The Journal of the Acoustical Society of America

MARCH 08 2024

Discrimination and sensorimotor adaptation of self-produced vowels in cochlear implant users

Agudemu Borjigin ; Sarah Bakst ; Katla Anderson; Ruth Y. Litovsky ; Caroline A. Niziolek

Check for updates

J. Acoust. Soc. Am. 155, 1895–1908 (2024)
https://doi.org/10.1121/10.0025063

View Online
Export Citation

14 August 2024 20:29:38

ASA
LEARN MORE
Advance your science and career as a member of the
Acoustical Society of America

ASA ACOUSTICAL SOCIETY OF AMERICA

JASA ARTICLE

Discrimination and sensorimotor adaptation of self-produced vowels in cochlear implant users

Agudemu Borjigin,[1,a)] Sarah Bakst,[1,b)] Katla Anderson,[1] Ruth Y. Litovsky,[1,2] and Caroline A. Niziolek[1,2]
[1]*Waisman Center, University of Wisconsin–Madison, Madison, Wisconsin 53705, USA*
[2]*Department of Communication Sciences and Disorders, University of Wisconsin–Madison, Madison, Wisconsin 53705, USA*

ABSTRACT:
Humans rely on auditory feedback to monitor and adjust their speech for clarity. Cochlear implants (CIs) have helped over a million people restore access to auditory feedback, which significantly improves speech production. However, there is substantial variability in outcomes. This study investigates the extent to which CI users can use their auditory feedback to detect self-produced sensory errors and make adjustments to their speech, given the coarse spectral resolution provided by their implants. First, we used an auditory discrimination task to assess the sensitivity of CI users to small differences in formant frequencies of their self-produced vowels. Then, CI users produced words with altered auditory feedback in order to assess sensorimotor adaptation to auditory error. Almost half of the CI users tested can detect small, within-channel differences in their self-produced vowels, and they can utilize this auditory feedback towards speech adaptation. An acoustic hearing control group showed better sensitivity to the shifts in vowels, even in CI-simulated speech, and elicited more robust speech adaptation behavior than the CI users. Nevertheless, this study confirms that CI users can compensate for sensory errors in their speech and supports the idea that sensitivity to these errors may relate to variability in production. © *2024 Acoustical Society of America*.
https://doi.org/10.1121/10.0025063
(Received 13 September 2023; revised 12 December 2023; accepted 11 February 2024; published online 8 March 2024)
[Editor: Christopher A. Brown] Pages: 1895–1908

14 August 2024 20:29:38

I. INTRODUCTION

The production of clear speech requires maintaining distinct contrasts between neighboring phonemes. The perception of the sound of one's own speech, or auditory feedback, plays an ongoing role in maintaining these phoneme contrasts (Cowie *et al.*, 1982; Lane and Webster, 1991). When speakers' perception of their own speech is manipulated experimentally by altering the auditory feedback, they tend to make adjustments to their speech to reduce the perceived auditory error (Houde and Jordan, 1998). Although these adjustments are largest when the altered feedback gives rise to a phonemic error, humans can perceive feedback alterations even within the same phoneme category and make changes in their production accordingly (Niziolek and Guenther, 2013). Additionally, these compensatory changes are greater in individuals with higher auditory acuity (Villacorta *et al.*, 2007; Martin *et al.*, 2018; Nault and Munhall, 2020), which may provide evidence that auditory acuity directly supports the sensitivity to auditory feedback required to maintain accurate speech. This relationship between auditory acuity and sensitivity to the error in auditory feedback can be seen as a reflection of the more general perception-production link, which has been extensively studied in the speech domain. For example, speakers who more accurately discriminate between two phonemes also produce greater acoustic and articulatory contrasts between those phonemes, for both vowels (Perkell *et al.*, 2004) and consonants (Brunner *et al.*, 2011). Similarly, in second-language learners, gains in perception of non-native contrasts transfer to production, with clearer and more accurate utterances following a perceptual training program (Bradlow *et al.*, 1997).

This perception-production interplay is absent in individuals who lack auditory input. However, cochlear implants (CIs) restore auditory feedback to individuals who are deaf, improving their perception of speech. More specifically, CI users showed better-defined phonemic categories in perception tasks 1 year after implantation than at 1 month, and, like acoustic hearing (AH) individuals, are much better at differentiating sounds that span across phonemic boundaries than sounds within a single phoneme category (Lane *et al.*, 2007b). As a consequence of these perceptual improvements, their speech tends to become more intelligible after receiving CIs (Vick *et al.*, 2001; Abbs *et al.*, 2020; Cychosz *et al.*, 2021; Gautam *et al.*, 2019; Priner *et al.*, 2021; Svirsky *et al.*, 1992; Svirsky *et al.*, 1998; Ubrig *et al.*, 2010). CI users show improved contrasts in their voiced vs voiceless consonants after getting their CIs (Lane and Perkell, 2005) and demonstrate increased phoneme contrast in their speech when their CIs are turned on as opposed to turned off (Bharadwaj *et al.*, 2006; Bharadwaj *et al.*, 2007; Matthies *et al.*, 1996; Richardson *et al.*, 1993; Lane *et al.*, 2007a). The phoneme contrast in CI users' speech also improves over time with longer exposure to auditory feedback. There is significant improvement after just 1 month post-implantation and more after 1 year post-implantation as

[a)]Email: dagu@wisc.edu
[b)]Current address: SRI International, Menlo Park, California 94025, USA

 © 2024 Acoustical Society of America

compared to the pre-implantation status (Ménard *et al.*, 2007). Additionally, similar to AH individuals, CI users also rely on auditory feedback in ongoing speech, producing more errors and speaking at a slower speech rate under conditions of a feedback delay (Taitelbaum-Swead *et al.*, 2019) and changing the pitch of their voice in response to a shift in the fundamental frequency (f0) of their auditory feedback (Loucks *et al.*, 2015; Gautam *et al.*, 2020).

Despite the aforementioned benefits from the restored auditory input after implantation, there are still challenges in both speech perception and production. CIs coarsely approximate the frequency analysis that is achieved by thousands of sensory cells in the peripheral auditory system by using a filter bank of only 12–22 frequency channels. The specific number of frequency channels depends on the type of electrode array and the number of electrodes available and activated along the implanted array. CIs discard fine-grained spectral and temporal information in the speech input, limiting access to rich acoustic information. Furthermore, due to anatomical restrictions, the electrode array cannot be fully inserted into the cochlea to cover the entire frequency range perceived by humans. It often takes some time for CI users to adjust to the new and coarse tonotopic (or frequency-to-place) map presented by their CIs (Ito and Sakakihara, 1994; Svirsky *et al.*, 2001). This mismatched frequency allocation with poor resolution limits CI users' access to cues that are available to AH individuals, such as fundamental frequency and vocal tract length (Gaudrain and Başkent, 2018). Although past research shows that CI users are capable of perceiving perturbations to their fundamental frequency, there is conflicting evidence as to how sensitive they are to small changes. Gfeller *et al.* (2007) reported that some CI users have the same sensitivity as AH individuals, while other research suggests that CI users are much less sensitive (Gautam *et al.*, 2020). Further, no study with CI users has yet investigated perturbations to formant frequencies, the broad peaks in the spectrum shaped by filtering from the vocal tract, which provide the primary perceptual cues to vowels and some consonants. Given this, it is unclear to what degree self-perception with a CI may allow for corrective changes to production that impact intelligibility.

The goal of this study is to investigate at what resolution CI listeners can perceive formant changes in their self-produced speech and whether they can use this information during speech production. Knowing the extent to which CI users use auditory feedback through electrical stimulation is crucial to understanding the mechanisms underlying deficits in speech clarity in some CI users. In this study, the first aim is to measure CI users' thresholds for detecting acoustic differences in self-produced vowels and, specifically, whether these CI users are sufficiently sensitive to detect subphonemic differences. Given this sensitivity to formant frequency differences in their speech, the second aim is to examine whether CI listeners can use this information to adjust their speech online and over time. AH participants were also recruited in these two experiments to serve as a control and validation of the measurements. In the first experiment, which measured thresholds for detecting acoustic differences in self-produced vowels, AH listeners' speech recordings were vocoded to simulate CI listening. The hypothesis is that CI listeners may not be sensitive to subphonemic acoustic differences unless these differences span multiple electrode boundaries. Similarly, AH listeners may not be sensitive to such differences in the vocoded speech unless the subphonemic differences cross multiple filter boundaries (Casserly, 2015; Casserly *et al.*, 2018). We further hypothesize that listeners with better sensitivity to small changes will demonstrate a greater ability to adapt their speech in response to detected errors.

II. METHODS

A. Participants

Fifteen participants with CIs (4 males and 11 females, 20–80 years old, average = 59 years) and 10 self-reported typical hearing participants (4 males and 6 females, 23–68 years old, average = 51 years) participated in the first experiment. Of these individuals, eight participants with CIs (3 males and 5 females, 42–80 years old, average = 61 years) and six self-reported typical hearing participants (4 males and 2 females, 45–68 years old, average = 56 years) also participated in the second experiment. Participants were age-matched in both experiments (experiment 1, $p = 0.2$; experiment 2, $p = 0.7$). All participants spoke North American English as a first language, with the exception of one self-reported typical hearing participant who learned Japanese and Mandarin from birth and began learning English at age 4.

All CI users were implanted with Cochlear, Ltd., CIs (Cochlear, Ltd., Sydney, Australia) that were programmed with the advanced combination encoder (ACE) stimulation strategy (Vandali *et al.*, 2000). All except one were bilaterally implanted. Demographic and implant details for CI users are listed in Table I. One CI user's data were excluded from experiment 2 due to the participant being "hyper-aware" of the speech manipulation and consciously changing their voice. Excluding this participant, experiment 2 included seven participants with CIs (2 males and 5 females, 42–80 years old, average = 61 years). All self-reported typical hearing participants underwent pure-tone audiometry to measure hearing sensitivity as a confirmation; here, we report measured hearing thresholds in the range of 250–4000 Hz, reflecting information relevant for vocalic sounds. All had thresholds of $\leq$10 dB hearing level (HL) in the range of the formant frequencies manipulated in the study (1000 Hz and below). Three of these ten participants had mild hearing loss (thresholds between 26 and 40 dB HL) in at least one ear at 2000 and/or 4000 Hz. The hearing loss at these frequencies is unlikely to affect perception of vowel formants in the range of 500–1000 Hz; we verified that there were no systematic differences in our dependent measures (formant discrimination thresholds and adaptation magnitude) between individuals with and without hearing loss at these higher frequencies. These self-reported typical hearing

14 August 2024 20:29:38

https://doi.org/10.1121/10.0025063

TABLE I. Demographic and implant information for CI users.[a]

Participant	Gender	Age at testing (years)	Age at hearing loss (years)	Age at implantation (years) (left, right)	Electrodes (left, right)
IBK*	Male	80	53	63, 69	22, 22
IDM*	Female	42	5	33, 35	20, 22
IBZ*	Female	52	38	40, 38	22, 22
IBY*	Female	56	41	43, 48	22, 19
IAU*	Male	71	3	50, 56	22, 21
IDK*ˆ	Male	64	16	50, 53	22, 14
ICJ*	Female	70	25	60, 60	22, 22
IZD*	Female	55	46	48, –	22, –
ICD	Female	61	3	50, 44	22, 22
ICC	Female	74	9	63, 60	22, 22
IBO	Female	70	23	45, 42	22, 22
ICI	Female	61	31	50, 51	22, 22
IDH	Male	20	3	5, 5	20, 20
IDA	Female	52	8	47, 46	22, 22
ICM	Female	66	23	57, 58	22, 22

[a]The last column indicates the total number of active electrodes in each ear. * indicates individuals who participated in both experiments 1 and 2. All others completed experiment 1 only. ˆ indicates the CI user whose data were excluded from experiment 2. IZD had only one implant.

individuals will be referred to as AH individuals for the rest of the text.

B. Experiment 1: Vowel discrimination in self-produced speech

In experiment 1, CI users and AH controls took part in an auditory discrimination task designed to assess their ability to discriminate formant differences in their own speech when perceived through a CI processor or a vocoder.

1. Stimuli

a. Speaker-specific formant continua. The stimuli for experiment 1 were the monosyllabic words "Ed" and "oh," which were chosen to align with previous studies investigating feedback-related compensatory changes to produced vowels (Bakst and Niziolek, 2019; Niziolek *et al.*, 2013). In a baseline recording phase at the beginning of experiment 1, participants were asked to speak each word aloud 30–60 times while their speech was recorded with a head-mounted microphone (AKG C520, sampling rate: 48 000 Hz). Participants received immediate feedback about the duration of each utterance, indicated by the length and color of a bar on the screen, and were instructed to keep their duration within a target range of 200–550 ms. For each word ("Ed" and "oh"), the experimenter chose one of the utterances with a clear recording (no creaky voice, coughs, etc.) that best matched the desired duration of 400 ms to be used as that word's baseline recording for experiment 1. The original recordings were padded with silence of equal length before the onset and after the offset to result in a stimulus duration of 400 ms. All other experiment 1 stimuli were generated by shifting the formants in these two baseline utterances (one for "Ed" and one for "oh"). Formant shifting was applied using AUDAPTER in offline mode [Fig. 1(a)] (Cai *et al.*, 2008; Tourville *et al.*, 2013). After the recordings were downsampled from 48 000 Hz to 16 000 Hz, the first two formants (F1 and F2) were estimated in the vowel portion of the baseline recordings using linear predictive coding (LPC), and the signal was filtered to transform the LPC coefficients and generate stimuli with the desired formants. Each participant's own formants were altered offline to pre-generate the following stimulus continuum: For each word, either the F1 (for "Ed") or the F2 (for "oh") was shifted in 1-Hz steps, spanning 500 Hz below to 500 Hz above the original recording. These stimuli were played through headphones as part of a discrimination paradigm (see Sec. II B 2).

b. Vocoding. A 16-channel vocoder (Leclère *et al.*, 2018) was used to simulate CI listening for the AH participants in experiment 1. Vocoding is a signal processing strategy adopted by CIs to transform the acoustic signal into several frequency bands of signals for electrical stimulation. To approximate CI users' frequency allocation tables, the signal was bandpass filtered into 16 frequency bands spanning the same frequency range as CIs (50–8000 Hz). The frequency boundaries for all filters are listed in Table II. The envelope of the signal from each frequency channel was then extracted through half-wave rectification and low-pass filtering with a cutoff frequency of 50 Hz. The envelope from each frequency channel was used to modulate a pure-tone signal at the center frequency of the same channel. The envelope-modulated sine waves from all 16 channels were finally combined to generate the vocoded signals. For more details on vocoder processing, please refer to Dorman *et al.* (1997).

c. Electrodogram. To estimate what CI users were receiving through their processors and implants while recorded stimuli were delivered through headphones, pulsatile stimulation patterns were simulated retrospectively using the ACE processing strategy (Holden *et al.*, 2002). This simulation produces a visualization of the electrical stimulation called an electrodogram. The audio recordings of the produced speech (i.e., "Ed" and "oh") were processed and turned into a series of pulses for all frequency channels according to the ACE stimulation strategy by using NUCLEUS MATLAB TOOLBOX (NMT [developed by Cochlear, Ltd.]). For more details on the generation of electrodograms, please refer to Peng *et al.* (2019). The electrodogram provides a visualization of how acoustic changes are represented in the electrical stimulation pattern and whether the shifts in stimulation pattern can reflect the perceptual thresholds. For direct visualization of the difference between electrodograms of the baseline and threshold recordings, each of the original electrodograms was first "smoothed" by counting the number of pulses in a sliding 10-ms window along the time axis. This process reduced noise in the difference electrodogram by minimizing small differences in the precise timing of pulses. The difference in pulse counts was then plotted with a color map. To extract a single quantitative

14 August 2024 20:29:38

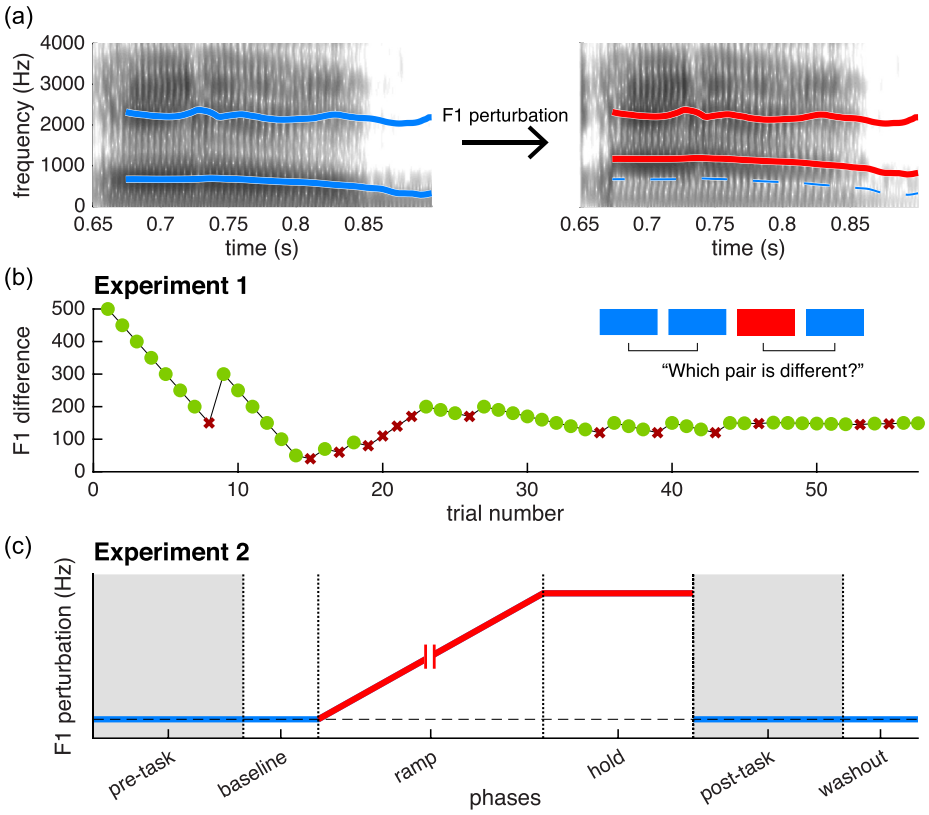

FIG. 1. (Color online) (a) Spectrogram showing a participant's production of "Ed" with formants overlaid (blue, original unaltered formants; red, altered formants). (b) Response track from a participant. Green round markers show correct responses, and red crosses show incorrect responses. (c) Phases of experiment 2. The gray sections indicate phases where the participant did not hear auditory feedback. For CI users, this meant turning their CIs off. For AH participants, this meant hearing accompanying masking noise.

value reflecting the difference between the two "smoothed" electrodograms, the cross correlation coefficient was also calculated using the MATLAB function corrcoef, which converts each 2-dimensional "smoothed" electrodogram into its vector representation and calculates the cross correlation of the two 1-dimensional vector arrays.

2. Procedure

The stimuli generated from each participant's baseline recordings were used in an auditory discrimination task designed to measure just-noticeable differences in formant frequencies of participants' own speech when perceived through a CI processor or vocoder. Participants wore circumaural headphones (Beyerdynamic DT 770; Beyerdynamic, Heilbronn, Germany) during the experiment. After participants' speech was recorded and formant-shifted stimuli were generated (as described in Sec. II B 1), thresholds were measured via a 4-interval, 2-alternative forced choice task using these stimuli [Fig. 1(b)]. The first and last intervals were the unaltered baseline recording, while either the second or third interval had an alteration to the formant of interest (F1 for "Ed" or F2 for "oh"). Loudness was set to a comfortable volume determined by the participant. To simulate the output from CIs, AH participants were presented with 16-channel vocoded versions of their speech (see Sec. II B 1). Participants' discrimination thresholds were determined using an adaptive procedure, starting with a maximal shift of 500 Hz (i.e., the relevant formant differed by 500 Hz from the baseline). The formant shift decreased by a set step size with each correct answer and increased by three times that step size with each incorrect answer (i.e., a weighted up-down procedure) in order to converge to an approximately 80% performance level on the psychometric function. The step size started at 50 Hz and was decreased to 10 Hz after 3 reversals or if the formant shift size reached 50 Hz (whichever happened first). The step size further decreased to 1 Hz if the formant shift size reached 10 Hz. The procedure terminated after 21 reversals. The formant shift sizes of the final 6 reversals were averaged to determine each participant's discrimination threshold.

TABLE II. Channel frequency information for the 16-channel vocoder.

	Cutoff frequency (Hz)		
Channel	Lower	Center	Upper
1	50	76	101
2	101	134	166
3	166	207	248
4	248	300	351
5	351	417	482
6	482	564	646
7	646	750	854
8	854	986	1117
9	1117	1283	1448
10	1448	1657	1866
11	1866	2130	2394
12	2394	2728	3061
13	3061	3482	3902
14	3902	4434	4965
15	4965	5636	6306
16	6306	7153	8000

14 August 2024 20:29:38

https://doi.org/10.1121/10.0025063

Four thresholds were measured for each participant: both formants (F1 for "Ed" and F2 for "oh") crossed with both shift directions (upward and downward formant shifts). The thresholds for each formant were collected in the same experimental run, interleaving the trials for upward and downward shifts. The order of the two formants was counterbalanced across participants.

3. Analysis

The purpose of experiment 1 was to determine CI users' perceptual thresholds for detecting formant shifts in their own speech. The thresholds were analyzed in relation to the number of frequency channels they spanned, as determined by each individual's processing strategy (for CI users) or by the vocoder channel cutoffs (for AH participants). Analyzing the thresholds in terms of the number of frequency channels allows for a perspective on how the poor frequency resolution in CIs would impact the listener's ability to perceive the shift in formants. A linear mixed effects model (LME4 package in RSTUDIO, R version 4.3.1) was used to analyze the effects of group (CI vs AH), formant (F1 vs F2), and the direction of the formant shifts (up vs down) on the thresholds: $\text{model} = \text{lmer}(\text{threshold} \sim \text{formant} + \text{direction} + \text{group} + 1|\text{subject})$. Before model analysis, thresholds in Hz were first mapped onto the mel scale, which accounts for the non-linearity of the cochlear tonotopic map by log-transforming frequencies such that equal steps sound equally different in pitch. The equation of the conversion is $\text{mel} = 2595 * \log_{10}(1 + \text{Hz}/700)$. The comparison of the quantiles from model residuals and a sample normal distribution verified that the residuals roughly follow a normal distribution. The homogeneity was checked by plotting the residuals against the model predictions. Across-individual variability between CI and AH groups was compared by an F test. Note that the thresholds from all 4 tests (2 formants $\times$ 2 directions) were merged for each group for this comparison analysis of across-individual variability. Finally, the F1-up and F1-down thresholds were averaged to give an F1 threshold that could be correlated with F1 production variability measured in the baseline recording phase. We limited our analysis to F1 in order to relate these data to experiment 2, in which only F1 is manipulated. The production variability for each individual was quantified by the mean difference from the average of the F1 estimates of all 30–60 baseline recordings (see Sec. II B 1). F1 estimates were made from the middle 50% of each recording in order to avoid dynamic formant transitions at the onset and offset of the vowel. Note that, in addition to regular linear regression [lm() function in R], the correlation was also calculated using the "robust linear regression" [rlm() function from R's MASS library], to investigate the possible impact of extreme values (i.e., potential outliers). For robust linear regression, we used three different estimators, including Huber, Hampel, and Tukey bisquare. Note that to test the significance of the correlation computed by the rlm models, we used the robust F test [or Wald test; f.robftest() function from the sfsmisc library in R] instead of the regular F test through analysis of variance (ANOVA). In addition to robust linear regression analyses, we also re-ran the regular linear regression analysis with the extreme values removed. All these measures were taken to see if the result would stay consistent with different approaches to handling potential outliers.

C. Experiment 2: Sensorimotor adaptation

In experiment 2, participants received altered auditory feedback in real time while speaking. We measured how much participants changed their speech in opposition to the feedback shift in order to assess sensorimotor adaptation.

1. Stimuli

In experiment 2, only the word "Ed" was used. This word appeared on a computer screen as a visual prompt to produce the word aloud. As in experiment 1, participants wore circumaural headphones and a head-mounted microphone. AUDAPTER was used to record speech at a sampling rate of 48 000 Hz, downsample it to 16 000 Hz, and, on some trials, apply a shift to the F1 in real time. For all trials, whether or not a formant shift was applied, participants' feedback was played back with a processing delay of $\sim$17 ms. As in experiment 1, after each trial, a visual bar indicated the duration of the participant's speech, with the aim of keeping duration within a target range of 200–550 ms. This range maximized the duration of auditory feedback received while maintaining naturalness.

2. Procedure

On each trial, speakers said the word "Ed" when prompted on the screen. The participant received a self-timed break after every 20 trials throughout experiment 2. Experiment 2 consisted of the following phases [Fig. 1(c)]:

a. Pre-task (100 trials). Participants spoke without auditory feedback: CI users were instructed to remove their implant(s), and AH participants heard masking noise at 80 dB sound pressure level (SPL). As participants received no feedback on the volume of their voice, the microphone gain was decreased to avoid clipping, precluding the analysis of amplitude across phases. At the end of the pre-task phase, speakers with CIs re-attached their devices.

b. Baseline (50 trials). All speakers heard their feedback through their headphones without any alteration to the formants. Both CI and AH participants heard accompanying speech-shaped noise at 60 dB SPL to mask bone-conducted sound so that speakers primarily hear the air-conducted altered feedback (Max and Maffett, 2015).

c. Ramp (variable number of trials, subject dependent: 90–194 trials). Speakers' auditory feedback was altered gradually: F1 was increased by 2 Hz on each successive trial. As in the baseline phase, all participants heard accompanying speech-shaped noise at 60 dB SPL to mask bone-conducted sound. To elicit the maximum adaptation from

14 August 2024 20:29:38

each participant, the ramp phase was of variable length that depended on the adaptive (F1-lowering) behavior of the participant. During the ramp phase, the baseline F1 values were plotted on the experimenter's screen for reference. As the ramp phase progressed, the F1 from each successive trial was overlaid on the baseline for comparison. For the first 75 ramp trials, the F1 was increased regardless of the speaker's behavior, corresponding to an F1 increase of 150 Hz. From this point onward, the speaker's productions were evaluated for a saturation point in their adaptive behavior. Every time the participant lowered their F1 in compensation, a 15-trial *observation* period was scheduled, in which the F1 shift continued to increase. As soon as the participant lowered their F1 again in that period, a new 15-trial observation period started. If the participant did not lower their F1 further within the 15-trial period, the ramp phase concluded, and the experimenter selected a maximum perturbation level corresponding to the trial where the participant showed the greatest adaptive behavior. This process (ramp + observation) was not automated to allow the experimenter to exclude trials with speech errors or non-speech sounds (such as coughing), which could result in erroneous F1 tracking or other obvious tracking errors. The experimenter's discretion was required to ensure that the tracking reflected intended productions.

d. Hold (100 trials). The alteration to speakers' F1 was held at the maximum perturbation level as defined by the ramp phase. As in the baseline and ramp phases, all participants heard accompanying speech-shaped noise at 60 dB SPL to mask bone-conducted sound.

e. Post-task (100 trials). Identical to the pre-task phase, participants with CIs removed their device(s) and AH participants heard accompanying masking noise at 80 dB SPL.

f. Washout (50 trials). Identical to the baseline phase, CI users re-attached their device(s) and all participants were re-acclimated to normal feedback.

3. Analysis

Formants were estimated from each spoken trial using WAVE_VIEWER (Niziolek, 2015), a MATLAB interface for formant tracking using PRAAT (Boersma and Weenink, 2019). Each trial was visually reviewed to ensure correct formant tracking, with analysis parameters (LPC order, pre-emphasis) adjusted if necessary. Within each trial, the formant tracks for the middle 40% of the vowel were averaged to produce an average F1 for that trial.

For each phase of interest, we calculated a single F1 measurement for each individual by averaging across the last ten trials of that phase. To measure potential aftereffects at the beginning of the post-task phase, the first ten trials of this phase were additionally analyzed. To visualize each participant's data, formant values were normalized by subtracting the average F1 value in the baseline phase, yielding the change in F1 from baseline. Although pre- and post-task phases did not provide participants with auditory feedback as the baseline and other phases did, resulting in changes to produced F1 values as a consequence of the Lombard effect, we decided to use the baseline phase as a single reference point for all phases for simplified visualization.

We report two metrics of adaptation: First, in order to compare adaptation magnitudes across participants, we measured the F1 change in a fixed set of trials in which the perturbation size was constant across all individuals. As the length of the ramp phase varied across participants from 90 to 194 trials, we defined this fixed set as ramp trials 81–90, the last ten trials before the perturbation schedules diverged. Second, and important for our purposes in assessing the maximum capacity for adaptation, we measured the maximum speech adaptation, defined as the F1 change in the last ten trials of the ramp phase, when each individual's adaptation response was observed to be at a plateau.

For each of these metrics, in addition to the absolute adaptation in Hz, we also calculated the percent adaptation, i.e., the adaptation as a percent of the corresponding perturbation. During the fixed set of ramp trials 81–90, the average applied perturbation was 171 Hz for all participants, so these percentages simply reflect a linear scaling of the adaptation. Conversely, the maximum speech adaptation occurred at a different set of trials for each participant, and therefore the applied perturbation magnitude differed. As a result, the percentage for these trials at the end of the ramp phase reflects the maximum speech adaptation as a percentage of the maximum applied perturbation. Importantly, percent adaptation has been shown to decrease as perturbation size increases (Katseff *et al.*, 2012; MacDonald *et al.*, 2010), and so this percentage is not appropriate for comparing adaptation across individuals, as it disadvantages individuals (mostly CI users) who took a longer time to reach the plateau. Therefore, whenever we compare adaptation across individuals (e.g., in a correlation), we use percent adaptation during fixed ramp trials. Specifically, we looked at the correlation between percent adaptation (i.e., fixed ramp trials 81–90) and other speech production and perception measures, including production variability and discrimination thresholds. Production variability was measured as in experiment 1 (see Sec. II B 3), but using the utterances from the baseline phase of the adaptation experiment. F1 discrimination thresholds were taken from experiment 1 (average of F1 thresholds measured with both shift directions).

One-tailed paired t tests were used to compare phases within each group (CI and AH). First, the last ten trials of the baseline phase were compared with the last ten trials of the ramp phase (at the plateau of the response) and the last ten trials of the hold phase (after maximum exposure to the perturbation) to assess formant changes in the presence of altered feedback. Similarly, the last ten trials of the pre-task phase were compared with ten-trial increments at the beginning and end of the post-task phase to assess learned changes to produced formants in the absence of ongoing feedback shifts. Finally, the first ten trials of the washout

14 August 2024 20:29:38

https://doi.org/10.1121/10.0025063

phase were compared with the last ten trials of the baseline phase as a metric of retained formant change after the reintroduction of normal feedback. Note that although both pre- and post-task phases were referenced to the baseline phase for visualization, statistical comparison was never conducted between a phase with feedback and a phase without feedback. For comparisons between CI and AH groups, pooled, two-tailed, two-sample *t* tests were used.

III. RESULTS

A. Experiment 1: Vowel discrimination in self-produced speech

1. Sensitivity to formant differences

A linear mixed-effects model revealed that CI listeners had higher discrimination thresholds than AH participants ($F_{1,94} = 7.4, p = 0.006$), despite the fact that AH participants listened to simulated CI stimuli. Figure 2 shows individual as well as group-level discrimination thresholds across all 4 test conditions: F1 up, F1 down, F2 up, and F2 down. F2 discrimination thresholds were higher than F1 discrimination thresholds ($F_{1,94} = 4.1, p = 0.04$). In general, participants with CIs showed higher average thresholds and more between-individual variability ($F_{59,37} = 2.1, p = 0.02$) than AH participants.

To determine the spectral resolution needed for detecting formant differences, the threshold was calculated in terms of the number of frequency channels it spanned for both CI and AH participants. The electrode/channel numbers containing the baseline and threshold formant frequencies were determined using the frequency allocation table (Table II). Some participants had baseline and threshold formant frequencies that fell within the same frequency band, i.e., they could detect within-channel differences. Others could only detect formant differences that spanned multiple channels. Seventy-nine percent of the thresholds from the AH participants fell within the same channel as the baseline, while the other 21% crossed only one channel. For CI users, on the other hand, 45% of the thresholds fell within the same channel as the baseline, while 55% of the thresholds crossed one or more channel boundaries. Regardless, CI users had better thresholds than hypothesized: many were sensitive to small formant changes, despite reduced spectral resolution provided by the CI. This result suggested that CI listeners might be able to use their auditory feedback to adapt their speech in opposition to the formant shifts applied in experiment 2.

2. Electrodogram analysis

To validate the observations above, electrodograms were generated for CI users. Electrodograms represent the electrical stimulation patterns/energy in each channel (see Sec. II B 1). Figure 3 shows examples of simulated electrodograms of both baseline and threshold recordings from a CI user with low discrimination thresholds (ICC; see Table I) and a CI user with high thresholds (IZD). Although CI users had higher discrimination thresholds in general, some of their thresholds were in the AH range, illustrated by the almost identical electrodograms of the baseline and threshold recordings, shown in panels (a) and (b) in Fig. 3. The cross correlation coefficient was used to quantify the difference between electrodograms of baseline and threshold recordings (see Sec. II B 1). The cross correlation coefficients between the baseline and threshold electrodograms correlated significantly with the thresholds in Hz for all experiments—F1 up ($r = -0.9, p < 0.0001$), F1 down ($r = -0.86, p < 0.0001$), F2 up ($r = -0.64, p = 0.01$), F2 down ($r = -0.9, p < 0.0001$)—suggesting that the simulated stimulation pattern reflects the measured perceptual thresholds. This *post hoc* electrodogram simulation therefore complements the analysis that considered how thresholds related to frequency channel boundaries.

3. Relationship between auditory discrimination and production variability

To determine whether these perceptual thresholds were related to participants' speech production, the correlation between the F1 threshold and F1 production variability was calculated, using the regular linear regression function (Fig. 4).

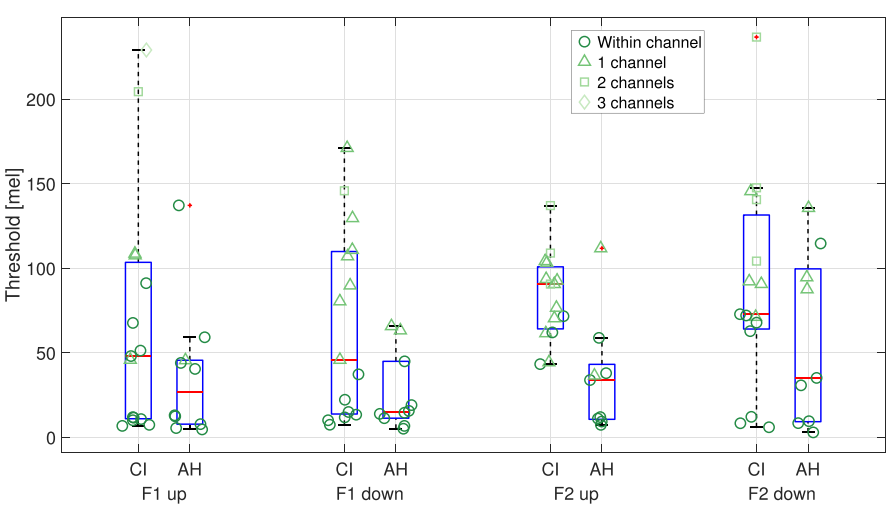

FIG. 2. (Color online) Discrimination thresholds for AH and CI participants. The box plots indicate the median and 25th to 75th percentiles. Note that some high thresholds are still "within channel," as some participants had baseline productions that fell in a broader frequency band or near the edge of a frequency band (e.g., baseline at the lower end in the frequency band and the threshold for detecting upward formant shift at the higher end of the same frequency band).

14 August 2024 20:29:38

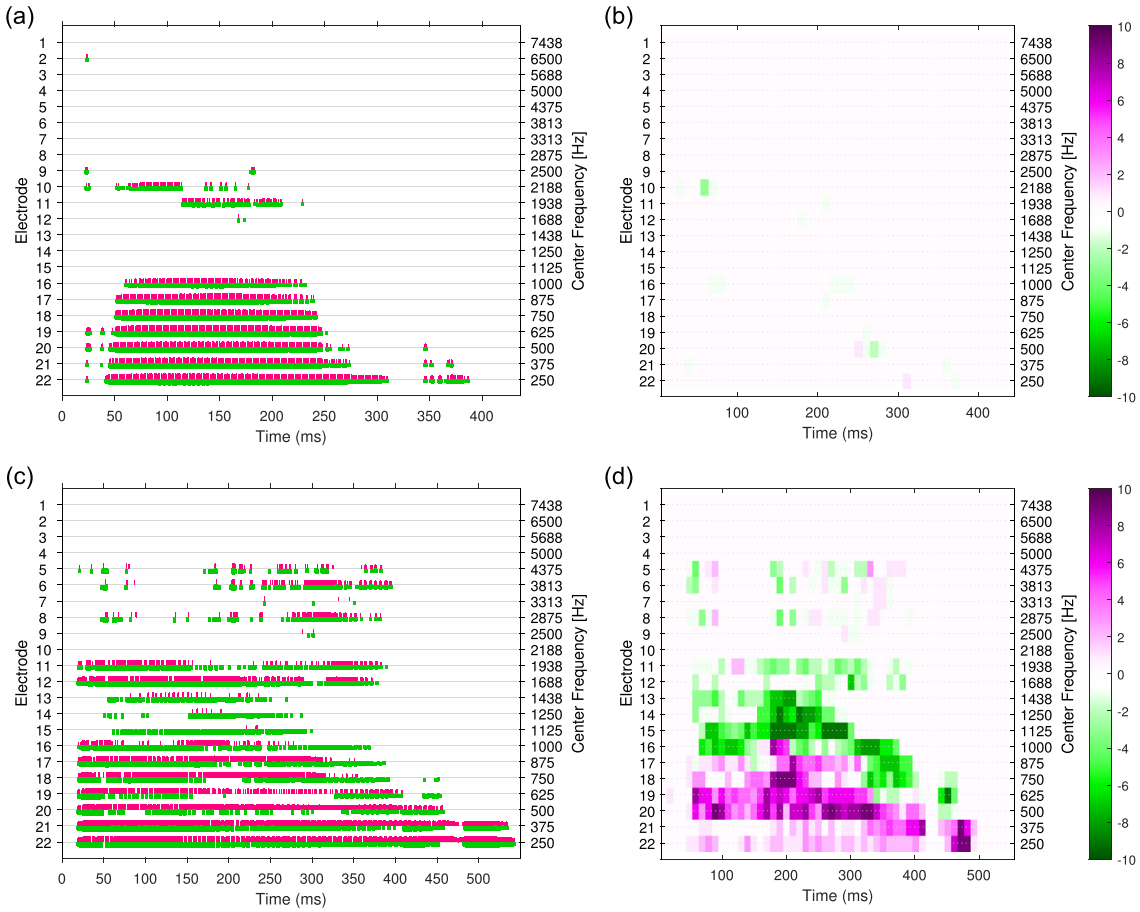

FIG. 3. (Color online) Electrodogram analyses from two participants: (a) and (b) from a participant with low discrimination thresholds (ICC) and (c) and (d) from a participant with high discrimination thresholds (IZD). Left panels (a) and (c) show electrodograms from the baseline (green) and threshold (pink) stimuli superimposed. Right panels (b) and (d) show the difference between baseline and threshold stimuli in the number of electrical pulses in a 10-ms window. The green end of the spectrum (negative values on color thermometer) indicates more pulses in the baseline electrodogram; the pink end (positive values) indicates more pulses in the threshold electrodogram.

Although the correlation was not significant for the CI ($r = 0.36, p = 0.19$) and AH ($r = 0.17, p = 0.63$) groups separately, when the groups were combined, a significant correlation was found: Participants with lower discrimination thresholds had less production variability ($r = 0.45, p = 0.025$). To account for several extreme data points, as shown in Fig. 4, robust linear regression was also conducted (see Sec. II B 3). The significant correlation still holds when the two groups were combined with the Huber and Hampel estimators ($F = 5.5$, $p = 0.028$; $F = 5.031$, $p = 0.035$, respectively). However, the correlation did not reach statistical significance when the bisquare estimator was used ($F = 0.043$, $p = 0.81$). Similarly, when the extreme data points were removed for the regular linear regression analysis, the correlation also failed to reach statistical significance.

B. Experiment 2: Sensorimotor adaptation

1. Adaptation to altered feedback

Figure 5 shows the produced F1 for an example CI user [Fig. 5(a)] and an example AH participant [Fig. 5(b)] over the course of the experiment. Both participants showed adaptive behavior, decreasing their F1 in response to the increased F1 in their auditory feedback and partially maintaining this decrease in the post-task phase relative to the pre-task phase (both without auditory feedback). The CI user showed a gradual decrease in F1 that lasted 194 trials,

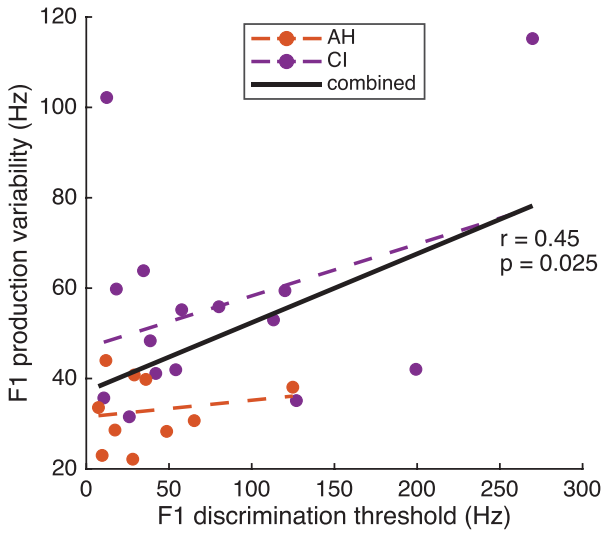

FIG. 4. (Color online) Participants with lower discrimination thresholds had less variability in F1 production.

14 August 2024 20:29:38

https://doi.org/10.1121/10.0025063

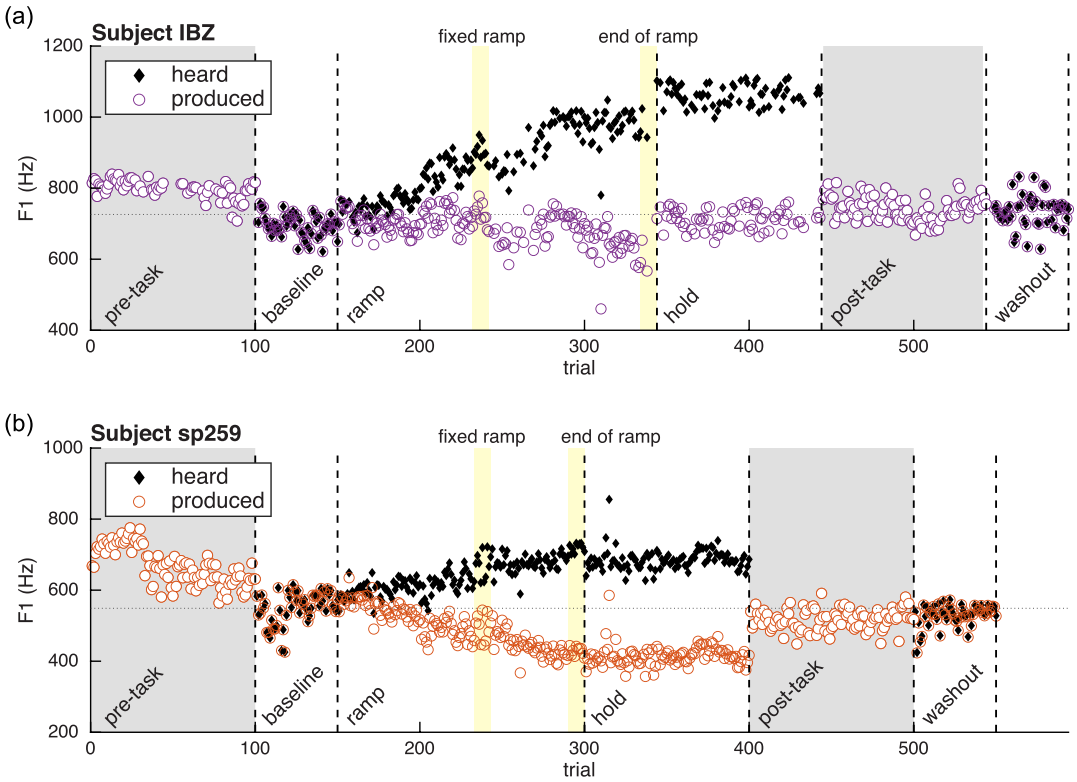

FIG. 5. (Color online) (a) F1 adaptation for an example CI user (IBZ) with a ramp phase lasting 194 trials. (b) F1 adaptation for an example AH participant with a ramp phase lasting 150 trials.

compared to the AH participant, whose F1 reached a plateau after 150 trials.

Figure 6 shows the average F1 adaptation for each group. Paired t tests showed that only the AH group had a statistically significant compensatory change in F1 with the fixed ramp trials (i.e., 81–90) [AH, $t(5) = 4.6, p = 0.003$; CI, $t(6) = 1.6, p = 0.081$]. This corresponded to average percent adaptations of 43.87% for AH participants and 13.28% for CI users. CI users had significantly lower percent adaptation than AH participants [$t(11) = -2.44, p = 0.033$]. However, both CI [$t(6) = 2.7, p = 0.017$] and AH [$t(5) = 4.6, p = 0.003$] groups showed a significant compensatory change in F1 in the last ten ramp trials, showing clear evidence of adaptation when their maximum adaptation was analyzed. CI users averaged a maximum F1 change of 41.7 Hz, and AH individuals averaged a maximum F1 change of 87.2 Hz. The difference between these formant changes at the group level did not reach statistical significance in a two-sample t test [$t(11) = 1.9, p = 0.08$]. These maximum F1 changes corresponded to average percent adaptations of 44.95% for AH participants and 16.84% for CI users. Note that CI users on average needed 31 more trials than AH participants (equivalent to a perturbation 62 Hz larger, as each ramp trial increased the perturbation by 2 Hz) to reach their maximum adaptation. AH participants maintained this F1 change at the end of the hold phase [$t(5) = 5.4, p = 0.002$], while the end of the hold phase did not differ from baseline for CI users [$t(6) = 0.9, p = 0.21$].

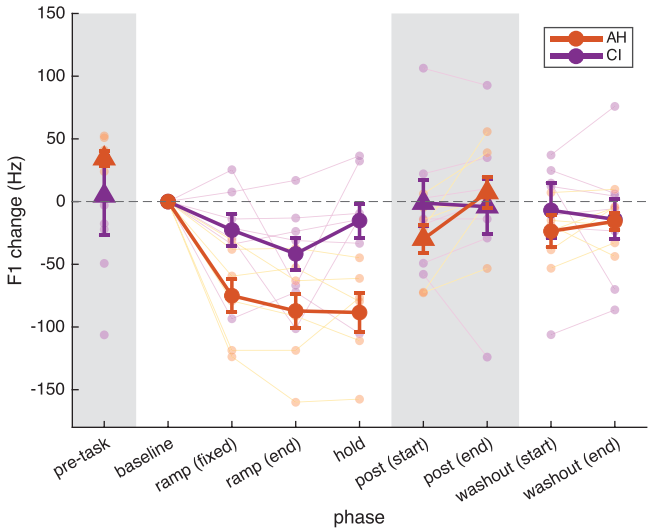

FIG. 6. (Color online) Mean F1 change throughout experiment 2 for CI and AH groups was plotted in thick lines overlaid on top of individual data (in thin lines). Both groups adapted to the F1 auditory shift by lowering their F1 production. The shaded areas represent phases where participants did not have access to auditory feedback.

In order to assess learned speech changes in the absence of the formant shift, the last ten trials of the pre-task phase were compared with ten-trial increments at the beginning and end of the post-task phase, as both phases were completed without auditory feedback. Paired t tests revealed that AH participants showed adaptation at the beginning of the post-task phase [$t(5) = 3.6, p = 0.008$] which decreased by the end of that phase [$t(5) = 1.6, p = 0.081$]. CI users did

14 August 2024 20:29:38

not show significant adaptation at either the beginning [$t(6) = 0.3, p = 0.38$] or the end [$t(6) = 0.5, p = 0.33$] of the post-task phase. Two-sample t tests showed that the difference in F1 values (post-task phase referenced to pre-task phase) between CI users and AH individuals was statistically significant in the beginning of the post-task [$t(11) = 2.31, p = 0.041$: more change in AH individuals], but not at the end of the post-task [$t(11) = 0.72, p = 0.48$].

In addition, to observe the degree to which adaptation persisted when normal auditory feedback was reintroduced, the ten-trial increment at the *beginning* of the washout phase was compared with the baseline phase (last ten trials). Paired t tests showed that AH participants more or less maintained adaptation at the beginning of the washout phase by sustaining their shifted F1 productions compared to their F1 productions in the baseline [$t(5) = 2.8, p = 0.019$]. However, CI users did not show the same adaptation at the beginning of the washout [$t(6) = 0.4, p = 0.36$]. Both groups' F1 returned to the level of baseline by the end of the washout phase [AH, $t(5) = 1.9, p = 0.056$; CI, $t(6) = 0.7$, $p = 0.26$]. Two-sample t tests showed that the difference in F1 values (washout phase referenced to baseline phase) between CI users and AH individuals was not statistically significant at the beginning of the washout [$t(11) = 0.79$, $p = 0.45$], or the end of the washout [$t(11) = 0.07$, $p = 0.95$].

2. Relationship between adaptation and speech production and perception measures

Past research has suggested that greater variability in F1 production is associated with greater speech compensation (Nault and Munhall, 2020), although evidence for this relationship has been mixed (MacDonald *et al.*, 2011). We determined the participants' variability in F1 production in the baseline phase by calculating the mean of the distances between each trial's F1 to the F1 value averaged across trials. These variances were then correlated with the participants' percent adaptation at fixed ramp trials (81–90). No correlation was found between production variability and speech adaptation in CI users ($r = 0.25, p = 0.59$), in AH participants ($r = 0.052, p = 0.92$), or when the two groups were combined ($r = -0.23, p = 0.45$). The correlation between participants' discrimination thresholds in experiment 1 and their percent adaptation at fixed ramp trials (81–90) in experiment 2 also failed to reach statistical significance for either group (CI users, $r = 0.68, p = 0.093$; AH participants, $r = 0.11, p = 0.83$) or for both groups combined ($r = 0.12, p = 0.69$).

IV. DISCUSSION

Speech production is not an isolated ability that only depends on the speech motor system. Like many other abilities, speech production requires exquisite multi-sensory integration. Auditory feedback plays a critical role for error detection and therefore online control of speech production. The goal of this study was to see if CI users can make use of auditory feedback for speech motor control. Auditory perception with CIs is degraded due to the poor spectral resolution in CI sound processing: Thousands of sensory cells encode different frequencies in a typical auditory system, whereas CIs reduce this information down to 12–22 frequency channels. In this study, two experiments were conducted to understand if CI users can use auditory feedback delivered by electrical stimulation to guide their speech production. The first experiment investigated the smallest change in formant frequencies that CI users might be able to detect in their own vowel productions, as well as how CI users compare to AH participants in detecting those changes. The second experiment investigated whether and how CI users might adapt their speech in the presence of altered auditory feedback. The knowledge from these two experiments has important implications for developing customized speech-production rehabilitation/optimization solutions for CI users. An effective speech rehabilitation solution is especially important for CI users who lost their hearing prelingually before the typical speech acquisition process.

In experiment 1, we found that CI users showed higher formant discrimination thresholds on average and more variability than the AH control group. However, CI listeners had much better discrimination sensitivity than hypothesized: Almost half of the group could detect formant differences within the same frequency channel as their baseline recording. These CI users were sensitive to small, within-category differences such as those introduced to their feedback during the second experiment, suggesting that they may use their auditory feedback to self-correct in running speech. In the second experiment, CI users did adapt their speech production under conditions of altered auditory feedback, although they did not change their F1 as much nor maintain this change as long as AH participants did. While AH participants showed the expected effects of longer-term learning, CI users did not.

It is encouraging to see that CI users can detect small fluctuations in auditory feedback from their own speech, given that their spectral resolution is considerably poor compared to AH listeners. CI users struggle to adapt to the shorter tonotopic range simulated through the CI, which results from the limited implantation depth of the electrode array into the cochlea. Further, the limited number of frequency channels coarsely encodes sound information (Ito and Sakakihara, 1994; Svirsky *et al.*, 2001). CI listeners usually cannot label vowels that they hear upon the initial activation of the implant. However, they seem to be able to fine-tune this skill over time with auditory feedback provided by their CI processors as they become more accustomed to the new frequency mapping (Kishon-Rabin *et al.*, 1999; Svirsky *et al.*, 2001). Nevertheless, it is a challenging and time-consuming process for CI users to adapt to the new frequency range in their auditory feedback. Some CI users' high sensitivity to small changes can be explained by the abundance of information in a vowel. The formants in a vowel are not fixed at a constant frequency, but rather

14 August 2024 20:29:38

change in frequency over time (Strange *et al.*, 1983; Hillenbrand, 2013). In this study, the formants were estimated at many time points throughout the vowel, but only the midpoint estimate (average of the middle 40%) was extracted as the reference frequency for the vowels at baseline and threshold. The single threshold point itself, as in many cases, does not give a full picture. Different electrodes may be stimulated over the course of the vowel, depending on the interaction between channel bandwidths and the extent to which the formants change over time. Such time-dependent channel stimulation is apparent in electrodograms, shown in Fig. 3, which highlights the importance of including an electrodogram analysis. Further, the formant values we report from the single midpoint estimate are the frequencies at the estimated *maxima* of the spectral peaks. Formants reflect the amplification of several harmonics that coincide at or near resonant frequencies of the vocal tract. Depending on whether the adjacent amplified harmonics fall in the same CI frequency channel as the spectral peak (i.e., formant frequency), as well as how much these adjacent harmonics are amplified relative to the peak (i.e., whether they hit the threshold for stimulation determined by the processing strategy), the within-channel sensitivity to small changes in *center* frequency may also reflect the activation of several adjacent stimulation channels.

In addition to showing higher average thresholds and more variability overall than the AH group, the results from experiment 1 also demonstrated higher thresholds in detecting changes in F2 than F1, even when using a logarithmic (mel) frequency scale. In fact, as a solution for faster rehabilitation, Svirsky *et al.* (2015) stimulated the cochlea only with a lower range of speech frequencies that sounded more "normal" first, rather than exposing the listeners to the higher-frequency spectrum right after implantation. Over time, more high-frequency channels were gradually introduced to include the full speech spectrum. This method of perceptual training was shown to be much more effective for improving speech understanding than the immediate exposure to the typical stimulation range containing higher frequency components. Considering the significant correlation between lower perceptual discrimination thresholds and less production variability (at the group level where data from CI and AH are combined), this more effective perceptual training paradigm from Svirsky *et al.* (2015) may also lead to more stable speech production. Perkell *et al.* (2008) have shown a similar finding on the relationship between perception and production as in our study, where participants who discriminated vowel contrasts more accurately also produced formants in vowels with less variability (specifically, task-relevant variability, as a consistent vowel quality would be expected across multiple repetitions of a vowel in the identical context). Although our results support this previous finding, the relationship between perceptual acuity and production variability has not been consistently reported. Franken *et al.* (2017) show only weak evidence for this relationship. In addition, it is not definitively true that the less variability in production, the better. Task-relevant motor variability can benefit the motor learning process in humans (Wu *et al.*, 2014). In everyday speech, variability in vowel duration or quality can convey critical information about the intended meaning of an utterance, marking stress, lexical focus, and segmental focus (de Jong, 2004), as well as phrasal position (Klatt, 1975). In addition, the correlation between discrimination thresholds and production variability should be interpreted with caution, considering the small number of samples involved and a few extreme data points that may have driven the correlation results. As mentioned in Secs. II and III, we used several methods to control for these potential outliers, and the significance of the correlation depended on the method used, highlighting the sensitivity of the results to how outlying observations are handled. Finally, the amount of difference in the stimulation patterns shown by the electrodograms for recordings at the baseline and threshold correlates with participants' sensitivity to the acoustic changes. It highlights the potential of the translation of this simulation into a clinical assessment metric that can help clinicians understand and monitor patients' perceptual outcomes, which is especially critical for assessing individuals who face communication barriers such as infants with CIs.

Experiment 1 showed that CI users were able to perceive small alterations in their auditory feedback, and experiment 2 showed that they were able to integrate information from that feedback into speech production: CI users did show compensatory speech behavior in response to altered auditory feedback. Considering that CI users had higher discrimination thresholds than AH participants in experiment 1, it was not surprising to see that CI users on average had less speech adaptation than AH participants at a fixed perturbation size (ramp trials 81–90). However, our goal in the current study was to maximize the opportunity to observe adaptation through a variable ramp that increased the feedback perturbation until participants stopped adapting. With prolonged exposure and increased magnitude of the perturbation, CI users' adaptation continued to increase until it was not significantly different from AH participants' adaptation at the group level. This underscores that the capacity to adapt to altered feedback is present even when discrimination thresholds are relatively high, given a large enough error in the feedback signal.

In addition, we observed substantial between-individual variability in adaptation from both CI and AH groups. In certain phases of the experiment, the between-individual variability in the AH group is comparable to or even greater than that observed in the CI group. This wide individual variability in speech adaptation behavior is not uncommon and has been reported by many other studies (Parrell and Niziolek, 2021; Martin *et al.*, 2018; Munhall *et al.*, 2009; Parrell *et al.*, 2017; Villacorta *et al.*, 2007). This variability was also reported in CI users in fundamental frequency (f0) control (Gautam *et al.*, 2020). At the individual level, some CI users showed even more speech adaptation behavior than some AH participants. No correlation was found between percent adaptation (at fixed ramp trials, 81–90) and within-

14 August 2024 20:29:38

subject variability in baseline production, suggesting that the variability in adaptation is not primarily due to overall individual differences in baseline production variability. We did not observe effects of learning in the CI group in this experiment. It may be the case that the high rate of variability, combined with a very small sample size, may have obscured a relatively small learning effect. Similarly, due to the large individual variance and small sample size, the correlation between the two experiments, i.e., perceptual discrimination thresholds and speech adaptation in the production, was not conclusive, although Martin *et al.* (2018) showed that auditory acuity strongly predicts speech adaptation with a larger population sample ($n = 36$).

V. CONCLUSION

CI users are sensitive to small changes in their auditory feedback and use this information in speech motor control, like AH speakers. Some CI users were able to detect small, within-channel differences in their auditory feedback. Further, many CI users were able to integrate this auditory information into their speech production and adapt to real-time alterations in their feedback. Although we investigated this speech adaptation phenomenon in highly constrained settings (i.e., within single vowels), the ultimate goal is to apply the findings to speech in natural settings. Lametti *et al.* (2018) showed similar adaptation in sentences, indicating that the interpretation of our finding could be extended to natural connected speech. However, it is important to note that some CI users could only detect very large perturbations in their feedback, and some did not adapt their speech in response to altered feedback. Further investigation into the mechanisms underlying these individual differences would be the next step towards developing customized solutions based on unique individual profiles.

ACKNOWLEDGMENTS

Research reported in this paper was supported by NIH NIDCD F32 Grant No. F32Dc017653 to S.B., NIH Grant No. R01DC003083 to R.Y.L., NIH Grant No. R00 DC014520 to C.A.N., NIH T32 training Grant No. 5T32DC005359-13 to UW–Madison, and by NIH NICHD Grant No. U54HD090256 to the Waisman Center. Many thanks to Stephen R. Dennison for help with electrodogram analysis and manuscript revision.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

Participants provided consent and were compensated financially for their time. All procedures were approved by the Institutional Review Board of University of Wisconsin–Madison.

Author Contributions

Agudemu Borjigin: Data curation (supporting); Formal analysis (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (lead). **Sarah Bakst:** Conceptualization (lead); Formal analysis (supporting); Funding acquisition (equal); Investigation (lead); Methodology (lead); Writing – original draft (supporting); Writing – review & editing (supporting). **Katla Anderson:** Data curation (lead); Formal analysis (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Ruth Y. Liovsky:** Conceptualization (supporting); Formal analysis (supporting); Funding acquisition (equal); Methodology (supporting); Supervision (equal); Writing – review & editing (supporting). **Caroline A. Niziolek:** Conceptualization (supporting); Formal analysis (supporting); Funding acquisition (equal); Methodology (lead); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting).

DATA AVAILABILITY

The data that support the findings of this study are openly available in the Open Science Framework (OSF) project at https://osf.io/zchma/?view_only=0cbbe7f6f4cb419cb562280878f310e7.

Abbs, E., Aronoff, J. M., Kirchner, A., O'Brien, E., and Harmon, B. (**2020**). "Cochlear implant users' vocal control correlates across tasks," J. Voice **34**(3), 490.e7–490.e10.

Bakst, S., and Niziolek, C. A. (**2019**). "Self-correction in L1 and L2 vowel production," in *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, edited by S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Australasian Speech Science and Technology Association Inc., Canberra City, Australia), pp. 3185–3189.

Bharadwaj, S. V., Graves, A. G., Bauer, D. D., and Assmann, P. F. (**2007**). "Effects of auditory feedback deprivation length on the vowel /e/ produced by pediatric cochlear-implant users," J. Acoust. Soc. Am. **121**(5), EL196–EL202.

Bharadwaj, S. V., Tobey, E. A., Assmann, P. F., and Katz, W. (**2006**). "Effects of auditory feedback on fricatives produced by cochlear-implanted adults and children: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **119**(3), 1626–1635.

Boersma, P., and Weenink, D. (**2019**). "Praat: Doing phonetics by computer (version 6.4.05) [computer program]," https://www.fon.hum.uva.nl/praat/ (Last viewed February 17, 2024).

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/. IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**(4), 2299–2310.

Brunner, J., Ghosh, S., Hoole, P., Matthies, M., Tiede, M., and Perkell, J. (**2011**). "The influence of auditory acuity on acoustic variability and the use of motor equivalence during adaptation to a perturbation," J. Speech. Lang. Hear. Res. **54**(3), 727–739.

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (**2008**). "A system for online dynamic perturbation of formant trajectories & results from perturbations of the Mandarin triphthong /iau/," in *Proceedings of the 8th International Seminar on Speech Production, ISSP 2008*, Strasbourg, France, edited by R. Soook, S. Fuchs, and Y. Laprie (INRIA, Rocquencourt, France), pp. 65–68.

Casserly, E. D. (**2015**). "Effects of real-time cochlear implant simulation on speech production," J. Acoust. Soc. Am. **137**(5), 2791–2800.

Casserly, E. D., Wang, Y., Celestin, N., Talesnick, L., and Pisoni, D. B. (**2018**). "Supra-segmental changes in speech production as a result of

14 August 2024 20:29:38

spectral feedback degradation: Comparison with Lombard speech," Lang. Speech **61**(2), 227–245.

Cowie, R., Douglas-Cowie, E., and Kerr, A. G. (**1982**). "A study of speech deterioration in post-lingually deafened adults," J. Laryngol. Otol. **96**(2), 101–112.

Cychosz, M., Munson, B., Newman, R., and Edwards, J. R. (**2021**). "Auditory feedback experience in the development of phonetic production: Evidence from preschoolers with cochlear implants and their normal-hearing peers," J. Acoust. Soc. Am. **150**(3), 2256–2271.

de Jong, K. (**2004**). "Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration," J. Phon. **32**(4), 493–516.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**(4), 2403–2411.

Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., and Hagoort, P. (**2017**). "Individual variability as a window on production-perception interactions in speech motor control," J. Acoust. Soc. Am. **142**(4), 2007–2018.

Gaudrain, E., and Başkent, D. (**2018**). "Discrimination of voice pitch and vocal-tract length in cochlear implant users," Ear Hear. **39**(2), 226–237.

Gautam, A., Brant, J. A., Ruckenstein, M. J., and Eliades, S. J. (**2020**). "Real-time feedback control of voice in cochlear implant recipients," Laryngoscope Investig. Otolaryngol. **5**(6), 1156–1162.

Gautam, A., Naples, J. G., and Eliades, S. J. (**2019**). "Control of speech and voice in cochlear implant patients," Laryngoscope **129**(9), 2158–2163.

Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., and Olszewski, C. (**2007**). "Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise," Ear Hear. **28**(3), 412–423.

Hillenbrand, J. M. (**2013**). "Static and dynamic approaches to vowel perception," in *Modern Acoustics and Signal Processing: Vowel Inherent Spectral Change*, edited by G. S. Morrison and P. F. Assmann (Springer, Berlin), pp. 9–30.

Holden, L. K., Skinner, M. W., Holden, T. A., and Demorest, M. E. (**2002**). "Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy," Ear Hear. **23**(5), 463–476.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**(5354), 1213–1216.

Ito, J., and Sakakihara, J. (**1994**). "The mechanism of speech perception in patients with a multichannel cochlear implant," Clin. Otolaryngol. **19**(4), 346–349.

Katseff, S., Houde, J., and Johnson, K. (**2012**). "Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback?," Lang. Speech **55**(2), 295–308.

Kishon-Rabin, L., Taitelbaum, R., Tobin, Y., and Hildesheimer, M. (**1999**). "The effect of partially restored hearing on speech production of postlingually deafened adults with multichannel cochlear implants," J. Acoust. Soc. Am. **106**(5), 2843–2857.

Klatt, D. H. (**1975**). "Vowel lengthening is syntactically determined in a connected discourse," J. Phon. **3**(3), 129–140.

Lametti, D. R., Smith, H. J., Watkins, K. E., and Shiller, D. M. (**2018**). "Robust sensorimotor learning during variable sentence-level speech," Curr. Biol. **28**(19), 3106–3113.e2.

Lane, H., Denny, M., Guenther, F., Hanson, H., Marrone, N., Matthies, M., Perkell, J., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (**2007a**). "On the structure of phoneme categories in listeners with cochlear implants," J. Speech Lang. Hear. Res. **50**, 2–14.

Lane, H., Matthies, M. L., Guenther, F. H., Denny, M., Perkell, J. S., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (**2007b**). "Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts," J. Speech Lang. Hear. Res. **50**(4), 913–927.

Lane, H., and Perkell, J. S. (**2005**). "Control of voice-onset time in the absence of hearing," J. Speech Lang. Hear. Res. **48**(6), 1334–1343.

Lane, H., and Webster, J. W. (**1991**). "Speech deterioration in postlingually deafened adults," J. Acoust. Soc. Am. **89**(2), 859–866.

Leclère, T., Kan, A., and Litovsky, R. (**2018**). "Binaural intelligibility level difference with a mixed-rate strategy simulation," J. Acoust. Soc. Am. **143**(3), 1941.

Loucks, T. M., Suneel, D., and Aronoff, J. M. (**2015**). "Audio-vocal responses elicited in adult cochlear implant users," J. Acoust. Soc. Am. **138**(4), EL393–EL398.

MacDonald, E. N., Goldberg, R., and Munhall, K. G. (**2010**). "Compensations in response to real-time formant perturbations of different magnitudes," J. Acoust. Soc. Am. **127**(2), 1059–1068.

MacDonald, E. N., Purcell, D. W., and Munhall, K. G. (**2011**). "Probing the independence of formant control using altered auditory feedback," J. Acoust. Soc. Am. **129**(2), 955–965.

Martin, C. D., Niziolek, C. A., Duñabeitia, J. A., Perez, A., Hernandez, D., Carreiras, M., and Houde, J. F. (**2018**). "Online adaptation to altered auditory feedback is predicted by auditory acuity and not by domain-general executive control resources," Front. Hum. Neurosci. **12**, 91.

Matthies, M. L., Svirsky, M. A., Perkell, J. S., and Lane, H. (**1996**). "Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant," J. Speech Lang. Hear. Res. **39**(5), 936–946.

Max, L., and Maffett, D. G. (**2015**). "Feedback delays eliminate auditory-motor learning in speech production," Neurosci. Lett. **591**, 25–29.

Ménard, L., Polak, M., Denny, M., Burton, E., Lane, H., Matthies, M. L., Marrone, N., Perkell, J. S., Tiede, M., and Vick, J. (**2007**). "Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants," J. Acoust. Soc. Am. **121**(6), 3790–3801.

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**(1), 384–390.

Nault, D. R., and Munhall, K. G. (**2020**). "Individual variability in auditory feedback processing: Responses to real-time formant perturbations and their relation to perceptual acuity," J. Acoust. Soc. Am. **148**(6), 3709–3721.

Niziolek, C. (**2015**). "wave_viewer," https://doi.org/10.5281/zenodo.13839 (Last viewed February 17, 2024).

Niziolek, C. A., and Guenther, F. H. (**2013**). "Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations," J. Neurosci. **33**(29), 12090–12098.

Niziolek, C. A., Nagarajan, S. S., and Houde, J. F. (**2013**). "What does motor efference copy represent? Evidence from speech production," J. Neurosci. **33**(41), 16110–16116.

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., and Ivry, R. B. (**2017**). "Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration," J. Neurosci. **37**(38), 9249–9258.

Parrell, B., and Niziolek, C. A. (**2021**). "Increased speech contrast induced by sensorimotor adaptation to a nonuniform auditory perturbation," J. Neurophysiol. **125**(2), 638–647.

Peng, Z. E., Hess, C., Saffran, J. R., Edwards, J. R., and Litovsky, R. Y. (**2019**). "Assessing fine-grained speech discrimination in young children with bilateral cochlear implants," Otol. Neurotol. **40**(3), e191–e197.

Perkell, J., Lane, H., Ghosh, S., Matthies, M., Tiede, M., Guenther, F., and Ménard, L. (**2008**). "Mechanisms of vowel production: Auditory goals and speaker acuity," in *Proceedings of the 8th International Seminar on Speech Production, ISSP 2008*, Strasbourg, France, edited by R. Soook, S. Fuchs, and Y. Laprie (INRIA, Rocquencourt, France), pp. 29–32.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (**2004**). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**(4), 2338–2344.

Priner, R., Cranial, C., Chayat, C., Fraenkel, R., and Brand, D. (**2021**). "Effect of auditory feedback on speech intelligibility of adults with cochlear implants," Eur. Arch. Otorhinolaryngol. **279**(9), 4345–4351.

Richardson, L. M., Busby, P. A., Blarney, P. J., DoweII, R. C., and Clark, G. M. (**1993**). "The effects of auditory feedback from the nucleus cochlear implant on the vowel formant frequencies produced by children and adults," Ear Hear. **14**(5), 339–349.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**(3), 695–705.

Svirsky, M., Silveira, A., Suarez, H., Neuburger, H., Lai, T., and Simmons, P. (**2001**). "Auditory learning and adaptation after cochlear implantation: A preliminary study of discrimination and labeling of vowel sounds by cochlear implant users," Acta Oto-Laryngologica **121**, 262–265.

Svirsky, M. A., Jones, D., Osberger, M. J., and Miyamoto, R. T. (**1998**). "The effect of auditory feedback on the control of oral-nasal balance by pediatric cochlear implant users," Ear Hear. **19**(5), 385–393.

14 August 2024 20:29:38

Svirsky, M. A., Lane, H., Perkell, J. S., and Wozniak, J. (**1992**). "Effects of short-term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**(3), 1284–1300.

Svirsky, M. A., Talavage, T. M., Sinha, S., Neuburger, H., and Azadpour, M. (**2015**). "Gradual adaptation to auditory frequency mismatch," Hear. Res. **322**, 163–170.

Taitelbaum-Swead, R., Avivi, M., Gueta, B., and Fostick, L. (**2019**). "The effect of delayed auditory feedback (DAF) and frequency altered feedback (FAF) on speech production: Cochlear implanted versus normal hearing individuals," Clin. Linguist. Phon. **33**(7), 628–640.

Tourville, J. A., Cai, S., and Guenther, F. (**2013**). "Exploring auditory-motor interactions in normal and disordered speech," Proc. Mtgs. Acoust. **19**(1), 060180.

Ubrig, M. T., Goffi-Gomez, M. V., Weber, R., Menezes, M. M., Nemr, N. K., Tsuji, D. H., and Tsuji, R. K. (**2010**). "Voice analysis of postlingually deaf adults pre- and postcochlear implantation," ScienceDirect **25**(6), 692–699.

Vandali, A. E., Whitford, L. A., Plant, K. L., and G. M. Clark (**2000**). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system," Ear Hear. **21**(6), 608–624.

Vick, J., Lane, H., Perkell, J., Matthies, M., Gould, J., and Zandipour, M. (**2001**). "Covariation of cochlear implant users' perception and production of vowel contrasts and their identification by listeners with normal hearing," J. Speech Lang. Hear. Res. **44**(6), 1257–1267.

Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (**2007**). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," J. Acoust. Soc. Am. **122**(4), 2306–2319.

Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., and Smith, M. A. (**2014**). "Temporal structure of motor variability is dynamically regulated and predicts motor learning ability," Nat. Neurosci. **17**(2), 312–321.

14 August 2024 20:29:38