JASA ARTICLE

Strategic perceptual weighting of acoustic cues for word stress in listeners with cochlear implants, acoustic hearing, or simulated bimodal hearing

Justin T. Fleming[a] and Matthew B. Winn
Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, Minnesota 55455, USA

ABSTRACT:
Perception of word stress is an important aspect of recognizing speech, guiding the listener toward candidate words based on the perceived stress pattern. Cochlear implant (CI) signal processing is likely to disrupt some of the available cues for word stress, particularly vowel quality and pitch contour changes. In this study, we used a cue weighting paradigm to investigate differences in stress cue weighting patterns between participants listening with CIs and those with normal hearing (NH). We found that participants with CIs gave less weight to frequency-based pitch and vowel quality cues than NH listeners but compensated by upweighting vowel duration and intensity cues. Nonetheless, CI listeners' stress judgments were also significantly influenced by vowel quality and pitch, and they modulated their usage of these cues depending on the specific word pair in a manner similar to NH participants. In a series of separate online experiments with NH listeners, we simulated aspects of bimodal hearing by combining low-pass filtered speech with a vocoded signal. In these conditions, participants upweighted pitch and vowel quality cues relative to a fully vocoded control condition, suggesting that bimodal listening holds promise for restoring the stress cue weighting patterns exhibited by listeners with NH. © *2022 Acoustical Society of America.*
https://doi.org/10.1121/10.0013890
(Received 27 January 2022; revised 8 August 2022; accepted 16 August 2022; published online 1 September 2022)
[Editor: Matthew J. Goupell] Pages: 1300–1316

I. INTRODUCTION

Comprehending spoken language requires more than accurate recognition of phonemes and words. A great deal of speech content is conveyed through prosody, which includes qualities such as the durations of speech units and pauses between them, intensity, and voice pitch. Prosody is used to signal information about the talker—including their gender, degree of certainty or uncertainty, and emotional state—and serves suprasegmental linguistic functions, such as highlighting key information and differentiating questions from statements. In situations where there are multiple possible meanings of the same utterance, prosodic cues can resolve syntactic ambiguities (Garro and Parker, 1982; Price *et al.*, 1991). Therefore, an individual's ability to comprehend the various complex layers of speech cannot be assessed by testing recognition of words of phonemes alone (as is commonly done in assessments of speech recognition by people with hearing difficulty); the ability to encode and interpret prosodic cues must be considered as well.

Prosodic differences can be meaningfully contrastive at the level of the individual word. For instance, placing emphasis (stress) on one syllable vs another can produce different words, such as the difference between contract (a written agreement) and contract (to squeeze or make smaller). In English, such stress-contrastive word pairs typically differentiate a noun from a verb, which appear in distinct contexts and are therefore not particularly confusable in a practical sense. These word pairs are used for demonstration and for controlled experimental purposes in this and other studies, but word stress perception likely holds practical importance in speech processing beyond these isolated contrastive cases. For instance, priming and eye tracking studies have shown that listeners use word stress to narrow down candidate words in real time, such that only words consistent with the perceived stress pattern are considered, regardless of phonetic (dis)similarity (Cooper *et al.*, 2002; van Donselaar *et al.*, 2005; Kong and Jesse, 2017). However, when constrained to the goal of recognizing individual words, other studies have shown limited influence of stress among native speakers of English, as reviewed by Cutler and Jesse (2021). For example, cross-splicing misstressed syllables does not ultimately prevent correct word recognition (Small *et al.*, 1988), and even unambiguous stress acoustics do not prevent lexical activation of both members of a stress-contrastive pair (Cutler, 1986). Still, the goals of perception and lexical competition extend far beyond recognition of individual words, and the apparent benign effect of stress ambiguity for a single word could impair the perception of speech in context and have downstream consequences. Toward understanding the auditory factors that play a role in stress perception, the current study focuses on the perceptual weighting of acoustic cues within single words—not for their elevated status or

[a]Electronic mail: jtf@umn.edu

https://doi.org/10.1121/10.0013890

communicative load, but because they provide a constrained platform to address the experimental question.

Word stress is conveyed via the acoustic features of fundamental frequency (F0, i.e., pitch), intensity, and duration, with stressed syllables generally being higher pitched, louder, and longer (Lehiste, 1970). Additionally, word stress can be signaled by vowel quality (VQ), which refers to whether a vowel is fully realized with the tongue reaching its target position, as is typically the case in stressed syllables, or reduced to be similar to the neutral vowel /ə/ (which results from tongue position undershoot), as often occurs in unstressed syllables (Fry, 1958). For native English speakers, VQ is particularly important for correct perception of lexical stress. Previous research has shown that switching a full vowel with a reduced vowel (or vice versa) negatively affects subjective ratings of word acceptability (Fear *et al.*, 1995), the accuracy of speech shadowing (Bond and Small, 1983), and word recognition (Cutler and Clifton, 1984). While VQ effects were dominant in these reports, the latter two studies also showed detrimental effects of incorrect stress patterns in conditions without any alteration to VQ. Consistent with this, participants can use differences in F0 and duration between syllables to determine stress in word fragments removed from the surrounding lexical context (Mattys, 2000). Similarly, reaction times are slowed when stress is misplaced in a word, even when no VQ errors are made, suggesting a cost of mentally repairing incorrect stress patterns (Slowiaczek, 1990). In these and most other prior studies on this topic, tokens with incorrect stress patterns were spoken naturally rather than created synthetically, so stress was likely produced with some combination of pitch, intensity, and duration modulations.

Chrabaszcz *et al.* (2014) directly investigated the relative weighting of four acoustic cues to word stress in normal hearing (NH) listeners and found that in English, VQ was the most highly weighted cue, followed by pitch, intensity, and duration. However, listeners who use cochlear implants (CIs) have impaired access to pitch and other frequency-based cues, potentially demanding cue weighting strategies that differ from those used by NH listeners. Additionally, the importance of the cues might vary depending on the particular words and vowels carrying the cues. The current study examines how the perceptual strategies used to perceive word stress differ based on hearing status, the availability of specific acoustic cues, and the relative contrastiveness of cues within word judgments.

A. CI processing creates challenges for stress perception

Perception of prosodic features is likely to be very difficult for people who use CIs, because these devices have severe limitations in encoding pitch and other frequency-based properties of sound (Moore and Carlyon, 2005). Frequency coding—particularly the precise harmonic cochlear tonotopy needed for harmonic pitch—is distorted by spread of electrical excitation within an implanted cochlea, limiting the accuracy of the "place code" for pitch (Nelson *et al.*, 1995). In addition, most modern CI processors represent the temporal envelope in each frequency band at a constant pulse rate rather than the repetition rate (F0) of the actual incoming sound, preventing pitch from being conveyed via temporal fine structure cues (Oxenham, 2008). While some low F0s can be transmitted via amplitude modulations in the speech envelope, this type of pitch cue is weaker or entirely absent for high-pitched voices (Gaudrain and Başkent, 2018) and would be easily corrupted by noise that fills in essential valleys in the modulated envelope.

The challenges in encoding F0 result in weaker pitch perception among CI users as compared to NH listeners, which manifests as poorer discrimination thresholds for pure and complex tones, impaired detection of frequency modulation, and other pitch-related deficits (Moore and Carlyon, 2005; Won *et al.*, 2010). This in turn leads to impaired perception of the pitch contour in speech (Holt and McDermott, 2013) and prosodic information that depends on it. CI users often struggle to classify and accurately produce emotional speech; this is true both for adults (Agrawal *et al.*, 2013; Jiam *et al.*, 2017) and children (Barrett *et al.*, 2020; Chatterjee *et al.*, 2015; Nakata *et al.*, 2012; Wang *et al.*, 2013). Distinguishing questions from statements can also present problems for CI users, particularly in noisy environments (Meister *et al.*, 2009; Van Zyl and Hanekom, 2013). NH listeners can attend the cues to word stress in a talker-specific manner—such as selectively using intensity or F0 (Severijnen *et al.*, 2021)—but this ability has not yet been explored in people who use CIs. In addition to effects on accuracy at the time of perception, spectral degradation can also inhibit perceptual learning of patterns, such as those driven by word segmentation (Grieco-Calub *et al.*, 2017).

Although many studies have demonstrated weaker prosody perception in CI listeners than in NH participants, much remains unknown about the acoustic cues CI listeners use to achieve prosody perception. The relatively few investigations into this topic have revealed different perceptual strategies employed by NH participants and listeners with CIs. For instance, Peng *et al.* (2009) found that NH listeners rely primarily on the pitch contour to discriminate questions from statements, whereas CI users are heavily influenced by intensity information, which is normally redundant with the pitch cue. Intensity and duration cues are conveyed via the speech envelope, potentially allowing them to be preserved more faithfully by a CI compared to the frequency-based cues like F0 and VQ. In the present study, we tested the hypothesis that participants with CIs rely more on intensity and duration and less on frequency-based cues relative to NH listeners when perceiving word stress.

As with other forms of prosody perception, previous work has shown poorer word stress perception among CI listeners than NH controls (D'Alessandro and Mancini, 2019). This difficulty extends to perceiving stress within sentences as well (Meister *et al.*, 2009). Nonetheless, CI users appear to rely on regularities in word stress to determine the segmentation and sequencing of words in a sentence. In an analysis of error types made by CI users in a speech

recognition task, Perry and Kwon (2015) found that a majority of perceptual errors by CI listeners were consistent with a strategic assumption of trochaic (strong-weak pattern) word forms, which dominate the English lexicon (Cutler and Carter, 1987). CI users tended to split iambic words (weak-strong stress pattern, like "platoon") into two words ("the tunes"), and to combine separate words into a single strong-weak trochaic word (e.g., perceiving "wrote this" as "open"). These results point to practical issues stemming from impaired word stress perception among listeners with CIs, underscoring the importance of an improved understanding of stress perception in this population.

B. The potential contribution of bimodal hearing

Bimodal hearing (the combination of a CI and residual aided or unaided acoustic hearing, i.e., electric-acoustic hearing) may hold promise for improving stress perception relative to CI-only hearing. This is because residual low-frequency hearing would improve access to VQ and pitch cues, which strongly influence word stress perception in NH listeners. Bimodal hearing might be particularly effective at restoring the pitch cue, as F0 contour is preserved even within a relatively scant amount of residual hearing (Gifford *et al.*, 2010a). Indeed, several studies have shown better speech recognition among bimodal than CI-only listeners in various types of noise (Dorman *et al.*, 2008; Gifford *et al.*, 2013; Kong *et al.*, 2005; Woodson *et al.*, 2010), a benefit that has been linked to residual access to the pitch of the target talker's voice (Zhang *et al.*, 2010). Bimodal hearing may also restore some VQ information, provided that the bandwidth of the residual hearing extends into the range of vowel formant frequencies. This degree of residual hearing has become more common among individuals receiving CIs as candidacy criteria for unilateral CI implantation have continued to relax (Gifford *et al.*, 2010b; Perkins *et al.*, 2021).

In terms of prosody perception specifically, many studies have demonstrated benefits of bimodal as compared to CI-only hearing. Previous reports have shown better syllable stress perception, sentence focus identification, and question-statement discrimination among bimodal listeners (Marx *et al.*, 2015; Most *et al.*, 2011). Additionally, Spitzer *et al.* (2009) demonstrated that bimodal and NH listeners make similar use of F0 to perceive stress for speech segmentation, whereas use of F0 was less clear among listeners with only CIs (though note that the CI sample size in that study may have been too small to detect effects). Similar results have been shown in studies that simulated some aspects of bimodal listening among NH participants by combining a vocoded signal in one ear with a low-pass filtered "acoustic" signal in the other ear. In simulated bimodal as compared to fully vocoded conditions, such studies have found improved processing of talker-specific acoustic features (Başkent *et al.*, 2018; Krull *et al.*, 2012), lexical stress (Kong and Jesse, 2017), and recognition of Mandarin tones, phonemes, and sentences (Luo and Fu, 2006). However, not all studies have shown clear differences in prosody perception between bimodal and CI-only listeners. For instance, a study by Cullington and Zeng (2011) found that emotion and sarcasm detection were slightly but not significantly better for bimodal than CI-only listeners. Another study showed that bimodal listening improves question-statement discrimination only when pitch contours are stretched to the extremes of natural speech production (Straatman *et al.*, 2010). Nonetheless, similar percent correct scores on prosody classification tasks could result from different strategies in prosodic cue weighting between bimodal, CI-only, and NH listeners, under investigation in the present work.

C. Summary and hypotheses

In the current study, we use a cue weighting paradigm to compare how NH and CI listeners prioritize the four cues to word stress: VQ, pitch, duration, and intensity. The influence of each of these cues was assessed in a stress-contrastive word judgment task, which constrained phonetic environments and required the listener to rely on stress perception—a skill that is not directly tapped by traditional measures of speech perception, such as word recognition. Stress-contrastive word pairs were used as an experimental tool to study auditory perceptual abilities that are potentially applicable to stress processing in a variety of other listening tasks (e.g., lexical segmentation, sentence-level stress) in which the complexity of language processing might complicate auditory assessment. Acoustic cues to stress manifest differently depending on the prosodic intent (segmentation, discourse prominence, emotion, etc.), but we are primarily interested in the basic low-level encoding and weighting of these acoustic dimensions. Consistent with the approach taken by Peng *et al.* (2009), we use stimuli controlled at the single-word level to examine the specific cues in question in a way more reminiscent of psychoacoustics than lexical activation. This approach has the limitation of excluding any stress contours that might emerge only in longer utterances but has the advantage of intentionally avoiding the complexity of sentence-level stress that would be entangled with the listener's ability to comprehend meaning as it unfolds across the utterance.

Guided by previous experiments on phonetic cue weighting (Moberly *et al.*, 2014; Winn *et al.*, 2012), we hypothesized that CI listeners would weight the duration and intensity cues more highly than NH listeners, while downweighting the frequency-dependent pitch and VQ cues. The weighting of VQ was expected to also depend on the extent of vowel reduction between stressed and unstressed syllables, so we explored the extent of VQ weighting using two pairs of real, stress-contrastive English words. One of these word pairs (desert vs dessert) featured a greater VQ difference between trochaic and iambic word forms than the other (subject vs subject), which we hypothesized would further modulate the use of this cue.

As a preliminary investigation into the influences of bimodal hearing on stress perception, we conducted

https://doi.org/10.1121/10.0013890

additional online experiments with NH listeners and stimulus manipulations designed to mimic some critical features of bimodal listening that are relevant for stress perception. We compared stress cue weighting in a spectrally unprocessed condition to mixed vocoded-unprocessed conditions—in which the stimuli were vocoded only above a cutoff frequency—as well as fully vocoded listening. Similar to stress cue weighting with a CI, we expected that fully vocoding the stimuli would disrupt access to frequency information, causing listeners to downweight the VQ and pitch cues. We further hypothesized that VQ and pitch cues could be selectively restored to their normal weightings in the simulated bimodal conditions, provided that sufficient low-frequency acoustic information was present in the unprocessed portion of the signal. As will be illustrated below, the results confirmed the above hypotheses but also revealed substantial use of VQ and pitch cues among CI users, as well as different cue weighting strategies for each word pair that were consistent between NH and CI listeners.

II. MATERIALS AND METHODS

A. Participants

Twenty-one CI users (11 female, 10 male) and 75 NH listeners (38 female, 35 male, one non-binary, one who chose not to report) were included in the final dataset. The CI data were collected in person at the University of Minnesota, while the NH data were collected across three online experiments, each with 25 unique participants. A total of 83 NH listeners completed the online experiments, but eight participants met one or more of the exclusion criteria described below and were not included in the dataset.

CI participants ranged in age from 32 to 84 years [mean = 60.4, standard deviation (SD) = 16.1 years) and had between 1 and 30 years of experience using their device (mean = 7.7, SD = 6.6 years). All CI participants were native speakers of American English. No adjustments were made to the CI settings these participants used for everyday listening, other than that unilateral implanted listeners with a contralateral hearing aid (7 of 21 CI participants) turned off their hearing aid and wore an earplug in the acoustic ear for testing so that their pitch perception would be more fairly representative of electric hearing. Demographic and device information for the CI participants can be found in supplementary Table I.[1]

Across the sample of 75 NH participants in the final dataset, ages ranged from 18 to 63 years (mean = 30.9, SD = 10.3 years). Sixty-three participants identified as White, four as having more than one race, three as African American, three as Asian, and one as "other." All participants reported having no hearing difficulties and no language-related disorders on their Prolific demographic questionnaires. In an effort to ensure that all participants would be familiar with how stress was conveyed in our talker's voice, we only recruited participants who reported that they were born in and currently reside in the United States, acquired English as their first language, are fluent in English, and are an English-speaking monolingual. Note that the Prolific questionnaire does not currently distinguish between American English and other dialects. All participants gave informed consent, and all study procedures were approved by the University of Minnesota Institutional Review Board.

B. Stimuli

Stimuli were variations of two stress-contrastive word pairs: desert ("the Sahara is the world's largest desert") vs dessert ("let's have cake for dessert") and subject ("my favorite subject is math") vs subject ("subject the coal to heat and pressure"). For each stimulus, the four main acoustic cues to word stress in English—VQ, pitch, duration, and intensity—were independently manipulated to be consistent with either the trochaic or iambic word from a given pair. Acoustic targets for cue manipulation were based on recordings of the words desert, dessert, subject, and subject, each spoken naturally in isolation (i.e., citation form). All stimuli were spoken by the same male talker, who is a native speaker of American English trained in phonetics. We first measured the pitch contours, durations, and intensities of stressed and unstressed phonemes in each of these natural recordings. While normative data on stress cue acoustics are sparse, our stressed and unstressed acoustic measurements are in line with a previous study that used one of the same word pairs [desert vs dessert; see Table I of Zhang and Francis (2010)]. In our recordings, we found that stress was expressed primarily in the acoustics of the vowels, with minimal change in the consonant acoustics. The one exception was that vowel pitch contours carried over to the /b/, /dʒ/, and /z/ phonemes between vowels, so the pitch contours of the vowels were extrapolated out to the consonants. Otherwise, cue manipulation was restricted to the vowel segments of each stimulus, with pitch, duration, and intensity targets derived from the original recordings.

Acoustic cues to word stress were then manipulated in these same recordings, with the exception of VQ, which was the only cue that was maintained from the original recordings without any modification. The trajectory of VQ was highly affected by stress pattern for desert-dessert but changed only slightly for subject-subject [Fig. 1(A)], leading to the prediction that VQ should be a less useful cue for judgments between subject and subject compared to desert and dessert.

Next, two versions of each vowel from the previous step were made: one with a strong-weak pitch contour and one with a weak-strong pitch contour [Fig. 1(B)]. To facilitate pitch contour transplantation, we temporarily equated the durations of vowels from the same word pair and syllable position and buffered with 500 ms of silence (to avoid artifacts of pitch estimation at the edges of the signals) before replacing the pitch contours using the pitch-synchronous overlap-add synthesis method in Praat. After pitch re-synthesis, the buffer silence before and after each vowel was removed.

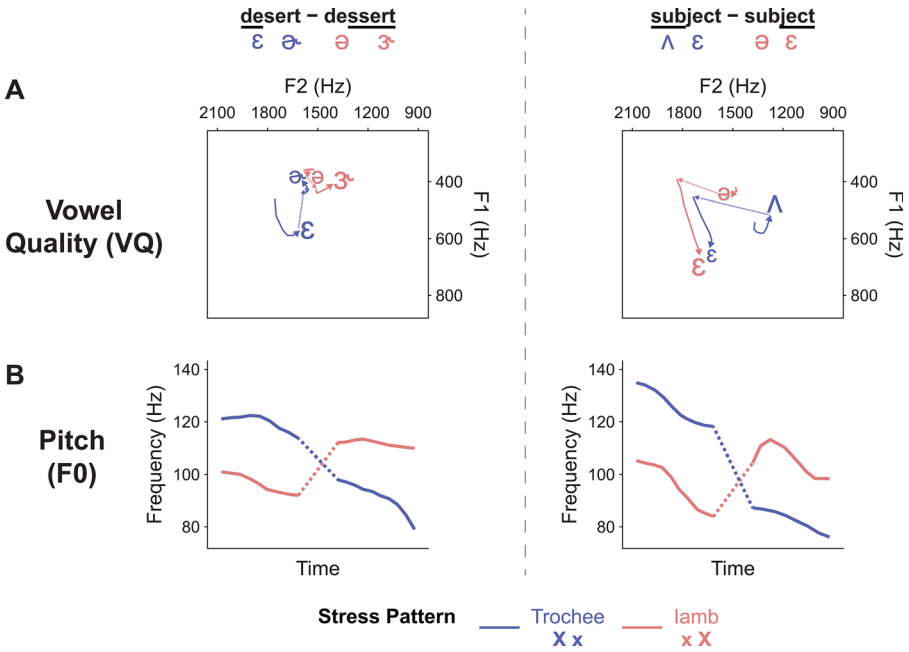

FIG. 1. (Color online) Details of voice quality and pitch cues. (A) F1–F2 plots depict vowel formant trajectories in each stimulus syllable to illustrate VQ differences between trochaic and iambic word forms. Arrows indicate time, and dashed lines indicate consonant segments in between vowels. International Phonetic Alphabet symbols for each vowel are plotted at the end point of its trajectory. (B) Pitch trajectories are shown for the voiced segments of all syllables. Dashed lines connect pitch time series belonging to the same word. These pitch series have been expanded in time for visualization (as vowel duration was later manipulated independently of pitch), so the time axis should be considered arbitrary.

The durations of each vowel were then modified to match durations of the corresponding stressed and unstressed vowels in the natural recordings. Each of those modified vowels was then scaled to match the root mean square intensity of the corresponding stressed and unstressed vowels. Duration and intensity values for all vowels can be found in Table I.

Finally, all the modified vowels were concatenated with the appropriate consonants, with a 2-ms crossfade between adjacent segments to avoid transient artifacts. Within a given syllable, each cue indicated either strong or weak stress; there were no neutral or ambiguous cue levels. Further, within each stimulus, each cue was contrastive across syllables (strong-weak or weak-strong, but not strong-strong or weak-weak). Put another way, each cue suggested either a trochaic (strong-weak) or iambic (weak-strong) stress pattern across the two syllables in the word. However, the cues could be (and often were) in conflict with each other. Cue levels were fully crossed across each of the four cues, resulting in 16 unique stimuli for each word pair: two stimuli in which all four cues were in agreement, eight stimuli in which one cue suggested a different stress pattern than the other three, and six stimuli in which two cues suggested one stress pattern and two suggested the other.

TABLE I. Stressed and unstressed values of the vowel duration and vowel intensity cues.

	Desert		Dessert		Subject		Subject	
	ɛ	ɚ	ə	ɝ	ʌ	ɛ	ə	ɛ
Duration (ms)	100	85	55	130	110	90	60	130
Intensity (dB SPL[a])	69	60	65	65	71	59	65	65

[a]Sound pressure level (SPL).

C. Vocoding (listening conditions)

Listeners with CIs only heard the modified natural-sounding version of each stimulus described above [henceforth referred to as the "unprocessed" condition; Fig. 2(A)]. In addition to these unprocessed stimuli, NH listeners (online experiments) heard one of three spectrally degraded stimulus variations. For experiment 1, the variation was basic sinewave vocoding [Fig. 2(B)]. Importantly, the goal of this was not to fully simulate the experience of CI hearing, but to restrict the availability of some spectral cues that could influence stress perception, as CI listening is marked by a notorious degradation in spectral resolution. Relative to this vocoded condition, potential restoration of stress cues could then be assessed when more unprocessed signal content was added in the simulated bimodal conditions.

Each unprocessed stimulus was vocoded using eight channels spanning frequencies between 100 and 8000 Hz. The signal was filtered into these channels using Hann filters with 25-Hz symmetrical sidebands, with center frequencies spaced logarithmically between 160 and 6399 Hz. The temporal envelope was calculated in each channel using the Hilbert transform and sampled at a rate of 600 Hz to ensure that the F0 was encoded in the envelope. Each envelope was then used to modulate a corresponding sinewave matched to the analysis band; all carrier bands were summed together to create the final vocoded stimulus. A sinewave vocoder was selected because it should preserve cues to voice pitch in a way that is similar to what is available in a CI; the F0 is encoded as the rate of amplitude modulations in the channel envelopes. Additionally, sidebands in the spectrum are created that correspond to the F0 of the voice, which should in theory preserve information about voice pitch that may be lost with other vocoder types (Souza and Rosen, 2009; Whitmal *et al.*, 2007).

For individuals with bimodal hearing, high-frequency signal content is typically conveyed through the CI, whereas

https://doi.org/10.1121/10.0013890

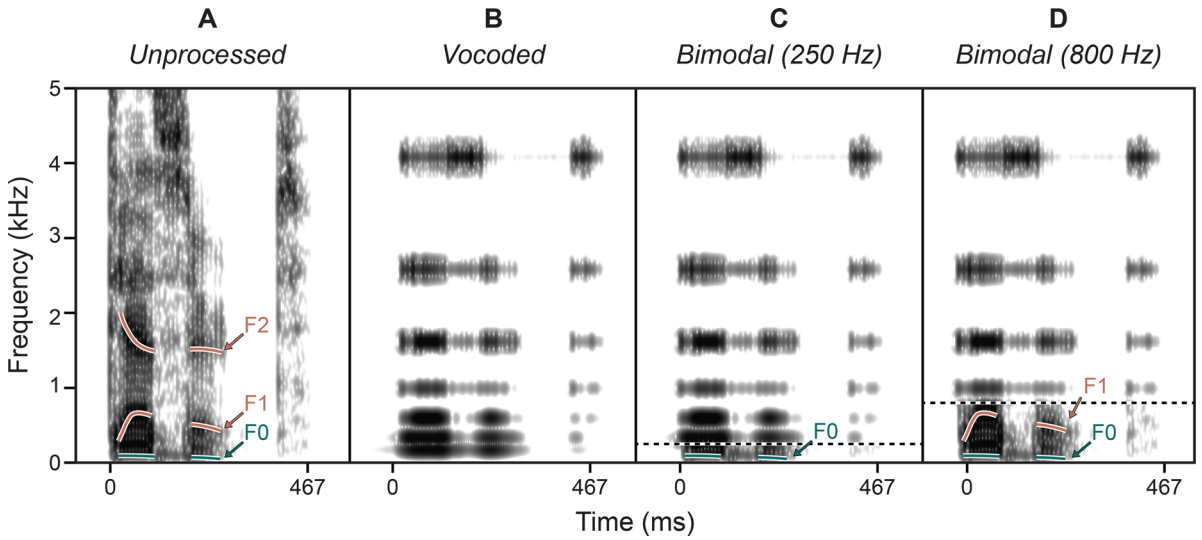

FIG. 2. (Color online) Vocoding and listening conditions. (A) A spectrogram of the word "desert" in the unprocessed condition, with all four stress cues indicating a trochaic stress pattern. F0 and first and second vowel formant (F1 and F2) trajectories are labeled. The same stimulus spectrogram is shown in the fully sine wave vocoded (B), vocoded only above 250 Hz (bimodal 250 Hz) (C), and vocoded only above 800 Hz (bimodal 800 Hz) (D) conditions.

residual hearing contributes predominantly low-frequency signal content. Mirroring this aspect of bimodal listening, two other stimulus variations were made that reintroduced unmodified low-frequency information to complement the vocoded channels. As with the fully vocoded condition, the goal was not to simulate the full experience of bimodal listening, but to capture how the availability of low-frequency residual hearing influences weighting of acoustic cues to word stress. In experiment 2, the vocoded stimuli were high-pass filtered above 250 Hz (removing just the lowest channel) and combined with versions of the corresponding unprocessed stimuli that were low-pass filtered up to 250 Hz, simulating a very small amount of residual hearing [i.e., a "corner audiogram"; Fig. 2(C)]. The 250 Hz cutoff frequency was chosen in particular because (1) it was equidistant between the center frequencies of two vocoder channels, and (2) it preserved F0 (as well as the second harmonic of the relatively low-pitched male voice used in this study) in the unprocessed range but excluded F1 for all stimuli. Thus, we expected this condition might restore the pitch cue to some extent but not convey any added information about vowel reduction.

In experiment 3, the cutoff between unprocessed and vocoded frequency regions was raised to 800 Hz, simulating a greater extent of residual hearing that is becoming increasingly common in bimodal listeners (Hughes *et al.*, 2014; Leigh *et al.*, 2016). The 800 Hz cutoff resulted in unprocessed stimulus content replacing the lowest three vocoder channels, which preserved F1 in the unprocessed range for all stimuli [Fig. 2(D)]. We expected that this would provide at least partial information about vowel reduction (F2 was still vocoded for all stimuli), in addition to preserving the pitch contour. For both experiments 2 and 3, filtering was done using a Hann filter with a transition width of 30 Hz centered on the cutoff frequency. Stimuli were presented diotically over headphones, with both the low-frequency unprocessed and higher-frequency vocoded stimulus components presented to both ears.

D. Procedure

1. Online screening and setup tasks

Before starting any of the online experiments, NH participants performed two preliminary tasks to check that they were wearing headphones and to set the sound level for stimulus presentation. Participants first performed a headphone screening task based on the Huggins Pitch phenomenon. In this effect, a noise stimulus is presented diotically, except that the noise is phase-inverted in a narrow frequency band. This results in an illusory pitch percept that is perceptible over headphones but unlikely to be heard using a loudspeaker due to destructive interference in the free-field (Milne *et al.*, 2020). Next, an auditory stimulus from the main experiment was used to let participants set their computer volume to a comfortable listening level. The sound could be replayed as many times as the listener required. After this task, participants were asked not to adjust their computer volume or audio setup for the remainder of the experiment.

2. The main perceptual test

The stimuli and structure of the main task were similar for the online (NH) and in-person (CI) experiments. On each trial, participants started stimulus presentation by clicking a "play" button or pressing the spacebar. After 150 ms, a single stress-manipulated word was presented auditorily, and participants indicated the word closest to what they heard by clicking one of four buttons labeled with the words desert, dessert, subject, and subject. To alleviate confusion, pictures of a desert, a dessert, and textbooks to represent subject were included next to these three words;

no picture was included for the verb subject. The button positions remained fixed throughout the experiment.

The main task was preceded by an instruction sequence that included eight practice trials. The first two trials were unambiguous (fully trochaic or fully iambic) unprocessed stimuli, followed by two potentially ambiguous (conflicting stress cues) unprocessed stimuli. Next, the fully vocoded or bimodal stimuli (depending on which experiment the listener was in) were introduced, with two practice trials in which the stress cues were all in agreement followed by two in which the stress cues were conflicting. Thus, there were four exposure trials to the fully or partially vocoded stimuli. During practice with the vocoded stimuli, participants rarely clicked on a word from the opposite word pair as the stimulus (never in experiments 1 and 3, 1.6% of practice trials in experiment 2). This indicates that participants were clearly able to distinguish the two word pairs despite the spectral degradation, even upon first encountering the degraded condition. Participants were instructed to choose whichever word was closest to what they heard if the percept was ambiguous.

Each block of the main experiment contained 32 randomized trials: one for each combination of trochaic and iambic stress cues for both word pairs. CI participants performed five such blocks with only the "unprocessed" stimuli, and thus each unique combination of stress cues and word pair was presented five times (160 total trials). The NH (online) participants completed five such blocks for each of two conditions—unprocessed and spectrally degraded (vocoded or one of the bimodal conditions, depending on the experiment)—making for a total of 10 blocks. The NH participants therefore performed 320 total trials, with five repetitions of each unique combination of stress cues, word pair, and the two listening conditions in which they were tested. Blocks alternated between unprocessed and vocoded/bimodal, with the condition of the first block randomized and counterbalanced across participants.

3. Experiment platforms

For the three online experiments, participants were recruited, screened, and compensated using the Prolific study recruitment platform.[2] Stimulus presentation and data collection for the main experiments, as well as an additional headphone screening task, were implemented on the Gorilla Experiment Builder platform.[3] Only participants using a laptop or desktop computer (no tablets or phones) were recruited, and the experiment had to be completed using either Google Chrome or Microsoft Edge due to audio playback issues with other web browsers. The task and participant interface were replicated in MATLAB using custom stimulus presentation functions for in-person collection of the CI data. The in-person experiment was carried out in a sound-attenuating booth with sounds presented at approximately 65 dB SPL using a free-field loudspeaker (Eris E5, PreSonus, Baton Rouge, LA).

E. Data exclusion and analysis

In the online experiments, participants had 6 s from the onset of the response screen to click on a word before the trial automatically advanced. Trials on which this time limit was reached were removed from further analysis. In addition, for the NH listeners, trials on which participants selected a word from the opposite word pair were discarded, as these responses likely represented lapses in attention or erroneous button clicks. A participant was excluded from further analysis if ten or more trials (of 320) were dropped from their data set due to any combination of reaching the time limit and choosing words from the opposite word pair. Across the three online experiments, this criterion resulted in the rejection of five participants. Of the included data, 0.25% of trials (58 of 24 000) were dropped due to trial timeouts. An additional 0.44% of trials (106 of 24 000) were dropped due to the participant selecting a word from the incorrect word pair. A more detailed analysis of these incorrect word pair responses can be found in supplementary Fig. 1.[1] Although such trials were rare overall, they were more common in the vocoded and bimodal conditions than in the unprocessed condition. For the in-person CI experiment, there was no response time limit, and participant responses never corresponded to the opposite word pair.

A key outcome measure throughout this study is the "weight" of each stress cue, defined as the proportion of trials on which the participant chose the trochaic word form (desert or subject) when a given cue was trochaic (pooled across levels of the other cues), minus the proportion of trochaic responses when that cue was iambic. For example, if a listener responded subject 100% of the time when pitch cue was strong-weak (proportion of 1) and responded subject 20% of the time the pitch cue was weak-strong, the cue weight would be 0.8 (1 – 0.2). Cue weight was therefore bounded between −1 and 1, with values near zero indicating minimal influence of the cue on stress perception and values closer to one indicating that stress perception was strongly influenced by the cue. In practice, cue weights were rarely negative, indicating that the stress cue manipulations influenced perception in the expected direction.

The cue-specific average weights described above were used primarily for data visualization; for statistical analysis, cue weighting data were analyzed at the individual-trial level using binomial (logistic) mixed-effects models. In all models, the outcome variable (response) reflects whether the response was trochaic (1) or iambic (0). Factors for each stress cue were coded with centered levels: –0.5 for trochaic and 0.5 for iambic. Fixed and random-effects terms depended on the analysis and are described in Sec. III. Specific conditions were compared by reporting the corresponding model terms. For comparisons that did not include the default "baseline" condition of NH unprocessed, the baseline level was manually changed, and the model was recomputed to obtain the required model term (e.g., for comparing cue weighting between the vocoded and bimodal

https://doi.org/10.1121/10.0013890

conditions). All analysis was carried out in R, with statistical modeling using the lme4 and lmerTest packages.

In a separate analysis, we summed the weights of the VQ, pitch, duration, and intensity cues to determine the extent to which the *combination* of cues influenced stress perception. Regardless of which cues participants used, we always expected this summed cue weight to be substantially above zero for NH participants in the unprocessed condition; values near zero could indicate that none of the cues influenced stress perception, that one or more stress cues influenced perception in the wrong direction, or that participants were choosing words at random. Thus, we excluded data from any participants whose summed cue weight was less than or equal to zero. This resulted in the exclusion of three additional online participants.

III. RESULTS

A. Patterns of stress cue weighting in participants with NH and CIs

Participants who use CIs gave markedly different weights to the four acoustic word stress cues than NH listeners. To capture these trends, we used a mixed-effects model with fixed-effects terms for each of the four stress cues (VQ, pitch, vowel duration, and vowel intensity, each with trochaic and iambic levels), condition (with levels of NH unprocessed, CI, and NH fully vocoded), and the interactions between each cue and the listening condition factor. The NH unprocessed data were collapsed across the three online experiments ($n = 75$ participants), whereas only participants in the first online experiment heard the fully vocoded condition ($n = 25$). Thus, the condition factor was partially within-subjects, with a subset of the NH participants overlapping between conditions.

The random-effects structure consisted of participant-specific intercepts and slopes for each of the four cues and the listening condition. A random interaction between pitch and listening condition was also included, as listening condition had particularly strong effects on usage of the pitch cue. Ideally, we would have been able to include random interactions between the other cues and the listening condition as well, but models with these terms failed to converge, suggesting that their increased complexity was not justified by the dataset. Separate models of the following structure were computed for each word pair:

$$\begin{aligned}&\text{logit}(\textit{Response}) \sim \textit{VQ} + \textit{Pitch} + \textit{Duration} + \textit{Intensity} + \textit{condition} + \\ &\textit{VQ:condition} + \textit{Pitch:condition} + \textit{Duration:condition} + \textit{Intensity:condition} + \\ &(1 + \textit{VQ} + \textit{Pitch} + \textit{Duration} + \textit{Intensity} + \textit{condition} + \textit{Pitch:condition} \mid \textit{Listener})\end{aligned}$$

As hypothesized, listeners with CIs tended to give less weight to the frequency-based stress cues—VQ and pitch—than listeners with NH (Fig. 3). For desert-dessert judgments, these differences in cue weighting were clear for both VQ and pitch ($\beta = -2.32$, $z = -3.86$, $p = 1.16 \times 10^{-4}$ for VQ; $\beta = -1.52$, $z = -4.71$, $p = 2.47 \times 10^{-6}$ for pitch). For subject-subject, CI users gave significantly less weight to pitch than NH participants ($\beta = -1.91$, $z = -3.84$, $p = 1.21 \times 10^{-4}$), but the difference between the two groups' use of VQ was marginal ($\beta = -0.44$, $z = -2.04$, $p = 0.04$); listeners in both groups tended not to rely on VQ for this word pair.

CI users compensated for their reduced access to frequency-based stress cues by increasing their reliance on duration and intensity cues. The CI group used vowel duration to a greater extent than NH listeners in the spectrally unprocessed condition for both word pairs ($\beta = 2.03$, $z = 5.02$, $p = 5.22 \times 10^{-7}$ for desert-dessert; $\beta = 1.70$, $z = 6.15$, $p < 10^{-9}$ for subject-subject). Participants with CIs also used the intensity cue more than NH listeners for subject-subject judgments ($\beta = 0.85$, $z = 3.17$, $p = 0.0015$), but there was not strong evidence for a difference between

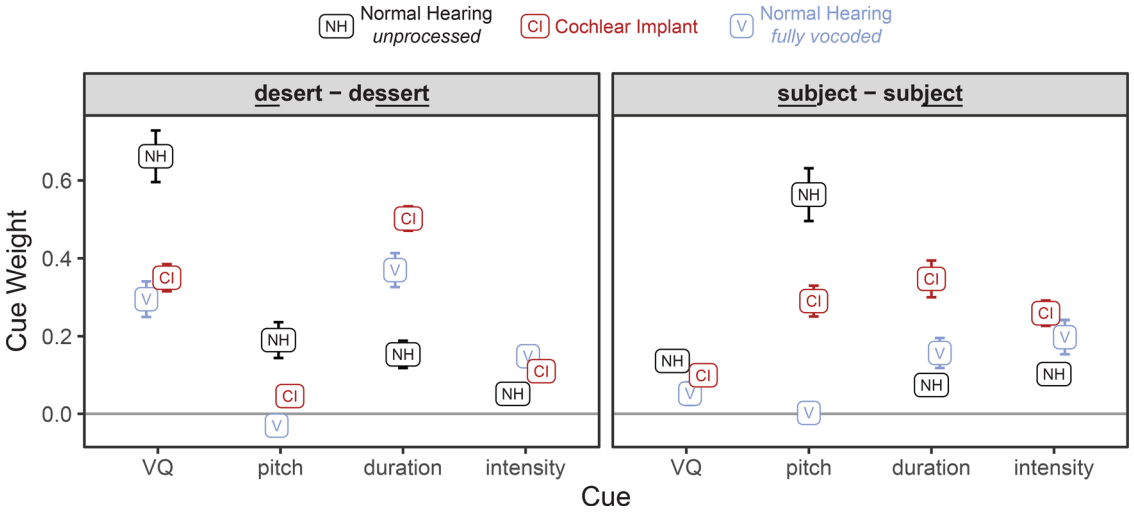

FIG. 3. (Color online) Cue weighting in the NH unprocessed, CI, and NH fully vocoded listening conditions. Condition labels are centered on the mean cue weight for each cue and word pair judgment. Data in the NH unprocessed condition are combined across the three online experiments ($N = 75$). Error bars represent standard error of the mean (SEM).

the groups in the use of intensity for desert-dessert judgments ($p = 0.14$; note that both groups made little use of the intensity cue for this word pair).

Figure 3 also shows cue weighting for NH listeners in the fully vocoded condition. We expected vocoding to reduce access to frequency-based stress cues, causing cue weighting patterns to shift to be more similar to those observed in the CI group. Indeed, vocoding reduced NH listeners' reliance on VQ ($\beta < -0.84$, $z < 5.74$, $p < 10^{-8}$ relative to the unprocessed condition for both word pairs) and effectively eliminated their ability to use the pitch cue. NH listeners compensated for the vocoding by upweighting duration ($\beta = 0.75$, $z = 3.95$, $p = 8.00 \times 10^{-5}$) and intensity cues ($\beta = 0.53$, $z = 3.27$, $p = 0.0011$) for desert-dessert judgments, similar to what was observed in the CI listeners. Although these patterns were present for subject-subject judgments as well, they did not reach the conventional criterion for statistical significance.

The mixed-effects models were next recomputed with the vocoded condition set to be the default condition against which other conditions were compared, allowing for direct comparison of cue weighting between CI and NH vocoded listening. This analysis revealed two important aspects of stress cue weighting with a CI that were not captured in the vocoded condition. First, participants with a CI used duration to an even greater extent than NH listeners in the vocoded condition ($\beta > 1.28$, $z > 3.01$, $p < 0.003$ for both word pairs). This finding, combined with the fact that NH listeners did not significantly upweight duration in the vocoded condition for subject-subject judgments, suggests a possible role of extended experience using a CI in developing the tendency to rely on temporal cues to word stress. Second, CI listeners used the pitch cue more than NH listeners in the vocoded condition. This effect was present in desert-dessert judgments ($\beta = 0.81$, $z = 3.01$, $p = 0.003$) but especially pronounced in subject-subject judgments ($\beta = 2.07$, $z = 5.87$, $p = 4.29 \times 10^{-9}$). Thus, CI listeners were able to use pitch contour differences—in addition to the expected reliance on duration and intensity—to judge contrastive word stress.

B. Listeners with NH and CIs adjusted cue weighting depending on the word pair

Patterns of cue weighting differed substantially between the two word pairs tested in this study. In the NH unprocessed data, this was especially evident in the trading between the VQ and pitch cues; VQ was the most highly weighted cue for stress judgments on the desert-dessert word pair, but far less influential for subject-subject. Conversely, pitch was the dominant cue for subject-subject but played a smaller role for desert-dessert (see Fig. 3). Therefore, these patterns validate the hypothesis that the weighting of cues is dependent partially on the specific word pair judgment, consistent with the degree of cue contrastiveness in the natural productions. Figure 4 shows that this word-specific pattern of acoustic cue weighting also reliably emerges in the CI listener group. Despite generally using the frequency-based cues to a lesser extent (large gray points lower than open white points for VQ and pitch), the CI users consistently exhibited the same shift in cue weighting patterns between the two word pairs. In fact, this pattern was so consistent that all but one CI listener showed downweighting of VQ for subject-subject, and all but one CI listener showed upweighting of F0 for subject-subject (left two panels of Fig. 4).

This pattern of trading between the VQ and pitch cues was captured using binomial models designed to directly compare cue weighting between word pairs. The data were first filtered to contain only the NH unprocessed and CI conditions (making the HearingStatus term fully between-subjects in these models). A separate model was computed for each of the four stress cues, as the analysis was focused on the effects of hearing status (NH or CI), word pair, and their interaction, within each cue. Each of these models was

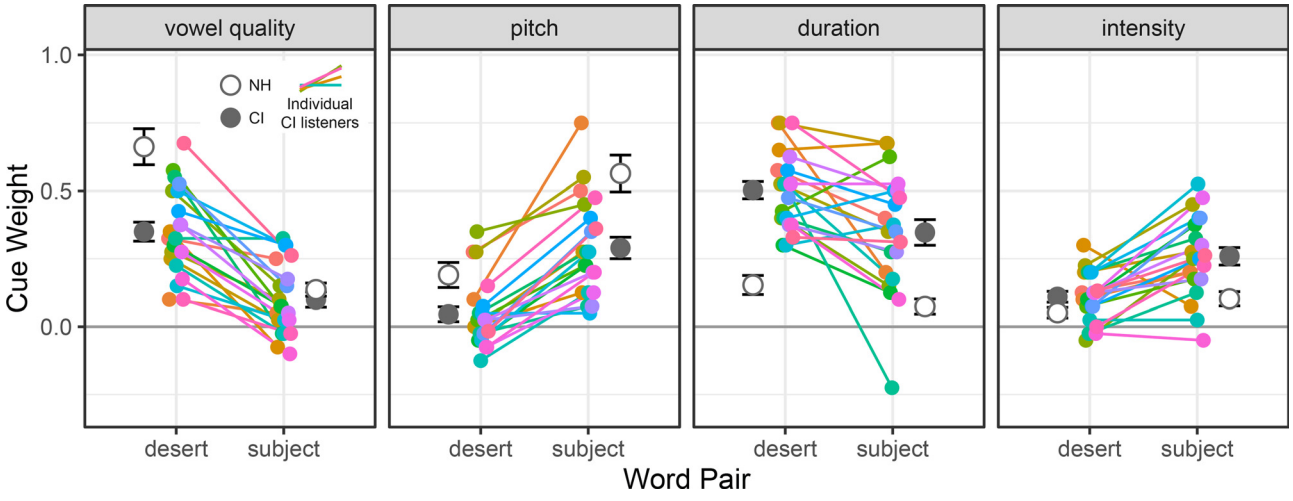

FIG. 4. (Color online) Differences in cue weighting by word pair for NH and CI listeners. Each panel shows the weighting of one of the four stress cues compared between desert-dessert (labeled "desert") and subject-subject ("subject") judgments. Colored points and lines represent individual CI listeners, and large points on the edges represent the grand average NH unprocessed (open circles) and CI (gray filled circles) cue weights. Error bars represent SEM.

https://doi.org/10.1121/10.0013890

of the following form, where [*Cue*] represents either VQ, pitch, vowel duration, or vowel intensity:

$$\text{logit}(Response) \sim [Cue] + WordPair + HearingStatus + [Cue]{:}WordPair + [Cue]{:}HearingStatus + [Cue]{:}WordPair{:}HearingStatus + (1 + \text{WordPair} \mid \text{Listener})$$

Participants with NH and CI users both gave significantly more weight to VQ for the desert-dessert word pair than for subject-subject ($\beta > 1.14$, $z > 7.7$, $p < 10^{-9}$ for both groups). Conversely, listeners in both groups gave more weight to the pitch cue for subject-subject than desert-dessert judgments ($\beta > 1.06$, $z > 7.3$, $p < 10^{-9}$ for both). Cue weighting differences between word pairs were also observed for the vowel duration and intensity cues. Both NH and CI listeners were more influenced by duration for desert-dessert than subject-subject judgments ($\beta > 0.32$, $z > 4.4$, $p < 10^{-5}$ for both groups). Conversely, CI users were generally more influenced by vowel intensity for subject-subject than desert-dessert judgments ($\beta = 0.64$, $z = 4.52$, $p = 6.23 \times 10^{-6}$). Although more subtle, this effect was also observed in the NH data ($\beta = 0.21$, $z = 2.86$, $p = 0.004$; note that the NH data were pooled across the three online experiments for this analysis, so comparisons within this group have greater statistical power).

In addition, the cue-specific models for VQ, pitch, and intensity all showed a significant three-way interaction between the effects of the cue, word pair, and hearing status. These interactions can be interpreted as larger word-specific shifts in the weights of VQ and pitch in the NH group and a larger word-specific shift in the weight of the intensity cue in the CI group (absolute values of $\beta > 0.48$ and $z > 2.7$, $p < 0.01$ for all interactions). A similar three-way interaction was trending toward significance in the duration model, suggesting a larger shift for participants with CIs ($p = 0.06$). These results suggest that whichever group (NH or CI) made greater use of a cue also showed a larger weight change for that cue between word pairs, even though NH and CI listeners always shifted cue weights in the same direction. In other words, these interactions reflect differences in the absolute magnitude of word-specific cue weight shifts, but the proportional changes in cue weights were similar between listeners with NH and CIs.

C. Simulations of residual hearing selectively restored frequency-based stress cues

Results from the first experiment (included in Sec. III A) showed that vocoding caused NH listeners to down-weight the frequency-based cues to word stress (VQ and pitch) and upweight the temporal cues (vowel duration and intensity), similar to participants with CIs. In experiments 2 and 3, we restored access to the detailed original signal below specific cutoff frequencies, as a preliminary examination of potential effects of residual low-frequency hearing on stress cue weighting patterns. Individual participant data from each of these experiments can be found in supplementary Fig. 2.[1] These data were analyzed using mixed-effects models of the same form described in Sec. III A, except that the condition factor had levels of unprocessed, fully vocoded, bimodal 250 Hz, and bimodal 800 Hz. Again, this factor was partially within-subjects, as all participants heard the unprocessed condition, while the other three conditions were presented in separate experiments to non-overlapping subsets of the NH participants.

In experiment 2, the cutoff frequency was set to 250 Hz, such that the F0 contour but no vowel formant information was included in the unprocessed range. As predicted, this manipulation resulted in significantly more use of the pitch cue than in the vocoded condition ($\beta > 2.57$, $z > 7.5$, $p < 10^{-9}$ for both word pairs; Fig. 5). In fact, this modest amount of unprocessed low-frequency stimulus content led to pitch cue usage that was not significantly different from the unprocessed condition ($p > 0.2$ for both word pairs). Although duration and intensity were upweighted on

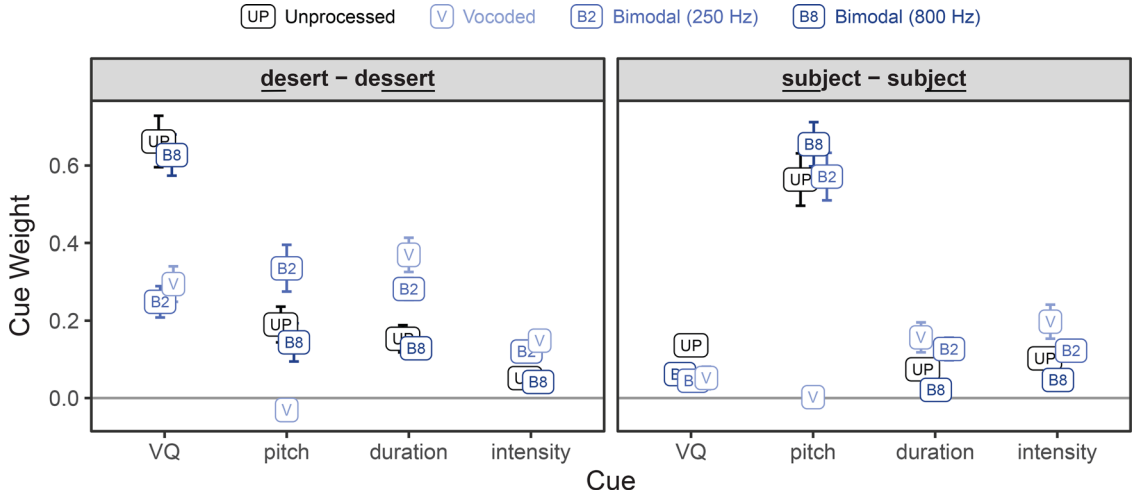

FIG. 5. (Color online) Stress cue weighting among NH participants in the unprocessed, vocoded, and simulated bimodal listening conditions. Condition labels are centered on the mean cue weight for each cue and word pair judgment. Data in the NH unprocessed condition are combined across the three online experiments ($N = 75$). Error bars represent SEM.

average in the bimodal 250 Hz vs the unprocessed condition, this effect did not reach significance for either cue or word pair. The restoration of the pitch cue likely reduced the necessity to upweight vowel duration and intensity, in turn reducing the magnitude of these effects.

In experiment 3, the cutoff frequency was set such that the 0–800 Hz range of the stimulus was left unprocessed, which included F0 and F1 trajectories for all stimuli. Qualitatively, this amount of simulated residual hearing was sufficient to restore a similar pattern of stress cue weighting as in the unprocessed condition (compare UP and B8 across cues in Fig. 5). However, reliance on VQ was slightly reduced in the bimodal 800 Hz condition as compared to the unprocessed condition, and this effect was statistically detectable ($\beta < -0.45$, $z < -2.78$, $p < 0.006$ for both word pairs), perhaps due to the exclusion of F2 from the unprocessed range. With the original signal restored up to 800 Hz, there was no longer any compensatory upweighting of duration and intensity relative to the unprocessed condition.

For conditions in which the stress cues were less reliable, we reasoned that listeners might be generally more likely to respond with the trochaic word form, reflecting a default assumption based on the fact that English words most commonly have first-syllable stress (Clopper, 2002; van Leyden and van Heuven, 1996). Supplementary Fig. 3[1] shows the proportion of trochee responses collapsed across all stress cue levels. For subject-subject judgments, listeners showed a trochee bias in the vocoded condition (62.4% of responses overall), in which VQ information was minimal and pitch cues were severely degraded. The trochee bias was reduced in the bimodal 250 Hz (55.7% trochee responses) and bimodal 800 Hz (52.2%) conditions, in which the pitch cue was restored. In contrast, however, there was a slight iambic bias in all conditions for desert-dessert judgments.

D. Overall utility of the cues in making stress judgments

Listening under spectrally degraded conditions (when the stimuli were vocoded or when using a CI) led participants to adopt perceptual strategies different from those used by listeners in the NH unprocessed condition. We wondered whether stress judgments were also *impaired* in these degraded conditions or if participants' compensatory upweighting of some cues was commensurate with the downweighting of other cues that were degraded. As a cursory exploration of this, we summed the weights across all four stress cues and analyzed whether these summed weights were reduced in the vocoded or CI conditions relative to the NH unprocessed condition. Note that any single cue can have a maximum weight of 1, but because the four stress cues were manipulated independently of one another, the summed weight across cues can exceed 1, as the weight missing from one cue can be attributed to *all* the remaining cues that were consistent with the participant's responses. For example, a trial on which the response is inconsistent with VQ could be consistent with *all* of F0, duration, and intensity, so the three latter cues could contribute more summed weight than the weight lost by VQ. The maximum summed value of 1.5 was determined via simulations that sampled the full set of possible responses in 16-trial blocks.

A linear mixed-effects model was used to statistically analyze the summed cue weight data. The model had a single fixed-effects term, *condition*, with levels of NH unprocessed, NH vocoded, and CI listening. A preliminary version of this model with another fixed-effects term for word pair showed no significant effects involving this factor, so the data were collapsed across desert-dessert and subject-subject judgments. The final model, shown below, was recomputed with cycled baseline levels to obtain the necessary comparisons of condition levels.

$$SummedWeight \sim condition + (1\,|Listener)$$

Within the NH group, fully vocoding the stimuli resulted in a significant decrease in summed cue weight relative to the unprocessed condition (Fig. 6; $\beta = -0.38$, $t = -7.21$, $p < 10^{-9}$). Some compensatory upweighting of the duration and intensity cues did occur in the vocoded condition, but this was not sufficient to restore overall combined use of the stress cues to that of the unprocessed condition. While Fig. 6 suggests that substantial compensation in the vocoded condition was possible for some NH participants (near the top of the plot), most experienced a steep decline in their ability to reliably use the available acoustic cues to perceive stress.

Although CI users employed a different pattern of cue weighting than NH controls, their combined use of the cues

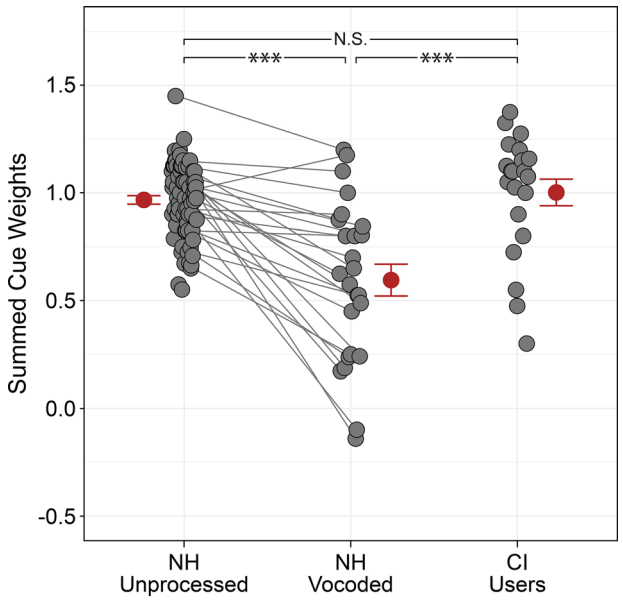

FIG. 6. (Color online) Summed weights across all four acoustic stress cues. Data are collapsed across word pairs. Gray circles represent individual participants, red circles to the side indicate group averages, and error bars represent SEM. The NH unprocessed data are pooled across the three online experiments ($n = 75$). Gray lines connect NH participants who completed both the unprocessed and vocoded conditions (experiment 1; 25 of the 75 NH participants). ***, $p < 0.001$; **, $p < 0.01$; N.S., not significantly different.

https://doi.org/10.1121/10.0013890

to make lexical stress judgments was not statistically different from that of the NH participants ($\beta = 0.03$, $t = 0.58$, $p = 0.57$). Interestingly, CI users had significantly higher summed cue weights than NH participants in the vocoded condition ($\beta = 0.41$; $t = 5.74$; $p = 7.60 \times 10^{-8}$). This again suggests that the vocoded condition in this study did not completely capture CI users' ability to utilize the four combined stress cues.

IV. GENERAL DISCUSSION

Across a series of experiments, we characterized the different acoustic cue weighting strategies used by NH participants and CI users to judge contrastive word stress. NH listeners gave more weight to VQ and pitch cues than CI users, while participants with CIs compensated for degraded frequency information by relying more on intensity and, in particular, duration cues. However, CI users also demonstrated substantial use of VQ and pitch, and they modulated their weightings of these cues depending on the specific word judgment in a manner similar to NH listeners. Conditions designed to imitate some aspects of bimodal listening, in which the stimuli were vocoded above a cutoff frequency but maintained below that frequency, restored the use of frequency-based cues (VQ and pitch) for NH listeners. This suggests that even a relatively modest amount of residual low-frequency hearing could be sufficient to support typical patterns of stress cue weighting.

The current study used stress-contrastive word pairs as an experimental tool to examine the perception of acoustic features relevant to stress perception. In real speech, contextual cues can be used to disambiguate such word pairs; if the phrase "I want to eat…" precedes an ambiguous production of desert/dessert, the listener will be able to reconstruct that the intended word was "dessert." Importantly, however, if a lack of robust stress perception forces a reconstruction of the stress pattern from context, there is almost certain to be a perceptual cost. Processing words with an ambiguous stress pattern may require more mental effort, which could impair the perception of later speech—such downstream consequences of contextual reconstruction have been shown previously in studies with noise masking (Winn and Teece, 2021). Further, stress patterns help NH listeners narrow lexical activation to only potential words matching the perceived stress pattern (Cooper *et al.*, 2002; van Donselaar *et al.*, 2005) and segment word boundaries (Perry and Kwon, 2015), so lexical processing costs may be incurred. Detrimental effects of impaired stress perception may also extend all the way to intelligibility. For instance, transplantation of an unnatural prosodic contour has been shown to reduce speech intelligibility scores (Preece-Pinet and Iverson, 2008). Thus, the perceptual consequences of failing to perceive stress are likely to reach beyond the relatively sterile context of single stress-contrastive words used in the current study. It is possible that the reliance on specific cues when perceiving stress in isolated words would not generalize to other uses of prosody, such as narrow focus, contrastive focus, indicating a question, or expressing an emotion (where the acoustics are not the same as for lexical stress). Bimodal CI listeners tend to show variation in the extent of their use of F0 particularly, with no clear relation between performance on various tasks (Cullington and Zeng, 2011). To better understand the full importance of word stress perception in everyday communication, future studies may probe stress cue weighting using longer utterances or running speech and impose processing time constraints like those faced during real speech processing.

Stress perception may play an even larger role in word recognition for listeners with CIs than NH listeners. Whereas access to phonetic information is degraded for individuals with CIs, the present results show that these listeners can achieve lexical stress perception [this is corroborated in Fig. 2 of Jiam *et al.* (2017)], albeit mainly using different acoustic cues than NH listeners. Future studies could explore differences in the reliance on stress between NH and CI listeners using eye gaze as a real-time index of word recognition in the visual world paradigm (Tanenhaus *et al.*, 1995). Experiments could test whether CI listeners use stress to suppress candidate words with non-matching stress patterns more reliably than NH listeners, similar to results reported by Kong and Jesse (2017) using vocoded speech.

An important consideration in interpreting the present results is that there was a substantial average age difference between the NH and CI groups, on the order of 30 years. This is largely because the data were collected as the laboratory was just reopening during the Covid-19 pandemic, so we were not actively recruiting new participants (especially older listeners). A problem would arise if there were strong reason to believe that the perception of voice pitch is substantially impaired because of age independently of hearing status, because that would imply that the pattern that we interpret as an effect of using a CI could be at least partially an effect of age. Some previous studies suggest that certain types of prosody perception change with age irrespective of hearing status, but the details of the evidence do not paint a clear picture about a possible age confound. For example, Clinard *et al.* (2010) found weaker F0 discrimination in older listeners but used F0s of 500 and 1000 Hz, which are not representative of those found in the human voice. Similarly, Sheft *et al.* (2012) found weaker performance by older listeners when discriminating F0 in a tone complex centered at 1 kHz with a stochastic 5 Hz low-pass noise modulator; it is not clear that this stimulus reflects the acoustics of F0 contours in harmonic stimuli like voices. Hee Lee and Humes (2012) found that older listeners benefit from voice F0 separation as well as younger listeners for a sentence-onset word, but this benefit for older listeners shrank for later words in the sentence—possibly implying a mechanism of semantic processing or memory rather than pure auditory ability. Older listeners suffer a greater cost of syntactic-prosodic mismatch when recalling words (Wingfield *et al.*, 1992), though this might result from a ceiling effect in NH listeners and might also involve a memory

component separate from auditory perception, which would be more demanding than the single-word design in the current study.

A study by Shen *et al.* (2016) tested perception of F0 contours in speech sounds while controlling realistic contours of F0 and also formant contours. Those authors found that some older listeners performed worse than the younger listeners, but they did not find statistical separation between the age groups due to the wide range of variability in the older group and substantial overlap with the younger group. Conversely, *all but one* of the CI listeners in the current study showed pitch cue weighting for subject-subject that was below the NH listeners' group average, and *all but one* CI listener showed VQ cue weighting for dessert-desert that was below the NH average (see Fig. 4). Considering the uncertainty of age effects from the literature and considering the dramatic difference in the mechanism of F0 perception in CI vs acoustic hearing, we are comfortable interpreting the difference in F0 cue weighting as an effect of cochlear implantation rather than an effect of age. This interpretation is even stronger for the finding of increased reliance on durational cues, which should be perceived more *weakly* with age (Gordon-Salant *et al.*, 2006) yet were more *heavily* weighted by our older CI listeners in the current study. Despite this argumentation favoring interpretation as primarily an effect of CI, future studies comparing stress perception between NH and CI listeners would benefit from age-matched participant groups, allowing the effects of age and hearing status to be better separated [as done by Bhargava *et al.* (2016) and O'Neill *et al.* (2019)]. This would avoid the need to interpret possible confounds among a web of related studies.

The current study expands on the conclusions made by Chrabaszcz *et al.* (2014), who measured the weighting of acoustic stress cues in speakers of three languages. In their study, stimuli contained full or reduced forms of only one vowel, and that difference in VQ was the most influential of all the cues that were available to the listeners. However, the current study suggests that the weight of the VQ cue is dependent on which vowel is being spoken, with the relative contrast of the full/reduced form appearing to drive the perceptual weight. In the case of a full reduction from /ɛ/ in the first vowel of desert to /ə/ in the first vowel dessert, the VQ cue is more influential than when the vowel reduction creates only a slight difference, such as the difference between /ʌ/ in subject and /ə/ in subject. For subject-subject, the VQ difference was present and used but was perceptually subordinate to the pitch cue. It is important to note that several previous studies that collapse across larger word sets point toward VQ as the most important cue to English stress perception (Bond and Small, 1983; Cutler and Clifton, 1984; Fear *et al.*, 1995; Ghosh and Levis, 2021). While there is no doubt that VQ is generally critical, the present results indicate that stress cue weighting may be dynamic across words in a language. It is also possible, however, that knowing the word possibilities in advance allowed participants in this study to deploy word-specific cue weighting strategies in a way that would not be possible in real speech perception. This further highlights the importance of future work extending cue weighting paradigms into more naturalistic stimulus sets and tasks.

In addition to the primary goal of exploring patterns of stress cue weighting across listener groups and spectral degradation conditions, the summed weights across the four cues were analyzed to determine how consistently listeners were using the combination of cues to perceive stress patterns. This analysis revealed that overall cue usage did not differ significantly between the NH unprocessed and CI listening conditions. While this result demonstrates an impressive degree of adaptation among the CI listeners, it should also be interpreted carefully. Full compensation by the CI users does not indicate lack of any difficulty; nor does it indicate that the ability to perform well in this task will generalize to perception of continuous running speech. The fact that NH listeners gave the most weight to VQ and pitch hints that, when accessible, these frequency-based cues are typically the most reliable sources of word stress information in English. The upweighting of duration and intensity cues by CI users might appear to be a fully successful compensation in this laboratory task, but duration and intensity might be affected by a wide range of other contextual factors that limit their utility outside the confines of this task. For instance, in addition to word stress, speech emotion, sentence focus, and syntactic boundaries can all be conveyed with duration cues in a way that would not be represented in the current study. VQ cues, on the other hand, may more selectively convey information about lexical stress. Thus, despite the apparent full compensation strategy in the current task, improving access to the frequency-based stress cues NH listeners rely on is likely beneficial for CI listeners, particularly in challenging listening conditions outside the laboratory.

A. CI participants are able to use pitch to make word stress judgments

One of the more surprising results from the current study was that CI users were significantly influenced by voice pitch contour in their perception of word stress. Due to a limited number of electrodes, channel interactions, and CI coding strategies that are incompatible with conveying rate-pitch, CI users generally face difficulties with pitch perception (Gfeller *et al.*, 2007; Looi *et al.*, 2004; Oxenham, 2008). In terms of judging voice pitch contours, Meister *et al.* (2009) showed that CI listeners are worse than NH participants at using voice pitch to distinguish questions from statements and to determine which word in a sentence was stressed. However, while participants with CIs in that study never reached complete consistency in their judgments, psychometric functions were qualitatively similar between the NH and CI groups, indicating that CI participants could access a weakened form of the pitch cue. Children with CIs also modulate voice pitch (as well as duration) to *produce* word stress (Mahshie and Larsen, 2021). These findings are consistent with our finding that CI

https://doi.org/10.1121/10.0013890

users used pitch to judge word stress, albeit to a lesser extent than listeners with NH.

The dynamic pitch contours in the current speech stimuli may have been more salient to CI users than the static pitch stimuli often used in classic psychophysical pitch perception experiments. Additionally, the fact that pitch contours unfold through time gives listeners a longer duration to sample them. CI users' perception of melodic contours has been shown to improve with the duration of the contour, up to at least 500 ms, when contours are presented directly to the CI via current steering (Luo *et al.*, 2010). In addition, CI users can integrate such place-pitch information with temporal pitch cues (amplitude fluctuations in the envelope), yielding more accurate pitch contour perception provided these cues are in agreement (Carlyon *et al.*, 2016; Landsberger *et al.*, 2018; Luo *et al.*, 2012). This envelope pitch is thought to be especially important for listeners with CIs (Laneau *et al.*, 2004), but it is only perceptible when the F0 is relatively low (becoming less perceptible as this limit is approached) and is notoriously susceptible to noise and room reverberation (Qin and Oxenham, 2005; Sayles and Winter, 2008). We used a single male voice with a relatively low F0 in this experiment and presented stimuli in quiet listening conditions (at least for the CI participants, for whom we could control this in the laboratory). Thus, the circumstance of the current experiment was ideal for CI users to be able to use pitch information. Whether CI participants can still use the pitch stress cue for talkers with higher-pitched voices (e.g., women or children) and in more challenging listening conditions remain unanswered questions for future research.

It is worth noting that, in contrast to the CI listeners, NH listeners were not able to use the pitch cue in the fully vocoded condition. Sinewave or noise-band vocoding is a common tool for studies that explore the mechanisms of speech perception in CI listeners, but the current study is one of many that demonstrate a divergence of vocoder results from the results of actual CI listeners (e.g., Laneau *et al.*, 2006; Winn, 2020). We used an eight-channel vocoder in this study, but previous speech intelligibility work has shown that CI listeners have more than eight functional listening channels (Berg *et al.*, 2019). That said, it is unlikely that adding more vocoder channels would have brought NH vocoded and CI stress cue weighting patterns into better alignment. This is because (1) an unreasonably high number of channels would be required for denser frequency sampling to improve the pitch cue; (2) while more channels could improve access to VQ, NH vocoded and CI listeners already made similar use of this cue; and (3) duration and intensity cues are preserved irrespective of the number of vocoder channels. The divergence between vocoded and CI results is likely due in part to CI listeners' extensive experience using cues such as duration and intensity, whereas NH listeners may be used to relying heavily on VQ. Understanding vocoded speech also requires a perceptual learning phase, but in this study, this learning was likely facilitated by the relatively small number of potential words and the fact that they were all known in advance (Davis *et al.*, 2005). Still, extended vocoder experience—beyond what was practical in the online setting used here—might result in increased usage of some stress cues. In general, it is important to recognize that while vocoders are successful at reproducing the overall intelligibility scores of better-performing CI users, they do not accurately convey the exact properties of the auditory stimulus that is received with the implant. Conclusions or clinical interventions motivated solely by the vocoder condition in this study would be imprudent, considering the meaningful differences between the NH vocoded and CI data, especially with regard to pitch perception.

B. Clinical relevance and caveats of the bimodal simulations in this study

In stimulus conditions designed to mimic bimodal listening, patterns of stress cue weighting indicated that normal word stress perception is highly dependent on low-frequency hearing. When the speech was vocoded above 250 Hz, leaving just F0 and the first harmonic of the talker's voice in the unprocessed range, weighting of the pitch cue was restored to at least the same level as when the stimuli were fully unprocessed. This finding is in line with previous research showing that the ability to track voice pitch contour is unaffected by severe disruption of harmonic spacing, suggesting that access to the full harmonic spectrum is not required for this ability (McPherson and McDermott, 2018). Similar findings were also reported in clinical populations by Sheffield and Gifford (2014). With a broader low-frequency range (0–800 Hz) left unprocessed in the current study, the use of VQ was also largely restored, even though only F1 was preserved in the unprocessed range. These results complement previous work showing varying amounts of improvement that result from the addition of low-frequency cues (Cullington and Zeng, 2010), including a simplified tonal representation of F0 alone (Sheffield and Zeng, 2012).

As CI candidacy continues to expand to include patients with increasing amounts of residual hearing, a greater proportion of the CI population will have access to residual low-frequency cues that potentially benefit prosody perception. Cue weighting paradigms like the one used here may have clinical potential for assessing the individual benefits of preserving this low-frequency hearing before and during cochlear implantation. In the bimodal simulations of this study, we observed substantial individual variability in the weights participants gave to the VQ and pitch cues, which could be restored via bimodal hearing [see supplementary Figs. 1(B) and 1(C)].[1] Such individual differences are likely present in actual patients with hearing loss as well, opening the possibility that they could be used to predict which individuals would benefit from bimodal listening. For instance, patients who rely heavily on VQ and pitch might benefit from the amplification of low-frequency acoustic hearing in addition to a CI. On the other hand, patients who give less weight to VQ and pitch when the stimuli are spectrally

unprocessed—or who can readily adapt to use duration and intensity in vocoded listening—might stand to gain more from bilateral cochlear implantation [for a review of bilateral CI benefits, see Brown and Balkany (2007), Litovsky *et al*. (2006), Schafer *et al*. (2011), and van Hoesel (2004)]. That said, unilateral CI users are adept at judging for themselves whether they would benefit more from maintaining their residual low-frequency hearing or receiving a second CI (Gifford and Dorman, 2019). Perhaps patients who would prefer and benefit from a second CI are those who can no longer utilize low-frequency cues because of suprathreshold distortion. For example, Grant (1987) showed that pitch contours need to be exaggerated for listeners with hearing loss to successfully identify them. If a CI candidate's residual hearing is such that they can no longer take advantage of informative pitch contours in real speech, they may be more likely to benefit from bilateral CIs than bimodal hearing.

There are some important differences to consider between the bimodal simulations in this study and real bimodal listening. First and foremost, the quality of the residual hearing in a bimodal CI user is unlikely to be fairly represented by a simple low-pass filter; there would likely be distortions in frequency selectivity and loudness growth consistent with severe cochlear damage that is typical of CI candidates. We presented stimuli diotically, with both the unprocessed and vocoded frequency ranges transmitted to both ears. In the typical bimodal arrangement (a unilateral CI paired with a hearing aid in the opposite ear), acoustic and electric hearing must be perceptually integrated across ears, a problem we did not address in the present study. This challenge is further complicated by frequency mismatches across ears, which can be caused when the broadband frequency spectrum is mapped onto a CI electrode array of incomplete insertion depth (Francart and McDermott, 2013). That said, several studies have demonstrated that residual hearing is often preserved after cochlear implantation in the implanted ear (Di Nardo *et al*., 2007; Hodges *et al*., 1997; Moteki *et al*., 2018; Sierra *et al*., 2019). This opens the possibility for electric-acoustic hearing within the same ear, an arrangement more similar to the bimodal conditions in the current study (Scheperle *et al*., 2017). However, we also simulated a hard cutoff between "acoustic" and "electric" hearing, whereas real residual hearing is more likely to slope gradually into higher frequencies, causing significant overlap with frequencies represented by the CI. The most straightforward way to address these shortcomings would be to replicate the results from our simulated bimodal conditions with actual bimodal listeners.

V. CONCLUSIONS

Listeners with CIs can reliably judge word stress but use different weighting of acoustic cues compared to listeners with NH. Participants with NH relied more on the frequency-based cues (VQ and pitch) compared to listeners with CIs, who compensated by giving more weight to temporal cues (duration and intensity) compared to NH participants. Despite these different cue weighting patterns, participants with CIs still made substantial use of VQ and pitch cues, shifted their cue weightings between word judgments in a manner similar to NH listeners, and used the full combination of stress cues to the same extent as NH listeners. Simulations of bimodal hearing indicated that even modest low-frequency acoustic hearing could restore use of the frequency-based cues. Future work should determine the importance of preserving these NH cue weighting patterns in realistic listening environments and leverage individual differences in cue weighting to evaluate the relative gains from bimodal hearing or a second CI.

ACKNOWLEDGMENTS

The authors would like to thank Katherine Teece, Emily Hugo, and Siuho Gong for collecting data from participants with CIs. This work was supported by National Institutes of Health, National Institute on Deafness and Other Communication Disorders, Grant No. R01 DC017114. J.T.F. and M.B.W. both contributed to conceptualization and design of the experiments, data analysis, and interpretation. J.T.F. also coded the experiments, collected the online experiment data, and wrote the manuscript; M.B.W. edited the manuscript. The authors declare no competing interests. The University of Minnesota stands on Miní Sóta Makhóčhe, the homelands of the Dakhóta Oyáte.

[1]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0013890 for supplementary tables and figures described in the paper.

[2]The Prolific study recruitment platform is available at https://www.prolific.co (Last viewed August 24, 2022).

[3]The Gorilla Experiment Builder platform is available at https://www.gorilla.sc (Last viewed August 24, 2022).

Agrawal, D., Thorne, J. D., Viola, F. C., Timm, L., Debener, S., Büchner, A., Dengler, R., and Wittfoth, M. (**2013**). "Electrophysiological responses to emotional prosody perception in cochlear implant users," Neuroimage Clin. **2**, 229–238.

Barrett, K. C., Chatterjee, M., Caldwell, M. T., Deroche, M. L. D., Jiradejvong, P., Kulkarni, A. M., and Limb, C. J. (**2020**). "Perception of child-directed versus adult-directed emotional speech in pediatric cochlear implant users," Ear Hear. **41**, 1372–1382.

Başkent, D., Luckmann, A., Ceha, J., Gaudrain, E., and Tamati, T. N. (**2018**). "The discrimination of voice cues in simulations of bimodal electro-acoustic cochlear-implant hearing," J. Acoust. Soc. Am. **143**, EL292–EL297.

Berg, K. A., Noble, J. A., Dawant, B. M., Dwyer, R. T., Labadie, R. F., and Gifford, R. H. (**2019**). "Speech recognition as a function of the number of channels in perimodiolar electrode recipients," J. Acoust. Soc. Am. **145**, 1556–1564.

Bhargava, P., Gaudrain, E., and Başkent, D. (**2016**). "The intelligibility of interrupted speech: Cochlear implant users and normal hearing listeners," J. Assoc. Res. Otolaryngol. **17**, 475–491.

Bond, Z. S., and Small, L. H. (**1983**). "Voicing, vowel, and stress mispronunciations in continuous speech," Percept. Psychophys. **34**, 470–474.

Brown, K. D., and Balkany, T. J. (**2007**). "Benefits of bilateral cochlear implantation: A review," Curr. Opin. Otolaryngol. Head Neck Surg. **15**, 315–318.

Carlyon, R. P., Cosentino, S., Deeks, J. M., Parkinson, W., and Bierer, J. A. (**2016**). "Limitations on temporal processing by cochlear implant users," J. Acoust. Soc. Am. **139**, 2154–2154.

Chatterjee, M., Zion, D. J., Deroche, M. L., Burianek, B. A., Limb, C. J., Goren, A. P., Kulkarni, A. M., and Christensen, J. A. (**2015**). "Voice emotion recognition by cochlear-implanted children and their normally-hearing peers," Hear. Res. **322**, 151–162.
Chrabaszcz, A., Winn, M., Lin, C. Y., and Idsardi, W. J. (**2014**). "Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers," J. Speech. Lang. Hear. Res. **57**, 1468–1479.
Clinard, C. G., Tremblay, K. L., and Krishnan, A. R. (**2010**). "Aging alters the perception and physiological representation of frequency: Evidence from human frequency-following response recordings," Hear. Res. **264**, 48–55.
Clopper, C. G. (**2002**). "Frequency of stress patterns in English: A computational analysis," Indiana Univ. Linguist. Club Work. Pap. Online **2**(2), 1–9.
Cooper, N., Cutler, A., and Wales, R. (**2002**). "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," Lang. Speech **45**, 207–228.
Cullington, H. E., and Zeng, F.-G. (**2010**). "Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear," Ear Hear. **31**, 70–73.
Cullington, H. E., and Zeng, F.-G. (**2011**). "Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification," Ear Hear. **32**, 16–30.
Cutler, A. (**1986**). "Forbear is a homophone: Lexical prosody does not constrain lexical access," Lang. Speech **29**, 201–220.
Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.
Cutler, A., and Clifton, C. E. (**1984**). "The use of prosodic information in word recognition," in *Attention and Performance X: Control of Language Processes*, edited by H. Bouma and D. G. Bouwhuis (Erlbaum, Hillsdale, NJ), pp. 183–196.
Cutler, A., and Jesse, A. (**2021**). "Word stress in speech perception," in *The Handbook of Speech Perception*, 2nd ed. (Wiley, New York), pp. 239–265.
D'Alessandro, H. D., and Mancini, P. (**2019**). "Perception of lexical stress cued by low-frequency pitch and insights into speech perception in noise for cochlear implant users and normal hearing adults," Eur. Arch. Otorhinolaryngol. **276**, 2673–2680.
Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. **134**, 222–241.
Di Nardo, W., Cantore, I., Melillo, P., Cianfrone, F., Scorpecci, A., and Paludetti, G. (**2007**). "Residual hearing in cochlear implant patients," Eur. Arch. Otorhinolaryngol. **264**, 855–860.
Dorman, M. F., Gifford, R. H., Spahr, A. J., and McKarns, S. A. (**2008**). "The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies," Audiol. Neurotol. **13**, 105–112.
Fear, B. D., Cutler, A., and Butterfield, S. (**1995**). "The strong/weak syllable distinction in English," J. Acoust. Soc. Am. **97**, 1893–1904.
Francart, T., and McDermott, H. J. (**2013**). "Psychophysics, fitting, and signal processing for combined hearing aid and cochlear implant stimulation," Ear Hear. **34**, 685–700.
Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang. Speech **1**, 126–152.
Garro, L., and Parker, F. (**1982**). "Some suprasegmental characteristics of relative clauses in English," J. Phon. **10**, 149–161.
Gaudrain, E., and Başkent, D. (**2018**). "Discrimination of voice pitch and vocal-tract length in cochlear implant users," Ear Hear. **39**, 226–237.
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**, 412–423.
Ghosh, M., and Levis, J. M. (**2021**). "Vowel quality and direction of stress shift in a predictive model explaining the varying impact of misplaced word stress: Evidence from English," Front. Commun. **6**, 628780.
Gifford, R. H., and Dorman, M. F. (**2019**). "Bimodal hearing or bilateral cochlear implants? Ask the patient," Ear Hear. **40**, 501–516.
Gifford, R. H., Dorman, M. F., and Brown, C. A. (**2010a**). "Psychophysical properties of low-frequency hearing: Implications for perceiving speech and music via electric and acoustic sstimulation," Adv. Otorhinolaryngol. **67**, 51–60.
Gifford, R. H., Dorman, M. F., Shallop, J. K., and Sydlowski, S. A. (**2010b**). "Evidence for the expansion of adult cochlear implant candidacy," Ear Hear. **31**, 186–194.
Gifford, R. H., Dorman, M. F., Skarzynski, H., Lorens, A., Polak, M., Driscoll, C. L. W., Roland, P., and Buchman, C. A. (**2013**). "Cochlear implantation with hearing preservation yields significant benefit for speech recognition in complex listening environments," Ear Hear. **34**, 413–425.
Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Barrett, J. (**2006**). "Age-related differences in identification and discrimination of temporal cues in speech segments," J. Acoust. Soc. Am. **119**, 2455–2466.
Grant, K. W. (**1987**). "Encoding voice pitch for profoundly hearing-impaired listeners," J. Acoust. Soc. Am. **82**, 423–432.
Grieco-Calub, T. M., Simeon, K. M., Snyder, H. E., and Lew-Williams, C. (**2017**). "Word segmentation from noise-band vocoded speech," Lang. Cogn. Neurosci. **32**, 1344–1356.
Hee Lee, J., and Humes, L. E. (**2012**). "Effect of fundamental-frequency and sentence-onset differences on speech-identification performance of young and older adults in a competing-talker background," J. Acoust. Soc. Am. **132**, 1700–1717.
Hodges, A. V., Schloffman, J., and Balkany, T. (**1997**). "Conservation of residual hearing with cochlear implantation," Am. J. Otol. **18**, 179–183.
Holt, C. M., and McDermott, H. J. (**2013**). "Discrimination of intonation contours by adolescents with cochlear implants," Int. J. Audiol. **52**, 808–815.
Hughes, M. L., Neff, D. L., Simmons, J. L., and Moeller, M. P. (**2014**). "Performance outcomes for borderline cochlear implant recipients with substantial preoperative residual hearing," Otol. Neurotol. **35**, 1373–1384.
Jiam, N. T., Caldwell, M., Deroche, M. L., Chatterjee, M., and Limb, C. J. (**2017**). "Voice emotion perception and production in cochlear implant users," Hear. Res. **352**, 30–39.
Kong, Y.-Y., and Jesse, A. (**2017**). "Low-frequency fine-structure cues allow for the online use of lexical stress during spoken-word recognition in spectrally degraded speech," J. Acoust. Soc. Am. **141**, 373–382.
Kong, Y.-Y., Stickney, G. S., and Zeng, F.-G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.
Krull, V., Luo, X., and Iler Kirk, K. (**2012**). "Talker-identification training using simulations of binaurally combined electric and acoustic hearing: Generalization to speech and emotion recognition," J. Acoust. Soc. Am. **131**, 3069–3078.
Landsberger, D. M., Marozeau, J., Mertens, G., and Van de Heyning, P. (**2018**). "The relationship between time and place coding with cochlear implants with long electrode arrays," J. Acoust. Soc. Am. **144**, EL509–EL514.
Laneau, J., Moonen, M., and Wouters, J. (**2006**). "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," J. Acoust. Soc. Am. **119**, 491–506.
Laneau, J., Wouters, J., and Moonen, M. (**2004**). "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," J. Acoust. Soc. Am. **116**, 3606–3619.
Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).
Leigh, J. R., Moran, M., Hollow, R., and Dowell, R. C. (**2016**). "Evidence-based guidelines for recommending cochlear implantation for postlingually deafened adults," Int. J. Audiol. **55**(2), S3–S8.
Litovsky, R. Y., Johnstone, P. M., and Godar, S. P. (**2006**). "Benefits of bilateral cochlear implants and/or hearing aids in children," Int. J. Audiol. **45**(1), 78–91.
Looi, V., McDermott, H., McKay, C., and Hickson, L. (**2004**). "Pitch discrimination and melody recognition by cochlear implant users," Int. Congr. Ser. **1273**, 197–200.
Luo, X., and Fu, Q.-J. (**2006**). "Contribution of low-frequency acoustic information to Chinese speech recognition in cochlear implant simulations," J. Acoust. Soc. Am. **120**, 2260–2266.
Luo, X., Landsberger, D. M., Padilla, M., and Srinivasan, A. G. (**2010**). "Encoding pitch contours using current steering," J. Acoust. Soc. Am. **128**, 1215–1223.
Luo, X., Padilla, M., and Landsberger, D. M. (**2012**). "Pitch contour identification with combined place and temporal cues using cochlear implants," J. Acoust. Soc. Am. **131**, 1325–1336.
Mahshie, J. J., and Larsen, M. D. (**2021**). "Contrastive stress production by children with cochlear implants: Accuracy and acoustic characteristics," JASA Express Lett. **1**, 115201.

Marx, M., James, C., Foxton, J., Capber, A., Fraysse, B., Barone, P., and Deguine, O. (**2015**). "Speech prosody perception in cochlear implant users with and without residual hearing," Ear Hear. **36**, 239–248.
Mattys, S. L. (**2000**). "The perception of primary and secondary stress in English," Percept. Psychophys. **62**, 253–265.
McPherson, M. J., and McDermott, J. H. (**2018**). "Diversity in pitch perception revealed by task dependence," Nat. Hum. Behav. **2**, 52–66.
Meister, H., Landwehr, M., Pyschny, V., Walger, M., and von Wedel, H. (**2009**). "The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients," Int. J. Audiol. **48**, 38–48.
Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., and Chait, M. (**2020**). "An online headphone screening test based on dichotic pitch," Behav. Res. Methods **53**, 1551–1562.
Moberly, A. C., Lowenstein, J. H., Tarr, E., Caldwell-Tarr, A., Welling, D. B., Shahin, A. J., and Nittrouer, S. (**2014**). "Do adults with cochlear implants rely on different acoustic cues for phoneme perception than adults with normal hearing?," J. Speech. Lang. Hear. Res. **57**, 566–582.
Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch*, edited by C. J. Plack, R. R. Fay, A. J. Oxenham, and A. N. Popper (Springer, New York), pp. 234–277.
Most, T., Harel, T., Shpak, T., and Luntz, M. (**2011**). "Perception of suprasegmental speech features via bimodal stimulation: Cochlear implant on one ear and hearing aid on the other," J. Speech. Lang. Hear. Res. **54**, 668–678.
Moteki, H., Nishio, S.-Y., Miyagawa, M., Tsukada, K., Noguchi, Y., and Usami, S.-I. (**2018**). "Feasibility of hearing preservation for residual hearing with longer cochlear implant electrodes," Acta Otolaryngol. **138**, 1080–1085.
Nakata, T., Trehub, S. E., and Kanda, Y. (**2012**). "Effect of cochlear implants on children's perception and production of speech prosody," J. Acoust. Soc. Am. **131**, 1307–1314.
Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., and Levine, S. (**1995**). "Electrode ranking of 'place pitch' and speech recognition in electrical hearing," J. Acoust. Soc. Am. **98**, 1987–1999.
O'Neill, E. R., Kreft, H. A., and Oxenham, A. J. (**2019**). "Cognitive factors contribute to speech perception in cochlear-implant users and age-matched normal-hearing listeners under vocoded conditions," J. Acoust. Soc. Am. **146**, 195–210.
Oxenham, A. J. (**2008**). "Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants," Trends Amplif. **12**, 316–331.
Peng, S.-C., Lu, N., and Chatterjee, M. (**2009**). "Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners," Audiol. Neurotol. **14**, 327–337.
Perkins, E., Dietrich, M. S., Manzoor, N., O'Malley, M., Bennett, M., Rivas, A., Haynes, D., Labadie, R., and Gifford, R. (**2021**). "Further evidence for the expansion of adult cochlear implant candidacy criteria," Otol. Neurotol. **42**, 815–823.
Perry, T. T., and Kwon, B. J. (**2015**). "Amplitude fluctuations in a masker influence lexical segmentation in cochlear implant users," J. Acoust. Soc. Am. **137**, 2070–2079.
Preece-Pinet, M., and Iverson, P. (**2008**). "Segmental and supra-segmental contributions to cross-language speech intelligibility," J. Acoust. Soc. Am. **123**, 3071.
Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (**1991**). "The use of prosody in syntactic disambiguation," J. Acoust. Soc. Am. **90**, 2956–2970.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Sayles, M., and Winter, I. M. (**2008**). "Reverberation challenges the temporal representation of the pitch of complex sounds," Neuron **58**, 789–801.
Schafer, E. C., Amlani, A. M., Paiva, D., Nozari, L., and Verret, S. (**2011**). "A meta-analysis to compare speech recognition in noise with bilateral cochlear implants and bimodal stimulation," Int. J. Audiol. **50**, 871–880.
Scheperle, R. A., Tejani, V. D., Omtvedt, J. K., Brown, C. J., Abbas, P. J., Hansen, M. R., Gantz, B. J., Oleson, J. J., and Ozanne, M. V. (**2017**). "Delayed changes in auditory status in cochlear implant users with preserved acoustic hearing," Hear. Res. **350**, 45–57.
Severijnen, G. G. A., Bosker, H. R., Piai, V., and McQueen, J. M. (**2021**). "Listeners track talker-specific prosody to deal with talker-variability," Brain Res. **1769**, 147605.
Sheffield, B. M., and Zeng, F.-G. (**2012**). "The relative phonetic contributions of a cochlear implant and residual acoustic hearing to bimodal speech perception," J. Acoust. Soc. Am. **131**, 518–530.
Sheffield, S. W., and Gifford, R. H. (**2014**). "The benefits of bimodal hearing: Effect of frequency region and acoustic bandwidth," Audiol. Neurotol. **19**, 151–163.
Sheft, S., Shafiro, V., Lorenzi, C., McMullen, R., and Farrell, C. (**2012**). "Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception," Ear Hear. **33**, 709–720.
Shen, J., Wright, R., and Souza, P. E. (**2016**). "On older listeners' ability to perceive dynamic pitch," J. Speech. Lang. Hear. Res. **59**, 572–582.
Sierra, C., Calderón, M., Bárcena, E., Tisaire, A., and Raboso, E. (**2019**). "Preservation of residual hearing after cochlear implant surgery with deep insertion electrode arrays," Otol. Neurotol. **40**, e373–e380.
Slowiaczek, L. M. (**1990**). "Effects of lexical stress in auditory word recognition," Lang. Speech **33**, 47–68.
Small, L. H., Simon, S. D., and Goldberg, J. S. (**1988**). "Lexical stress and lexical access: Homographs versus nonhomographs," Percept. Psychophys. **44**, 272–280.
Souza, P., and Rosen, S. (**2009**). "Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech," J. Acoust. Soc. Am. **126**, 792–805.
Spitzer, S., Liss, J., Spahr, T., Dorman, M., and Lansford, K. (**2009**). "The use of fundamental frequency for lexical segmentation in listeners with cochlear implants," J. Acoust. Soc. Am. **125**, EL236–EL241.
Straatman, L. V., Rietveld, A. C. M., Beijen, J., Mylanus, EaM., and Mens, L. H. M. (**2010**). "Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid," J. Acoust. Soc. Am. **128**, 1884–1895.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (**1995**). "Integration of visual and linguistic information in spoken language comprehension," Science **268**, 1632–1634.
van Donselaar, W., Koster, M., and Cutler, A. (**2005**). "Exploring the role of lexical stress in lexical recognition," Q. J. Exp. Psychol. A **58**, 251–273.
van Hoesel, R. J. M. (**2004**). "Exploring the benefits of bilateral cochlear implants," Audiol. Neurotol. **9**, 234–246.
van Leyden, K., and van Heuven, V. J. (**1996**). "Lexical stress and spoken word recognition: Dutch vs. English," Linguist. Neth. **13**(1), 159–170.
Van Zyl, M., and Hanekom, J. J. (**2013**). "Perception of vowels and prosody by cochlear implant recipients in noise," J. Commun. Disord. **46**, 449–464.
Wang, D. J., Trehub, S. E., Volkova, A., and van Lieshout, P. (**2013**). "Child implant users' imitation of happy- and sad-sounding speech," Front. Psychol. **4**, 351.
Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.
Wingfield, A., Wayland, S. C., and Stine, E. A. L. (**1992**). "Adult age differences in the use of prosody for syntactic parsing and recall of spoken sentences," J. Gerontol. **47**, P350–P356.
Winn, M. B. (**2020**). "Accommodation of gender-related phonetic differences by listeners with cochlear implants and in a variety of vocoder simulations," J. Acoust. Soc. Am. **147**, 174–190.
Winn, M. B., Chatterjee, M., and Idsardi, W. J. (**2012**). "The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing," J. Acoust. Soc. Am. **131**, 1465–1479.
Winn, M. B., and Teece, K. H. (**2021**). "Listening effort is not the same as speech intelligibility score," Trends Hear. **25**, 23312165211027688.
Won, J. H., Drennan, W. R., Kang, R. S., and Rubinstein, J. T. (**2010**). "Psychoacoustic abilities associated with music perception in cochlear implant users," Ear Hear. **31**, 796–805.
Woodson, E. A., Reiss, L. A. J., Turner, C. W., Gfeller, K., and Gantz, B. J. (**2010**). "The hybrid cochlear implant: A review," Adv. Otorhinolaryngol. **67**, 125–134.
Zhang, T., Dorman, M. F., and Spahr, A. J. (**2010**). "Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation," Ear Hear. **31**, 63–69.
Zhang, Y., and Francis, A. (**2010**). "The weighting of vowel quality in native and non-native listeners' perception of English lexical stress," J. Phon. **38**, 260–271.