

# Perception of Sung Speech in Bimodal Cochlear Implant Users

Trends in Hearing  
2016, Vol. 20: 1–15  
© The Author(s) 2016  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/2331216516669329  
tia.sagepub.com



Joseph D. Crew<sup>1</sup>, John J. Galvin III<sup>2</sup>, and Qian-Jie Fu<sup>2</sup>

## Abstract

Combined use of a hearing aid (HA) and cochlear implant (CI) has been shown to improve CI users' speech and music performance. However, different hearing devices, test stimuli, and listening tasks may interact and obscure bimodal benefits. In this study, speech and music perception were measured in bimodal listeners for CI-only, HA-only, and CI + HA conditions, using the Sung Speech Corpus, a database of monosyllabic words produced at different fundamental frequencies. Sentence recognition was measured using sung speech in which pitch was held constant or varied across words, as well as for spoken speech. Melodic contour identification (MCI) was measured using sung speech in which the words were held constant or varied across notes. Results showed that sentence recognition was poorer with sung speech relative to spoken, with little difference between sung speech with a constant or variable pitch; mean performance was better with CI-only relative to HA-only, and best with CI + HA. MCI performance was better with constant words versus variable words; mean performance was better with HA-only than with CI-only and was best with CI + HA. Relative to CI-only, a strong bimodal benefit was observed for speech and music perception. Relative to the better ear, bimodal benefits remained strong for sentence recognition but were marginal for MCI. While variations in pitch and timbre may negatively affect CI users' speech and music perception, bimodal listening may partially compensate for these deficits.

## Keywords

cochlear implants, pitch perception, atypical speech, electro-acoustic stimulation, bimodal perception

Date received: 23 February 2016; revised: 22 August 2016; accepted: 23 August 2016

## Introduction

Cochlear implants (CIs) lack the spectral resolution and fine structure cues available to normal hearing (NH) listeners. Because only coarse spectral envelope and weak temporal cues are available, CI users have greater difficulty with perception of speech in noise (Friesen, Shannon, Baskent, & Wang, 2001; Fu & Nogaki, 2005; Luo, Fu, & Galvin, 2007), “atypical speech” (Ji, Galvin, Chang, Xu, & Fu, 2014; Li et al., 2011; Luo et al., 2007), and melodic pitch (Galvin, Fu, & Shannon, 2009; Gfeller et al., 2002; McDermott, 2004). In NH listeners, pitch is extracted from the fine structure and harmonic content in a sound. Because these cues are typically removed in CI signal processing, CI users must primarily use spectral envelope information to extract pitch (Crew, Galvin, & Fu, 2012). In NH listeners, timbre is a complex, multi-dimensional percept that is often considered to be independent of pitch. Timbre perception has been shown to depend strongly on attack time and spectral centroid in both NH (e.g., Elliott, Hamilton, & Theunissen, 2013; Grey, 1977) and CI listeners (Mullangi, Marozeau, & Epstein, 2011; Macherey & Delpierre, 2013). These

previous studies used multidimensional scaling to determine the perceptual space of timbre and its correlates to the acoustic stimuli. However, it is unclear how the multidimensional scaling space translates to intelligibility and identification. Also, it is unclear whether such a space would be maintained in the presence of dynamic changes in pitch or timbre. NH listeners perceive pitch according to harmonic and fine structure cues and timbre according to attack and spectral envelope cues. CI users largely depend on spectral or temporal envelope cues for both pitch and timbre perception, which may give rise to potential confusion depending on the listening task (speech vs. music perception). This potential confounding of pitch and timbre cues has implications for real-

<sup>1</sup>University of Southern California, Los Angeles, CA, USA

<sup>2</sup>University of California-Los Angeles, CA, USA

## Corresponding author:

Joseph D. Crew, Department of Biomedical Engineering, University of Southern California, 1042 Downey Way, Denney Research Center (DRB) 140, Los Angeles, CA 90089, USA.  
Email: jcrew.research@gmail.com



world speech and music perception by CI users. In everyday speech, voice pitch can vary depending on the type of communication (e.g., asking a question, expressing an emotion). For example, Su, Galvin, Zhang, Li, and Fu (2016) found that sentence intelligibility was poorer with emotional and shouted speech (both of which involve changes in voice pitch) relative to normal speech. If changes in pitch are perceived as changes in timbre or if articulation is affected by changes in voice pitch, speech intelligibility may worsen. Likewise, if changes in timbre are perceived as changes in pitch, melodic pitch perception may worsen.

Reintroducing fine structure cues may allow CI users to perceptually separate these potentially confounded pitch and timbre cues. Bimodal listening—combined use of a hearing aid (HA) with a CI—has been shown to improve both speech (Brown & Bacon, 2009; Dorman & Gifford, 2010; Gifford, Dorman, McKarns, & Spahr, 2007; Mok, Galvin, Dowell, & McKay, 2010; Mok, Grayden, Dowell, & Lawrence, 2006; Turner, Gantz, Vidal, & Behrens, 2004; Yoon, Li, & Fu, 2012; Zhang, Spahr, & Dorman, 2010) and music perception (Crew, Galvin, Landsberger, & Fu, 2015; Dorman, Gifford, Spahr, & McKarns, 2008; Kong, Cruz, Jones, & Zeng, 2004; Kong, Stickney, & Zeng, 2005). The HA provides fundamental frequency (F0) cues and possibly harmonic information, depending on the audibility and resolution of acoustic hearing. Many CI users report that sound quality is more natural and pleasant when listening with both devices (Armstrong, Pegg, James, & Blamey, 1997; Looi, McDermott, McKay, & Hickson, 2007, 2008; Tyler et al., 2002). While these previous studies have shown results in which CI+HA performance is better than with the CI-only, bimodal benefits have been inconsistent across studies and have depended to some extent on the performance measure and stimuli and whether calculated relative to the better ear (which might be the CI or the HA) or to the CI alone. Because one device may better convey a particular cue that is important for a particular task, it may be difficult to ascertain the overall bimodal benefit, as listeners may only focus on the better ear. One alternative would be to use the same stimuli to measure both speech and music perception. This way, it may be easier to observe how dynamic changes in pitch and timbre cues contribute to speech and music perception when only the spectral envelope cues are available, as in the CI case.

To address these concerns, we recently developed the Sung Speech Corpus (SSC; Crew, Galvin, & Fu, 2015). The SSC consists of monosyllabic words produced for a range of F0s. The SSC can be used to measure speech intelligibility when pitch cues are held constant or varied across words in a sentence. Similarly, the SSC can be used to measure music perception when timbre (i.e., words) is held constant or varied across a melodic

contour. As such, the SSC may provide insight regarding how pitch perception may be influenced by dynamic changes in timbre and vice versa, and how the availability of residual acoustic hearing might mitigate perceptual confusion between pitch and timbre cues in CI users. As a precursor to the present study, Crew, Galvin, and Fu (2015) measured sentence recognition and Melodic contour identification (MCI) in NH musicians and nonmusicians using the SSC stimuli. Results showed near-perfect speech performance in both NH musicians and nonmusicians whether pitch cues were held constant or varied across the words in sentences, suggesting that perception of large timbre variations associated with different words was not susceptible to variations in pitch. For NH musicians, MCI performance was nearly perfect, whether timbre was held constant or varied across notes in the contours. MCI performance was poorer for NH nonmusicians, especially when timbre cues were varied across notes, suggesting that spectral envelope cues may have played a stronger role in pitch percepts or judgments. The results also suggested that extensive training may have allowed NH musicians to better extract melodic pitch from complex stimuli. However, Allen and Oxenham (2014) measured F0 and spectral envelope difference limens in musicians and nonmusicians; changes in F0 and timbre were either congruent or incongruent across stimuli. Their results showed that while overall performance was better for musicians, nonmusicians and musicians were similarly susceptible to incongruent F0 and timbre cues. The difference in findings between Crew, Galvin, and Fu (2015) and Allen and Oxenham (2014) might be due to the extent of timbre dissimilarities (e.g., variable changes in timbre across words vs. orderly shifts in spectral envelope) or ceiling performance effects in Crew, Galvin, and Fu (2015).

In this study, speech and music perception were measured in bimodal CI subjects with the CI-only, the HA-only, and both CI+HA using the SSC stimuli. Sentence recognition in quiet was measured with a matrix test paradigm using sung speech as well as spoken speech. For sung speech, F0 was held constant or varied across the words in sentences. We hypothesized that performance with the CI or HA would worsen from spoken to sung speech and from constant to varying pitch cues in the sung speech, and that the bimodal listening would offset these deficits. MCI was measured using sung speech in which timbre cues (words) were held constant or varied across the notes in the contours. Again, we hypothesized that performance with the CI or HA would worsen from constant to varying timbre cues, with bimodal listening offsetting the deficit. Bimodal benefits were calculated relative to the CI alone or to the better ear, which may have been the CI or the HA, depending on the subject or listening task.

## Methods

### Subjects

Seven adult postlingually deafened bimodal CI users (CI in one ear and HA in the opposite ear) participated in this study. Table 1 shows relevant subject information, including age at the time of testing, years of combined device use, and the CI and HA manufacturers. All subjects had more than one year of combined device use, which was the only inclusion criteria for participation in the study; no subjects were excluded on the basis of speech or music performance, music experience, and so forth. All subjects except for C10 had participated in a previous study (Crew, Galvin, Landsberger, et al., 2015). Informed consent was obtained from each subject, and all procedures were approved by the local

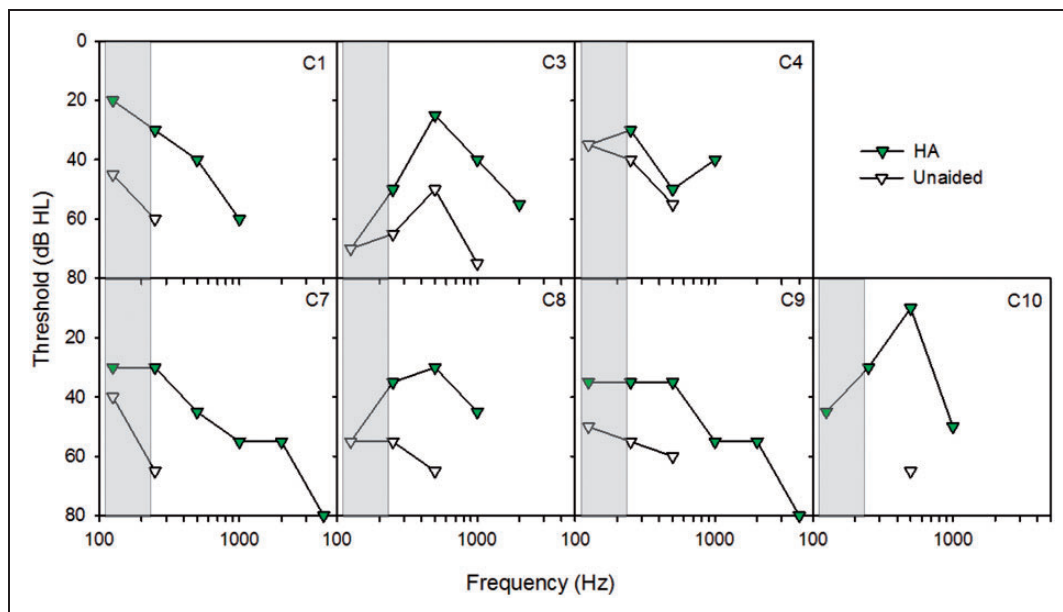
institutional review board. Subjects were paid for their participation.

In this study, speech and music performance was measured using subjects' everyday devices and settings. As such, bimodal subjects were tested with their clinical HA and CI devices and settings, which differed across subjects. We did not change or modify the HA or CI parameters for any subject, and there were no controls for HA prescription (e.g., half-gain rule, frequency transposition) or CI signal processing (e.g., noise reduction, preprocessing schemes) across subjects. Figure 1 shows aided (HA-only) and unaided (no CI and no HA) warble-tone thresholds measured in sound field for each subject. CI-only and CI+HA thresholds were not collected. The shaded area in each panel represents the maximum extent of F0 for the SSC stimuli (110 Hz to 220 Hz). HA-aided thresholds indicate that F0 for the

**Table 1.** Subject Demographic Information.

Subject	Age	Onset of hearing loss (Years)	CI experience (Years)	CI	HA	Etiology of hearing loss
C1	81	17	16	Advanced bionics	Phonak	Sudden sensorineural
C3	78	34	6	Cochlear	Resound	Noise exposure
C4	46	25	5	Cochlear	Phonak	Sensorineural genetic
C7	61	13	11	Advanced bionics	Oticon	Ototoxicity
C8	67	37	11	Advanced bionics	Widex	Cochlear otosclerosis
C9	81	16	2	Cochlear	Oticon	Familial
C10	54	31	2	Advanced bionics	Phonak	Infection

Note. HA: hearing aid; CI: cochlear implant.



**Figure 1.** HA (aided; green symbols) and unaided warble-tone thresholds (white symbols) measured in sound field (dB HL). The shaded area shows the maximum range of F0 for the sung speak stimuli, from 110 Hz to 220 Hz.

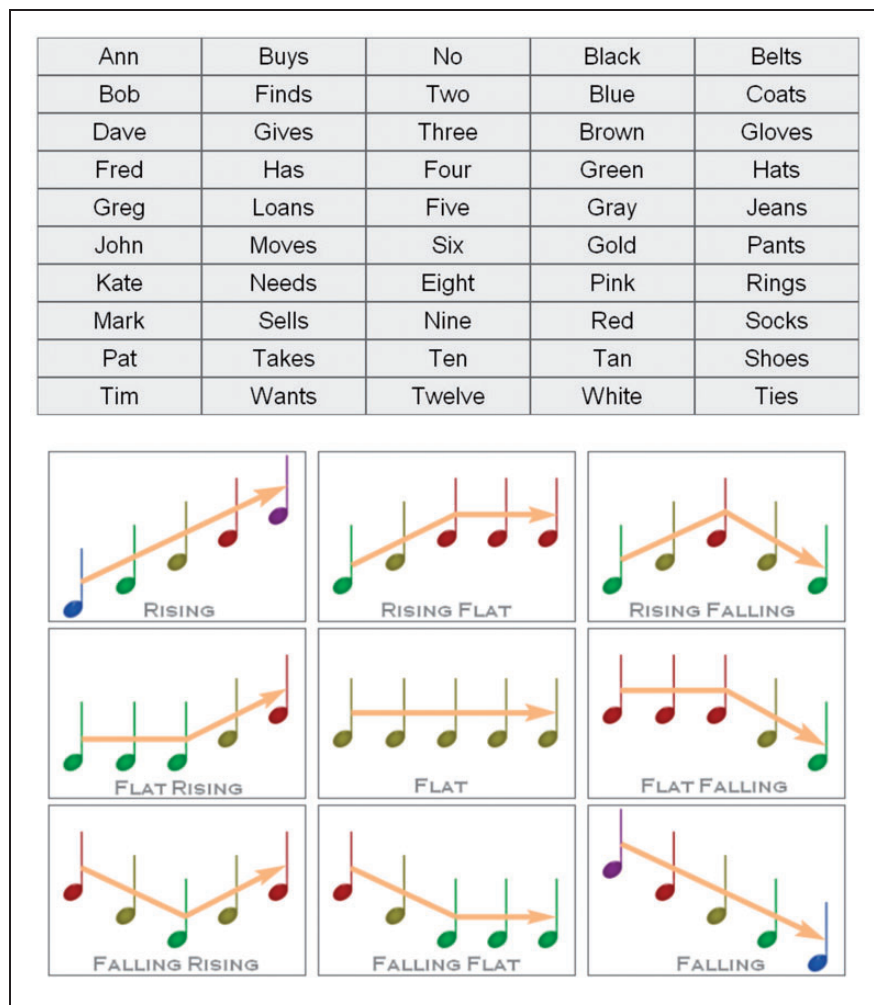
SSC stimuli were audible for most subjects (potential exceptions being C3 and C8). There was also great variability in the HA prescription among the subjects, with some receiving considerable low-frequency amplification (C1, C9, and C10) and others receiving relatively little low-frequency amplification (C4 and C7). Some subjects (C3, C8, C9, and C10) also exhibited HA-aided audibility for first formant (F1) frequency ranges (average F1 values for English vowels ranged from 235 Hz for /y/ to 850 Hz for /a/). Only Subject C10 had frequency transposition in the HA.

### Stimuli

The SSC (Crew, Galvin, & Fu, 2015) was used for testing in all conditions. The SSC stimuli consist of recordings of 50 words naturally produced and sung over a 1-octave range. The SSC stimuli were designed to be used in a matrix test for speech testing. Thus, there were 10 words each in five categories (name, verb, number,

color, and clothing), each produced at all 13 F0s in semi-tone steps between 110 Hz (A2) and 220 Hz (A3); these stimuli were used to measure perception of sung speech. Spoken speech was also measured using the same words produced in a clear speaking manner. The sung speech stimuli were also used to measure melodic pitch perception using an MCI task. The duration of each word was 500 ms and the amplitude was normalized to have the same long-term root mean square amplitude. Please refer to Crew, Galvin, and Fu (2015) for further details regarding the SSC stimuli. Figure 2 shows the stimuli and the response screen used to test speech (top panel) and music perception (bottom panel).

Three conditions were tested for speech perception: (a) “Spoken”—in which the speech stimuli were natural utterances, (b) “Constant Pitch”—in which sung speech stimuli had the same F0 across all words in a sentence, and (c) “Variable Pitch”—in which one of the nine contours used for MCI was randomly selected and applied to the sentence (i.e., F0 changed across words).

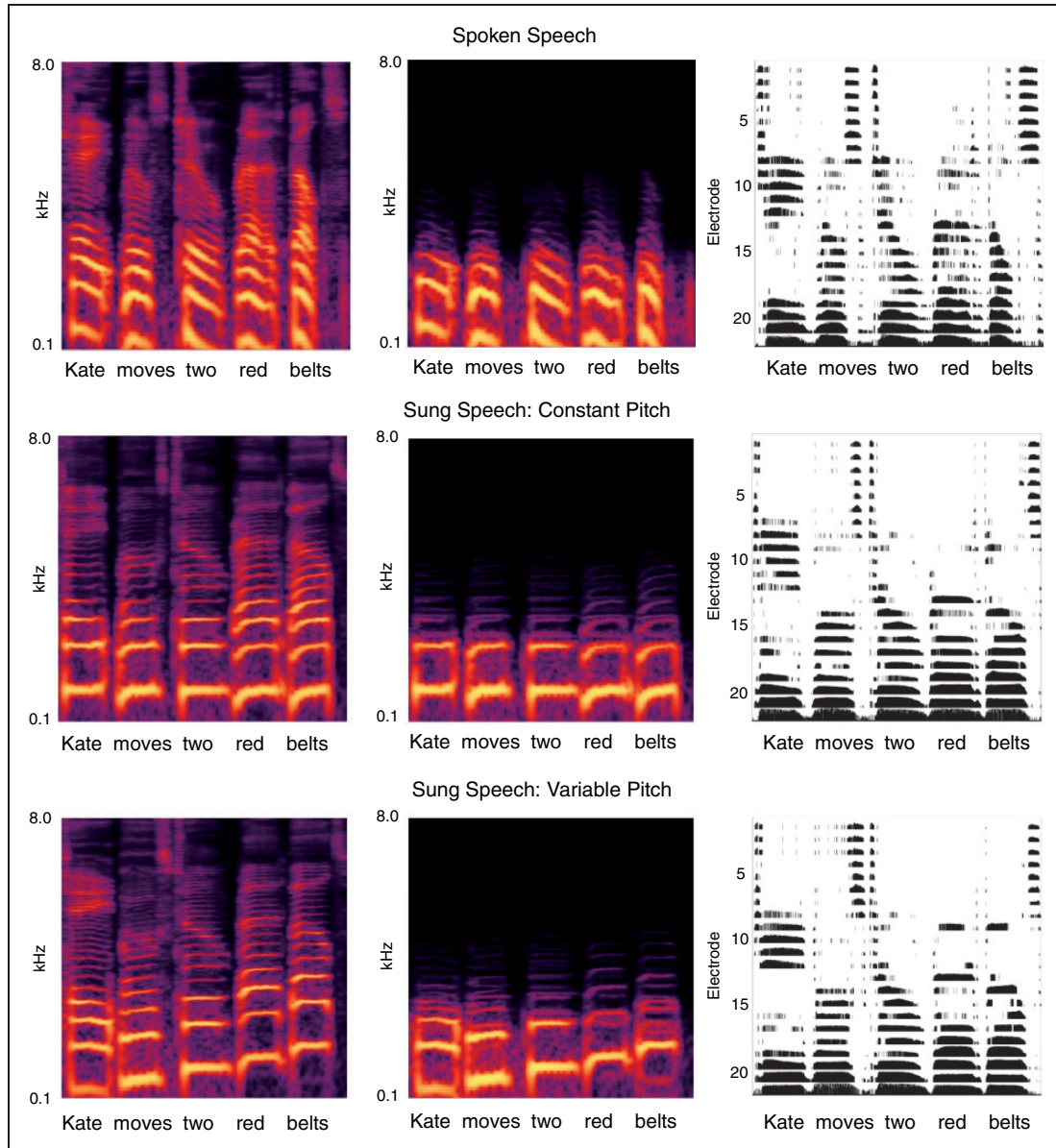


**Figure 2.** Top panel: Matrix sentence test stimuli and response screen. Bottom panel: MCI test stimuli and response screen.



These conditions allowed for examination of the influence of vocal production (i.e., spoken vs. sung) and changes in voice pitch (steady or dynamic) on speech intelligibility. In the Constant Pitch condition, the F0 applied to each of the words in a sentence was fixed at 155 Hz. In the Variable Pitch condition, the F0 difference between consecutive notes in the contour was 1, 2, or 3 semitones. Thus, the maximum F0 range within the Random Contour condition was between 110 Hz and 220 Hz.

Figure 3 shows full bandwidth spectrograms (left column), simulated HA spectrograms (center column), and electrodograms (right column) for the sentence “Kate moves two red belts” in the Spoken, Constant Pitch, and Variable Pitch conditions; a rising contour with three-semitone spacing was used for the Variable Pitch example. The HA condition simulated an audiogram with good low-frequency thresholds but a steeply sloping loss starting at 500 Hz. The electrodogram



**Figure 3.** Example spectrograms and electrodograms for stimuli used for speech testing; the same sentence was used for all panels. Spoken speech is shown in the top row, sung speech with Constant Pitch is shown in the middle row, and sung speech with Variable Pitch is shown in the bottom row. The left column shows full-band spectrograms, the middle column shows spectrograms with a HA simulation (steeply sloping hearing loss beyond 500 Hz), and the right column shows electrodograms generated using default stimulation parameters for Cochlear Corporation devices. For all panels, the x-axis shows time. For the spectrograms, the y-axis shows frequency in kHz; for the electrodograms, the y-axis shows electrode number from most apical (22) to most basal (1).

simulated CI signal processing using default parameters for Cochlear Corporation devices (e.g., 900 pps/ch, eight spectral maxima, input frequency range of 188–7938 Hz, default frequency allocation, etc.). Clear differences between spoken (top row) and sung speech (middle and bottom rows) can be observed in the full band (left column) and the simulated HA spectrograms (middle column). For sung speech, the flat or rising F0 contours across words can be easily observed; much of the harmonic information and consonant information is lost in the HA simulation, suggesting that speech perception, particularly consonant perception, would be poor with only the HA. For spoken speech, there is a downward trajectory in F0 within each word, due to the production of each word in isolation. Also, the vowel portion of each word appears to be longer with sung than with spoken speech. The electrodograms reveal few differences in stimulation patterns between spoken and sung speech and nearly no difference between the Constant and Variable Pitch conditions. Because the analysis filters used for the CI signal processing are typically quite broad (e.g., 1 octave or more in the low-frequency range), the changes in F0 would mostly occur within a channel of a CI, resulting in little change to stimulation pattern across electrodes. However, the extent of the stimulation pattern includes vowel and consonant information in the upper frequency that is not available with the HA.

Two conditions were tested for music perception: (a) “Constant Timbre”—in which the same word was used for every note in each melodic contour during a test run and (b) “Variable Timbre”—in which a random sentence (i.e., different words) was used for each melodic contour during a test run. Note that the Constant Timbre descriptor signifies only that the same word was used for testing. In producing the different F0s for the melodic contours, some variability in spectral envelope might be expected due to differences in articulation. However, such a change in spectral envelope would be much smaller than if producing a different word, as in the Variable Timbre condition. These conditions allowed for examining the influence of timbre (static vs. dynamic) on melodic pitch perception. In the Constant Timbre condition, a word was randomly selected from the sung speech stimuli and used for each note of the contour during MCI testing; this same word was used for each trial. In the Variable Timbre condition, a sentence was randomly generated from the sung speech stimuli and used for the target contour, with a new sentence generated for each trial.

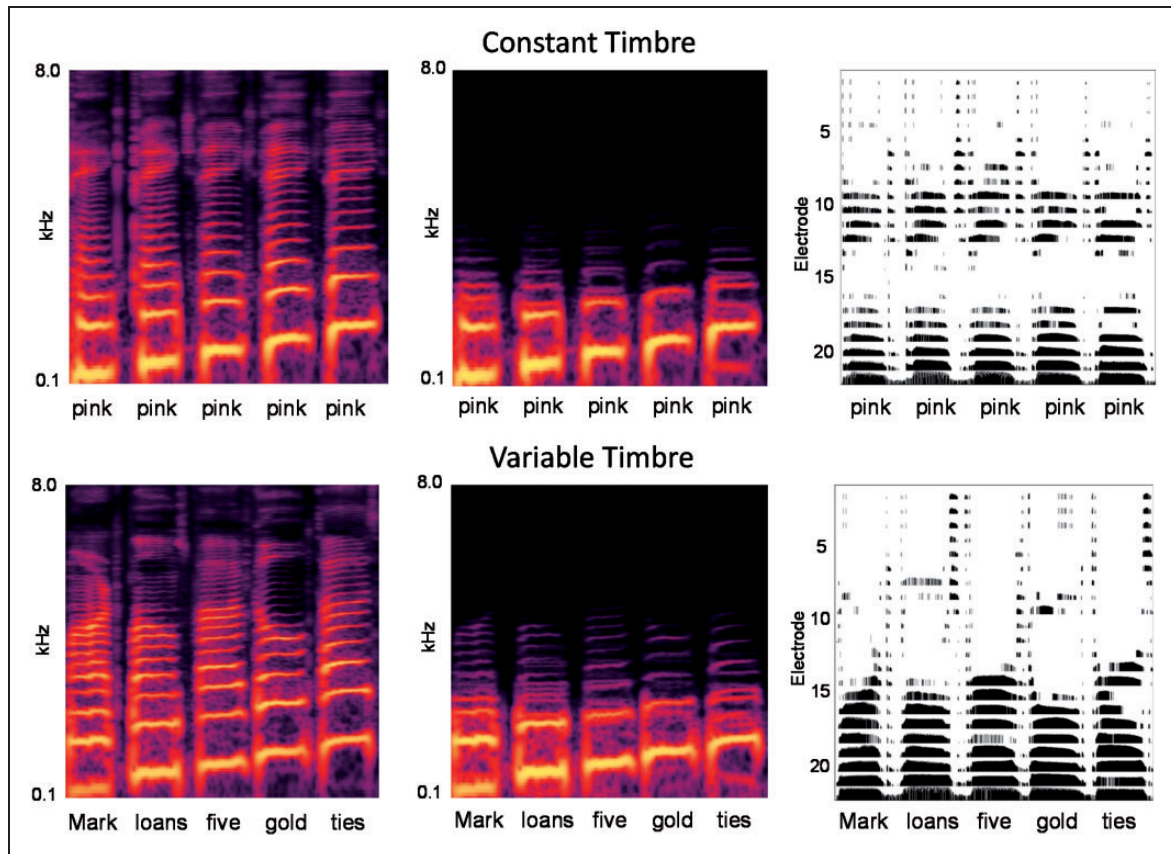
Figure 4 shows full bandwidth spectrograms (left column), simulated HA spectrograms (center column), and electrodograms (right column) for a rising contour with 3-semitone spacing, for the Constant (“pink”) and Variable Timbre (“Mark loans five gold ties”)

conditions. In both the full bandwidth and simulated HA spectrograms, the rising F0 in the contour can be easily observed; much of the higher frequency harmonic information is absent in the HA spectrogram. For the Constant Timbre electrodogram, there is very little change in the stimulation pattern across the notes in the contour, and it is difficult to observe changes in F0 across notes. For the Variable Timbre electrodogram, the stimulation pattern changes across notes, but in a way that relates to the spectrum of each word more than to changes in F0. As such, the stimulation pattern and F0 contour may be in conflict if a CI user were to attend to the spectral envelope.

### Testing

Subjects were tested using their everyday clinical devices and settings throughout the experiment. For all conditions, testing was performed with the CI-only, the HA-only, and with the CI+HA. Testing was always performed first with the CI+HA stimuli, which allowed subjects to familiarize themselves with the stimuli and test procedures. All stimuli were presented in the sound field at 65 dBA in a sound-treated booth. Subjects were seated directly facing a single loudspeaker 1 m away.

Sentence recognition was measured in quiet using a Matrix Sentence Test procedure with 5 categories and 10 items within each category. During each trial of sentence testing, a target sentence was generated by randomly selecting a word from each category. For each trial in the Spoken condition, the words were selected from the spoken speech stimuli. For each trial in the Constant Pitch condition, words were randomly selected from the sung speech stimuli, all with the same F0 (155 Hz). For each trial in the Variable Pitch condition, words were randomly selected from the sung speech stimuli and one of the nine contours used for MCI testing (see Figure 2) was randomly selected and applied to the sentence. During testing in all three conditions, the target sentence was presented to the subject who responded by clicking on one of the words in each of the categories. Subjects were allowed to repeat the sentence before they completed their response by pressing on the “Next” button, after which a new target sentence would be generated and presented. There were 27 trials in each test run, and a minimum of three test runs were performed for each condition. Scoring was based on correct identification of all five words in a sentence. Thus, if a subject correctly identified only four out of five words in a sentence, it was scored as 0% correct. This served to expand the upper range of performance as word identification with a closed-set list of only 10 items would be too easy in quiet. Performance was first measured with the Spoken condition (again, to allow for better familiarization with the test procedures), after which performance



**Figure 4.** Example spectrograms and electrograms for stimuli used for music testing. A rising melodic contour with three-semitone spacing is shown for the Constant Timbre (top row) and Variable Timbre conditions (bottom row). The left column shows full-band spectrograms, the middle column shows spectrograms with a HA simulation (steeply sloping hearing loss beyond 500 Hz), and the right column shows electrograms generated using default stimulation parameters for Cochlear Corporation devices. For all panels, the x-axis shows time. For the spectrograms, the y-axis shows frequency in kHz; for the electrograms, the y-axis shows electrode number from most apical (22) to most basal (1).

was measured with the Constant and Variable Pitch conditions, which were randomly ordered within and across subjects. No preview or trial-by-trial feedback was provided.

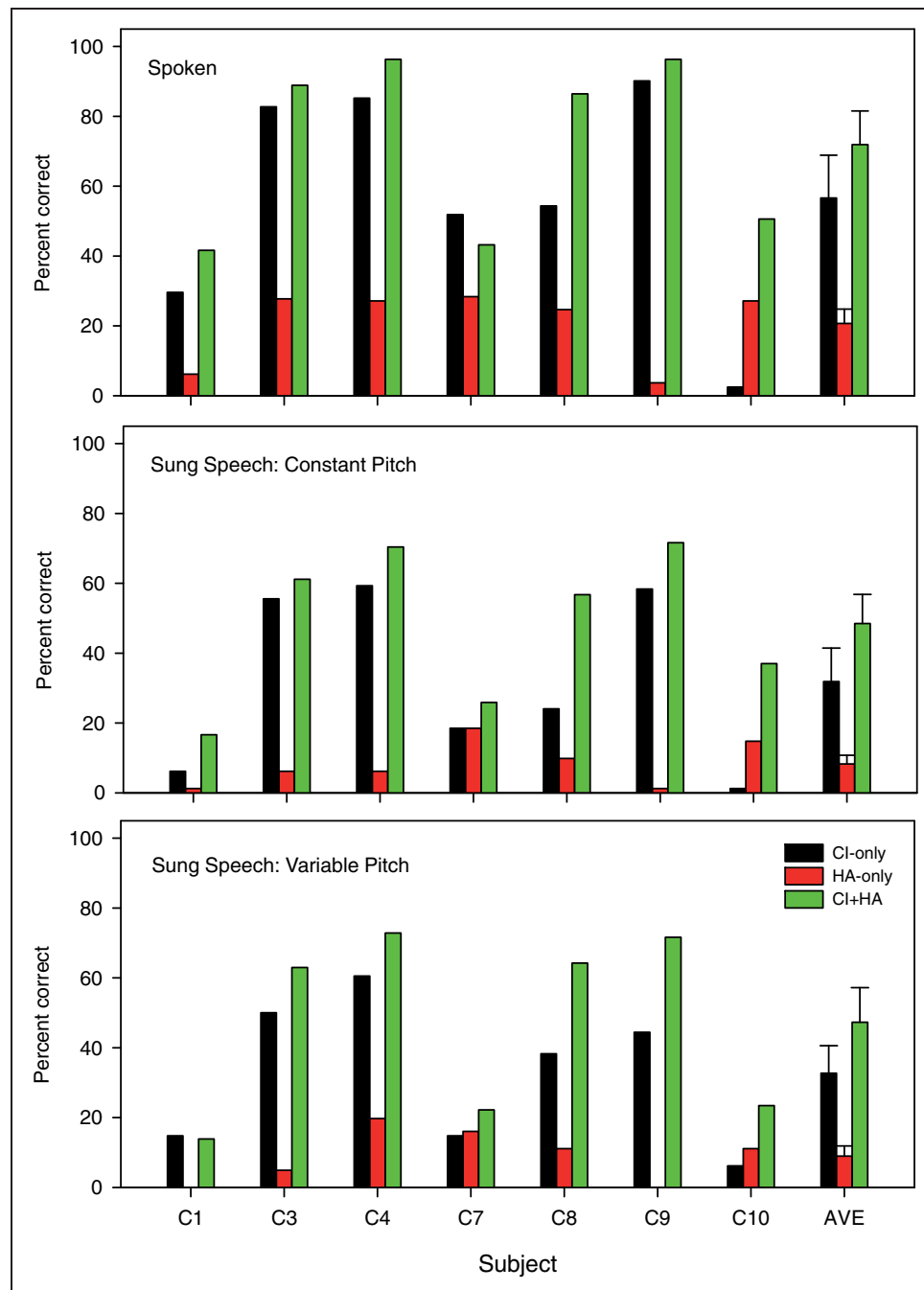
Music perception was measured using an MCI task (Crew, Galvin, Landsberger, et al., 2015; Crew, Galvin, & Fu, 2015; Galvin et al., 2009). During each trial of MCI testing, a target contour was randomly selected from among the nine contours (see Figure 2); the spacing between notes in the contour was 1, 2, or 3 semitones. For the Constant Timbre condition, a word was randomly selected from among the sung speech stimuli and used for each note of the contour throughout the entire test run. For the Variable Timbre condition, a unique sentence was generated for each trial during a test run and applied to the target contour. During testing, a target contour was presented to the subject, who responded by clicking on one of the nine choices shown onscreen (see Figure 2). Performance was scored in terms of percent correct identification. There was a total of 27 trials in each test run, and a minimum of three test runs

were run for each condition. Music testing conditions were interleaved with speech testing conditions and randomized within and across subjects.

## Results

### *Sentence Identification*

Figure 5 shows mean performance for individual subjects for the three listening modes and the three speech testing conditions, as well as average performance across all subjects. Overall, CI performance was much better than HA performance for all speech testing conditions, and CI + HA performance was typically better than CI-only performance. Interestingly, Subject C10 performed better with the HA-only than with the CI-only for all speech tests. With sung speech, performance was similarly poor with the CI-only and the HA-only for Subject C7; performance was markedly better with spoken speech and better with the CI-only than with the HA-only in this case. For the remaining subjects, sentence

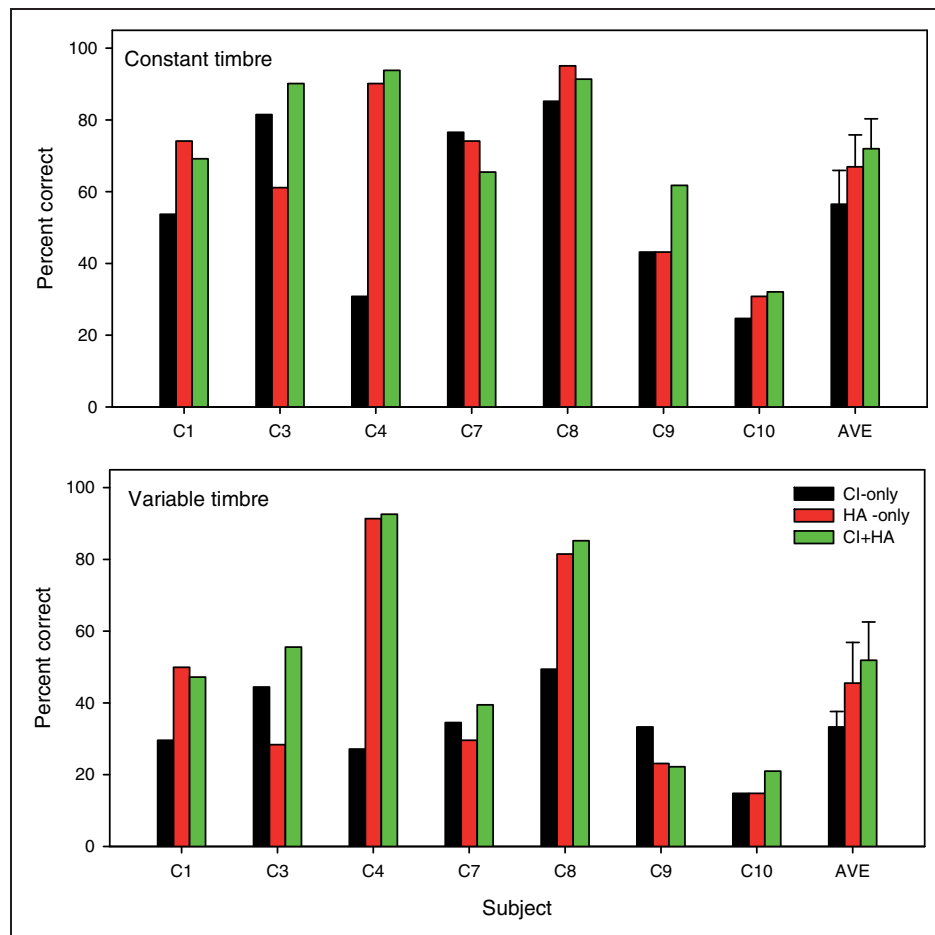


**Figure 5.** Individual and mean sentence recognition for Spoken (top panel), Constant Pitch (middle panel), and Variable Pitch (bottom panel) speech with the CI-only (black bars), the HA-only (red bars), and the CI + HA (green bars). For mean performance across subjects, the error bars indicate the standard error.

recognition was largely driven by the CI, with the addition of the HA further improving performance. Mean performance dropped sharply from the spoken to sung speech, with little difference between the Constant and Variable Pitch conditions. A two-way repeated-measures analysis of variance (RM ANOVA), with listening mode (CI-only, HA-only, and CI + HA) and test condition (Spoken, Constant Pitch, Variable Pitch) as factors,

showed significant effects for listening mode,  $F(2, 24) = 12.6, p = .001$ , and test condition,  $F(2, 24) = 67.8, p < .001$ , as well as a significant interaction,  $F(4, 24) = 3.1, p = .033$ . Post hoc Bonferroni-corrected pairwise comparisons revealed significant differences between the Spoken versus the Constant and Variable Pitch conditions ( $p < .05$  in both cases), and between the HA-only versus the CI-only and CI + HA conditions





**Figure 6.** Individual and mean MCI performance for Constant (top panel) and Variable Timbre stimuli with the CI-only (black bars), the HA-only (red bars), and the CI + HA (green bars). For mean performance across subjects, the error bars indicate the standard error.

( $p < .05$  in both cases). There were no significant differences among the remaining conditions.

### Melodic Contour Identification

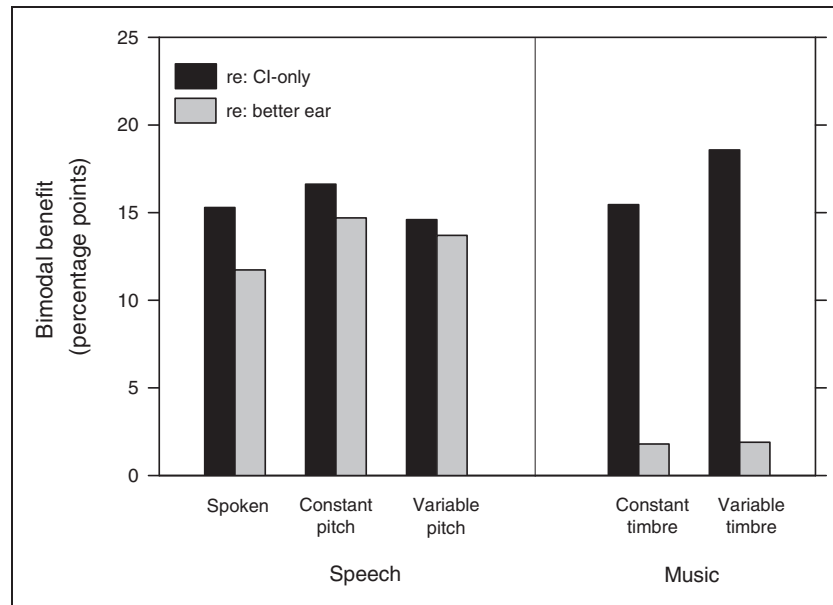
Figure 6 shows mean performance for individual subjects (averaged across the 1-, 2-, and 3-semitone spacings) for the three listening modes and the two music testing conditions, as well as average performance across all subjects. Overall, mean performance was better with the HA-only than with the CI-only, and slightly better with the CI + HA than with the HA-only. For some subjects (C1, C4, and C8), performance was largely driven by the HA. Subject C3 performed better with the CI-only than with the HA-only, and performance with the CI + HA was better than with either device alone. CI + HA scores relative to either device alone were variable with some subjects seeming to attend to the HA only (i.e., CI + HA  $\approx$  HA). Other subjects (C7 and C9) did not show a consistent advantage with either device alone. Mean performance dropped sharply between the Constant and Variable Timbre conditions. A two-way

RM ANOVA, with listening mode and test condition (Constant Timbre, Variable Timbre) as factors, showed a significant effect for test condition,  $F(1, 12) = 20.7$ ,  $p = .004$ , but not for listening mode,  $F(2, 12) = 2.4$ ,  $p = .131$ ; there was no significant interaction,  $F(2, 12) = 0.18$ ,  $p = .837$ .

### Bimodal Benefits

Figure 7 shows the mean bimodal benefit (CI + HA) relative to performance with the CI-only or to performance with the better ear (the CI or the HA, depending on the subject or test) for speech and music perception. Relative to the CI-only, the mean bimodal benefit was 15.5 percentage points across all speech measures and 17.0 percentage points across all music measures. Relative to the better ear, the mean bimodal benefit was 13.4 percentage points across all speech measures and 1.9 percentage points across all music measures.

Because of difference in scoring procedures across the speech and music tests, and because of differences when calculating the bimodal benefit relative to the CI-only or



**Figure 7.** Mean bimodal benefit (across subjects) relative to performance with the CI-only (black bars) or to the better ear (gray bars) for speech and music perception.

to the better ear, a number of RM ANOVAs were performed to compare performance with the CI + HA to that with the CI-only or with the better performing device. Relative to the CI-only, a two-way RM ANOVA, with listening mode (CI-only, CI + HA) and speech type (Spoken, Constant Pitch, Variable Pitch) as factors, showed significant effects for both listening mode,  $F(1, 12) = 11.6$ ,  $p = .014$ , and speech type,  $F(2, 12) = 53.4$ ,  $p < .001$ , with no significant interaction,  $F(2, 12) = 0.1$ ,  $p = .918$ . Post hoc Bonferroni-corrected pairwise comparisons revealed that performance was significantly better with the CI + HA than with the CI alone, and significantly better with Spoken speech than with the Constant or Variable Pitch speech ( $p < .05$  in all cases); there were no significant differences among the remaining stimuli. Relative to the CI-only, a two-way RM ANOVA, with listening mode and music type (Constant Timbre, Variable Timbre) as factors, showed a significant effect for music type,  $F(1, 6) = 22.9$ ,  $p = .003$ , but not for listening condition,  $F(1, 6) = 4.1$ ,  $p = .090$ ; there were no significant interactions,  $F(1, 6) = 0.2$ ,  $p = .664$ . Note that power was low for the listening mode (0.3). Post hoc Bonferroni pairwise comparisons revealed that performance was significantly better with the Constant than with the Variable Timbre ( $p < .05$ ).

Relative to the better ear, a two-way RM ANOVA, with listening mode (better ear, CI + HA) and speech test as factors, showed significant effects for both listening mode,  $F(1, 12) = 14.2$ ,  $p = .009$ , and speech type,  $F(2, 12) = 83.3$ ,  $p < .001$ ; there were no significant interactions,  $F(2, 12) = 0.3$ ,  $p = .738$ . Post hoc Bonferroni pairwise comparisons revealed that performance was significantly better with the CI + HA than with the better

ear alone, and significantly better with Spoken speech than with the Constant or Variable Pitch speech ( $p < .05$  in all cases); there were no significant differences among the remaining stimuli. Relative to the better ear, a two-way RM ANOVA, with listening mode and music-type condition as factors, showed a significant effect for music type,  $F(1, 6) = 16.7$ ,  $p = .006$ , but not for listening mode,  $F(1, 6) = 1.1$ ,  $p = .339$ ; there were no significant interactions,  $F(1, 6) = 0.1$ ,  $p = .981$ . Post hoc Bonferroni pairwise comparisons revealed that performance was significantly better with the Constant than with the Variable Timbre ( $p < .05$ ).

### Correlations to Audiometric Thresholds

Correlational analyses were performed between MCI performance and aided or unaided HA thresholds. Aided thresholds were not correlated with Constant Timbre scores ( $r^2 = .133$ ,  $p = .421$ ) or Variable Timbre scores ( $r^2 = .0565$ ,  $p = .608$ ). Unaided thresholds were better fit to MCI performance, but neither Constant Timbre scores ( $r^2 = .365$ ,  $p = .151$ ) or Variable Timbre scores ( $r^2 = .498$ ,  $p = .076$ ) were significantly correlated.

### Discussion

The present data show that CI speech performance generally worsened from spoken to sung speech; similarly, MCI performance worsened from the Constant Timbre to the Variable Timbre conditions. Consistent with our hypothesis, bimodal listening offset these deficits; indeed, bimodal listening generally improved speech and music perception relative to the CI alone. As in Crew, Galvin,

Landsberger, et al. (2015), the present results showed bimodal speech perception was largely driven by the CI, and that bimodal melodic pitch perception was largely driven by the HA. Later we discuss the results in greater detail.

### *Intelligibility of Spoken Versus Sung Speech*

Whether with the CI alone, the HA alone, or the CI+HA, there was little difference in sentence recognition scores with sung speech when the pitch was fixed or varied across words. However, there was a large drop in performance from spoken speech to sung speech, indicating that spoken speech was more intelligible than sung speech. In a related study, Crew, Galvin, and Fu (2015) found that NH listeners had little difficulty recognizing sentences with the same spoken or sung speech stimuli used in this study. Indeed, NH performance, whether for musicians or nonmusicians, was nearly 100% correct across the different speech stimuli conditions. None of the present CI subjects scored 100% correct in any of the speech stimuli conditions, in any of the listening conditions (although Subjects C4 and C8 scored better than 90% correct with spoken speech). Previous studies have shown that NH listeners are less susceptible to “atypical” speech (e.g., telephone speech, computer speech, fast speech, nonnative speech, etc.) than are CI users (e.g., Ji et al., 2014). Previous studies have also shown that CI users have greater difficulty with speaker normalization than do NH listeners (Chang & Fu, 2006). The present data support this notion that CI performance quickly degrades when speech signals are altered from the normal representation. NH listeners might also experience some deficit in word identification with atypical speech, but seem to be less impacted than CI users.

There was no significant difference in performance with sung speech when pitch cues were held constant or varied across words, suggesting that the present range of pitch variations may not have been adequate to affect sung speech perception. Note that the maximum range of pitch variation was quite large (1 octave) and greater than would be typically observed with conversational or emotional speech. As shown by the electrodiagrams in Figure 3, the stimulation patterns were quite similar for the Constant Pitch and Variable Pitch conditions. Because the CI frequency analysis bands in the low-frequency range are typically quite broad, changes in F0 most likely occurred within an analysis band, thereby stimulating the same electrode despite changes in F0. As such, CI users may rely on temporal pitch cues encoded in the modulation envelope to derive pitch. Because temporal envelope pitch is quite weak in the presence of a complex, multi-channel spectral envelope (Kreft, Nelson, & Oxenham, 2013), there was likely little

perceptual difference between the present Constant Pitch and Variable Pitch stimuli. According to Hillenbrand's (1995) analyses of vowels produced by men, women, and children (who exhibit strong categorical differences in terms of voice pitch), formant information varied by approximately 10% across talker groups, with the largest changes between men and women or children. In this study, where changes in F0 varied within a talker, formant information may have been lost with the HA for higher pitched words. The spectral envelope for the word “red” was analyzed for Spoken, Constant Pitch, and Variable Pitch productions (see Figure 3). F2 and F3 values were very similar across the three conditions. However, F1 values were 25% higher for the rising pitch than for the Flat pitch or spoken speech. Depending on the word, audibility may have been an issue for F1 with the HA. This may have limited the bimodal benefit for sung speech within the Variable Pitch condition.

The deficit with sung speech relative to spoken speech most likely was due to differences in articulation. In the example shown in Figure 3, the mean F0 across words was approximately 100 Hz for spoken speech and 155 Hz for sung speech with Constant Pitch; F0 ranged from 110 Hz to 220 Hz for sung speech with Variable Pitch. Producing speech at these various F0s most likely resulted in some changes in formant frequency and energy (see the harmonic structure for the full band spectrograms in Figure 3). However, these changes were not well represented by the CI (see the electrodiagrams in Figure 3). The deficit with sung speech may also have been due to altered consonant-vowel ratios in terms of energy and duration. The consonant-vowel ratios in terms of energy appear to be quite different between spoken and sung speech in Figure 3, with greater consonant energy for spoken speech. Vowel duration appears to be longer with sung speech than with spoken speech. These differences in speech production may have contributed to differences in intelligibility between spoken and sung speech.

### *Influence of Timbre on Melodic Pitch Perception*

For all three listening modes, mean MCI performance significantly dropped when timbre (words) were varied across notes, relative to the Constant Timbre stimuli. In the full-band and simulated HA spectrograms shown in Figure 4, the rising changes in F0 can be easily observed, suggesting that fine structure cues with normal or impaired hearing would be sufficient for the pitch ranges used in the MCI task. Indeed, MCI performance was largely driven by the HA. However, the electrodiagrams in Figure 4 show a marked difference for the rising contour between the Constant and Variable Timbre stimuli. While there are subtle changes in the stimulation

pattern for the Constant Timbre stimulus, the Variable Timbre stimulus gives rise to inconsistent changes in the stimulation pattern across the rising contour. If CI subjects attended to edge or the spectral centroid of the stimulation pattern, the pitch would seem to slightly rise, then fall, then rise again according to the spectral envelope of the different words. Thus, spectral envelope cues may be unreliable indicators of pitch direction under certain circumstances and may be perceptually confounded with changes in timbre. It is also possible that temporal envelope pitch cues (Zeng, 2002) may have contributed to the present pattern of results. In the Constant Timbre condition, the spectral envelope is largely identical across notes; thus, CI listeners might attend to a different cue (e.g., temporal envelope rate cues within channels) to determine pitch changes within the melodic contours. The present results suggest that temporal envelope cues were not robust to changes in timbre (i.e., the Variable Timbre condition).

Mean performance for all three listening modes was poorer than that observed with NH musicians in Crew, Galvin, and Fu (2015); however, some good performers (C4, C8 with the HA-only, or CI+HA) performed nearly as well as the NH musicians. Interestingly, mean performance with the CI+HA was comparable to that of NH nonmusicians in Crew, Galvin, and Fu (2015). Hearing impaired listeners and NH nonmusicians both seem to have difficulty attending to and extracting pitch cues from complex stimuli. The sources of difficulty might be different across these listener groups, as hearing impaired listeners must extract pitch from a spectrotemporally degraded signal and NH nonmusicians must learn to use all the fine structure cues available to extract pitch and to ignore changes in timbre. In both cases, music training may benefit melodic pitch perception, as has been shown in previous CI studies (e.g., Fu, Galvin, Wang, & Wu, 2015; Galvin et al., 2009; Gfeller, Witt, Stordahl, & Mehr, 2000).

Some subjects (C4 and C8) exhibited similar performance for the Constant Timbre and Variable Timbre conditions with the HA-only or with the CI+HA; these subjects had good residual acoustic hearing and may have successfully attended to the harmonic fine-structure cues during the MCI task. Subjects C3, C7, and C8 exhibited good CI-only performance (> 75%) for the Constant Timbre condition, but much poorer performance for the Variable Timbre condition, suggesting that pitch cues may have been largely derived from the spectral envelope. These results have some implications when testing pitch perception with CI users. Musical pitch is primarily based on harmonics and fine-structure cues that are not well conveyed by CIs. As such, CI users may rely on spectral envelope (and to some extent, temporal envelope cues) to make pitch judgments. Introducing some jitter to the spectral envelope (as in

the Variable Timbre condition) might reduce melodic pitch perception but might also be more akin to real world music listening in which timbres may change across notes, especially in vocal music. While testing with sung speech allows for some insights regarding how dynamic changes in pitch and timbre might affect speech and music perception, it may be preferable to explicitly control pitch and timbre information (e.g., synthesized stimuli with a fixed spectral envelope and varying F0/harmonics and vice versa) in future studies.

As seen in the electrodiagrams in Figures 3 and 4, CIs do not provide the fine-structure cues that are present in acoustic hearing. Engineering attempts to restore these cues (e.g., current shaping, explicit time coding for the apical channels) have had limited success thus far. Attempts to optimize the CI frequency allocation for music (e.g., Kasturi & Loizou, 2007) may have limited utility, as such allocations may not support multiple pitch ranges or speech recognition. Because CI electrode arrays typically do not extend to apical frequency regions, activation of these low F0 neurons might be better targeted with a HA if there is low-frequency residual acoustic hearing. Coordinated, optimized mapping between the CI and HA might maximize bimodal listening by reducing the frequency overlap between devices. Such an approach would also reduce the frequency-to-place mismatch in the apical region for the CI and allocate more channels for frequency regions beyond those targeted by the HA. Ultimately, CIs must restore perception of harmonic pitch to be sufficiently robust to changes in spectral envelope.

### *Bimodal Benefits for Speech and Music Perception*

As shown in Figure 7, large bimodal benefits were observed for sentence recognition with spoken or sung speech, relative to the CI-only or to the better ear. We had hypothesized that the bimodal benefit would increase as the speech stimuli became more difficult, with the greatest advantage for the Variable Pitch condition. This was not the case on average, although bimodal benefits were observed in some subjects (C8 and C9) for the Variable Pitch condition. It is unclear why adding the HA to the CI did not produce a consistent advantage. Acoustic hearing may have provided a useful overall voicing cue which helped in segmentation of the monosyllabic words in the sentence, rather than specific voice pitch information. The HA may have added useful formant information in some cases. The present pattern of results suggests that the bimodal benefit observed in previous studies for speech in noise may not be exclusively due to better segregation of speech and noise by tracking voice pitch cues.

Adding the HA to the CI greatly improved MCI performance, but adding the CI- to the HA-only marginally



improved performance. Again, contrary to our hypothesis, bimodal benefits did not increase from the Fixed Timbre to the Variable Timbre condition. The lack of bimodal benefit relative to the better ear (HA) may be idiosyncratic to the music stimuli, MCI procedure, and the pitch ranges used in this study. Many CI users report better music perception and appreciation with bimodal listening. While not a formal part of this study, a few subjects reported that music and voices sounded more natural when using the HA in combination with the CI. The present MCI task focused on functional pitch perception in a melodic context and did not capture other important aspects of music perception (e.g., emotional response). Also, for pitch ranges above 500 Hz, HAs may not provide much information to bimodal listeners. There were some cases in which performance worsened when both devices were used (C7 and C9), although this effect was not consistent across conditions. Such a pattern of results was also observed in some bimodal CI subjects in Crew, Galvin, Landsberger, et al. (2015). The present results showed that bimodal MCI performance was almost exclusively driven by the HA. CI settings may require optimization to better work with residual acoustic hearing. If this is not possible, preservation of acoustic hearing is vital to restoring good music perception to CI users.

One limitation to this study is that subjects were tested with their clinical HA and CI devices and settings, as we were interested in performance with their everyday listening configuration. As shown in Figure 1, there was some variability among subjects' aided and unaided HA thresholds, raising the question of whether audibility contributed to the bimodal benefit. One would expect that subjects that had lower aided thresholds would show better MCI scores; however, the results are less clear about such a relationship. For example, Subjects C3 and C8 had high-aided thresholds at 125 Hz but had good MCI performance with the HA; Subject C9 had relatively low-aided thresholds at 125 Hz, but had relatively poor MCI performance with the HA. This suggests that audibility alone (and by extension, the HA prescription) does not fully account for HA performance and bimodal benefit. It is possible that the spectral resolution of the acoustic ear (e.g., Zhang, Dorman, Fu, & Spahr, 2012) or central processing may have contributed to the present pattern of results. To better control for HA variability, it may be desirable to fit all bimodal subjects with the same prescription (e.g., half-gain rule) when measuring speech and music perception. However, the greater issue is that in clinical fitting of bimodal patients, there is little to no coordination of fitting between devices. Parameters between HAs and CIs (amplitude mapping, acoustic input frequency range, etc.) should be optimized to obtain the maximum bimodal benefit.

### *Sung Speech Perception Versus Previous Speech and Music Tasks*

Bimodal benefits were observed for sentence recognition in quiet, consistent with previous studies (Gifford et al., 2007; Neuman & Svirsky, 2013; Zhang et al., 2010) and in noise (Crew, Galvin, Landsberger, et al., 2015; Dorman et al., 2008; Kong et al., 2005). In the present study, the average bimodal benefit was similar across the three speech conditions, suggesting that the increased difficulty was not mitigated by bimodal listening. It could be that performance was related to the idiosyncratic speech productions in the SSC. Although the monosyllabic word duration (500ms) was similar to those in other databases (e.g., CNC words), speech production and the assembly of the monosyllabic words into sentences may have affected the overall intelligibility of the SSC sentences, even for the Spoken Speech condition. Another factor may have been the scoring system in the present study, where subjects were required to recognize every word in the sentence to receive credit. While some subjects performed well with spoken speech with the CI-only (C3, C4, and C9), mean CI-only performance was only 56.6% correct. Note that during testing, subjects could often identify four out of five words correctly, but receive no credit. The present scoring rule was adopted to be consistent with previous matrix test rules (Kollmeier et al., 2015) for testing in noise, but may have underestimated the word recognition (as opposed to sentence recognition) performance with the SSC. Nevertheless, the SSC stimuli allowed for speech intelligibility to be measured in quiet while explicitly manipulating the articulation (sung vs. spoken speech) and the variability in pitch cues across words. As such, it provides a glimpse into perception of atypical speech without the uncertain masking effects of noise.

The MCI task has been used in many previous studies to characterize CI users' melodic pitch perception (Crew et al., 2012; Galvin et al., 2009; Zhu, Chen, Galvin, & Fu, 2011). In Crew, Galvin, Landsberger, et al. (2015), MCI performance was evaluated in bimodal CI subjects listening to piano samples. Interestingly, mean CI-only performance with the piano in Crew, Galvin, Landsberger, et al. (2015) was poorer than that with the present Constant Timbre stimuli. Some subjects who participated in both studies exhibited much better performance with sung speech than with the piano when listening with the CI-only (C1, C7, and C8), while others did not (C3, C9). There were some differences between the previous piano stimuli and the present sung speech stimuli. The maximum F0 range was 220 Hz to 440 Hz in Crew, Galvin, Landsberger, et al. (2015) and 110 Hz to 220 Hz in the present study. Given that the low-frequency cutoff for the input

frequency range is approximately 200 Hz in many CI devices, this would suggest some advantage for the piano stimuli used in Crew, Galvin, Landsberger, et al. (2015). However, among subjects who participated in both the present study and the Crew, Galvin, Landsberger, et al. (2015) study, where the lowest note was 220 Hz, performance was better with sung speech. The better CI-only scores with sung speech than with piano notes may reflect the ability to use temporal pitch cues. Temporal pitch generally falls off around 300 Hz; F0s fell within this range for the SSC stimuli but not for the piano stimuli. It is also possible that optimization of CI signal processing for speech perception may have provided some advantage for the present sung speech stimuli.

## Conclusion

In this study, speech and music perception was measured in bimodal subjects listening with the CI-only, the HA-only, or with the CI+HA. Sentence recognition was measured in quiet with spoken and sung speech using a matrix test paradigm. For sung speech, F0 was held constant or varied across words and music perception was measured using sung speech in a MCI task; the timbre (words) was held constant or varied across notes in the contour. Major findings include:

1. Mean sentence recognition was poorer with sung speech than with spoken speech, most likely due to differences in production. There was no significant difference in sung speech performance when F0 was held constant or varied across words.
2. Mean MCI performance with sung speech worsened when timbre cues were varied across notes, suggesting that CI users have greater difficulty extracting melodic pitch from complex stimuli.
3. Speech performance was largely driven by the CI, while music performance was almost exclusively driven by the HA.
4. A strong benefit was observed for sentence recognition when listening with both devices, whether relative to the CI-only or to the better ear. However, the benefit was similar across the spoken and sung speech conditions, suggesting that the addition of acoustic to electric hearing had a global benefit (e.g., better overall perception of voicing or formant information) rather than a pitch-specific benefit.
5. There was a strong benefit for MCI when the HA was added to the CI, but not when the CI was added to the HA. Better optimization of both devices for bimodal listening may improve music perception (e.g., reducing frequency overlap or coordinating frequency allocation between the two devices).

## Acknowledgments

We thank all of the CI patients for their participation in this study. And we thank the reviewers and editor for their helpful comments and additions.

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the NSF GK-12 Body Engineering Los Angeles program and NIDCD R01-DC004993 and R01-DC004792.

## References

- Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America*, *135*, 1371–1379.
- Armstrong, M., Pegg, P., James, C., & Blamey, P. (1997). Speech perception in noise with implant and hearing aid. *American Journal of Otolaryngology*, *18*, S140–S141.
- Brown, C. A., & Bacon, S. P. (2009). Achieving electric-acoustic benefit with a modulated tone. *Ear and Hearing*, *30*, 489–493.
- Chang, Y.-P., & Fu, Q.-J. (2006). Effects of talker variability on vowel recognition in cochlear implants. *Journal of Speech Language and Hearing Research*, *49*, 1331–1341.
- Crew, J. D., Galvin, J. J. 3rd, & Fu, Q.-J. (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *The Journal of the Acoustical Society of America*, *132*, EL429–EL435.
- Crew, J. D., Galvin, J. J. 3rd, & Fu, Q.-J. (2015). Melodic contour identification and sentence recognition using sung speech. *The Journal of the Acoustical Society of America*, *138*, EL347–EL351.
- Crew, J. D., Galvin, J. J. 3rd, Landsberger, D. M., & Fu, Q.-J. (2015). Contributions of electric and acoustic hearing to bimodal speech and music perception. *PLoS One*, *10*, e0120279.
- Dorman, M. F., & Gifford, R. H. (2010). Combining acoustic and electric stimulation in the service of speech recognition. *International Journal of Audiology*, *49*, 912–919.
- Dorman, M. F., Gifford, R. H., Spahr, A. J., & McKarns, S. A. (2008). The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiology and Neurotology*, *13*, 105–112.
- Elliott, T. M., Hamilton, L. S., & Theunissen, F. E. (2013). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *The Journal of the Acoustical Society of America*, *133*, 389–404.
- Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, *110*, 1150–1163.

- Fu, Q.-J., Galvin, J. J. 3rd, Wang, X., & Wu, J. L. (2015). Benefits of music training in mandarin-speaking pediatric cochlear implant users. *Journal of Speech Language and Hearing Research, 58*, 163–169.
- Fu, Q.-J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology, 6*, 19–27.
- Galvin, J. J. 3rd, Fu, Q.-J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences, 1169*, 518–533.
- Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J., ... Stordahl, J. (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants International, 3*, 29–53.
- Gfeller, K., Witt, S., Stordahl, J., & Mehr, M. (2000). The effects of training on melody recognition and appraisal by adult cochlear implant recipients. *Journal of Academy of Rehabilitative Audiology, 33*, 115–138.
- Gifford, R. H., Dorman, M. F., McKarns, S. A., & Spahr, A. J. (2007). Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing. *Journal of Speech Language and Hearing Research, 50*, 835–843.
- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America, 61*, 1270–1277.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099–3111.
- Ji, C., Galvin, J. J. 3rd, Chang, Y.-P., Xu, A., & Fu, Q.-J. (2014). Perception of speech produced by native and non-native talkers by listeners with normal hearing and listeners with cochlear implants. *Journal of Speech Language and Hearing Research, 57*, 532–554.
- Kasturi, K., & Loizou, P. C. (2007). Effect of filter spacing on melody recognition: acoustic and electric hearing. *The Journal of the Acoustical Society of America, 122*, EL29–EL34.
- Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., ... Wagener, K. C. (2015). The multi-lingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology, 54*(Suppl 2), 3–16.
- Kong, Y.-Y., Cruz, R., Jones, J. A., & Zeng, F.-G. (2004). Music perception with temporal cues in acoustic and electric hearing. *Ear and Hearing, 25*, 173–185.
- Kong, Y.-Y., Mullangi, A., Marozeau, J., & Epstein, M. (2011). Temporal and spectral cues for musical timbre perception in electric hearing. *J Speech Lang Hear Res, 54*, 981–994.
- Kong, Y.-Y., Stickney, G. S., & Zeng, F.-G. (2005). Speech and melody recognition in binaurally combined acoustic and electric hearing. *The Journal of the Acoustical Society of America, 117*, 1351–1361.
- Kreft, H. A., Nelson, D. A., & Oxenham, A. J. (2013). Modulation frequency discrimination with modulated and unmodulated interference in normal hearing and in cochlear-implant users. *Journal of the Association for Research in Otolaryngology, 14*, 591–601.
- Li, Y., Zhang, G., Kang, H.-Y., Liu, S., Han, D., & Fu, Q.-J. (2011). Effects of speaking style on speech intelligibility for Mandarin-speaking cochlear implant users. *The Journal of the Acoustical Society of America, 129*, EL242–EL247.
- Looi, V., McDermott, H., McKay, C., & Hickson, L. (2007). Comparisons of quality ratings for music by cochlear implant and hearing aid users. *Ear and Hearing, 28*, 59S–61S.
- Looi, V., McDermott, H., McKay, C., & Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear and Hearing, 29*, 1–14.
- Luo, X., Fu, Q.-J., & Galvin, J. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends in Amplification, 11*, 301–315.
- Macherey, O., & Delpierre, A. (2013). Perception of musical timbre by cochlear implant listeners: A multidimensional scaling study. *Ear and Hearing, 34*, 426–436.
- McDermott, H. J. (2004). Music perception with cochlear implants: A review. *Trends in Amplification, 8*, 49–82.
- Mok, M., Galvin, K. L., Dowell, R. C., & McKay, C. M. (2010). Speech perception benefit for children with a cochlear implant and a hearing aid in opposite ears and children with bilateral cochlear implants. *Audiology and Neurotology, 15*, 44–56.
- Mok, M., Grayden, D., Dowell, R. C., & Lawrence, D. (2006). Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears. *Journal of Speech Language and Hearing Research, 49*, 338–351.
- Neuman, A. C., & Svirsky, M. A. (2013). Effect of hearing aid bandwidth on speech recognition performance of listeners using a cochlear implant and contralateral hearing aid (bimodal hearing). *Ear Hear, 34*, 553–561.
- Su, Q., Galvin, J. J., Zhang, G., Li, Y., & Fu, Q. J. (in press). Effects of within-talker variability on speech intelligibility in mandarin-speaking adult and pediatric cochlear implant patients. *Trends in Hearing*. doi:10.1177/2331216516654022.
- Turner, C. W., Gantz, B. J., Vidal, C., & Behrens, A. (2004). Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. *The Journal of the Acoustical Society of America, 115*, 1729–1735.
- Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., & Noble, W. (2002). Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization. *Ear and Hearing, 23*, 98–105.
- Yoon, Y.-S., Li, Y., & Fu, Q.-J. (2012). Speech recognition and acoustic features in combined electric and acoustic stimulation. *Journal of Speech Language and Hearing Research, 55*, 105–124.
- Zeng, F.-G. (2002). Temporal pitch in electric hearing. *Hearing Research, 174*, 101–106.
- Zhang, T., Dorman, M. F., Fu, Q.-J., & Spahr, A. J. (2012). Auditory training in patients with unilateral cochlear implant and contralateral acoustic stimulation. *Ear and Hearing, 33*, 70–79.
- Zhang, T., Spahr, A. J., & Dorman, M. F. (2010). Frequency overlap between electric and acoustic stimulation and speech-perception benefit in patients with combined electric and acoustic stimulation. *Ear and Hearing, 31*, 195–201.
- Zhu, M., Chen, B., Galvin, J., & Fu, Q.-J. (2011). Influence of pitch, timbre and timing cues on melodic contour identification with a competing masker. *The Journal of the Acoustical Society of America, 130*, 3562–3565.