Effects of real-time cochlear implant simulation on speech production^{a)}

Elizabeth D. Casserly^{b)}

Department of Psychological and Brain Sciences, Speech Research Laboratory, Indiana University,
 1101 East 10th Street, Bloomington, Indiana 47405

(Received 15 April 2014; revised 17 March 2015; accepted 24 March 2015)

Investigations using normal-hearing subjects listening to simulations of cochlear implant (CI) acoustic processing have provided substantial information about the impact of these distorted listening conditions on the accuracy of auditory perception, but extensions of this method to the domain of speech production have been limited. In the present study, a portable, real-time vocoder was used to simulate conditions of CI auditory feedback during speech production in NH subjects. Acoustic-phonetic characteristics of sibilant fricatives, aspirated stops, and F1/F2 vowel qualities were analyzed for changes as a result of CI simulation of acoustic speech feedback. Significant changes specific to F1 were observed; speakers reduced their phonological vowel height contrast, typically via talker-specific raising of the low vowels [æ] and [ɑ] or lowering of high vowels [i] and [u]. Comparisons to the results of both localized feedback perturbation procedures and investigations of speech production in deaf adults with CIs are discussed.

© 2015 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4916965>]

[JFC]

Pages: 2791–2800

I. INTRODUCTION

Simulation of cochlear implant (CI) signal processing has been used as a means of investigating these devices with normal-hearing subjects for nearly 20 years (Shannon *et al.*, 1995). Studies using this “vocoding” technique are ideal for determining the impact of processing factors on speech perception (Fu *et al.*, 1998a; Shannon *et al.*, 1995). They use listeners with relatively few demographic or etiological confounds to interfere with perceptual accuracy (e.g., history of deafness or hearing loss or performance limitations imposed by the need for plasticity or reorganization in auditory cortex). Any difficulties observed when these listeners respond to CI-simulated stimuli, therefore, can be attributed to either the novelty of the signal transformation (cf. Fu *et al.*, 2005; Loebach and Pisoni, 2008; Smalt *et al.*, 2013) or the degradation of spectral information introduced by the simulation itself (Fu and Nogaki, 2005; Fu *et al.*, 1998a; Shannon *et al.*, 1995). Accordingly, research using the CI simulation method has been far-ranging and prolific, studying such phenomena as speech perception accuracy in quiet, noise, and babble (Cullington and Zeng, 2008; Fu and Nogaki, 2005; Fu *et al.*, 1998a), accuracy of environmental sound recognition (Loebach and Pisoni, 2008; Shafiro, 2008), listeners’ ability to repeat novel word forms (Burkholder-Juhász *et al.*, 2007), the effects of target

language on speech recognition (e.g., Fu *et al.*, 1998b), the influence of particular signal processing algorithms (e.g., Fu *et al.*, 1998a), the benefits of additional frequency channels or electrodes (Shannon *et al.*, 1995), and the effectiveness of different training methods for promoting perceptual learning (Loebach and Pisoni, 2008; Loebach *et al.*, 2010).

As the preceding selection of topics implies, however, the vast majority of research related to signal properties from CIs has focused on perception. Relatively few studies have analyzed the speech production of individuals with CIs, and still fewer of those have attempted to separate out the influence of the CI signal degradation itself from the myriad other factors potentially at work (but see Tobey *et al.*, 2003; Tobey *et al.*, 2011). This focus on perception may seem warranted, given the perceptual nature of these individuals’ disturbances (i.e., deafness) and their CI devices, but extension into the realm of speech production is necessary because of two factors. First, accurate speech production is a critical component of the connection between talkers and listeners occurring in fluent spoken communication (Denes and Pinson, 1963). Second, there is evidence that performance accuracy in speech production is closely related to perceptual accuracy in CI users: For children with CIs, for example, early speech intelligibility is not only strongly correlated with concurrent perceptual and global language skills (Blamey *et al.*, 2001; Tobey *et al.*, 2003), but it also serves as a significant predictor of development in these areas following nearly 10 additional years of CI use (Casserly and Pisoni, 2013; Tobey *et al.*, 2011).

A link between success in speech production and the receptive linguistic abilities of a CI user should not be surprising; the importance of auditory feedback for initial acquisition of speech articulation and ongoing maintenance of adult production has been known for over a century (Lane *et al.*, 2007; Lane and Tranel, 1971; Oller and Eilers, 1988).

^{a)}Portions and preliminary versions of this work were presented at “Consonants vs vowels: Phonetic changes under acoustic feedback transformation” at the CUNY Conference on the Segment in Phonology, New York, NY, January 2012, and in “Speech production under real-time simulation of cochlear implant acoustic feedback,” at the 164th Annual Meeting of the Acoustical Society of America, Kansas City, MO, October 2012. It appeared also as part of the author’s doctoral thesis.

^{b)}Author to whom correspondence should be addressed. Present address: Department of Psychology, Trinity College, 300 Summit St., Hartford, CT 06106. Electronic mail: elizabeth.casserly@trincoll.edu

Without immediate perceptual feedback, infants do not babble like their NH peers (Oller and Eilers, 1988), and adults whose acoustic feedback is manipulated experimentally, e.g., by the addition of noise or the artificial shifting of properties such as f_0 or vowel formant frequencies, accordingly alter the speech they produce in response (Elman, 1981; Houde and Jordan, 1998; Reilly and Dougherty, 2013; Shiller *et al.*, 2009). Viewed from this perspective of feedback-based motor control, the fact that CI acoustic processing might impact speech production should be quite clear: CIs provide a distinctive type of auditory feedback, causing CI users to face a unique challenge in the monitoring and control of their own speech. Understanding this facet of the impact of CI signals on a user's experience, then, constitutes an important objective for research in the field.

A few studies have investigated the links between feedback and speech production for CI users (e.g., Lane *et al.*, 2007; Perkell *et al.*, 2007). Lane and colleagues recorded the speech of post-lingually deafened adults, for example, before they received CIs, immediately upon activation of their CIs, and after several months of use (Lane *et al.*, 2007). Speech gradually improved in these speakers as they gained experience with their devices; the dispersion of vowels increased, as did the contrast between the sibilant fricatives [s] and [ʃ], across 1 yr of experience with their CIs. To examine the impact of feedback over shorter periods of time, i.e., its role in the real-time control of speech production rather than in cumulative fine-tuning of articulatory plans, the speech of NH and deaf adults with CIs producing speech has been analyzed in a range of environmental noise conditions (Perkell *et al.*, 2007). Once again, the status of a speaker's feedback, on average, substantially impacted the acoustic dispersion and consistency of their speech, while their durational and amplitude changes mirrored those of NH speakers responding to noise (cf. Lane and Tranel, 1971).

These studies and others like them, however, have largely been limited to investigations of speech production in the clinical CI population. While this makes their results particularly direct in application, it also limits their interpretability; confounds such as an individual's age, duration of hearing loss, which can be substantial in post-lingually deafened adults, and possible concomitant diagnoses, i.e., in the case of lesion-based hearing loss or deafness due to ototoxic drugs, are extremely difficult to eliminate from experiments using clinical populations. Studies paralleling those in the perceptual literature, using NH subjects exposed to simulations of CI acoustic output, would clearly be useful in separating out the influence of peripheral, signal-driven effects on speech production from the effects of these and other confounds.

II. CURRENT INVESTIGATION

The research described in the following text represents an initial attempt at examining speech production in NH subjects experiencing CI-simulated acoustics. In typical CI-simulation procedures, subjects are exposed to pre-recorded stimuli that have been degraded according to the vocoding algorithm offline. By contrast, to observe potential effects of

vocoding on speech production, our subjects had to hear their own speech through the CI simulator, so that their acoustic feedback was similar to what a CI user might experience. This feedback transformation also had to occur with minimal delay so that speech fluency would not be disrupted (cf. Howell *et al.*, 1987). Kaiser and Svirsky (1999) succeeding in making early, desktop-based devices capable of the transformation more than a decade ago, but little work has followed up on their attempts until recently. The portable, real-time vocoder (PRTV) was developed as an updated version of Kaiser and Svirsky's devices (Casserly *et al.*, 2011; Smalt *et al.*, 2013). The PRTV (see Sec. III) performs continuous noise-vocoded CI simulation with less than a 10 ms delay and runs on a compact, highly portable platform. Using the PRTV, subjects in the present study were able not only to hear their own speech through a real-time transformation but also to hear the transformed speech of other talkers.

The aim of the present study was therefore twofold: First, to examine changes that might occur in a NH participant's speech as a result of experiencing CI-simulated feedback and, second, to observe whether these changes were affected by a short period of naturalistic learning during which participants were able to engage in (vocoded) unrestricted conversation with an interlocutor.

Changes in speech production as a result of CI simulation were assessed in terms of segmental acoustic phonetics, specifically sibilant fricative centroid frequencies, vowel formants (F_1 and F_2), and stop voice onset times (VOTs). The sibilants [s] and [ʃ] in American English rely on an acoustic contrast primarily loaded on their spectral centroid frequencies, in which [s] is typically higher in frequency than [ʃ] (Jongman *et al.*, 2000). This distinction is heavily loaded in the high-frequency acoustic domain, sampled coarsely or eliminated entirely by CI processing (see Table I), making them good candidates for feedback-based changes in production (but see Ghosh *et al.*, 2010; Perkell *et al.*, 2004b). Sibilant fricatives have also been shown to be vulnerable to real-time acoustic feedback perturbation in studies of NH speakers experiencing experimental upward or downward shifts of these centroid frequencies (Shiller *et al.*, 2009; Casserly, 2011). Similarly, vowel quality was targeted as an area that relies on fine-grained frequency contrasts and has been shown to be sensitive to localized perturbation in other feedback-alteration paradigms (e.g., Houde and Jordan, 1998, 2002; Purcell and Munhall, 2006; Reilly and Dougherty, 2013) and to change as a result of deafness and/or CI use (Lane *et al.*, 2007; Perkell *et al.*, 2007). VOT contrasts, on other hand, do not rely as heavily on spectral cues, hinging on large-scale changes in amplitude and periodicity over time (Stevens, 1998). Measurements of VOT were therefore included to provide a contrast with the primarily spectral fricative and vocalic quality cues constituting the bulk of the analysis. We did not predict changes in this aspect of speakers' segmental acoustic phonetics.

To determine whether these aspects of subjects' speech were altered as a result of CI-simulated feedback or later learning, participants were recorded across three "epochs": Baseline, initial simulation exposure, production following

TABLE I. Frequency boundaries for the band-pass filters used to separate acoustic signals into eight non-overlapping channels (Kaiser and Svirsky, 1999). The bandwidth of each channel is given on the right for convenience; these increase approximately logarithmically.

Spectral channel	Lowest frequency (Hz)	Cut-off frequency (Hz)	Channel bandwidth (Hz)
1	252	500	248
2	500	730	230
3	730	1020	290
4	1020	1500	480
5	1500	2000	500
6	2000	2600	600
7	2600	3800	1200
8	3800	7000	3200

learning period. Acoustic phonetic characteristics were compared across epochs 1 and 2 to determine initial, signal-driven effects and across epochs 2 and 3 to examine the effects of learning.

Finally, a small control group of talkers was also included in the study to examine the potential confound arising from comparing speech produced on the first, second, and third repetition of scripted stimuli (see Sec. III C, Stimulus materials). Repetition of items is known to affect speech production, both in terms of hyper- or hypo-articulation (Lindbloom, 1990) and in terms of prosody or emphasis (cf. Baker and Bradlow, 2009). We asked talkers in the control group to simply go through all three recording epochs with normal feedback intact, essentially eliciting these repetition-related changes in production so that they could be contrasted with the speech of our experimental group. Changes unique to the experimental group and its individual talkers, therefore, could be taken as products of the PRTV feedback processing as opposed to mere repetition effects.

III. DESIGN AND METHODS

A. Subjects

Monolingual native speakers of American English between 18 and 35 yr of age were recruited to participate in this study. To reach the target sample size of nine experimental participants and three controls, a total of 18 volunteers were evaluated. Six were eliminated prior to analysis: Two for failing to meet inclusionary criteria for age or monolingualism, two for technical failures in recording equipment, and two for deviations in experimental protocol. All participants reported no history of speech or hearing disorders. Nine (two male) completed the experimental protocol, while three participants (all female) served as controls.

B. Experimental design

Subjects produced speech in response to written prompts repeated across three 20-min recording epochs. No feedback manipulation was used during the first epoch, providing a baseline of normal speech for each subject. During epochs 2 and 3, the nine experimental subjects wore the PRTV and experienced real-time CI-simulation; controls did not

experience CI-simulated signals at any time. Speech production materials were presented in blocks by stimulus type (words, sentences, etc.—see Sec. III C), and all materials were repeated by subjects across all three epochs. Subjects in both conditions, therefore, produced the same set of words, e.g., in a different random order three times across the duration of the experiment.

C. Stimulus materials

Subjects were asked to read aloud a list of 114 isolated English words. One hundred stimuli were selected from the Hoosier Mental Lexicon database (Nusbaum *et al.*, 1984) to be balanced in frequency of occurrence: 50 words were high-frequency (occurrence of at least 195/million, Francis and Kucera, 1982) and 50 were relatively low-frequency (5 appearances per million, Francis and Kucera, 1982). Test words were also balanced for their phonetic content, such that the [i, æ, ɑ, u] English vowels were each included in 10 unique items (split between high- and low-frequency items), 16 items each contained the voiceless sibilant fricatives [s] and [ʃ], and the voiceless aspirated stops [p, t, k] were each present in eight unique items. Four filler words and 14 words with the American English monophthongs and diphthongs in a [h_d] frame were also included, bringing the total to 100.

In addition to the isolated word list, subjects also produced speech in response to prompts for short passages, meaningful English sentences, and words and sentences produced with particular prosodic structure. The order of presentation was always: Short passage, isolated words, sentences without prosodic direction, and sentences with prosodic focus. Speech from materials other than isolated words will not be analyzed in the following text.

D. Simulation of cochlear implant processing

Eight-channel noise vocoding was used to simulate CI acoustic processing (cf. Kaiser and Svirsky, 1999). Sampled acoustics were filtered to a window of 252–7000 Hz, then divided into eight non-overlapping frequency channels using band-pass filters of increasing width through the spectrum (see Table I for details). The signal amplitude from each channel was then used to filter broadband white noise. The eight amplitude-matched noise channels were then summed to create the noise-vocoded cochlear implant simulation.

This distortion of acoustic structure results in a substantial degradation of spectral resolution, as well as complete loss of direct information regarding the source signal above 7000 Hz or below 252 Hz. Speakers experiencing these acoustics as speech feedback, however, also inevitably received some low-frequency information via bone conduction (cf. Stenfelt and Håkansson, 2002). No attempts were made to mask this source of non-airborne feedback; adding noise to the signal, as has been done to mask bone-conducted feedback elsewhere (e.g., Purcell and Munhall, 2006), would have added a problematic element of challenging speech-perception-in-noise to the feedback perturbation being used here, particularly given the noise-based nature of the vocoded speech being heard by subjects.

The acoustic transformation simulating the effects of CI processing was completed using the PRTV, which consisted of a high-speed processor on a solid-state drive (8 GB iPod touch, model A1367), a head-mounted input microphone (Williams MIC090 mini lapel clip), output noise-isolating earphones with disposable single-use foam tips (Etymotic HF5), and a set of Elvex SuperSonic noise-attenuating headgear. Custom PRTV software took continuous acoustic input from the lapel microphone, worn on the attenuators clipped just above a subject's left ear, applied the noise-vocoding CI simulation, and then relayed the modified acoustic signal to the insert earphones with less than a 10 ms delay.

E. Procedure

During each recording epoch, subjects were seated in a sound-attenuating booth (Industrial Acoustics Co.) in front of a computer monitor and a condenser microphone (Audiotechnica AT2021) on a desk stand. Sixteen-bit, 44.1 kHz audio recordings were made of subjects' speech during each of three 20 min recording epochs.

In the first epoch, all subjects produced speech without wearing the PRTV or experiencing any manipulation of their acoustic perception. The PRTV was then introduced to subjects in the experimental condition ($n=9$) so that the first speech heard through the CI simulation occurred during the epoch 2 recording session. That is, when the PRTV was fitted on the subject following epoch 1, no further spoken interactions occurred until after epoch 2 was complete. At that point, the experimenter engaged the subject in conversation, outside the booth, for a 15 min "break"; due to the portable nature of the PRTV, CI-simulation of acoustics was continuous during this period and the transitions to and from the recording booth. Epoch 3 was then recorded following this period of potential naturalistic learning.

The entire protocol lasted approximately 90 min with experimental subjects experiencing 55 min of continuous real-time CI simulation. Control subjects ($n=3$) completed epochs 2 and 3, as well as the 15-min breaks spent in conversation with a researcher, under identical conditions to those of experimental subjects without ever wearing the PRTV or experiencing altered feedback.

F. Data analysis

Each isolated-word token containing a target sibilant fricative was analyzed for the segment's spectral center of gravity or centroid frequency (Jongman *et al.*, 2000). Centroids were calculated over the spectrum of the entire fricative regardless of its duration (delineated by the onset and offset of turbulent noise without simultaneous periodic voicing). Fricative contrast distances were also calculated, finding the difference in Hz between centroid frequencies of [s] and [ʃ] token pairs within each recording session. VOTs were measured as the duration, in milliseconds, between the presence of aspiration in the acoustic signal and the onset of periodic voicing for word stimuli containing target [p, t, k] in appropriate phonological context for aspiration to occur.

Speakers' production of vowel quality was analyzed using measurements of F1 and F2 formant frequencies taken

from the midpoints of stressed [i, æ, ɑ, u] vowels located in target isolated words. Automated PRAAT formant tracking (Boersma and Weenink, 2008) was used to extract F1 and F2 frequencies with corrections for tracking errors where consecutive tracking points jumped discontinuously by 50 Hz or more or were absent altogether for periods of at least 50 ms (e.g., due to glottalization or breathiness in the signal). Manual measurements were made in these regions by individuals trained in acoustic-phonetic research methods and blind to the hypotheses of the experiment. Inter-coder agreement was examined for approximately 10% of the total data (two epochs of each vowel for one male and one female talker); within each of these blocks, all of coders' mean F1 and F2 values for each vowel were found to be within one standard deviation of one another, the criteria set during initial experimental design (SD range 28.8 Hz for F1 of [æ] in the male talker to 313.6 Hz for F2 of [u] in the female talker).

These acoustic-phonetic data were then analyzed using repeated measures analysis of variance (RM-ANOVAs) with within-subject comparisons across variables such as Epoch (1, 2, 3), Fricative ([s, ʃ]), Vowel ([i, æ, ɑ, u]) and StopType ([p, t, k]), and between-subjects fixed effects of Group (experimental, control). Separate RM-ANOVAs were conducted for each segment type; in the case of F1 and F2 formant frequencies, analyses included both a multivariate omnibus RM-ANOVA test of both factors and the accompanying sub-analyses of each univariate effect. Other ANOVA were univariate. In all cases, if significant effects involving group were found, sub-analyses examining each group and the patterns of individual talkers were then conducted. If no differences were found, observed effects in subjects' speech were still reported, but no claims can be made as to the source or cause for these effects.

IV. RESULTS

A. Fricative center of gravity

RM-ANOVA (Huynh-Feldt corrected) conducted on spectral center of gravity measures for [s] and [ʃ] revealed significant effects of Epoch [$F(2,349)=12.88$, $p<0.001$, $\eta^2=0.066$] and Fricative [$F(1,182)=2799.33$, $p<0.001$, $\eta^2=0.939$] with no other significant main effects or interactions (p 's >0.05 , η^2 's <0.01). The two groups, therefore, did not differ in terms of the sibilant fricative production over the course of the experiment.

The significant effect of Fricative reflects the spectral contrast between [s] and [ʃ] in the phonology of American English with [s] having a higher average center of gravity (8989.4 Hz) than [ʃ] (5071.1 Hz; cf. Jongman *et al.*, 2000). Significant changes across Epoch were the result of significant raising between epoch 1 and epochs 2 and 3; *post hoc* comparisons were significant for these pairs (p 's <0.01 , Bonferroni correction) but not between the two non-initial epochs ($p>0.05$). The average sibilant centroid frequencies in epoch 1 were 6859.6 Hz (2241.5 Hz SD), with epochs 2 and 3 means at 7151.7 Hz (2175.4 Hz SD) and 7079.5 Hz (2210.5 Hz SD), respectively. Broken down by sibilant, the average centroid frequencies for [s] were 8799.2 Hz

(1122.9 Hz SD), 9033.9 Hz (1087.1 Hz SD), and 8980.7 Hz (1233.5 Hz SD), for epochs 1–3, respectively, while for [j] they were 4762.0 Hz (812.1 Hz SD), 5186.8 Hz (962.4 Hz SD), and 5109.9 Hz (901.5 Hz SD).

Measures of contrast distance between [s] and [ʃ], also analyzed with RM-ANOVA, did not yield significant effects of Epoch or Group, and no significant interaction was observed (all p 's > 0.05, η^2 's < 0.01). No evidence was obtained, therefore, that speakers treated their sibilant fricatives differently (i.e., by collapsing or expanding the contrast between them) or that this contrast responded in any way to the introduction of CI-simulated acoustic feedback. Changes were observed across time in the spectral characteristics of the sibilants [s] and [ʃ], but they do not appear to be distinguishable from those occurring in repetition alone.

B. Voice onset time

RM-ANOVA of aspirated stop VOTs revealed their production to be similarly consistent across groups, and also across the three recording epochs: No significant effects or interactions with either Epoch or Group were found (all p 's > 0.05, Huynh–Feldt corrected, η^2 's < 0.025). There was a significant main effect of StopType [$F(2,180) = 51.851$, $p < 0.001$, $\eta^2 = 0.366$], reflecting a well-documented correlation between place of articulation in stops and VOT (Cho and Ladefoged, 1999). As predicted, therefore, no evidence of changes in response to CI simulation was observed for this temporally-based phonetic cue.

C. Vowel quality

Although multivariate RM-ANOVA was planned for F1 and F2 vowel quality measurements, these data failed Box's M-test of homogeneity of covariance-variance ($p < 0.001$) and therefore violated assumptions of the multivariate test that are critical for analysis of unequal groups. Univariate analyses of F1 and F2 were therefore calculated to determine the effect of Group on vowel production.

RM-ANOVA of F1 revealed significant main effects of Vowel [$F(2.03,221.6) = 1079.156$ with Greenhouse–Geisser correction, $p < 0.001$, $\eta^2 = 0.908$] and Group [$F(1,109) = 10.267$, $p = 0.002$, $\eta^2 = 0.086$], along with the following two- and three-way interactions: Vowel $\times$ Group [$F(2.03,221.6) = 6.36$ with Greenhouse–Geisser correction, $p = 0.002$, $\eta^2 = 0.055$] and Vowel $\times$ Epoch $\times$ Group [$F(5,541.7) = 3.88$ with Huynh–Feldt correction, $p = 0.002$, $\eta^2 = 0.034$]. Significant group differences were observed, therefore, in speakers' production of F1; all other main effects and interactions were non-significant (p 's > 0.05, η^2 's < 0.025).

RM-ANOVA of F2 also revealed significant effects of Vowel [$F(2.07,231.4) = 669.750$ with Greenhouse–Geisser correction, $p < 0.001$, $\eta^2 = 0.857$] and Group [$F(1,112) = 5.862$, $p = 0.017$, $\eta^2 = 0.050$] with no other significant effects or interactions (all p 's > 0.05, η^2 's < 0.025).

The group differences observed in F1 and F2 above, however, were partially confounded with gender differences across the experimental and control groups: Two of the nine experimental talkers were male with correspondingly low vowel formants, while none of the three control talkers were

male. To control for the effects of gender on the preceding results, therefore, the RM-ANOVA were conducted again on data from only the female talkers in both groups. The results were unchanged in F1 [Group $F(1,96) = 4.825$, $p = 0.030$, $\eta^2 = 0.048$], but in F2, the main effect of Group was no longer significant [$F(1,97) = 0.635$, $p = 0.427$, $\eta^2 = 0.007$], leaving only a significant effect of Vowel. The group effect in F2, therefore, appears to have been caused by the confounding gender imbalance, while the differences observed in F1 across recording epochs and groups were robust to this factor.

Group differences in F1 production were therefore explored further via separate RM-ANOVA for each group to determine their differing patterns of response. In these within-group analyses, Subject was added as a between-subjects factor to investigate potential individual differences in the changes of F1 across vowels and epochs. RM-ANOVA of the experimental group included both male and female talkers because group differences were no longer being assessed. Analysis of female-only data from the experimental group resulted in qualitatively identical results in terms of significant and non-significant factors.

RM-ANOVA of F1 data (summarized in Table II) revealed the expected significant effects of Vowel and Subject in both the experimental and control groups, reflecting differences in formant frequency across the [i, æ, ɑ, u] vowels and across individuals' unique vowel spaces. A Vowel $\times$ Subject interaction was also observed for both groups, reflecting the ways in which individual differences impacted the realization of phonetic contrasts (cf. Peterson and Barney, 1952). These effects will not be discussed further.

In the experimental group, significant Epoch $\times$ Subject and Vowel $\times$ Epoch interactions were observed (Table II), indicating that different talkers and vowels, respectively, responded differently across the three recording epochs. A three-way Vowel $\times$ Epoch $\times$ Subject interaction revealed that these two factors also combined: Vowel-specific responses across epochs varied across experimental speakers. In the control group, only a significant main effect of Epoch and an Epoch $\times$ Subject interaction were found. Based on these results, we can conclude that subjects in the experimental group altered their F1 production in vowel-specific ways across recording epochs (regardless of whether we examine female-only or female/male data sets), while subjects in the control group only produced significant changes that were common across all four sampled vowels.

To illustrate these effects, average vowel spaces across the three recording epochs are plotted in Fig. 1 for the experimental group and Fig. 2 for the control group. As can be seen in these figures, speakers in the experimental group appear to have collapsed their F1 contrast over time, producing phonological raising (F1 decreases) in the low vowels [æ] and [ɑ], and lowering (F1 increases) in the high vowels [i] and [u]. Specifically, the average F1 of low vowels was significantly lower in epoch 1 baseline (mean F1 840.6 Hz) than in either epoch 2 (817.2 Hz) or epoch 3 (808.8 Hz), according to *post hoc* analysis (Bonferroni correction, p 's < 0.01). Simultaneously, the average F1 of high vowels was significantly higher in epochs 2 and 3 (mean F1

TABLE II. Results from two RM-ANOVA conducted on sub-components of F1 data from talkers in the experimental ($n = 9$) and control ($n = 3$) groups. Main effects and interactions which do *not* involve epoch, the factor related to changes from the introduction of CI simulation or multiple repetitions, are given in italics. Statistical results (F -values, p -values, and η^2 estimates of effect size) are shown in bold black font if they are significant at the 0.05 level. Non-significant results are in unbold gray font.

Effect	Experimental			Control		
	$F(\nu)$	p	η^2	$F(\nu)$	p	η^2
<i>Vowel</i>	<i>1267.3(2.3,173.8)</i>	<i><0.001</i>	<i>0.945</i>	<i>674.1(1.9,46.6)</i>	<i><0.001</i>	<i>0.964</i>
<i>Subject</i>	<i>34.28(8,74)</i>	<i><0.001</i>	<i>0.787</i>	<i>18.17(2,25)</i>	<i><0.001</i>	<i>0.592</i>
<i>Vowel $\times$ subject</i>	<i>6.99(18.8,173.8)</i>	<i><0.001</i>	<i>0.430</i>	<i>6.14(3.8,46.6)</i>	<i>0.001</i>	<i>0.329</i>
Epoch	0.172(2,148)	0.842	0.002	5.08(2,50)	0.010	0.169
Epoch $\times$ subject	7.46(16,148)	<0.001	0.446	3.57(4,50)	0.012	0.222
Vowel $\times$ epoch	8.52(4.2,309.5)	<0.001	0.103	1.03(3.9,99.1)	0.393	0.040
Vowel $\times$ epoch $\times$ subject	2.01(33.5,309.5)	<0.001	0.179	1.07(8,99.1)	0.390	0.079

410.3 Hz) than at baseline in epoch 1 (mean F1 383.2 Hz; *post hoc* Bonferroni comparisons, p 's < 0.01). Speakers in the control group, by contrast, produced small amounts of phonological lowering (F1 raising) between epochs 1 and 3 across all four vowels (mean difference of 17.5 Hz, significant Bonferroni comparison at $p < 0.05$). In summary, talkers who experienced CI-simulation of their feedback responded to this change by selectively collapsing their F1 vowel height contrast such that low vowels were phonologically raised and high vowels were phonologically lowered. Talkers who merely repeated the materials under normal conditions, on the other hand, shifted their entire vowel spaces downward, via small but consistent F1 raising in [i], [æ], [a], and [u].

The significant interactions involving the Subject factor, however, also indicate that there were individual differences in how these patterns were realized. The Epoch $\times$ Subject interactions in the experimental and control groups, for example, indicates that some experimental subjects did produce vowel-general changes in F1, and not every control talker produced the F1 shift so in the same way. Repeated-measures analysis of within-subject data for individual talkers supported this interpretation: In the experimental group, three of the nine talkers had significant effects of epoch

(after adjustment for multiple comparisons, see Table III), while in the controls, one of the three talkers did. Interestingly, the three experimental talkers with significant epoch effects were split in the direction of those changes: One talker increased F1 between the baseline and CI-simulated epochs, while the other two decreased average F1. This difference in the direction of F1 movement was also seen in those subjects with non-significant trends for change across epochs (Table III).

The combination of these contrasting global shifts and the significant three-way Vowel $\times$ Epoch $\times$ Subject interaction in experimental subjects likely resulted in the full-group observation of both significant low vowel raising and high vowel lowering discussed in the preceding text. As the results of the within-subject analyses for experimental talkers in Table III indicate, only 3/9 experimental talkers had Vowel $\times$ Epoch interactions that were significant after correction for multiple comparisons, and these stemmed from both raising and lowering behaviors. However, both types of change did not appear within single individuals. The vowel spaces of two of these talkers (S1 and S7), plotted in Fig. 3, demonstrate this specificity: The talker in panel A produced a large difference in the F1 of low vowels [æ, a] between baseline and CI-simulated speaking conditions, but very little difference in the F1 of high vowels [i, u]. The talker in

FIG. 1. (Color online) F1/F2 average [i, æ, a, u] qualities for speakers in the experimental group ($n = 9$) across recording epochs. Baseline, unperturbed productions are shown in black with a solid interpolation mapping the boundaries of their vowel space. Vowels produced under initial CI-simulation of acoustic feedback (epoch 2) have a large-dash interpolation, and those produced after increased experience with the real-time CI simulation (epoch 3) have a small-dash vowel space boundary.

FIG. 2. (Color online) Average F1/F2 vowel productions for speakers in the control group. No feedback perturbation was present in any of the recording epochs; materials were repeated three times under identical conditions. First repetition is shown in black (solid line), second and third repetitions are in dark and light gray, respectively (dashed lines). A global phonological lowering (increase in F1 frequencies) was observed.

the bottom panel, conversely, altered his production of F1 with the opposite specificity: Height in low vowels changed only minimally, while the F1 of [i, u] altered under CI-simulation. These statistically significant, opposite patterns held true for other experimental talkers as well, and together they demonstrate that, while all experimental talkers responded to CI simulation by reducing their F1 contrast, it was not the case that all talkers achieved this effect via over-all centralization and collapse. Rather, individual talkers produced vowel space reduction by selectively manipulating the height of subsets of vowels (see Table III).

V. DISCUSSION

In this investigation, we applied real-time CI simulation to the acoustic feedback of normal-hearing subjects in an effort to isolate and quantify the effects of CI signal degradation on speech production. Although we predicted that acoustic-phonetic characteristics of both sibilant fricatives and vowels would alter in response to the feedback degradation, significant changes above and beyond those that could be expected from procedural effects alone were observed only in vowel production, and there only in the domain of F1 or phonological height.

The lack of changes due to real-time CI simulation of sibilant fricatives stands in contrast both to studies of feedback perturbation in NH talkers and to studies of fricative production in deaf CI users. That is, localized perturbation of fricative frequency has been shown to cause compensatory effects in production of [s] and [ʃ] (Casserly, 2011; Shiller et al., 2009), and post-lingually deafened CI users’ contrast between the two sibilants is sensitive to the status of their auditory feedback (Lane et al., 2007; Perkell et al., 2007). Our NH subjects, on the other hand, did not show any change in the dispersion of their sibilant fricatives or in their absolute centroid frequencies. This difference in behavior could be due to the availability of bone-conducted acoustic feedback for our speakers, although similar conditions did

TABLE III. Summary of results from within-subject RM-ANOVA of F1 production for each talker in the experimental (CI-simulation) group. Effect size estimates (η^2) and p -values are given in lieu of full statistical results; α -criterion for significance, with Bonferroni correction for multiple comparisons, is $\alpha = 0.0055$. Significant results at this level are shown in bold black font. Directions of change in vowel quality are given for those speakers whose results had $\eta^2 \geq 0.20$. “Raising” indicates phonological raising (F1 decreases), while “lowering” indicates phonological lowering (F1 increases).

Subject	Epoch main effect			Vowel $\times$ epoch		
	η^2	p	Direction	η^2	p	Direction
S1	0.302	0.040	Raising	0.361	<0.001	[æ, ɑ] raising
S2	0.519	0.006	Raising	0.318	0.028	[æ, ɑ] raising
S4	0.551	0.018	Raising	0.394	0.014	[æ, ɑ] raising
S6	0.060	0.505		0.084	0.428	
S7	0.537	0.010	Lowering	0.386	0.005	[i, u] lowering
S10	0.673	<0.001	Raising	0.486	<0.001	[æ, ɑ] raising
S11	0.612	<0.001	Lowering	0.207	0.044	[i] lowering
S12	0.455	0.004	Lowering	0.212	0.038	[i, u] lowering
S13	0.254	0.072	Lowering	0.044	0.716

FIG. 3. (Color online) Mean F1/F2 vowel qualities for [i, æ, ɑ, u] produced by two experimental talkers with different realizations of individual significant vowel $\times$ epoch interactions in F1 (p ’s ≤ 0.005 ; see Table III). S1 (top panel) shows significant isolated raising of low vowels in CI-simulated (dashed lines) versus unperturbed speech, while S7 (bottom panel) shows similar isolated lowering of high vowels.

not prevent responses to perturbation in Casserly (2011) in [ʃ] perturbation, or it could be symptomatic of a group of talkers with particularly strong kinesthetic cues for fricative production, who were therefore unlikely to be particularly swayed by alteration of their acoustic feedback (Ghosh et al., 2010; Perkell et al., 2004b). Related to this possibility of individual differences, which will also be discussed in terms of vowel production in the following text, there is also the possibility that the current investigation was simply under-powered relative to the magnitude of individual differences, as a large degree of inter-talker variability has been observed in the segmental feedback perturbation literature (e.g., Casserly, 2011; Katseff et al., 2012; Villacorta et al., 2007). A third alternative, however, is that speakers in our study were able to somehow “disengage” with their feedback to a greater degree than is typically seen in localized perturbation research (Munhall et al., 2009). We will return to this possibility in the following text.

In terms of vowel quality, speakers significantly altered their speech production as a result of real-time CI simulation, as initially predicted. Given the number of phonological contrasts in the American English vowel space and the degree of spectral detail needed to maintain these segments, the loss of information in this domain represented a substantial problem for subjects attempting to monitor their production, particularly because kinesthetic feedback for vowels is relatively poor (cf. Perkell et al., 2004a). The specific patterns of response observed in vowels, however—the

specificity of responses to F1 and the idiosyncratic nature of individuals' responses—were difficult to predict from causal first principles.

Patterns of change seen in our talkers were consistent with findings from deaf CI users' vowel production (Lane *et al.*, 2007) but not with results from the NH localized feedback perturbation literature. That is, in formant-perturbation studies of vowel feedback, talkers consistently alter their production to partially compensate for experimental changes, either in F1 or F2 alone or in both formants simultaneously (e.g., Houde and Jordan, 1998, 2002; Purcell and Munhall, 2006; Reilly and Dougherty, 2013). Instead of such a regular, signal-driven response, however, our speakers showed impacts of altered feedback only in the first formant, produced changes that were often limited to particular vowels despite a global perturbation of spectral resolution, and demonstrated substantial individual differences in the direction, specificity, and degree of response to feedback alterations.

Moreover, these changes do not appear to be driven directly by properties of the CI signal processing transformation itself: There is no straightforward relationship between the frequency of a speaker's vowels and the boundaries of the CI spectral degradation. For example, it might have been the case that speakers whose vowels tended to fall in the lower half of frequency channels would have responded to the degradation by lowering F1 and those whose vowels fell in the top half of a channel would tend to raise F1—compensating for the average raising or lowering introduced by the smearing of frequency channels. But no such patterns emerged in the speaker/simulator relationships; as Fig. 4 shows for the vowel [æ], the relationship between acoustic starting point and eventual end point appeared to be random, at least with respect to the CI simulation transformation itself. S12 and S2 in Fig. 4, for example, have nearly identical values of F1 in [æ] at baseline, but under CI simulation, their behaviors are strikingly divergent. The same divergence from a common baseline production of F1 can also be seen in S11 and S10 and for a variety of other talkers in the group.

Overall, therefore, the changes in vowel quality observed here appear to be similar to the effects reported in the investigation of speech production and auditory feedback in deaf CI users (e.g., Lane *et al.*, 2007), rather than standard feedback perturbation responses (e.g., Purcell and Munhall, 2006). In Lane *et al.* (2007), when CI users' acoustic feedback was eliminated, they showed reductions in the distinctiveness of their vowel articulation, just as was observed here. The total lack of feedback achieved by turning off a CI is more extreme than the present simulation degradation, but it is possible that our degradation was sufficiently severe for our naive speakers that they effectively could not (or did not) maintain their use of feedback in real-time control whatsoever. This would be consistent with the lack of changes observed in fricative production; as discussed in the preceding text, it is possible the speakers were able to essentially disengage from their acoustic feedback and operate via a strictly feedforward control process.

FIG. 4. Average F1 frequencies of the low front vowel [æ] for talkers in the experimental group ($n=9$) across recording epochs. Those talkers whose within-subject data showed significant global changes in F1 or significant vowel-specific changes involving [æ] are plotted with solid lines; those without significant changes related to [æ] are dashed. Channel boundaries in the spectral processing of the CI simulator are also indicated. No straightforward relationship is apparent between a talker's F1 at baseline relative to the CI-simulator channels and the speaker's subsequent production behavior.

Our subjects did have an advantage in feedback use over the majority of CI users with the availability of an alternative source of information: Bone-conducted acoustic feedback. As mentioned in Sec. III, masking noise was not used to limit the effects of bone conduction in our subjects. Theoretically, therefore, they could have relied more heavily on this source of unperturbed feedback to maintain speech production accuracy, but it is not clear that they, as a group, did so—particularly given the similarity of the vowel production effects seen here and in deaf CI users with and without acoustic feedback.

If this parallel with the clinical speech production literature is correct and our talkers were severely disengaging with their acoustic feedback under CI simulation, then a critical question remains concerning the specificity of responses to the F1 dimension and the individual differences in the realization of that reduction in height contrast. There is no signal-driven motive for greater changes in F1 than F2, particularly because spectral resolution is actually better in these lower-frequency domains than in high-frequency F2 ranges (Table I), yet this specificity was consistent across subjects. Why, then, did speakers alter their production of F1 in particular? Why was the average tendency across speakers to collapse this contrast in height rather than to expand it or simply maintain it? What motivated the particular contrasting patterns of collapse seen, for example, in Fig. 3? And moreover, why alter vowel production at all if speakers were not using acoustic feedback correctively?

These questions are ripe for further exploration, but evidence suggests that the answers may lie in articulatory mechanics and individual differences in that domain (Perkell *et al.*, 2004b) and, e.g., auditory sensitivity (e.g., Shiller *et al.*, 2009; Villacorta *et al.*, 2007). If speakers were attempting to maximize the availability of somatosensory

feedback for speech produced under CI simulation, for instance, one strategy might have been to reduce the degree of openness used in the oral cavity. Depending on a speaker's particular vocal tract morphology, this reduction might have effect of lowering high vowels or raising low vowels, but it would almost certainly reduce the range of F1s produced, given its relationship with jaw height and the size of the oral resonating chamber (Stevens, 1998). Recent evidence has also shown that individuals are differentially sensitive to changes in their acoustic versus kinesthetic speech feedback and that these sensitivities affect the ways speakers respond to perturbations of either type (Lametti *et al.*, 2012). Such individual differences could clearly be at play in the present study even in the specifics of the vowel production response as well as potentially impacting the outcomes of deaf CI users or their success with articulation therapies (e.g., Pomaville and Kladopoulos, 2013). These differences therefore need to be studied more thoroughly to fully inform research on speech production under atypical auditory conditions and other domains in the study of perception/production interaction.

VI. CONCLUSIONS

Real-time CI simulation appears to be a viable means of investigating the effects of cochlear implant acoustic processing on speech production. In this preliminary exploration of the domain, changes in vocalic F1 production were observed in response to CI simulation that included reduction of the phonological height contrast, global shifts in F1 frequency, and selective changes to the height of classes of high or low vowels across individuals. These changes do not appear to be driven by the specifics of the CI signal transformation or to mirror the compensatory responses to feedback perturbation seen in localized manipulations but could reflect strategies to maximize kinesthetic feedback. Future studies using this method of probing the links between the perceptual consequences of cochlear implantation and subsequent speech production would have the potential to investigate these and a variety of questions capable of improving our understanding of the relationship between hearing and speech.

- Baker, R. E., and Bradlow, A. R. (2009). "Variability in word duration as a function of probability, speech style, and prosody," *Lang. Speech* **52**, 391–413.
- Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., Wright, M., Psarros, C., Rattigan, K., and Tooher, R. (2001). "Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing," *J. Speech Lang. Hear. Res.* **44**, 264–285.
- Boersma, P., and Weenink, D. (2008). "PRAAT: Doing phonetics by computer," Version 5.0.40.
- Burkholder-Juhasz, R. A., Dillon, C., Levi, S. V., and Pisoni, D. B. (2007). "Nonword repetition with spectrally reduced speech: Some developmental and clinical findings from pediatric cochlear implantation," *J. Deaf Studies Deaf Ed.* **12**, 472–485.
- Casserly, E. (2011). "Speaker compensation for local perturbation of fricative acoustic feedback," *J. Acoust. Soc. Am.* **129**, 2181–2190.
- Casserly, E., and Pisoni, D. B. (2013). "Nonword repetition as a predictor of long-term speech and language skills in children with cochlear implants," *Otol. Neurol.* **34**, 460–470.
- Casserly, E., Pisoni, D. B., Smalt, C. J., and Talavage, T. (2011). "A portable, real-time vocoder: Technology and preliminary perceptual learning findings," *J. Acoust. Soc. Am.* **129**, 2527A.
- Cho, T., and Ladefoged, P. (1999). "Variations and universals in VOT: Evidence from 18 languages," *J. Phonet.* **27**, 207–229.
- Cullington, H. E., and Zeng, F. (2008). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects," *J. Acoust. Soc. Am.* **123**, 450–461.
- Denes, P., and Pinson, E. (1963). *The Speech Chain* (Anchor/Doubleday, Garden City, NY), pp. 1–10.
- Elman, J. L. (1981). "Effects of frequency-shifted feedback on the pitch of vocal productions," *J. Acoust. Soc. Am.* **70**, 45–50.
- Francis, W. N., and Kucera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar* (Houghton Mifflin, Boston, MA), pp. 1–561.
- Fu, Q.-J., and Nogaki, G. (2005). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," *J. Assoc. Res. Otol.* **6**, 19–27.
- Fu, Q.-J., Nogaki, G., and Gavin, J. J. (2005). "Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation," *J. Assoc. Res. Otolaryngol.* **6**, 180–189.
- Fu, Q.-J., Shannon, R. V., and Wang, X. (1998a). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," *J. Acoust. Soc. Am.* **104**, 3586–3596.
- Fu, Q.-J., Zeng, F., Shannon, R. V., and Soli, S. D. (1998b). "Importance of tonal envelope cues in Chinese speech recognition," *J. Acoust. Soc. Am.* **104**, 505–510.
- Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., Guenther, F. H., Lane, H., and Perkell, J. S. (2010). "An investigation of the relation between sibilant production and somatosensory and auditory acuity," *J. Acoust. Soc. Am.* **128**, 3079–3087.
- Houde, J. F., and Jordan, M. I. (1998). "Sensorimotor adaptation in speech production," *Science* **279**, 1213–1216.
- Houde, J. F., and Jordan, M. I. (2002). "Sensorimotor adaptation of speech. I: Compensation and adaptation," *J. Speech Hear. Res.* **45**, 295–310.
- Howell, P., El-Yaniv, N., and Powell, D. J. (1987). "Factors affecting fluency in stutterers when speaking under altered auditory feedback," in *Speech Motor Dynamics in Stuttering*, edited by H. Peters and W. Hulstijn (Springer, New York), pp. 361–369.
- Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.* **108**, 1252–1263.
- Kaiser, A. R., and Svirsky, M. (1999). "A real time PC based cochlear implant speech processor with an interface to the Nucleus 22 electrode cochlear implant and a filtered noiseband simulation," Progress Report No. 23, Research on Spoken Language Processing, Indiana University, Bloomington, IN, pp. 417–428.
- Katseff, S., Houde, J. F., and Johnson, K. (2012). "Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback?," *Lang. Speech* **55**, 295–308.
- Lametti, D. R., Nasir, S. M., and Ostry, D. J. (2012). "Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback," *J. Neurosci.* **32**, 9351–9358.
- Lane, H., Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (2007). "Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts," *J. Speech Lang. Hear. Res.* **50**, 913–927.
- Lane, H., and Tranel, B. (1971). "The Lombard sign and the role of hearing in speech," *J. Speech Hear. Res.* **4**, 677–709.
- Lindbloom, B. (1990). "Explaining phonetic variation: A sketch of the H and H theory," in *Speech Production and Speech Modeling*, edited by W. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.
- Loebach, J. L., and Pisoni, D. B. (2008). "Perceptual learning of spectrally degraded speech and environmental sounds," *J. Acoust. Soc. Am.* **123**, 1126–1139.
- Loebach, J. L., Pisoni, D. B., and Svirsky, M. (2010). "Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant," *J. Exp. Psychol. Human Percept. Perform.* **36**, 224–234.
- Munhall, K. G., MacDonald, E. N., Byrne, S. K., and Johnsrude, I. (2009). "Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate," *J. Acoust. Soc. Am.* **125**, 384–390.

- Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (1984). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," Progress Report No. 10, Research on Speech Perception, Indiana University, Bloomington, IN, pp. 357–375.
- Oller, D. K., and Eilers, R. E. (1988). "The role of audition in infant babbling," *Child Dev.* **59**, 441–449.
- Perkell, J. S., Denny, M., Lane, H., Guenther, F., Matthies, M. L., Tiede, M., Vick, J., Zandipour, M., and Burton, E. (2007). "Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and post-lingually deafened cochlear implant users," *J. Acoust. Soc. Am.* **121**, 505–518.
- Perkell, J. S., Gunether, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (2004a). "The distinctness of speakers' production of vowel contrasts is related to their discrimination of the contrasts," *J. Acoust. Soc. Am.* **116**, 2338–2344.
- Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. H. (2004b). "The distinctness of speakers' /s-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," *J. Speech Hear. Res.* **47**, 1259–1269.
- Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* **24**, 175–184.
- Pomaville, F. M., and Kladopoulos, C. N. (2013). "The effects of behavioral speech therapy on speech sound production with adults who have cochlear implants," *J. Speech Hear. Res.* **56**, 531–541.
- Purcell, D. W., and Munhall, K. G. (2006). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," *J. Acoust. Soc. Am.* **120**, 966–977.
- Reilly, K. J., and Dougherty, K. E. (2013). "The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback," *J. Acoust. Soc. Am.* **134**, 1314–1323.
- Shafiro, V. (2008). "Identification of environmental sounds with varying spectral resolution," *Ear Hear.* **29**, 401–420.
- Shannon, R. V., Zeng, F., Kamath, V., and Ekelid, M. (1995). "Speech recognition with primary temporal cues," *Science* **270**, 303–304.
- Shiller, D. M., Sato, M., Gracco, V. L., and Baum, S. R. (2009). "Perceptual recalibration of speech sounds following speech motor learning," *J. Acoust. Soc. Am.* **126**, 1103–1113.
- Smalt, C. J., Gonzalez-Castillo, J., Talavage, T. M., Pisoni, D. B., and Svirsky, M. A. (2013). "Neural correlates of adaptation in freely-moving normal hearing subjects under cochlear implant acoustic simulations," *NeuroImage* **82**, 500–509.
- Stenfelt, S., and Håkansson, B. (2002). "Air versus bone conduction: An equal loudness investigation," *Hear. Res.* **167**, 1–12.
- Stevens, K. N. (1998). *Acoustic Phonetics* (MIT Press, Cambridge, MA), pp. 243–255.
- Tobey, E. A., Geers, A., Brenner, C., Altuna, D., and Gabbert, G. (2003). "Factors associated with development of speech production skills in children implanted by age five," *Ear Hear.* **24**, 36S–45S.
- Tobey, E. A., Geers, A., Sundarajan, M., and Shin, S. (2011). "Factors influencing speech production in elementary and high school-aged cochlear implant users," *Ear Hear.* **32**, 27S–38S.
- Villacorta, V. M., Perkell, J., and Guenther, F. H. (2007). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," *J. Acoust. Soc. Am.* **122**, 2306–2319.