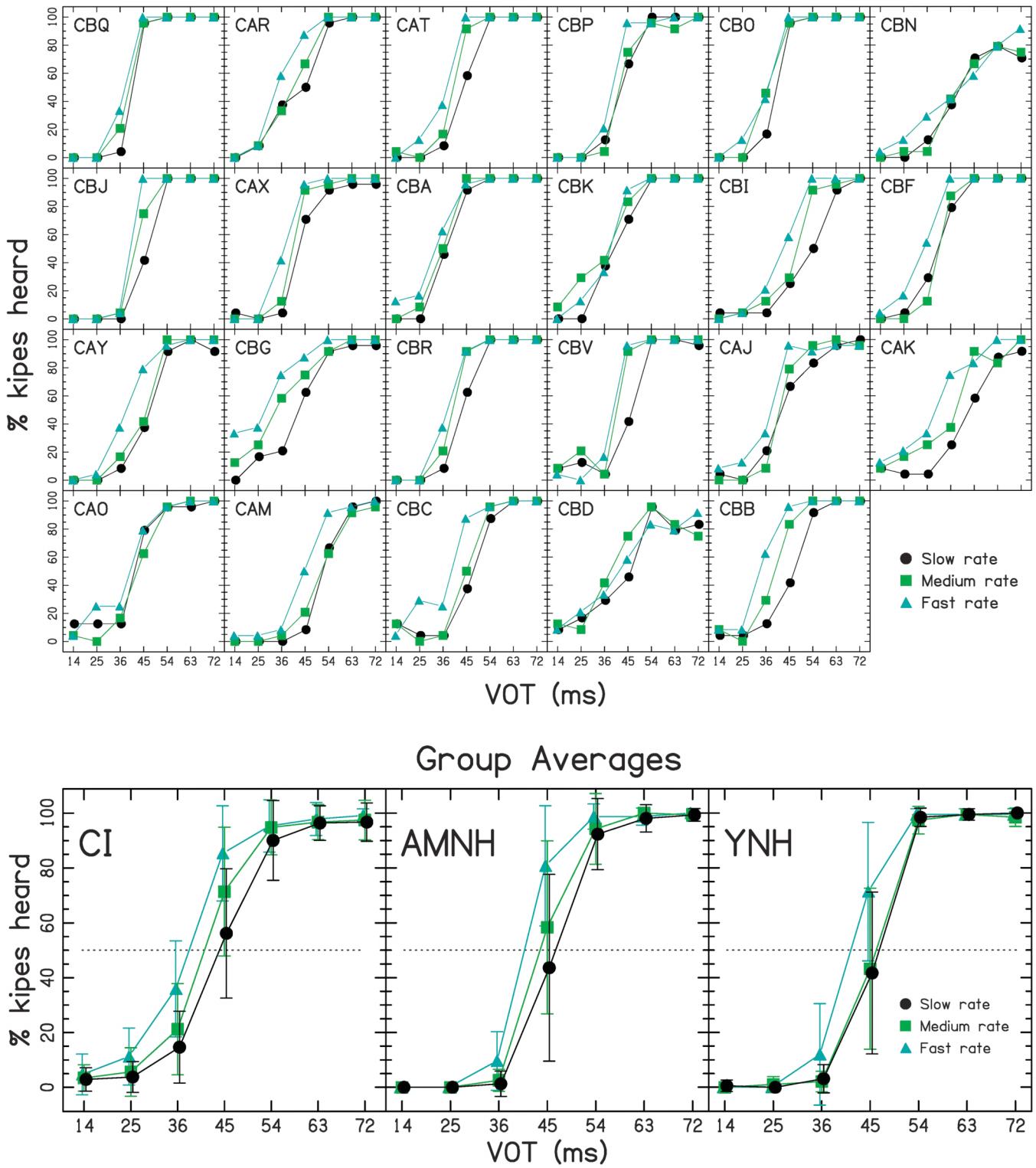


sigmoid function on the basis of maximum likelihood estimates. Psychometric functions were composed of percent *kipe* responses at Steps 2 through 8 of the *gipe*/*kipe* continuum. The first step, the veridical *gipe*, had the exact same VOT duration as the second step, which replaced the /g/ release with the /k/ release, and left VOT and syllable duration unaffected. Only scores from the second step in the continuum were retained for the analysis, to avoid a psychometric function with two data points at 14 ms. There were no significant differences in performance between Step 1 and Step 2 in any condition tested (all $p > .05$ using paired samples t tests).

Results

All CI participants had distinguishable *gipe* and *kipe* endpoints, meaning individual participants experienced a shift in perception from one phoneme to the other over the span of the continuum (see Figure 2). The percent *kipes* heard was approximately 0% at the 14-ms VOT and 100%

Figure 2. (Top) Individual results for percent *kipes* heard for the 23 CI participants, ordered by chronological age (youngest to oldest). Participant codes appear in the upper left corner. (Bottom) Average percent *kipes* heard in the unprocessed speech condition for the cochlear implant (CI) participants in Experiment 1 (left panel) and for the age-matched normal-hearing (AMNH; middle panel) and younger normal-hearing (YNH; right panel) participants in Experiment 2. Bars indicate ±1 *SD*. VOT = voice-onset time.



at the 72-ms VOT. The categorical nature of this perception was variable across CI participants, with some participants showing stark changes in perception at some point along the continuum (see participants CBQ and CBJ in Figure 2), and other participants showing more gradual changes (see participants CAK and CBN). The percent *kipes* heard increased with faster speech rates indicating a shift in the category boundary toward smaller VOT values. All CI participants rate normalized, per visual inspection of the individual and average plots in Figure 2.

For the first analysis, binary responses (*gipe* coded as 0 and *kipe* coded as 1) for all CI participants, from all speech rate conditions and test items, were fit with a generalized linear mixed-effects model with a logit link. The random effects were intercepts for participant and item. The fixed effects were speech rate (categorical, with slow as the referent), VOT (a continuous variable, centered at the average VOT value in the VOT continuum), chronological age (a continuous variable, centered at the average age of the group), and the interaction of VOT and age.

This optimal model for the data (see Table 3) was found by initially including all factors and interactions of interest in the regression model, and then removing the highest-order interaction that was least significant and rerunning the regression model until all remaining interactions were significant and all stand-alone main effects were significant. Nonsignificant main effects were kept in the model only if they were part of a significant interaction. The participant factors that were included in the initial model, but were found to be insignificant and thus removed, were: duration of deafness, IEEE score, and their interactions with the factors of speech rate and VOT.

Speech rate significantly affected *gipe* and *kipe* responses (see Table 3). CI participants were 1.5 times more likely to report *kipe* in the medium than in the slow speech rate conditions ($p = .001$) and 3.2 times more likely to report *kipe* in the fast than in the slow speech rate conditions ($p < .001$), holding other variables constant. VOT was also a significant factor, in that with each 1-ms increase in VOT, the number of *kipe* responses increased by a factor of 1.2 ($p < .001$), holding other variables constant. In other words, as VOT increased, CI participants were more likely to report hearing a *kipe* than a *gipe*. Chronological age was not a significant factor ($p = .197$), but the interaction of age and VOT was significant ($p = .001$), indicating that older CI participants were slightly less affected by changes in VOT than younger CI participants.

The second analysis investigated how VOT category boundaries and their slopes were affected by speech rate. A one-way repeated measures ANOVA analyzed the effects of speech rate for the CI participants (slow, medium, and fast; within subjects) on VOT category boundaries (see Figure 3). The effect of speech rate was significant, $F(2, 44) = 53.97$, $p < .001$, $\eta_p^2 = .71$, and post hoc paired-samples t tests, Bonferroni-corrected for multiple comparisons, revealed that VOT category boundaries decreased significantly (i.e., were located at shorter VOT durations) with each increase in speech rate ($p < .001$ for all comparisons). Because CI participants rate normalized, we measured the size of the *rate effect*, the difference in VOT category boundaries between the slow and fast speech rate conditions. The average rate effect was 6.6 ms ($SD = 3.7$) for CI participants (see Figure 4). The minimum rate effect observed was 1.7 ms, and the maximum rate effect was 15.8 ms.

A second one-way repeated measures ANOVA analyzed the effects of speech rate on psychometric function slopes, which appeared to be fairly similar across rates (see Figure 5). The effect of speech rate was not significant, $F(1.56, 34.33) = 0.30$, $p = .69$, $\eta_p^2 = .01$, Greenhouse–Geisser corrected.

Discussion

The effects of speech rate on VOT category boundaries indicated that CI participants rate normalized. Despite spectral and temporal degradations in the speech signal, CI participants gleaned rate information from the sentence context, applied this information in the remapping of phonemic category boundaries, and ultimately showed a change in their identification of phonemes. The pattern of performance was similar to rate normalization performance observed in NH participants from previous research (e.g., Newman & Sawusch, 2009). CI participants labeled certain identical acoustic stimuli differently on the basis of precursor sentence speech rate—that is, a change in the VOT category boundary location across different speech rate contexts meant that some stimuli that were identified as *gipe* in a slow rate context were identified as *kipe* in a fast rate context. Slopes, or the reliability or consistency with which CI participants identified *gipe* versus *kipe*, were generally shallow and unaffected by speech rate, but there was variability across participants (Figures 2 and 5). The regression analysis revealed that participant factors such as IEEE score and duration of deafness were not predictive of *gipe* and *kipe* responses, but that chronological age did seem to moderate the impact of VOT on responses. Therefore, whether age is also a significant factor in NH participants will be of interest in Experiment 2. The results from CI participants will be further evaluated below, where NH participant information can be included. In general, however, CI participants rate normalized, though slopes of the psychometric functions indicated a more continuous than categorical phonemic category boundary. CI users are able to utilize the information they do have to attempt to accommodate rate variability across sentences, and the lack of a reliable, categorical boundary did not disrupt the remapping process. Because certain identical stimuli were interpreted differently on the basis of the surrounding context, shallow slopes seem to indicate a difficulty with contrasts rather than a difficulty with hearing the targets themselves.

Experiment 2: Rate Normalization in NH Participants

The second experiment examined the extent to which NH participants rate normalized unprocessed stimuli and

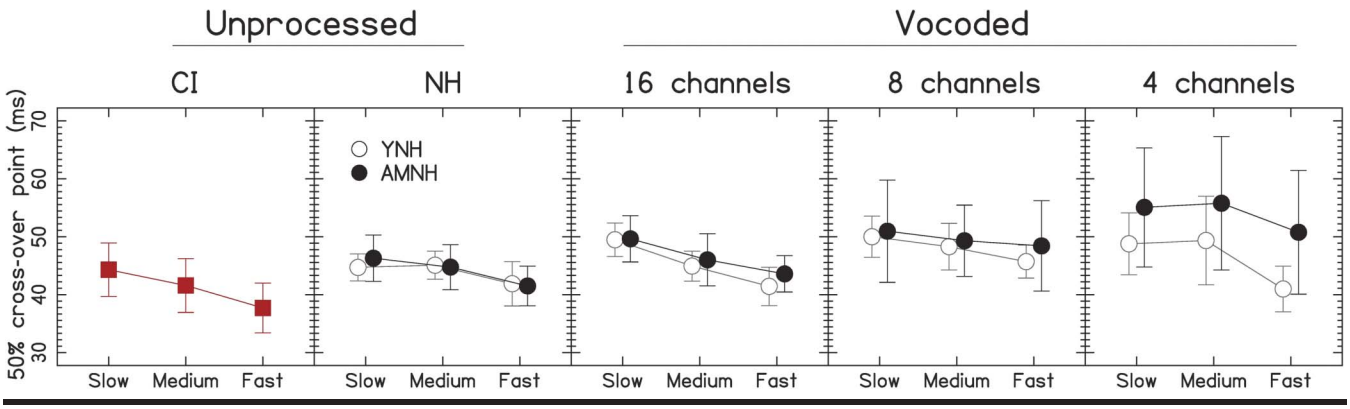
Table 3. Optimal logistic regression models for each analysis.

Model Terms	Coefficient	*SE*	*t*	*p*	Odds ratio
CI					
Intercept	–0.347	0.152	–2.28	.022	0.707
Speech rate					
Fast	1.151	0.135	8.52	<.001	3.161
Medium	0.423	0.122	3.45	.001	1.527
Age	–0.009	0.007	–1.29	.197	0.991
VOT	0.174	0.012	14.15	<.001	1.190
Age × VOT	–0.002	0.0007	–3.35	.001	0.998
NH					
Intercept	–1.850	0.253	–7.33	<.001	0.157
Speech rate					
Fast	1.350	0.179	7.55	<.001	3.857
Medium	0.257	0.157	1.63	.102	1.293
Signal degradation					
16-channel	–0.802	0.310	–2.59	.01	0.448
8-channel	0.279	0.292	0.96	.339	1.322
4-channel	0.784	0.319	2.46	.014	2.190
AMNH	–0.006	0.388	–0.02	.988	0.994
VOT	0.334	0.020	16.71	<.001	1.397
Speech rate × Signal degradation					
Fast, 16-channel	0.889	0.261	3.40	.001	2.433
Medium, 16-channel	1.130	0.261	4.34	<.001	3.096
Fast, 8-channel	–0.825	0.241	–3.42	.001	0.438
Medium, 8-channel	–0.022	0.185	–0.12	.905	0.978
Fast, 4-channel	–0.492	0.276	–1.78	.075	0.611
Medium, 4-channel	–0.255	0.208	–1.23	.221	0.775
AMNH × Signal degradation					
16-channel	–0.309	0.288	–1.07	.283	0.734
8-channel	–0.206	0.374	–0.55	.581	0.814
4-channel	–0.842	0.377	–2.23	.025	0.431
VOT × Signal degradation					
16-channel	–0.038	0.023	–1.68	.094	0.963
8-channel	–0.153	0.021	–7.29	<.001	0.858
4-channel	–0.200	0.022	–9.23	<.001	0.819
CI vs. AMNH					
Intercept	–0.342	0.139	–2.45	.014	0.710
Speech rate					
Fast	1.121	0.151	7.44	<.001	3.068
Medium	0.412	0.123	3.33	.001	1.510
Group					
AMNH, unprocessed	–1.875	0.376	–4.98	<.001	0.153
AMNH, 8-channel	–1.150	0.369	–3.12	.002	0.317
VOT	0.155	0.013	12.01	<.001	1.168
Group × Speech rate					
Fast, AMNH unprocessed	0.674	0.251	2.68	.007	1.962
Medium, AMNH unprocessed	0.192	0.248	0.77	.439	1.212
Fast, AMNH 8-channel	–0.928	0.253	–3.67	<.001	0.395
Medium, AMNH 8-channel	–0.369	0.207	–1.78	.075	0.691
Group × VOT					
AMNH unprocessed	0.194	0.043	4.56	<.001	1.214
AMNH 8-channel	0.004	0.020	0.19	.853	1.004
VOT × Speech rate					
Fast	0.011	0.008	1.40	.162	1.011
Medium	0.013	0.006	2.11	.035	1.013

Note. CI = cochlear-implant participants; VOT = voice-onset time; NH = normal-hearing participants; AMNH = age-matched normal-hearing participants.

stimuli processed through a vocoder, which simulated certain aspects of CI processing, to explore how degrading the speech signal might affect the process of rate normalization. Though the phonemic contrast in this experiment was thought to rely mostly on the temporal cue of duration (Klatt, 1976; Summerfield, 1981), NH participants likely additionally utilize cues besides durational information, such as spectral cues (Francis et al., 2008; Miller & Wayland, 1993; Stevens & Klatt, 1974; Winn et al., 2013). We might observe a different profile of rate normalization in NH participants compared to CI participants. With signal degradation, however, we expect NH

Figure 3. Average VOT category boundaries for cochlear-implant (CI) participants, and young normal-hearing (YNH; open circles) and age-matched normal-hearing (AMNH; closed circles) participants listening to four different spectral conditions. Bars indicate ±1 *SD*.



participants to show similar rate normalization behavior to CI participants. With the vocoding process reducing spectral resolution, we expected NH participants to rely more exclusively on duration cues (Donaldson, Rogers, Johnson, & Oh, 2015; Winn et al., 2012), showing a larger rate effect in vocoded conditions than in unprocessed conditions.

This experiment also examined effects of chronological age on rate normalization by testing two separate age groups of NH participants. First, word identification in older adults has been shown to be adversely affected by rate variability (Sommers, 1997), so older adults might perform less reliably in the present task. Second, the location of the category boundary along the VOT continuum may differ between the groups. Older adults with NH changed identifications of a word-initial /b/ to /p/ at longer VOTs compared to younger adults, when stimuli were presented at the end of a sentence (Gordon-Salant et al., 2008). This difference was attributed to age-related decline in the ability to detect short aspiration bursts. However, in the same study, older adults performed similarly to younger adults with a /gr/ and /kr/ VOT contrast. It is unclear if listeners in the present study will show an age difference in categorizing a /g/–/k/ VOT continuum, which has relatively shorter VOTs than the /gr/–/kr/ contrast used by Gordon-Salant et al. (2008). The above concerns were warranted because CI participants were on average middle-aged, and many participants were more than 65 years old. We wanted to rule out any confounds that would be a result of chronological age rather than use of a CI.

Figure 4. The mean change in category boundary, or rate effect, defined as the difference in VOT category boundaries between the slow and fast speech rate conditions, for cochlear-implant (CI), age-matched normal-hearing (AMNH), and younger normal-hearing (YNH) participants. NH participants were tested in four different spectral conditions. Bars indicate ±1 *SD*.

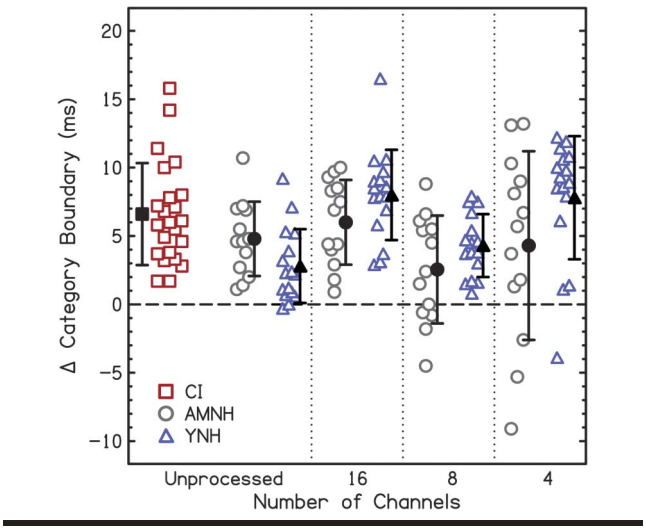
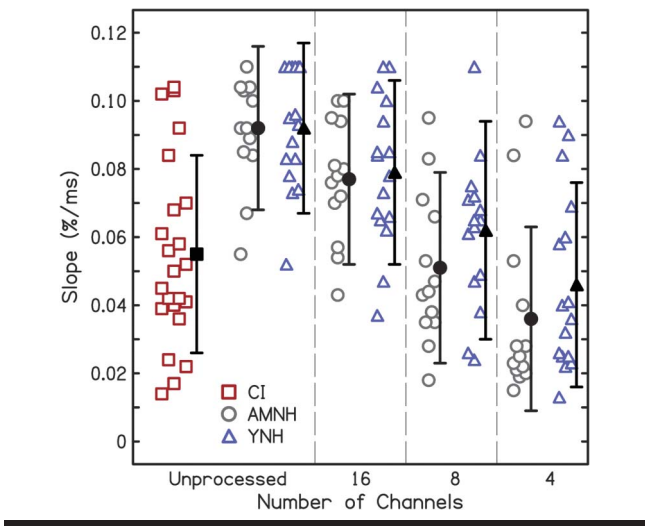


Figure 5. Average psychometric function slopes for cochlear-implant (CI), age-matched normal-hearing (AMNH), and younger normal-hearing (YNH) participants. NH participants were tested in four different spectral conditions. Bars indicate ±1 *SD*.



Method

Participants

Twenty-nine people with normal hearing participated in this study. Eligible participants were required to have hearing thresholds less than or equal to 25 dB HL at octave frequencies from 250 to 4000 Hz, differences in hearing threshold across ears at any tested frequency not exceeding 15 dB, and be self-reported native speakers of American English. Hearing screenings were administered with an audiometer (MAICO, MA-40, Eden Prairie, MN) in a double-walled sound-attenuating booth (Industrial Acoustics, Inc., Bronx, NY). Two separate groups of NH participants were recruited for this study: The young NH (YNH) group had 16 participants, with a mean age of 20.3 years (*SD* = 1.5 years, range 18 to 24 years), and the age-matched NH (AMNH) group had 13 participants, with a mean age of 56.9 years (*SD* = 10.8 years, range 35–73 years). The YNH group was similar in age to participants tested in most prior rate normalization studies (e.g., Newman & Sawusch, 2009). The AMNH group was recruited to match the average age of the CI participants in Experiment 1 at the group level. For the AMNH group, the average threshold for the left ear at 4 kHz was 10.4 dB HL (*SD* = 8.5, range −5 to 25 dB HL), and the average threshold at 8 kHz was 15.4 dB HL (*SD* = 11.8, range 5 to 50 dB HL). The average threshold for the right ear at 4 kHz was 7.3 dB HL (*SD* = 8.8, range −10 to 25 dB HL), and the average threshold at 8 kHz was 18.8 dB HL (*SD* = 12.3, range 0 to 45 dB HL). Comparing the upper range of AMNH participants' thresholds to those in YNH participants, YNH participants had maximum thresholds in the left ear of 20 dB HL at 4 kHz and 30 dB HL at 8 kHz, and maximum thresholds in the right ear of 20 dB HL at both 4 and 8 kHz. Thus, stimulus energy above 4 kHz might be less salient for some AMNH participants compared to the YNH participants, and could lead to slight decrements in performance.

Stimuli

NH participants heard the sentences and nonwords described in Experiment 1, which were *unprocessed*, containing their full spectral information. NH participants also heard vocoded versions of the original stimuli, which varied in spectral resolution (Dorman et al., 1997; Friesen et al., 2001; Loizou, 2006). The vocoding process consisted of the following procedure. Stimuli were bandpass filtered into 16, eight, or four channels using third-order Butterworth filters, and forward–backward filtering was applied to better preserve temporal cues in the vocoded signal (slopes = –36 dB/octave). The bands were contiguous and logarithmically spaced. The lower frequency boundary of the signal was 200 Hz and the upper frequency boundary was 8000 Hz. From these bands, envelopes were extracted, half-wave rectified, and second-order low-pass forward–backward filtered, using a cutoff frequency of 400 Hz. A sine-tone carrier with a frequency equal to the geometric center frequency of its band was then modulated by the extracted envelope for its band, after which all bands (either 16, 8, or 4) were recombined. The final vocoded stimuli had the same RMS energy as the original unprocessed stimuli and had a sampling rate of 44.1 kHz prior to digital-to-analog conversion.

A sine carrier was used to best preserve the incoming speech signal's temporal information by preserving the temporal envelope in a way similar to a CI's speech processor (Dorman et al., 1997; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Souza & Rosen, 2009; Whitmal, Poissant, Freyman, & Helfer, 2007). In contrast, a noise carrier would be less able to preserve the temporal envelope of speech, as noise carriers can create superfluous envelope fluctuations (Whitmal et al., 2007). Thus, it was anticipated that the sine carrier would maintain much of the durational information indicating speech rate, but similar to the CI processor, be unable to present much spectral information. The number of bandpass-filtered frequency channels was varied to analyze how different levels of signal degradation affected rate normalization. With a greater number of channels, more spectral detail was available to the participant. On average, CI users perform as if they have access to only about eight spectral channels at any one time (Friesen et al., 2001).

The first formant frequencies of the vocoded stimuli largely followed the trajectories described for the unprocessed stimuli across the VOT continuum. In the 16-channel condition, the first formant frequency of the vowel onset (through 30-ms postvowel onset) was 630 Hz for *gipe* and 730 Hz for *kipe*; in the eight-channel condition, 600 Hz for *gipe* and 650 Hz for *kipe*; and in the four-channel condition, 730 Hz for *gipe* and 870 Hz for *kipe*. Waveforms and spectrograms of the four-channel vocoded *gipe* and *kipe* stimuli are presented in Figure 1.

Procedure

The testing procedure and environment for NH participants was identical to those in the previous experiment, except NH participants heard unprocessed and vocoded sentences intermixed, and listened to stimuli diotically using circumaural headphones (Sennheiser, HD 650, Old Lyme, CT). All stimuli were presented to NH participants at the level 65 dB-A.

The 96 items (3 Precursor sentence rates × 8 Target items with varied VOT × 4 Spectral conditions [unprocessed, 16-, eight-, and four-channel vocoded stimuli]) in random order composed a single block, and 12 blocks were presented. Thus, the NH participants heard a total of 1,152 trials. Again, two techniques were used to analyze the data: generalized linear mixed-effects models with logit links directly modeling *gipe* and *kipe* responses, and repeated measures mixed ANOVAs, which matched previous analyses of rate normalization in the literature.

Results

Both the YNH and AMNH participants could, on average, identify the *gipe* and *kipe* endpoints (see Figure 2). This indicated that participants experienced a shift in

perception from one phoneme to the other over the span of the VOT continuum. Furthermore, rate normalization was generally observed across all conditions, including those with signal degradation (see Figure 3).

Similar to the first analysis for the CI participant data, *gipe* (coded 0) and *kipe* (coded 1) responses for all NH participants, for all speech rate and signal degradation conditions and test items, were fit with a generalized linear mixed-effects model with a logit link. The random effects were intercepts for participant and item, and the fixed effects were speech rate (categorical, with slow as the referent), signal degradation (categorical, with unprocessed speech as the referent), VOT (continuous, centered), age group (categorical, with YNH as the referent), and the following interactions: Speech rate × Signal degradation, Group × Signal degradation, and VOT × Signal degradation. This optimal model for the data (see Table 3) was constructed using the same method used for the CI participant data.

The strongest effects of speech rate on *gipe*/*kipe* responses occurred in the unprocessed (fast vs. slow rates, $p < .001$) and 16-channel vocoded conditions (fast vs. slow rates, $p = .001$; medium vs. slow rates, $p < .001$), holding other variables constant. The effect of speech rate on responses was diminished in the eight-channel and four-channel vocoded conditions.

Responses from the AMNH participant group were significantly different from those in the YNH participant group only in the most degraded signal condition, four-channel vocoded speech ($p = .025$), indicating that overall AMNH participants reported a smaller proportion of *kipes* (and therefore a greater proportion of *gipes*) in that condition compared to YNH participants.

The effect of VOT on responses was dependent on signal degradation. In the unprocessed speech condition, holding other variables constant, with each 1-ms increase in VOT, reports of *kipe* increased by a factor of 1.4 ($p < .001$). There was no significant change in the effect of VOT in the 16-channel vocoded condition ($p = .094$). In the eight-channel condition, for each 1-ms increase in VOT, reports of *kipe* increased by a factor of 1.2 ($p < .001$), and in the four-channel condition, increased by a factor of 1.1 ($p < .001$). Thus, the effect of VOT on responses diminished with more signal degradation.

The second analysis utilized ANOVAs to investigate changes in NH participants' VOT category boundaries, slopes, and rate effects. A three-way repeated measures mixed ANOVA analyzed the effects of speech rate (slow, medium, and fast; within subjects), signal degradation (unprocessed, 16-, eight-, and four-channel vocoded; within subjects), and chronological age group (young and age-matched to the CI group; between subjects) on VOT category boundaries.[2] The effect of speech rate was significant, $F(2, 54) = 71.49$, $p < .001$, $\eta_p^2 = .73$, and further analyzed with post hoc paired-samples t tests, Bonferroni-corrected for three tests. VOT category boundaries were significantly different from one another at every rate tested, decreasing in VOT duration with each increase in speech rate ($p < .02$ for all comparisons).

The effect of signal degradation was also significant, $F(1.79, 48.22) = 16.05$, $p < .001$, $\eta_p^2 = .37$, Greenhouse–Geisser corrected. Post hoc paired-samples t tests, Bonferroni-corrected for six tests, revealed that category boundaries were generally located at shorter VOT durations with less signal degradation. All comparisons were significant at $p \leq .005$, except the eight- versus four-channel vocoded conditions comparison, which was not significant.

The effect of group was not significant, $F(1, 27) = 2.95$, $p = .098$, $\eta_p^2 = .10$, but there was a significant interaction of Group × Signal degradation, $F(1.79, 48.22) = 5.72$, $p = .008$, $\eta_p^2 = .18$. Thus, the extent to which signal degradation affected VOT category boundaries was dependent on participant group. Four post hoc independent-samples t tests compared the two groups' performances in each of the four spectral conditions. Only the group difference in the four-channel vocoded condition was significant, $t(16.37) = 2.42$, $p = .027$, with AMNH participants' category boundaries occurring at comparatively longer VOT durations. There were no group differences for the unprocessed ($p = .79$), 16-channel vocoded ($p = .32$), or eight-channel vocoded ($p = .49$) conditions. Thus, when the stimuli were most degraded, older participants experienced a larger shift in category boundary location compared to younger participants. Group differences in the four-channel vocoded condition were also observed in the regression analysis (see Table 3).

The Rate × Signal degradation interaction was also significant, $F(3.62, 97.60) = 9.29$, $p < .001$, $\eta_p^2 = .26$, indicating that the effect of rate on VOT category boundaries was dependent on the amount of signal degradation (see Figure 3). Post hoc paired t tests, Bonferroni-corrected for multiple tests, were performed comparing overall participant performance at each rate within each spectral condition. VOT category boundaries occurred at significantly shorter durations with each increase in speech rate for the 16- and eight-channel vocoded conditions only ($p \leq .006$). In the unprocessed and four-channel vocoded conditions, category boundaries' VOT significantly decreased between fast and medium rates ($p < .001$) but not between medium and slow rates. The difference between slow and fast rates

[2]Two participants in the AMNH group did not reach the 50% criterion in the four-channel vocoded condition over the course of the tested *gipe*/*kipe* continuum—that is, they perceived the majority of the items as *gipe* rather than *kipe* at each point along the VOT continuum. To calculate participants' VOT category boundaries in these conditions, an extrapolation method was used which extended the continuum to longer, but untested, VOT values, until a potential VOT category boundary could be obtained (First participant: 75.7 ms for the slow rate, 76.6 ms for the medium rate, and 72 ms for the fast rate; Second participant: 75.6 ms for the medium rate). The analyses reported in the main text include these two participants, with these extrapolated values. When the same analyses were run without these two participants in the dataset, nearly all significant findings were replicated. Thus, the two participants did not seem to be driving the effects.

was also significant ($p < .001$). No other interactions were significant for the VOT category boundaries ANOVA analysis (Rate × Group, $p = .18$ and Rate × Group × Signal degradation, $p = .15$).

The sizes of the rate effects (i.e., the differences in VOT category boundaries between the slow and fast speech rates) were not monotonically related to the amount of signal degradation (see Table 4; Figure 4). Rate effect was analyzed with a two-way repeated measures mixed ANOVA with the factors of signal degradation and group. The effect of signal degradation was significant, $F(1.92, 51.95) = 7.99$, $p = .001$, $\eta_p^2 = .23$. Post hoc paired-samples t tests revealed that the rate effect for the 16-channel vocoded condition was significantly larger than rate effects for the unprocessed and eight-channel vocoded conditions ($p < .001$ for both). No other comparisons were significant (see Figure 4).

There was no significant effect of group on the rate effect, $F(1, 27) = 1.81$, $p = .19$, $\eta_p^2 = .06$, but there was a significant Group × Signal degradation interaction, $F(1.92, 51.95) = 3.54$, $p = .038$, $\eta_p^2 = .12$. This indicated that the effect of signal degradation on rate effect was dependent upon group (see Table 4; Figure 4). To further analyze the interaction, a one-way repeated measures ANOVA with the variable of signal degradation was run for each group. The rate effect was significantly affected by signal degradation only for the YNH group, $F(1.85, 27.78) = 12.44$, $p < .001$, $\eta_p^2 = .45$. Post hoc paired-samples t-tests, Bonferroni-corrected for multiple tests, compared the YNH group's rate effect in each spectral condition. Four comparisons were significant: Four-channel vocoded rate effect was larger than unprocessed and eight-channel vocoded rate effects ($p < .05$ for both); 16-channel vocoded rate effect was larger than unprocessed and eight-channel vocoded rate effects ($p < .01$ for both). Again, for the AMNH group, the rate effect was not significantly affected by signal degradation, $F(3, 36) = 1.97$, $p = .14$, $\eta_p^2 = .14$, indicating that the size of the rate effect was fairly consistent despite varying levels of signal degradation in the speech signal.

A three-way repeated measures mixed ANOVA was performed to analyze the effects of speech rate, signal degradation, and group on psychometric function slopes. In general, slopes became shallower with increases in signal degradation for both groups, regardless of speech rate (see Figure 5). The main effects of speech rate ($p = .39$) and group ($p = .35$) were not significant. However, the effect of signal degradation was significant, $F(2.27, 61.27) = 59.95$, $p < .001$, $\eta_p^2 = .69$. Post hoc paired-samples t tests, Bonferroni-corrected for multiple comparisons, revealed that slopes were significantly different from one another at every spectral condition tested ($p < .01$ for all comparisons), becoming shallower with each increase in signal degradation. Shallower slopes implied that phoneme perception for /g/ and /k/ became less categorical and more continuous. There were no significant interactions ($p > .05$ for all). These findings align with the regression analysis, which revealed that *gipe* and *kipe* responses were less affected by changes in VOT as the amount of signal degradation increased ($p < .001$ for the eight-channel and four-channel vocoded conditions; $p = .094$ for the 16-channel vocoded condition; see Table 3).

Table 4. Mean rate effects (ms) for normal-hearing (NH) participants in each signal degradation condition.

Parameter	YNH (*SD*)	AMNH (*SD*)	All NH (*SD*)
Unprocessed	2.8 (2.7)	4.8 (2.7)	3.7 (2.8)
16-channel vocoding	8.0 (3.3)	6.0 (3.1)	7.1 (3.3)
8-channel vocoding	4.3 (2.3)	2.5 (3.9)	3.5 (3.2)
4-channel vocoding	7.8 (4.5)	4.3 (6.9)	6.2 (5.9)

Note. YNH = young normal-hearing participants; AMNH = age-matched normal-hearing participants; NH = normal-hearing participants.

Discussion

The NH group analysis ensured replication of previous literature (e.g., Newman & Sawusch, 2009), confirming that participants rate normalized to the unprocessed (or full spectral information) stimuli. An effect of speech rate was observed in the category boundary analyses, in that phonemic category boundary was mediated by rate, becoming significantly shorter in VOT with each increase in speech rate (see Figure 3).

The analysis also measured how NH participants were affected by vocoding, which reduced spectral resolution but generally preserved temporal envelope information. An effect of signal degradation was observed on every dependent variable (category boundary, rate effect, and slopes) in this study's NH analyses. First, the locations of the category boundaries shifted together to longer VOT durations with more signal degradation—that is, participants more often identified the target as *gipe* as fewer channels of spectral information were available (see Figure 3). Second, NH listeners rate normalized with vocoded speech, but rate effects did not increase monotonically with signal degradation (see Figure 4), as was predicted. Rate effects were relatively large for the 16- and four-channel vocoded conditions, but small and comparable to unprocessed speech in the eight-channel vocoded condition. Thus, it was unclear if signal degradations systematically altered the use of duration-based cues, at least in NH participants, or if there were other competing factors affecting results. Third, slopes became shallower (less reliable) with fewer spectral channels, meaning the mapping of acoustic cues to phoneme categories was less consistent and that participants had more difficulty identifying /g/ from /k/ (see Figure 5).

In terms of the effects of chronological age on rate normalization in NH participants, the AMNH group had a larger effect of signal degradation on category boundaries compared with the YNH group, with category boundaries appearing at longer VOT durations (see Figure 3). The signal degradation may have particularly taxed the temporal processing of auditory systems already made less effective by age, and made it more difficult for older participants to determine the duration of aspiration in the phoneme (Gordon-Salant et al., 2008). Although category

boundaries were affected by signal degradation in AMNH participants, rate effect was not, though performance was variable (see Figure 4). AMNH participants were able to perceive and adjust for variability in speech rate in a similar way regardless of the level of signal degradation.

VOT category boundary slopes did not appear to be affected by age (see Figure 5). Likewise, an interaction of VOT and age group was not observed in the regression analysis, as was observed in the CI analysis. This did not match findings in the literature that older NH listeners are likely to have shallower slopes on average than younger listeners for an initial stop consonant phoneme identification task (Gordon-Salant, Yeni-Komshian, Fitzgibbons, & Barrett, 2006). It should be noted, however, that the present study's AMNH group was not composed solely of older NH participants, as it included participants as young as 35 years old in order to match the average age and standard deviation of the CI group. Thus, any aging effects might be tempered by the fact that the AMNH participants were not uniformly above the age of 65 years, as they were in Gordon-Salant et al.'s (2006) work.

Results: Comparing CI and AMNH Participants

Differences in rate normalization between CI and AMNH participants listening to unprocessed and eight-channel vocoded speech were analyzed to determine if the AMNH participants' performances changed relative to CI participants when exposed to the same or less spectral information. The eight-channel vocoded condition was chosen because it was thought to best represent the average number of functional channels available at any one time to a CI participant (Friesen et al., 2001). There were 23 CI participants and 13 AMNH participants.

First, *gipe* and *kipe* responses from all AMNH and CI participants for all speech conditions and items were fit with a generalized linear mixed-effects model with a logit link. The random effects were intercepts for participant and item. The fixed effects were group (coded 0 = CI participant, 1 = AMNH participant hearing unprocessed speech, and 2 = AMNH participant hearing eight-channel vocoded speech), speech rate (categorical, with slow as the referent), VOT (continuous, centered), and the following interactions: Group × Speech rate, Group × VOT, and Speech rate × VOT. This optimal model for the data (see Table 3) was constructed using the method described previously.

In general, AMNH participants in both the unprocessed and eight-channel vocoded speech conditions heard significantly fewer *kipes* than CI participants ($p < .001$ and $p = .002$, respectively), holding other factors constant (see Table 3). The effects of speech rate were dependent on group. Compared to CI participants, AMNH participants hearing unprocessed speech had significantly larger differences in responses due to speech rate, and AMNH participants hearing eight-channel vocoded speech had significantly smaller differences in responses. This might indicate that degraded signals lead to reduced speech rate effects, but that with experience (similar to that of a CI participant), speech rate effects can become stronger.

The strength of the effect of VOT was also dependent on group. There was no significant difference in VOT effects on responses between CI participants and AMNH participants hearing eight-channel vocoded speech ($p = .853$), but the effect of VOT was significantly stronger for AMNH participants presented unprocessed speech ($p < .001$), where with each 1-ms increase in VOT, *kipe* responses increased by a factor of 1.4. The effect of VOT on responses was significantly stronger in the medium rate compared to the slow rate conditions ($p = .035$), but not significantly stronger in the fast rate compared to the slow rate ($p = .162$), holding other factors constant.

Second, two separate suites of two-way repeated measures mixed ANOVAs with factors hearing status group (CI or AMNH; between subjects) and speech rate (slow, medium, and fast; within subjects) were computed for the dependent variables of VOT category boundaries and slopes. The first ANOVA suite analyzed performance for CI participants and for AMNH participants presented unprocessed speech, to understand performance when groups were presented the same stimuli. The second ANOVA suite analyzed performance for CI participants and for AMNH participants presented eight-channel vocoded speech, to understand the effects of hearing history on rate normalization performance.

For VOT category boundaries, the rate normalization effect was again confirmed: The effect of speech rate was significant both when CI participants and AMNH participants were presented unprocessed speech, $F(1.69, 57.44) = 67.33$, $p < .001$, $\eta_p^2 = .66$, Greenhouse–Geisser corrected, and when CI participants were presented unprocessed speech and AMNH participants were presented eight-channel vocoded speech, $F(2, 68) = 31.97$, $p < .001$, $\eta_p^2 = .49$. In both cases, VOT category boundaries significantly decreased with each increase in speech rate ($p \leq .001$ for all comparisons; see Figure 3).

The effect of hearing status on VOT category boundaries was significant both in the unprocessed speech condition analysis, $F(1, 34) = 4.74$, $p = .037$, $\eta_p^2 = 0.12$, as well as in the analysis where AMNH participants instead heard eight-channel vocoded speech, $F(1, 34) = 19.04$, $p < .001$, $\eta_p^2 = .36$. CI participants had category boundaries at earlier VOT durations than AMNH participants in both cases (see Figure 3).

More importantly, how hearing status and speech rate interacted with respect to VOT category boundaries was different between the two ANOVA analyses. In the unprocessed speech analysis there was no significant interaction, $F(1.69, 57.44) = 1.71$, $p = .19$, $\eta_p^2 = .05$, and the change in VOT category boundaries between the slow and fast rate conditions (the rate effect) was not significantly different between CI and AMNH participants, $t(34) = 1.53$, $p = .14$. This implied that CI participants rate normalized in much the same way as NH participants with a similar range of ages, despite being presented a degraded speech signal by their sound processor. In contrast, in the

analysis in which AMNH participants listened instead to eight-channel vocoded speech, there was a significant Hearing status × Speech rate interaction, $F(2, 68) = 6.75$, $p = .002$, $\eta_p^2 = .17$. As presented in Figure 4 and Table 4, CI participants experienced a 6.6-ms ($SD = 3.7$) rate effect whereas AMNH participants listening to eight-channel vocoded speech experienced a 2.5-ms ($SD = 3.9$) rate effect, and this difference was significant, $t(34) = 3.07$, $p = .004$. Again, this result contrasted with findings from the first analysis—that is, with unprocessed speech, there was no significant difference in rate effect between groups ($p = .14$), with AMNH participants experiencing a 4.8-ms rate effect in the unprocessed condition.

For slopes, the effect of speech rate and the Hearing status × Speech rate interaction were not significant in either ANOVA analysis. However, there was a significant effect of hearing status when comparing CI and AMNH participants listening to unprocessed speech, $F(1, 34) = 20.22$, $p < .001$, $\eta_p^2 = .37$ (see Figure 5). This significant effect disappeared in the analysis where AMNH participants listened to eight-channel vocoded speech ($p = .64$). When listening to unprocessed speech, slopes were shallower in CI participants than in AMNH participants, meaning CI participants were less categorical and perhaps less certain in their identification of *gipe* versus *kipe*. When AMNH participants listened to eight-channel vocoded speech, slopes were shallow and more comparable to CI participants'. Thus, whereas category boundary location seemed to be dependent upon hearing status, slopes seemed to be dependent on the quality of the signal itself. This finding matched results from the regression analysis (see Table 3), where the interaction of hearing status (i.e., "group,") and VOT was significant, but only for differences between the AMNH participants listening to unprocessed speech versus CI participants ($p < .001$).

Discussion

Both CI and AMNH participants rate normalized to unprocessed stimuli, as was evident in the effects of rate on VOT category boundaries. In general, AMNH participants began identifying *kipes* at longer VOT durations compared to CI participants, and slopes were steeper and more categorical in AMNH participants listening to unprocessed speech. One prediction of the current study was that the CI participants would show a larger rate effect than the AMNH participants. Instead, CI and AMNH participants showed similar rate effects when presented unprocessed speech (see Figure 4), with the mean rate effects of 6.6 ms for CI participants and 4.8 ms for AMNH participants. When AMNH participants were presented eight-channel vocoded speech, which contained less spectral information, their rate effect actually decreased ($M = 2.6$ ms) and was significantly smaller than CI participants' rate effect (see Figure 4). This was an unexpected result. However, the change in the size of the rate effect was not uniform across all AMNH participants. Some participants showed more negative, or reverse, rate effects in eight-channel vocoded speech (see Figure 4). Negative rate effects could indicate that participants were remapping phonemic boundaries in response to changes in speech rate, but not in the predicted manner. Reverse rate normalization has been reported previously (Diehl, Souther, & Convis, 1980), but why it occurs is unclear.

When AMNH participants listened to eight-channel vocoded speech, the effect of hearing status on slopes was no longer significant, indicating that phoneme identification was less categorical than in the unprocessed speech condition when AMNH participants had less access to spectral information. This indicated that access to spectral information was important for categorical phoneme perception. Although slopes became more similar between the two groups as AMNH participants listened to eight-channel vocoded speech, the eight-channel vocoded condition did not bring the location of AMNH participants' VOT category boundaries closer to CI participants', but moved boundaries in the opposite direction along the VOT continuum (see Figure 3).

CI participants were able to perceive rate information without access to full spectral information, as they showed a clear rate effect. In a similar manner, AMNH participants continued to show comparable rate effects even as the speech signal was increasingly degraded. Thus, overall, whether speech is processed in a CI speech processor or through a vocoder, the effect of speech rate on phoneme perception appears to be rather robust.

General Discussion

The ways in which CI users can navigate speech's acoustic variability are not well understood. Although much research has measured word recognition in CI users (e.g., Liu et al., 2004; Nie et al., 2006), especially under laboratory conditions using only a single talker speaking at a single rate, these types of tests unfortunately appear to underestimate the actual word recognition ability of CI users (Kirk, Pisoni, & Miyamoto, 1997). To be specific, this study investigated how aspects of speech perception in CI participants were affected by rate, and if CI participants were able to rate normalize in response to speech rate changes, which can occur across and within talkers.

Using this study's test paradigm, CI participants rate normalized across three naturalistically produced sentences spoken at slow, medium, and fast speech rates (see Figure 2). The remapped phonemic category boundary between /g/ and /k/, a distinction that primarily is based on VOT duration (Jiang, Chen, & Alwan, 2006; Klatt, 1976; Summerfield, 1981), was mediated by speech rate, thus matching behavior observed previously in the literature in young NH listeners (Newman & Sawusch, 2009) and in both YNH and AMNH participants in the current study. In fact, compared to NH participants listening to unprocessed stimuli, CI participants showed a *larger* response to rate changes, or rate effect, in terms of differences in VOT category boundaries between the slow and fast speech rates (see Figure 4). NH participants may have shown a smaller average rate effect because they potentially had access to other acoustic cues in the

stimuli besides VOT durations to categorize /g/ and /k/ (Miller & Wayland, 1993). Previous work has shown that NH participants may use different strategies to identify /g/ and /k/ phonemes besides or in addition to VOT durations (Lisker, 1975). For example, Stevens and Klatt (1974) found that although some NH participants relied exclusively on VOT duration, others gave more weight to the presence of a first formant transition. Others have pointed towards the onset frequency of the first formant as being important for this distinction (Summerfield & Haggard, 1974) or the presence of low-frequency energy (Zhou, Xu, & Lee, 2010) as an important cue for voicing. These additional spectral cues from the first formant may be less available to CI participants, who in the absence of adequate spectral information likely rely more on temporal cues for phoneme perception (Moberly et al., 2014; Peng, Lu, & Chatterjee, 2009; Winn et al., 2012). The stimuli in the present experiment had first formant frequency onset cues as well as VOT cues. As VOT increased, so did first formant frequencies, though not in perfect correlation (see Experiment 1: Methods and Experiment 2: Methods). VOT values increased in a continuous fashion, at roughly 9-ms increases per continuum step, whereas first formant frequencies were largely stable for the first three and last three continuum steps, with increases occurring between the third and sixth continuum steps—that is, between the 36- and 54-ms steps on the VOT continuum.

Further evidence that NH participants may be using cues beyond duration to identify /g/ and /k/ in the context of speech rate is that with signal degradation (i.e., the vocoded conditions, which aimed to simulate possible spectral resolutions of the CI), category boundaries shifted to *longer* VOTs compared to the NH unprocessed conditions. This is unlike the CI participants, whose category boundaries were generally at *shorter* VOTs (see Figure 3) compared to the NH unprocessed conditions. It was expected that the vocoded stimuli would cause NH participants to perform more similarly to CI participants, so this opposite shift in NH participants' category boundaries was surprising. This need for longer VOT durations at lower spectral resolutions to perceive a voiceless stop such as /k/, which was observed particularly in the AMNH group, could partly be because of temporal processing deficits associated with age. Gordon-Salant and colleagues showed that older NH listeners needed longer VOTs to categorize phonemes with certain temporal duration contrasts (Gordon-Salant et al., 2008; Gordon-Salant et al., 2006). In the current study, although AMNH participants needed longer VOT durations to perceive the contrast in spectrally sparse conditions, this pattern was not mirrored in their CI participant cohort. Thus, again, it is possible that AMNH participants were using more than just temporal information to make their decisions about /g/ and /k/, whereas CI participants were solely relying on temporal information. Additional problems in CI processing such as current spread or frequency-to-place mismatch were also not captured by the current study's vocoder, and could be contributing to the incongruence between CI and NH participants' data.

Perhaps the lack of training or long-term experience with listening to vocoded speech could explain these AMNH participants' deviations in performance from actual CI users. For example, Munson and Nelson (2005) found that although CI participants had little difficulty perceiving differences between "say" and "stay" in quiet versus noise, NH participants in that study had difficulty distinguishing this pair when provided with fewer spectral bands. Because the cue to the difference between "say" and "stay" was temporal in nature, the loss of spectral information was not expected to change perceptions in NH participants. Munson and Nelson (2005) concluded that potentially a lack of training could be causing the discrepancy. Perhaps older NH participants exhibit less plasticity to learn to perceive and understand degraded speech patterns (Shannon, 2002), and the experience with vocoded stimuli in the present experiment was too short a time span for AMNH participants to determine the best cues for categorizing spectrally degraded phonemes (Rosen, Faulkner, & Wilkinson, 1999; Schvartz, Chatterjee, & Gordon-Salant, 2008; Sheldon, Pichora-Fuller, & Schneider, 2008). How NH participants' phoneme category boundary remapping evolves over the course of extended training with vocoded speech could be an area of future study.

Last, there are additional factors that could be causing performance variability among CI participants. Better speech perception outcomes in CI users (typically measured as percent words correct) have been shown to be correlated with increased neural recruitment in association auditory cortices during speech perception tasks (K. M. J. Green, Julyan, Hasting, & Ramsden, 2007), deeper array insertion depths (Skinner et al., 2002), and shorter durations of deafness before implantation (van Dijk et al., 1999). For phoneme perception specifically, longer durations of deafness for postlingually deafened Korean adult CI users (Oh et al., 2003) and English-speaking adult CI users (Budenz et al., 2011) had a negative impact, CI users' voicing perception in consonants was shown to improve with access to residual hearing (Zhang, Dorman, & Spahr, 2010), and increases in the number of active electrodes gradually improved consonant recognition (Zeng & Galvin, 1999). Although the present study's results found no association of CI participants' rate effects with durations of deafness, and CI participants were not able to utilize any residual hearing during the task, there were clearly uncontrolled factors in the CI participant sample.

The data support this study's alternative hypothesis: Rate normalization was possible in CI participants, and this rate effect was larger in comparison to NH participants. CI participants' average rate effect was 6.6 ms, whereas NH participants' average rate effect was 3.7 ms and the subgroup of AMNH participants' average was 4.8 ms (see Figure 4) for unprocessed speech. The larger rate effect observed in CI participants may have been a result of this group's access to primarily temporal-related cues, which can be affected by changes in speech rate, and lack of access to other cues that are more invariant to rate normalization. NH participants' performance in vocoded conditions revealed that signal degradations involving reduced spectral

resolution were likely driving the less categorical, more continuous identification of /g/ and /k/ (i.e., relatively shallow psychometric slopes; see Figures 2 and 5), but that greater signal degradation did not always result in larger rate effects (particularly in the eight-channel condition; see Figure 4). It is possible that the presence of fewer redundant acoustic cues in the speech signal as a result of signal degradation may have contributed to these changes in NH participants' slopes (Peng et al., 2009).

To summarize, CI participants could rate normalize, but did so differently from NH participants. On average, CI participants' remapped category boundaries were at shorter VOT durations, their category boundaries were less consistent and reliable (with changes in VOT becoming less influential on *gipe*/*kipe* identification with greater age), and they tended to have a larger rate effect than NH participants. Rate normalization appears to be an obligatory, highly context-based process (Miller & Dexter, 1988; Sawusch & Newman, 2000), and the CI appears to be able to transmit the necessary information for rate normalization to occur. The variability in speech perception outcomes reported for CI users (Liu et al., 2004; Nie et al., 2006) is likely not the result of an inability to rate normalize, but perhaps partly because of a difficulty with categorical phoneme identification, which is the result of the alteration and degradation of the incoming speech signals by the speech processor. Encoding degraded phonetic information can lead to a larger processing cost as the brain attempts to account for variability in the speech signal (Mullennix, Pisoni, & Martin, 1989), which, in the present study, is variability which could be introduced by changes in speech rate between sentences. Understanding how CI users recognize speech in the real world, and what kinds of capacities they have to perceive less idealistic phoneme productions and to adapt to a powerful source of speech variability such as speech rate, could provide an avenue for improving how clinicians approach aural rehabilitation for CI users.

Acknowledgments

This work was supported by National Institute of Health Grant R01-AG051603 (M.J.G.), P30-DC004664 (C-CEBH), a training award T32-DC000046E (B.N.J.), and the University of Maryland. Special thanks to Matan Simhi for his help with data collection, James R. Sawusch for recording the stimuli, and Christopher Cullen Heffner for help with the statistical analysis. Portions of this work have been presented at the 169th Meeting of the Acoustical Society of America and the 2015 Conference on Implantable Auditory Prostheses.

References

Blamey, P. J., & Clark, G. M. (1990). Place coding of vowel formants for cochlear implant patients. *The Journal of the Acoustical Society of America, 88,* 667–673.

Budenz, C. L., Cosetti, M. K., Coelho, D. H., Birenbaum, B., Babb, J., Waltzman, S. B., & Roehm, P. C. (2011). The effects of cochlear implantation on speech perception in older adults. *Journal of the American Geriatrics Society, 59,* 446–453.

Ching, T. Y. C., van Wanrooy, E., & Dillon, H. (2007). Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review. *Trends in Amplification, 11,* 161–192.

Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *The Journal of the Acoustical Society of America, 72,* 705–716.

Diehl, R. L., Souther, A. F., & Convis, C. L. (1980). Conditions on rate normalization in speech perception. *Perception & Psychophysics, 27,* 435–443.

Donaldson, G. S., Rogers, C. L., Johnson, L. B., & Oh, S. H. (2015). Vowel identification by cochlear implant users: Contributions of duration cues and dynamic spectral cues. *The Journal of the Acoustical Society of America, 138,* 65–73.

Dorman, M. F., Dankowski, K., McCandless, G., & Smith, L. (1989). Identification of synthetic vowels by patients using the Symbion multichannel cochlear implant. *Ear and Hearing, 10,* 40–43.

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processers using sine-wave and noise-band outputs. *The Journal of the Acoustical Society of America, 102,* 2403–2411.

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science, 209*(4461), 1140–1141.

Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America, 124,* 1234–1251.

Friesen, L. M., Shannon, R. V., Başkent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America, 110,* 1150–1163.

Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125.

Gordon-Salant, S., Yeni-Komshian, G., & Fitzgibbons, P. (2008). The role of temporal cues in word identification by younger and older adults: Effects of sentence context. *The Journal of the Acoustical Society of America, 124,* 3249–3260.

Gordon-Salant, S., Yeni-Komshian, G., Fitzgibbons, P., & Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *The Journal of the Acoustical Society of America, 119,* 2455–2466.

Green, K. M. J., Julyan, P. J., Hasting, D. L., & Ramsden, R. T. (2007). Auditory cortical activation and speech perception in cochlear implant users. *The Journal of Laryngology & Otology, 122,* 238–245.

Green, K. P., Stevens, E. B., & Kuhl, P. K. (1994). Talker continuity and the use of rate information during phonetic perception. *Perception & Psychophysics, 55,* 249–260.

Hillenbrand, J. (1984). Perception of sine-wave analogs of voice onset time stimuli. *The Journal of the Acoustical Society of America, 75,* 231–240.

IEEE Subcommittee on Subjective Measurements. (1969). IEEE recommended practices for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17,* 227–246.

Iverson, P. (2003). Evaluating the function of phonetic perceptual phenomena within speech recognition: An examination of the perception of /d/-/t/ by adult cochlear implant users. *The Journal of the Acoustical Society of America, 113,* 1056–1064.

Iwasaki, S., Ocho, S., Nagura, M., & Hoshino, T. (2002). Contribution of speech rate to speech perception in multichannel

cochlear implant users. *The Annals of Otology, Rhinology, and Laryngology, 111,* 718–721.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change, 21,* 233–256.

Jiang, J., Chen, M., & Alwan, A. (2006). On the perception of voicing in syllable-initial plosives in noise. *The Journal of the Acoustical Society of America, 119,* 1092–1105.

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 736–748.

Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C. (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research, 40,* 1395–1405.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59,* 1208–1221.

Lisker, L. (1975). Letter: Is it VOT or a first-formant transition detector? *The Journal of the Acoustical Society of America, 57,* 1547–1551.

Liu, S., Del Rio, E., Bradlow, A. R., & Zeng, F. (2004). Clear speech perception in acoustic and electric hearing. *The Journal of the Acoustical Society of America, 116,* 2374–2383.

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. *Advances in Oto-Rhino-Laryngology, 64,* 109–143.

Lotto, A. J., Kluender, K. R., & Green, K. P. (1996). Spectral discontinuities and the vowel length effect. *Perception & Psychophysics, 58,* 1105–1114.

Miller, J. L. (1981). Some effects of speaking rate on phonetic perception. *Phonetica, 38,* 159–180.

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 369–378.

Miller, J. L., & Grosjean, F. (1981). How the components of speaking rate influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 208–215.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25,* 457–465.

Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46,* 505–512.

Miller, J. L., & Wayland, S. C. (1993). Limits on the limitations of context-conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics, 54,* 205–210.

Moberly, A. C., Lowenstein, J. H., Tarr, E., Caldwell-Tarr, A., Welling, D. B., Shahin, A. J., & Nittrouer, S. (2014). Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing? *Journal of Speech, Language, and Hearing Research, 57,* 566–582.

Mullennix, J. W., Pisoni, D. P., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America, 85,* 365–378.

Munson, B., Donaldson, G. S., Allen, S. L., Collison, E. A., & Nelson, D. A. (2003). Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability. *The Journal of the Acoustical Society of America, 113,* 925–935.

Munson, B., & Nelson, P. B. (2005). Phonetic identification in quiet and in noise by listeners with cochlear implants. *The Journal of the Acoustical Society of America, 118,* 2607–2617.

Nagao, K., & de Jong, K. (2007). Perceptual rate normalization in naturally produced rate-varied speech. *The Journal of the Acoustical Society of America, 121,* 2882–2898.

Nakai, S., & Scobbie, J. M. (2016). The VOT category boundary in word-initial stops: Counter-evidence against rate normalization in English spontaneous speech. *Laboratory Phonology, 7,* 13.

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics, 37,* 46–65. https://doi.org/10.1016/j.wocn.2008.09.001

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

Nittrouer, S., & Lowenstein, J. H. (2014). Separating the effects of acoustic and phonetic factors in linguistic processing with impoverished signals by adults and children. *Applied Psycholinguistics, 35,* 333–370.

Oh, S. H., Kim, C. S., Kang, E. J., Lee, D. S., Lee, H. J., Chang, S. O., . . . Koo, J. W. (2003). Speech perception after cochlear implantation over a 4-year time period. *Acta Otolaryngologica, 123,* 148–153.

Peng, S.-C., Lu, N., & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiology and Neurotology, 14,* 327–337.

Rosen, S. (1989). Temporal information in speech and its relevance for cochlear implants. In B. Fraysse & N. Cochard (Eds.), *Cochlear implant: Acquisitions and controversies*. Basel, Switzerland: Cochlear A.G. (pp. 3–26).

Rosen, R., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *The Journal of the Acoustical Society of America, 106,* 3629–3636.

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II: Effects of signal discontinuities. *Perception & Psychophysics, 62,* 285–300.

Schvartz, K. C., Chatterjee, M., & Gordon-Salant, S. (2008). Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners. *The Journal of the Acoustical Society of America, 124,* 3972–3988.

Shannon, R. V. (2002). The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design. *American Journal of Audiology, 11,* 124–127.

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

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Effect of age, presentation method, and learning on identification of noise-vocoded words. *The Journal of the Acoustical Society of America, 123,* 476–488.

Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., . . . Blocker, B. (2002). CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients. *Journal of the Association for Research in Otolaryngology, 3,* 332–350.

Sommers, M. S. (1997). Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment. *The Journal of the Acoustical Society of America, 101,* 2278–2288.

Souza, P., & Rosen, S. (2009). Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech. *The Journal of the Acoustical Society of America, 126,* 792–805.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America, 55,* 653–659.

Storkel, H. L., & Hoover, J. R. (2010). An on-line calculator to compute phonotactic probability and neighborhood density based on child corpora of spoken American English. *Behavior Research Methods, 42,* 497–506.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074–1095.

Summerfield, Q., & Haggard, M. P. (1974). Perceptual processing of multiple cues and contexts: Effects of followed vowel upon stop consonant voicing. *Journal of Phonetics, 2,* 279–295.

Tyler, R. S., & Moore, B. C. J. (1992). Consonant recognition by some of the better cochlear-implant patients. *The Journal of the Acoustical Society of America, 92,* 3068–3077.

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). *Irvine phonotactic online dictionary* (Version 2.0). [Data file]. Available from http://www.iphod.com

Välimaa, T. T., Määttä, T. K., Löppönen, H. J., & Sorri, M. J. (2002). Phoneme recognition and confusions with multichannel cochlear implants: Consonants. *Journal of Speech, Language, and Hearing Research, 45,* 1055–1069.

van Dijk, J. E., van Olphen, A. F., Langereis, M. C., Mens, L. H. M., Brokx, J. P. L., & Smoorenburg, G. F. (1999). Predictors of cochlear implant performance. *Audiology, 32,* 109–116.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *The Journal of the Acoustical Society of America, 82,* 1152–1161.

Whitmal, N. A., Poissant, S. F., Freyman, R. L., & Helfer, K. S. (2007). Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *The Journal of the Acoustical Society of America, 122,* 2376–2388.

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

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *The Journal of the Acoustical Society of America, 131,* 1465–1479.

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). The roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. *Journal of Speech, Language, and Hearing Research, 56,* 1097–1107.

Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *The Journal of the Acoustical Society of America, 137,* 1430–1442.

Zeng, F., & Galvin, J. J. (1999). Amplitude mapping and phoneme recognition in cochlear implant listeners. *Ear and Hearing, 20,* 60–74.

Zhang, T., Dorman, M. F., & Spahr, A. J. (2010). Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. *Ear and Hearing, 31,* 63–69.

Zhou, N., Xu, L., & Lee, C.-Y. (2010). The effects of frequency-place shift on consonant confusion in cochlear implant simulations. *The Journal of the Acoustical Society of America, 128,* 401–409.