Published in final edited form as:
Ear Hear. 2012 September ; 33(5): 645–659. doi:10.1097/AUD.0b013e318252caae.

Timbre and Speech Perception in Bimodal and Bilateral Cochlear-Implant Listeners

Ying-Yee Kong[a),b),*)], **Ala Mullangi**[b)], and **Jeremy Marozeau**[a),c)]

[a)]Department of Speech Language Pathology & Audiology, Northeastern University, Boston, MA 02115, USA

[b)]Bioengineering Program, College of Engineering, Northeastern University, Boston, MA 02115, USA

[c)]The Bionic Ear Institute, East Melbourne, VIC 3002, Australia

INTRODUCTION

Timbre, by definition, is the perceptual attribute that distinguishes two sounds that have the same pitch, loudness, and duration (ANSI 1973). For musical timbre, this perceptual attribute distinguishes instruments (e.g., guitar vs. piano) playing the same note with the same loudness. Differences in timbre percepts can be evaluated as listeners' rating or overall appraisal of the sounds, including subjective characteristics such as "pleasantness" or "naturalness" (e.g., Von Bismark 1974; Pratt & Doak 1976; Kendall & Carterette 1993; Gfeller & Lansing 1991; Schulz & Kerber 1994; Looi et al. 2007), or described as sound quality on different perceptual dimensions, such as dull/brilliant, compact/scattered, and full/empty (Gfeller et al. 2002).

Some researchers (e.g., Grey 1977; McAdams et al. 1995; Marozeau et al. 2003; Caclin et al. 2005) used a different technique, known as multidimensional scaling (MDS) to investigate timbre perception and derive the timbre space for normal-hearing (NH) listeners. Based on the similarity/dissimilarity matrix, the MDS technique can detect the number of meaningful dimensions for the data and assign a location for each item (stimulus) in the N-dimensional space. A classic example of the utility of the MDS technique is that based on the distance matrix between American cities, cities can be visually represented on a two-dimensional map, with one dimension representing north and south and the other east and west. For timbre perception, a number of temporal and spectral acoustic properties that are associated with each of the perceptual dimensions in the timbre space have been previously investigated (e.g., Grey 1977; Krumhansl 1989; Krimphoff et al. 1994; McAdams et al. 1995; Marozeau et al. 2003). These acoustic properties included temporal envelope features (e.g., attack time, impulsiveness), spectral envelope features that represent the spectral shape (e.g., spectral central moments – spectral centroid, spectral spread, spectral skewness, spectral kurtosis), spectral fine structure features (e.g., spectral irregularity), and features that represent spectral changes over time (e.g., spectral flux). A comprehensive description of the acoustic features that are relevant to musical sounds classification is provided by Peeters (2004). In their latest findings, McAdams and colleagues (Caclin et al. 2005) demonstrated that timbre space obtained from NH listeners can be represented in three dimensions, with one dimension associated with the attack time, another dimension associated with the spectral centroid, and the last dimension associated with the spectral

[*)]Corresponding author: Address: Department of Speech Language Pathology & Audiology, 106A Forsyth Building, Northeastern University, Boston, MA 02115, USA; Tel: +1 (617) 373-3704; Fax: +1 (617)373-2239; yykong@neu.edu.

irregularity of the stimuli. These findings indicated that temporal envelope (attack time), spectral envelope (spectral centroid), and spectral fine structure (spectral irregularity) cues are the relevant cues for perceiving musical timbre differences in NH listeners. In addition, unlike pitch perception, for which temporal and spectral envelope cues elicit only non-salient pitch, temporal and spectral envelope cues are found to be the dominant cues compared to the spectral fine structure cues for musical timbre perception (e.g., Grey 1977; Wessel 1979; Krumhansl 1989; McAdams et al. 1995).

Using the same synthesized musical stimuli and a similar analysis method as in McAdams et al. (1995), Kong et al. (2011) recently examined the timbre space in both NH and cochlear-implant (CI) listeners. In their study, the three-dimensional (3D) timbre space described in McAdams et al. (1995) and Caclin et al. (2005) was first verified in a group of NH listeners. Similar to Caclin et al.'s findings, the first dimension of the NH timbre space in Kong et al. (2011) was highly correlated with the temporal envelope cues (i.e., attack time and impulsiveness), the second dimension was highly correlated with the spectral envelope cue (i.e., the spectral centroid), and the third dimension was weakly correlated with the spectral fine structure cue (i.e., spectral irregularity). Both temporal and spectral envelope cues were the more salient cues for timbre perception as reflected by the relatively higher perceptual weights on the first and second dimensions compared to the third dimension. Unlike NH results, Kong et al. (2011) reported that the CI timbre space was best fit with a 2D solution, in which the first dimension was strongly correlated with the temporal envelope of the stimuli (i.e., attack time and impulsiveness), and the second dimension was weakly correlated with the spectral envelope of the stimuli (spectral centroid). Other spectral and temporal features, including spectral flux (spectral variation over time) and spectral irregularity (spectral fine structure cue) were not significantly correlated with any of the perceptual dimensions for CI users. These results suggested that CI listeners primarily used the temporal envelope cue to perceive differences in musical timbre, and the spectral envelope was less salient or reliable for timbre perception for CI listeners compared to NH listeners. Kong et al.'s results were in good agreement with previous findings on CI timbre perception when CI listeners were asked to discriminate between musical instruments. (Gfeller et al. 2002; McDermott 2004; Nimmons et al. 2008; Kang et al. 2009). McDermott (2004) tested a group of CI listeners on a musical instrument identification task with 16 instruments from non-percussive and percussive families. They found that more confusions were made among the instruments within the same category (i.e., percussive or non-percussive) than between categories. Based on these patterns of confusions, they concluded that temporal envelope cues are more salient cues for musical timbre perception for CI listeners. The reduced reliance on the spectral envelope cues might have contributed to the lower sound quality rating assigned by CI listeners compared to their NH counterparts (Schulz & Kerber 1994; Gfeller & Lansing 1991; Gfeller et al. 1998, 2000, 2002).

The reliance on spectral cues for timbre perception may have been different when CI listeners received additional low-frequency acoustic information, such as electric-acoustic stimulation (EAS) in the same ear (hybrid hearing) or in opposite ears (bimodal hearing). Previous studies on music perception in EAS listeners showed improved pitch or melody perception compared to CI alone (e.g., Kong et al. 2005; Dorman et al. 2008), due to the better encoding of the spectral or temporal fine structure cues with low-frequency acoustic stimulation. However, as pointed out by Kong et al. (2005), there is a dichotomy between EAS benefit for speech and pitch perception. For speech perception in quiet, the EAS performance is greater than performance with CI alone or with low-frequency acoustic hearing alone, especially for word and sentence recognition, partly due to the speech-cue integration mechanism when the information from electric stimulation and acoustic stimulation is complementary to each other (e.g., Mok et al. 2006; Dorman et al. 2008; see Dorman and Gifford 2010 for review). For pitch perception, however, the EAS performance

is generally determined by the performance of the low-frequency acoustic hearing alone (Kong et al. 2005; Dorman et al. 2008). For timbre perception, a recent study by Gfeller et al. (2006) showed better instrument identification performance in listeners with hybrid hearing, suggesting an integration of cues to enhance timbre perception in EAS listeners. However, unlike EAS, Gfeller et al. (2008) did not find that having two CIs (bilateral CI) improved instrument identification performance or appraisal ratings of music compared to one CI. Similar to instrument identification, there was also a reduction or absence of combined benefit for speech recognition in quiet and in a non-spatial noise condition in bilateral CI listeners (Mok et al. 2010; Yoon et al. 2011b) compared to EAS listeners (Dorman and Gifford 2010). As pointed out by Dorman and Gifford (2010), the reduced bilateral benefit in quiet and in non-spatial noise conditions could be due to the fact that bilateral CI listeners receive the similar type of low spectral resolution signal from both implants. One could argue that if there is a considerable difference in electrode placement, the two ears could receive different information from the two implants, which would lead to greater bilateral benefit. However, a simulation study by Yoon et al. (2011a) showed that bilateral benefit in quiet and in a non-spatial noise condition actually increased when the interaural difference in insertion depth decreased.

To the authors' knowledge, there has not been a systematic investigation on the relative contribution of temporal and spectral cues for musical timbre perception in bimodal CI and bilateral CI listeners, nor the relationship between combined benefit for speech recognition and timbre perception in these two groups of listeners. The first goal of this study is to investigate the contribution of temporal and spectral cues for musical timbre perception in bimodal and bilateral CI listeners. Differences in the reliance on spectral cues for timbre perception in EAS, CI alone, low-frequency acoustic hearing alone, and bilateral CIs is the focus of this investigation. Based on the results reported by Gfeller et al. (2006), as well as the importance of temporal and spectral envelope cues for timbre perception discussed above, it was hypothesized that patterns of results in EAS benefit for timbre perception would be similar to those observed for speech perception. That is, there would be stronger associations between the timbre dimensions and their corresponding acoustic properties, particularly the correlation between dimension 2 and the spectral envelope cue for the bimodal condition compared to CI alone or low-frequency residual hearing alone. When comparing timbre perception results from bimodal CI listeners to those from bilateral CI listeners, it was hypothesized that the use of bilateral CIs provides less or no additional benefit for timbre perception. That is, the associations between the timbre dimensions and their corresponding acoustic properties remain similar for the bilateral CI condition and the better single CI condition. This is based on the assumption that each CI provides similar information in the bilateral CI condition and on previous findings which showed the lack of bilateral benefit for instrument identification (Gfeller et al. 2008).

The second goal of this study is to relate phoneme recognition and timbre perception performance in bimodal CI and bilateral CI listeners. As mentioned above, impulsiveness (one of the measures for the temporal envelope feature), and spectral centroid (one of the measures for the spectral envelope feature) are the dominant cues for timbre perception. Temporal envelope cues are sufficient to support high levels of speech recognition, especially consonant recognition in quiet as demonstrated by reports using a vocoder-processing technique (Shannon et al. 1995). Spectral envelopes, which represent the formant structure, are the primary cues for vowel recognition in quiet. It was hypothesized that combining wide-band electric cues and low-frequency acoustic cues provides similar benefit for speech and timbre perception, particularly in bimodal CI listeners when complementary cues could come from electric and acoustic stimulation. For consonant recognition, if there is a significant combined benefit for the perception of impulsiveness, it was expected that there would be a significant combined benefit for the manner of articulation perception

(Shannon et al. 1995; van Tasell 1987; 1992). Using a one-channel noise vocoder with the absence of temporal periodicity cues, Shannon et al. (1995) showed that NH listeners could achieve greater than 50 percent manner scores. For vowel recognition, an increase in the sensitivity of spectral centroid change in EAS would be related to the increase in the sensitivity of change of formant frequencies. The range of spectral centroids in the musical stimuli used in this study overlaps with the first and second formant frequency regions of the vowel stimuli. Thus, an individual who demonstrates a combined benefit in perception of spectral centroid for timbre perception could also show a combined benefit for first formant (F1) and second formant (F2) perception.

MATERIALS AND METHODS

Subjects

Two groups of subjects – bimodal and bilateral CI users were recruited for this study.

Bimodal CI Group—Seven bimodal CI users (C1-C7, 3 males, 4 females), ages 16 to 65 years (mean = 41.9 years) participated in this study. Table I shows detailed demographic information for each bimodal CI subject, including age, etiology of hearing loss, duration of severe-to-profound hearing loss prior to implantation, CI processor, number of years of CI use prior to the study, and pure-tone average (PTA) of unaided thresholds across four frequencies (250, 500, 1000, and 2000 Hz). Four of the subjects (C3, C4, C5, and C6) were pre-lingually or peri-lingually severely hard-of-hearing bilaterally. Subject C7 acquired a severe hearing loss in the right ear at age 2 due to meningitis. She had normal hearing in her left ear until she had a sudden hearing loss at age 43. Two subjects (C1 and C2) had post-lingual onset of hearing loss. All subjects wore a hearing aid (HA) in the non-implanted ear on a daily basis. Figure 1 shows the measurable (i.e., ≤ 110 dB HL) aided and unaided thresholds in the non-implanted ear. With amplification, the aided thresholds were at the mild to moderate (30-60 dB HL) hearing loss level up to about 2000 Hz for most of the subjects.

Bilateral CI Group—Five bilateral CI users (C2, C5, C8, C9, and C10, 4 females), ages 17 to 66 years (mean = 41.8 years) participated in this study. Subjects C2 and C5 were also in the bimodal CI group. They completed the study as bimodal CI subjects prior to receiving their second CI. After 6-8 months of experience with their second CI, they returned to be tested with their second CI and with bilateral CIs, and re-tested with their first CI. Subject 2 used Nucleus 5 processors in both ears after receiving the second implant. Four of the five subjects were implanted with the same device in both ears and used the same CI processor for both ears. Subject C10 was first implanted with a five-channel ineraid device and used a Geneva processor in the left ear. He then received a Clarion HiRes implant 20 years later and used a body-worn Harmony processor in the right ear. Table II shows detailed demographic information for each bilateral CI subject and information for each CI.

Stimuli

Timbre Perception—The stimuli were the digitally synthesized musical instruments used in McAdams et al. (1995). In order to reduce the session duration, only 16 out of the original 18 stimuli were selected. Among the 16 stimuli, 11 of them were designed to imitate traditional western instruments: bassoon (bsn), english horn (ehn), guitar (gtr), harpsichord (hcd), French horn (hrn), harp (hrp), piano (pno), bowed string (stg), trombone (tbn), trumpet (tpt), and vibraphone (vbs). The remaining five stimuli were hybrids of two instruments to contain the perceptual characteristics of both instruments: guitarnet (gtn) [combination of guitar and clarinet], obochord (obc) [oboe and harpsichord], striano (sno) [bowed string and piano], trumpar (tpr) [trumpet and guitar], and vibrone (vbn) [vibraphone

and trombone]. All stimuli had the same fundamental frequency (F0) of 311 Hz (E-flat above middle C). The durations of these sounds ranged from 495 ms to 1096 ms. According to McAdams et al. (1995), these physical differences in duration across stimuli were required to produce similar perceptual duration for NH listeners. The stimuli were scaled to have equal maximum level. These stimuli were recently used by Kong et al. (2011) to investigate timbre perception in NH and CI listeners.

Speech Perception—The stimuli included 16 consonants /p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, ʒ, m, n/ in the /aCa/ context recorded by five male and five female talkers by Shannon et al. (1999), and nine monophthongs /i, ɪ, ɛ, æ, ɝ, ʌ, u, ʊ, ɔ/ in the /hVd/ context recorded from five male and five female talkers in the first author's laboratory. All stimuli were scaled to the same overall root-mean-squared (RMS) level. For each stimulus set, recordings from two male and two female talkers were used in the practice sessions, and recordings from the remaining six talkers were used in the test sessions. Three utterances of each consonant and vowel from each talker were used in the test sessions.

Procedure

Each subject was tested on three listening conditions – individual device alone and combined use of both devices (CI-alone, HA-alone, CI+HA for the bimodal CI group; left CI, right CI, bilateral CIs for the bilateral CI group). Subjects who wore a HA in the non-implanted ear were instructed to turn-off their HA but leave their earmold in place in that ear when tested on the CI-alone condition in order to prevent any contribution from residual acoustic hearing. Bilateral CI subjects were instructed to turn off the non-test implant for the single CI conditions. A foam earplug was inserted in the implanted ear(s) for both the bimodal and bilateral CI groups during testing to prevent any potential acoustic stimulation in case residual hearing was preserved in the implanted ear(s).

Timbre Perception—The test procedure was similar to that described in Kong et al. (2011). All subjects were tested in a double-walled sound-attenuating booth (8'6" × 8'6"). Stimuli were presented from a loudspeaker (M-Audio Studiophile BX8a Deluxe) one meter in front of the subjects. This loudspeaker has a flat frequency response (±2 dB) from 40 to 22000 Hz. Each subject used his/her own CI and HA settings during this experiment, except for the volume control setting of the HA. Before testing, a check was performed to verify similar loudness across devices (between CI and HA for the bimodal CI group and between the two CIs for the bilateral CI group), and adjustments were made when necessary. First, a reference sound – bassoon was presented at a comfortable listening level for each subject when he/she listened with a CI alone (or the right CI for the bilateral CI group). Subjects then listened to the reference sound at the same level with their HA alone (or the left CI in the bilateral CI group) to confirm that the perceived loudness was "similar" as on the other side. Note that an exact match in loudness across devices could not be guaranteed because this task required subjects to turn their devices on and off and hold the perceived loudness in memory during the comparison. All bilateral CI subjects reported similar loudness across devices during this assessment. Only a few bimodal subjects reported slight loudness difference between their CI and HA. These subjects were asked to adjust the volume control of their hearing aid until the reference sound reached a similar loudness level as on the CI side. The same HA and CI settings were then used for the combined condition (CI+HA or bilateral CI) for each subject. Once the comfortable reference sound levels and the volume control settings for the HA were determined, subjects then adjusted the levels of the rest of the stimuli to match the loudness of the reference sound for each listening condition. After the adjustments for each individual stimulus, the subjects played the sounds one at a time and made further adjustments as necessary if any of the stimuli sounded noticeably louder or softer than the rest. Thus, the presentational levels were different for different stimuli and

for different subjects. Separate adjustments for equal loudness across stimuli were performed for the CI-alone, HA-alone, and CI+HA condition in the bimodal CI group, and for the individual CI alone and bilateral CI condition in the bilateral CI group. In the CI+HA and bilateral CI conditions, subjects confirmed that the perceived fused sound image came directly from the front.

Prior to the experiment, each subject was instructed that the goal of the study was to estimate the similarity of sound quality between sounds. They were told that each sound used in the experiment was the same musical note and that the loudness should be roughly the same across sounds given that they already adjusted the level of each sound to produce equal loudness prior to the experiment.

For each listening condition, subjects were first familiarized with the 16 stimuli by listening to each stimulus one at a time as many times as they wanted. Subsequently, all possible pairs of the 16 stimuli (16 × 16 = 256 pairs) were presented in random order. Subjects were instructed to judge the dissimilarity of the two sounds in each pair by adjusting a sliding bar on a computer screen. The sliding bar was marked with numbers from 1 to 10 in steps of 1, and with labels “most similar” next to the number “1” and “most different” next to the number “10.” Subjects were encouraged to use the whole range of the scale. For each trial, the subject could play the pair as many times as he/she wanted before making a dissimilarity rating. Each subject was allowed to take a break at any time during the experiment. Before data collection, each subject received a training session for each listening condition that consisted of 20 randomly selected pairs to ensure that he/she understood the task and to practice making the dissimilarity rating. Each subject performed the experiment two or three times for each listening condition depending on the availability of the subject. The order of testing for the listening conditions was randomized for each subject.

Speech Perception—The test procedure was similar to that described in Kong and Braida (2011). Three listening conditions were tested – individual device alone and combined use of both devices. All stimuli were presented from the same loudspeaker and one meter in front of the subject as in the timbre perception task at an RMS level of 65 dBA. Subjects used their own CI and HA settings during this experiment, except for the volume setting in the HA. Subjects adjusted the volume of their HAs until the presented stimuli reached their most comfortable listening level. The same HA and CI settings were used for the combined condition for each subject.

Phoneme recognition was evaluated on all bimodal CI subjects and on three of the five bilateral CI subjects (C2, C9, C10). Bimodal subject C1 was only tested on consonant recognition; he was unable to return for the vowel recognition test. Subjects C5 (after receiving the second implant) and C8 were unable to return for speech recognition testing after the timbre experiment. Subject C2 was included in both bimodal and bilateral CI groups. Phoneme recognition of the first CI was re-tested for this subject after she received a second implant. For each condition, subjects first received practice trials identifying the consonant and vowel with visual correct-response feedback provided. Performance usually reached a plateau (i.e., within 3 percentage points difference) within three blocks of practice. If not, additional practice was given until the criterion was met. Each subject was then tested in blocks of 96 trials (16 consonants × 6 talkers) for consonant identification, and 54 trials (9 vowels × 6 talkers) for vowel identification. Each utterance from each talker was used for three blocks of testing. Nine blocks of testing were presented in each test condition, yielding a total of 54 trials (6 talkers × 9 blocks) per stimulus per condition per subject. Subject C1 only received seven blocks of testing for consonant recognition with a total of 42 trials per stimulus per task, due to his time constraints. No feedback was provided during test

sessions. A list of 16 /aCa/ or nine /hVd/ syllables was displayed on a computer screen and subjects responded by clicking a button corresponding to the syllable they heard.

Analyses

Timbre Perception—The primary goal of this study is to investigate the relative contribution of temporal and spectral cues to timbre perception in bimodal and bilateral CI listeners. To achieve this goal, timbre space for each listening condition was first derived using the MDS technique. Similar to Kong et al. (2011), the group data was analyzed using a weighted Euclidean model – Individual Differences MDS (INDSCAL) analysis (Carroll & Chang, 1970) in SPSS version 17.0 on the dissimilarity matrices for each listening condition for the bimodal CI and bilateral CI group. Individual data was analyzed using the traditional MDSCAL procedure, implemented in MATLAB (R2009a) according to the SMACOF algorithm (Borg & Groenen, 1997). The MDSCAL analysis was performed on the averaged, folded matrix for each subject for each listening condition. Using the MDS technique, the number of dimensions that best fit the data for each listening condition was determined. The two-dimensional (2D) and three-dimensional (3D) fits were the focuses in this analysis because previous results from NH and CI listeners (McAdams et al. 1995; Calcin et al. 2005; Kong et al. 2011) showed that a 3D solution provided the best fit for NH timbre data and a 2D solution was best for unilateral CI data. Second, correlational analyses were performed between the coordinates of the 16 stimuli in each of the timbre dimensions and the values of the known acoustic features associated with these dimensions for the 16 stimuli (McAdam et al. 1995; Marozeau et al. 2003; Caclin et al. 2005; Kong et al. 2011). Based on Kong et al.'s findings on NH and unilateral CI listeners, as well as previous findings on NH listeners (McAdams et al. 1995; Marozeau et al. 2003; Caclin et al. 2005), temporal envelope (impulsiveness) and spectral envelope (spectral centroid) were the features focused upon in the analyses in this study. Impulsiveness is calculated as one minus the ratio of the duration during which the temporal envelope is above 50% of its maximum and the duration during which it is 10 % above (Marozeau et al. 2003). Spectral centroid is defined as the spectral center of gravity, which is calculated as the amplitude weighted mean of the harmonic peaks averaged over the sound duration (Kong et al. 2011).

Speech Perception—Percent correct scores were calculated for consonant and vowel recognition for each listening condition. Planned pairwise comparisons (t-tests) were performed to determine if performance in the combined condition (CI+HA or bilateral CI) was significantly better than that in the single device condition that produced the higher percent correct score (i.e., better ear condition) for the group data, as well as for individual subjects. Additionally, percent information transmission for consonant (voicing, manner, and place of articulation) and vowel (height and back) features was computed from confusion matrices combined across nine runs per subject per listening condition.

Relationship between Timbre and Speech Perception—A stronger correlation between the acoustic feature and its corresponding timbre dimension would indicate better perception of the acoustic cue for timbre perception. These timbre acoustic cues may also be relevant to speech perception. To evaluate the relationship between timbre and speech perception, correlational analyses were performed (1) to correlate the consonant recognition scores and the correlation values between impulsiveness and its corresponding dimension, and (2) to correlate the vowel recognition scores and the correlation values between spectral centroid and its corresponding dimension. To evaluate the relationship with regard to the combined benefit between timbre and speech perception, a binominal probability calculation was performed to determine the probability of obtaining a certain number of matched cases (i.e., presence or absence of combined benefit) between speech and timbre perception by

chance[1]. A small probability (e.g., less than 5%) would indicate that it is very unlikely that a certain number of matched cases observed in the data occurred purely by chance.

Relationship between Residual Hearing and Timbre and Speech Perception— Pure-tone averages of unaided thresholds across four frequencies (250, 500, 1000, and 2000 Hz) were calculated for each subject in the bimodal CI group. The frequency of 2000 Hz was included in the calculation due to its relevance to the spectral centroid values. In cases where a frequency did not have a measurable threshold (i.e., ≤110 dB), 120 dB was used in the calculation for that frequency. Correlational analyses were performed (1) to correlate the PTAs and the correlation values between acoustic features and their corresponding dimensions for timbre perception, and (2) to correlate the PTAs and the overall phoneme recognition scores for speech perception.

RESULTS

Temporal and Spectral Cues for Timbre Perception

Using the definition provided in the Analyses section, the calculated impulsiveness values were in the range of 0.03 to 0.5, and the spectral centroid values were in the range of 800 Hz to 2000 Hz in the stimulus set used in this study. Figure 2 shows the waveforms and spectra of two instruments – bsn and vbn which differ considerably in the values of the impulsiveness and spectral centroid features. The majority of the stimuli have one distinct peak in the spectrum and have a considerable decrease in amplitude in the high-frequency harmonics (>3000 Hz). The spectra for bsn and vbc shown in Fig. 2 represent the general spectral shape in the stimulus set used in this study.

The MDS analysis provided the 2D and 3D solutions for each subject group and for each listening condition. Results from the acoustical analyses on the temporal envelope (impulsiveness) and spectral envelope (spectral centroid) features of the stimuli described above were used to correlate with the coordinates of the 16 stimuli in each of the MDS dimensions for the group data. In general, two of the dimensions in a 3D solution correlated with the impulsiveness and spectral centroid features. Additionally, other spectral features that were investigated previously, including spectral spread (spectral envelope cue), spectral irregularity (spectral fine structure cue), and spectral flux (spectrotemporal cue) were measured (Krimphoff et al. 1994; McAdams et al. 1995; Marozeau et al. 2003; Caclin et al. 2005; Kong et al. 2011). Spectral irregularity and spectral flux did not significantly correlate with any of the three dimensions. Spectral spread was found to significantly correlate with Dim 1 in the CI-alone condition in the bimodal CI group. However, spectral spread was also significantly correlated with the temporal envelope feature – log-attack time as previously reported by Kong et al. (2011). Taken together, Dim 1 represents the temporal envelope dimension because it significantly correlated with impulsiveness, and Dim 2 represents the spectral envelope dimension because it significantly correlated with spectral centroid only. A 2D fit, on the other hand, provided a solution that in most cases enhanced the correlations between each dimension and its corresponding acoustic feature. These findings are similar to those reported by Kong et al. (2011) for unilateral CI listeners.

Timbre Perception in Bimodal CI Listeners—Results from the acoustical analyses on impulsiveness and spectral centroid were used to correlate with the coordinates of the 16

[1]Previous reports on bimodal and bilateral phoneme recognition demonstrated that not all listeners received combined benefit for phoneme recognition in quiet; and those who demonstrated a combined benefit showed only a few percentage points of improvement (e.g., Mok et al. 2006; 2010; Yoon et al. 2011b). Given the small amount of combined benefit for phoneme recognition, it is not possible to correlate the amount of benefit for phoneme recognition and the increase in correlation between acoustic features and their corresponding dimensions.

stimuli for Dim 1 and Dim 2 of the MDS dimensions for the bimodal CI data. Figure 3 shows the correlations between impulsiveness and Dim 1 (top panel) and correlations between spectral centroid and Dim 2 (bottom panel) for each listening condition for the individual and group data. As a group, Dim 1 was found to be highly correlated with impulsiveness (temporal envelope cue) for all three listening conditions (CI-alone: $r = 0.89$, $p \leq 0.0001$; HA-alone: $r = 0.83$, $p \leq 0.0001$; CI+HA: $r = 0.93$, $p \leq 0.0001$). Dim 2 was significantly correlated with spectral centroid (spectral envelope cue) for the CI-alone ($r = 0.57$, $p \leq 0.05$) and CI+HA ($r = 0.61$, $p \leq 0.05$) conditions, but not for the HA-alone condition ($r = 0.47$, $p > 0.05$). Correlations between Dim 2 and spectral centroid were considerably weaker than the correlations between Dim 1 and impulsiveness for all listening conditions. The degrees of correlation between Dim 2 and spectral centroid (paired-t(6) = 1.33, $p > 0.05$) and between Dim 1 and impulsiveness (paired-t(6) = 0.27, $p > 0.05$) were not significantly different between CI-alone and CI+HA conditions. Overall weights for each perceptual dimension were also obtained from the INDSCAL analysis. Given that the INDSCAL procedure implicitly assumes that the data are conditional (McCallum, 1977), the ratio of the weights on each dimension indicates which dimension is more salient. For bimodal users, the overall weights for the Dim 1 were considerably higher than those for Dim 2 for all listening conditions, suggesting that perceptual salience was higher for impulsiveness than for spectral centroid cues. The ratios of the weights between Dim 2 and Dim 1 (Dim2/Dim1) were considerably higher for the CI-alone (0.72) and CI+HA (0.64) conditions than for the HA-alone condition (0.38). This indicates a reduced perceptual salience of spectral centroid cues when subjects listened with only residual low-frequency hearing.

Individual data revealed similar patterns of results compared to the group data: (1) All subjects showed significant correlations between Dim 1 and impulsiveness for all listening conditions (except for C6 in the HA-alone condition); (2) Correlations between Dim 1 and impulsiveness were considerably stronger than correlations between Dim 2 and spectral centroid for all subjects and listening conditions. There were also differences in the patterns of results among subjects. First, unlike the majority of the subjects, subject C5 in CI-alone and CI+HA conditions showed strong correlations between impulsiveness and both Dim 1 and Dim 2, suggesting that both dimensions were temporal dimensions. Second, different from the rest of the group, subject C7 showed a significant correlation between Dim 2 and spectral centroid in the HA-alone condition. For this subject, the Dim 2 and spectral centroid correlation was considerably stronger for the HA-alone condition than for CI-alone condition. Third, while some subjects (e.g., C2, C3) showed similar degrees of correlation between Dim 2 and spectral centroid for the CI-alone and CI+HA conditions, others (C1, C4, C6, C7) showed considerably enhanced correlation coefficients by 0.1 or greater between Dim 2 and spectral centroid in the CI+HA condition compared to the CI-alone condition. Fourth, unlike the rest of the group, subject C1 showed a considerable decrease of correlation coefficient by 0.2 between Dim 1 and impulsiveness in the CI+HA condition compared to the HA-alone condition.

Correlational analyses (n = 7) did not reveal a significant ($p > 0.05$) relationship between the amount of residual hearing (i.e., PTA) in the non-implanted ear and the correlation values between the acoustic features and their corresponding dimension in the HA-alone and CI +HA conditions. Also, there seems to be a lack of relationship between PTA and bimodal benefit. That is, individuals who had a greater amount of residual hearing (e.g., C7) did not necessarily receive combined benefit compared to the better ear condition.

Timbre Perception in Bilateral CI Listeners—Figure 4 shows the correlations between impulsiveness and Dim 1 (top panel) and correlation between spectral centroid and Dim 2 (bottom panel) for each listening condition for the bilateral CI individual and group

data. For the bilateral CI group, correlation analyses between each dimension and the acoustic features were performed for the single CI and for the bilateral CI conditions. Subjects C2 and C5 were re-tested with their first CI. As a group, impulsiveness was found to be highly correlated with Dim 1 for both single CI and bilateral CI conditions (single CI: $r = 0.96$, $p \leq 0.0001$; bilateral CI: $r = 0.95$, $p \leq 0.0001$). There was a significant, but weak correlation between spectral centroid and Dim 2 for the single CI condition ($r = 0.56$, $p \leq 0.05$). Correlation between spectral centroid and Dim 2 for the bilateral CI condition was 0.46, which was similar to that observed in the single CI condition, but it did not reach statistical significance ($p = 0.06$). Similar to the bimodal CI group, the overall weights for Dim 1 were considerably higher than those for Dim 2 for all listening conditions, suggesting that perceptual salience was higher for temporal envelope than for spectral envelope cues in bilateral CI users. The Dim2/Dim1 weight ratios were about 0.3, independent of listening conditions.

Similar to the bimodal CI group, individual data revealed a similar pattern of results compared to the group data: (1) All subjects showed significant correlations between impulsiveness and Dim 1 for all listening conditions; (2) Correlations between impulsiveness and Dim 1 were considerably stronger than correlations between spectral centroid and Dim 2 for all subjects and listening conditions. There were also differences in the patterns of results among subjects. Unlike other subjects in the group, subject C5 showed strong correlations between impulsiveness and Dim 1 and Dim 2 for the single CI and bilateral CI conditions, suggesting that both dimensions were temporal dimensions. This pattern of results was similar to those seen in the CI-alone and CI+HA conditions described above for this subject. While other subjects showed similar degrees of correlation between spectral centroid and Dim 2 for the better single CI and bilateral CI conditions, subject C10 showed a considerably enhanced correlation coefficient by 0.11 between spectral centroid and Dim 2 in the bilateral CI condition compared to the better CI (2^{nd} CI) condition. On the other hand, subject C9 showed a considerable decrease in correlation coefficient by 0.23 between impulsiveness and Dim 1 in the bilateral CI condition compared to the better CI (1^{st} CI) condition.

Phoneme Recognition

Phoneme Recognition in Bimodal CI Listeners—Overall percent correct consonant and vowel recognition scores for each listening condition for each bimodal subject and the group are shown in Fig. 5. Note that the individual bimodal CI data have been reported previously in Kong and Braida (2011). As a group, there was no significant difference in performance between the CI+HA and CI-alone conditions for consonant (paired-$t(6) = 1.05$, $p > 0.05$) and vowel (paired-$t(5) = 1.76$, $p > 0.05$) recognition. Individually, none of the subjects showed bimodal benefit for consonant recognition, and subject C5 showed a significant decrease in performance in the CI+HA condition compared to CI-alone ($t(16) = 3.95$, $p \leq 0.005$). For vowel recognition, three of the six subjects tested showed significant bimodal benefit in which CI+HA performance was greater compared to the performance in the better ear (C2: $t(16) = 3.69$, $p \leq 0.005$; C4: $t(16) = 2.17$, $p \leq 0.05$; C6: $t(16) = 2.62$, $p \leq 0.05$). For subject C7, the CI+HA performance was not significantly different than the HA-alone performance for both consonant and vowel recognition. Correlational analyses did not reveal a significant ($p > 0.05$) relationship between the amount of residual hearing in the non-implanted ear and the overall percent correct scores for consonant ($n = 7$) and vowel ($n = 6$) recognition in the HA-alone condition. Additionally, there was a lack of relationship between residual hearing and the amount of bimodal benefit.

Percent information transmission for the features of voicing, manner, and place for consonant recognition, and for the features of height (related to F1) and back (related to F2)

for vowel recognition was computed for each subject and for each listening condition (see Fig. 6). For consonant recognition, manner of articulation was perceived more accurately than voicing and place of articulation with electric stimulation. In general, there was no considerable difference between the CI+HA condition and the better ear condition, except for subject C4 who showed a 17 percentage point improvement with EAS compared to CI alone for the perception of the manner cue, and subject C7 who showed a 10 point improvement with EAS compared to HA alone for the perception of the place cue. For vowel recognition, the HA provided greater vowel height information compared to vowel back information. The pattern of results was reversed for the CI alone condition. The three subjects (C2, C4, C6) who had a significant bimodal benefit in the overall percent correct score also showed an improvement in the perception of both vowel height (3 – 10 points) and vowel back (4 – 9 points) with EAS compared to CI alone.

Phoneme Recognition in Bilateral CI Listeners—Three of the five subjects participated in the phoneme recognition task in the bilateral CI group. Overall percent correct consonant and vowel recognition scores for each listening condition for individual subjects and the group are shown in Fig. 7. Among these three subjects, none of them showed significant bilateral benefit (i.e., higher performance with bilateral CI compared to the better CI) for consonant recognition, and only one subject (C10) showed significant combined benefit for vowel recognition ($t(16) = 2.13$, $p \leq 0.05$). Subject C2 was tested before and after her second implant. As a bimodal CI user, she showed significant bimodal benefit for vowel recognition, but bilateral benefit was not evident in this subject. This could be due to the ceiling effect for vowel recognition in quiet with a single CI. While vowel recognition performance was similar for the first CI in the bimodal (77%) and bilateral (80%) conditions, performance for the second CI approached ceiling with 88% correct.

Percent information transmission for each consonant and vowel feature is shown in Fig. 8. For consonant recognition, manner of articulation was perceived more accurately than voicing and place of articulation with electric stimulation, similar to the pattern of results found in the bimodal CI group. While subject C9 showed an increase (about 6 points) in the perception of voicing and place features with bilateral CI compared to the better CI, other subjects did not show improved perception of any features. For vowel recognition, the feature of vowel back was perceived more accurately than vowel height with electric stimulation. Subject C10 who had significant bilateral benefit in the overall percent correct score also showed improvement in perception of both vowel height (6 points) and vowel back (10 points) with bilateral CI compared to the better CI.

Relationship between Timbre and Speech Perception

Phoneme Recognition and Acoustical Correlates—Correlational analyses were performed between the overall consonant recognition scores and the correlation values of impulsiveness with Dim 1 (temporal dimension), and between the overall vowel recognition scores and correlation values of spectral centroid with Dim 2 (spectral dimension) for the CI-alone (including the 1st CI and 2nd CI alone in the bilateral CI group) and HA-alone conditions. No significant correlation ($p > 0.05$) was found between consonant recognition and correlation values of impulsiveness with Dim 1 (Fig. 9, left panel) ($n = 13$ for CI-alone; $n = 7$ for HA-alone), nor between vowel recognition and correlation values of spectral centroid with Dim 2 (Fig. 9, right panel) ($n = 12$ for CI-alone; $n = 6$ for HA-alone) for both the CI-alone and HA-alone conditions. The finding that there is a lack of significant correlation between phoneme recognition and the correlation values in the CI-alone condition was consistent with that reported in Kong et al. (2011). Furthermore, no significant correlation ($p > 0.05$) was found between consonant recognition and correlation

values of spectral centroid with Dim 2, or between vowel recognition and correlation values of impulsiveness with Dim 1 for both the CI-alone and HA-alone conditions.

Combined Benefit for Phoneme and Timbre Perception—Although there was no significant correlation between overall phoneme recognition scores and the correlation values between perceptual dimensions and the acoustic features, there appears to be a consistent pattern regarding the combined benefit (bimodal and bilateral benefit) between timbre perception and vowel recognition. Three (C2, C4, C6) of the six bimodal CI subjects and one (C10) of the three bilateral CI subjects showed significant combined benefit for vowel recognition. Improved perception of both F1 and F2 was also evident in these subjects. Except for C2 in the bimodal group, these subjects also showed a considerable increase in correlation between spectral centroid and Dim 2 with the combined use of devices compared to the single device conditions. The remaining subjects (C3, C5, C7 in the bimodal group; C2, C9 in the bilateral group) who did not demonstrate significant combined benefit for vowel recognition also did not show a considerable increase in correlation between spectral centroid and Dim 2 with the combined use of devices compared to the single device conditions. It is noted that subject C7 showed a stronger correlation between spectral centroid and Dim 2 for the HA-alone condition compared to the CI-alone condition. This subject also demonstrated better vowel recognition performance for the HA-alone condition than for the CI-alone condition. Using binominal probability calculation, the probability of getting eight matched cases between timbre and vowel perception with regard to combined benefit out of the total of nine cases by chance is 1.8%.

Unlike vowel recognition, there was a lack of a consistent pattern regarding the combined benefit between timbre perception and consonant recognition. For example, subject C6 showed a considerable increase in correlation (enhanced correlation coefficient by 0.1 or greater) between impulsiveness and Dim 1 with the combined use of devices compared to the single device conditions, but this subject did not show a significant combined benefit for the overall consonant recognition score, or an improved perception of manner of articulation. Subjects C1 and C9 showed a considerable decrease in correlation between impulsiveness and Dim 1 with the combined use of devices compared to the better single device conditions, but none of these subjects showed a significant decrease in consonant recognition.

DISCUSSION

Timbre Perception with a Single CI

The present results on the CI-alone condition with different groups of CI listeners are in agreement with Kong et al.'s (2011) findings, as well as results from previous studies which showed a reduced musical instrument identification ability in unilateral CI listeners compared to NH listeners (Gfeller et al. 1998; 2002; McDermott 2004; Nimmons et al. 2008; Kang et al. 2009). Kong et al. (2011) showed that the first dimension in the 2D timbre space in unilateral CI listeners was a temporal dimension, in which it correlated strongly with the temporal envelope characteristics (log-attack time and impulsiveness) of the stimuli, and the second dimension was a spectral dimension which correlated significantly, but weakly with the spectral envelope characteristics (spectral centroid) of the stimuli. McDermott (2004) and Looi et al. (2008) examined the confusion matrix from a group of unilateral CI listeners tested on a musical instrument identification task and found that more confusions were made among the instruments within the same percussion category or with similar temporal envelope characteristics (i.e., percussive or non-percussive instrument) than between categories. These findings suggest that temporal envelope cues are more salient and reliable cues than spectral cues for musical timbre perception for listeners who listened with

a single CI. As suggested by Kong et al. (2011), the weak correlation between spectral centroid and the second timbre dimension could be attributed to a combination of factors which include: (1) CI users may not be able to detect small perceptual differences between instruments due to reduced frequency resolution, which results from channel interaction and the temporal-envelope-based coding strategies used in the speech processor; and (2) the relative importance of spectral envelope cues for timbre perception is reduced in CI users. For example, for subjects C5 and C10, the Dim2/Dim1 perceptual weight ratios were less than 0.25 for the CI-alone condition.

Timbre Perception with Low-Frequency Residual Acoustic Hearing

Similar to electrical stimulation, listeners with low-frequency acoustic residual hearing (HA-alone condition) also relied mainly on the temporal envelope cues to perceive differences in musical timbre. For the bimodal hearing group, there was no significant difference between the HA-alone and the CI-alone condition with regard to the correlation between spectral centroid and Dim 2 (paired-$t(6) = 1.22$, $p > 0.05$), suggesting that similar to electrical stimulation, the spectral cues were less reliable or salient with low-frequency acoustic stimulation. Listeners with severe high-frequency hearing loss may not be able to distinguish small differences of spectral centroid among the instruments. The decrease in Dim2/Dim1 weight ratio in the HA-alone condition compared to the CI-alone and the CI+HA condition may also suggest that perceptual salience for spectral envelope cues was reduced in the HA-alone condition. These findings are in agreement with previous reports on the similar performance between electrical stimulation and low-frequency acoustic stimulation on musical instrument identification and sound quality rating tasks (Looi et al. 2007; 2008). Looi et al. (2008) compared instrument identification ability in three groups of listeners: (1) control group – unilateral CI; (2) control group – HA users who had moderately-severe to profound hearing loss and poor speech recognition scores which met the current CI candidacy requirements; and (3) experimental group – hearing-impaired listeners who were on the waiting list to receive a CI. This waitlist group was tested pre- and post-implantation. The average unaided pure-tone thresholds for the waitlist group were similar to the average unaided thresholds in the bimodal CI group in this study. Looi et al. (2008) found that (1) instrument identification performance was similar between the control unilateral CI group and the control HA group; and (2) instrument identification improved in the post-CI testing (with CI-alone) compared to the pre-CI testing (with HA-alone) in the waitlist group. However, they argued that this difference could be attributed to the learning effect. They reported that instrument identification scores were higher on the second test block than on the first test block in the unilateral CI and HA control groups.

In this study, there was a lack of relationship between the amount of low-frequency residual hearing and the correlation values between acoustic features and their corresponding dimensions. This could be due to a number of factors including: (1) the small sample size in the current study ($n = 7$); (2) the limited range of PTAs in the subject group (five out of the seven subjects were in the 80 – 100 dB range); and (3) the reduced perceptual weight on the spectral dimension. In regard to the relative salience of acoustic cues, an individual who has greater residual hearing could potentially be more sensitive to spectral changes; but at the same time, this individual could also place a lower perceptual weight on spectral cues and a much higher weight on temporal cues for timbre perception when multiple relevant cues are available.

Timbre Perception and Combined Benefit

As a group, combined benefit for timbre perception was not evident for the bimodal and bilateral CI listeners in the present study. The correlations between each dimension and its corresponding acoustic features were similar for the combined use of devices and the better

single device condition. This finding is consistent with Sucher and McDermott (2009) which showed no significant difference in sound quality rating between CI+HA and the better ear (i.e., CI-alone) condition, but it is not in agreement with Gfeller et al. (2006), which showed significantly better instrument identification performance in listeners with hybrid hearing (i.e., electric and acoustic stimulation in the same ear) than listeners with a standard long-electrode implant. The differences in results between the present study and Gfeller et al.'s could be attributed to a number of factors including, (1) differences in the amount of residual hearing between the hybrid listeners in Gfeller et al. and the bimodal listeners in this study. Although unaided thresholds were not reported in Gfeller et al. (2006), it is generally the case that hybrid users have a considerably greater amount of low-frequency residual hearing compared to bimodal users; (2) differences in test stimuli; and (3) differences in experimental design and tasks (dissimilarity judgment in this study as opposed to the identification task in Gfeller et al.).

Although there was no bimodal benefit in the group data, there were individual differences in the combined benefit for both groups. Three out of the seven bimodal subjects showed a considerably enhanced correlation between spectral centroid and Dim 2 for the CI+HA condition compared to the better ear condition. For the bilateral CI group, one subject (C10) showed a considerably improved correlation between spectral centroid and Dim 2 in the bilateral condition compared to the better CI condition. To understand other factors (e.g., perceptual weighting of cues) that might contribute to the combined benefit in these subjects, the effect of change of perceptual weight (i.e., Dim2/Dim1 weight ratio) between the combined and better ear conditions on spectral centroid perception was examined. No apparent relationship between change of perceptual weights and enhanced correlation values between spectral centroid and Dim 2 was observed. This may suggest that the improved timbre perception with the combined use of devices cannot be attributed to the fact that listeners placed higher perceptual weights on the spectral envelope cues in the combined condition compared to the single device conditions. Interestingly, C10 was implanted with different devices and used different processors between the two ears (5-electrode Ineraid device in one ear and 16-electrode Advanced Bionics HiRes 90K device in the other). Anecdotal reports by this subject indicated that for the same input stimuli (e.g., a musical note), he perceived the sound as having a higher pitch on one side compared to the other, suggesting this subject received very different information from the two devices. Taken together, the results from both the bimodal and bilateral CI groups, as well as findings from previous studies, it is possible that musical timbre perception could be enhanced when listeners receive different spectral cues from individual devices.

Dichotomy between Timbre and Pitch Perception

The most interesting and striking findings in the present study are (1) that there were differences in the patterns of results between timbre perception and pitch perception with regard to the combined benefit, and (2) that there was a close relationship between speech and timbre perception with regard to the combined benefit. Previously, studies showed the dichotomy of speech and pitch perception, in which temporal envelope cues are sufficient for speech recognition in quiet; temporal and spectral fine structure cues, however, are critical for pitch perception (e.g., Smith et al. 2002; Kong and Zeng 2006). The dichotomy of speech and pitch perception for EAS benefit was first reported by Kong et al. (2005). They found that (1) low-frequency residual acoustic hearing yielded better melody recognition than electric hearing, but the pattern of results was reversed for speech recognition; and (2) combined EAS benefit was found for speech recognition, but not for melody recognition, in which melody recognition performance in the EAS condition was determined by the performance of the low-frequency residual hearing alone. Similar to

speech perception, the finding of combined benefit for timbre perception in this study suggests that there is also a dichotomy between timbre and pitch perception.

Previous studies showed a relationship between speech perception and timbre perception for CI listeners (e.g., Gfeller et al. 2008; Nimmons et al. 2008; Kang et al. 2009; Won et al. 2010), in which the speech recognition and instrument identification scores were significantly correlated. In this study, a significant correlation between the overall speech recognition scores and the correlation values between perceptual dimensions and acoustic features with CI-alone or HA-alone was not observed. The discrepancy between the current study and previous studies (e.g., Nimmons et al. and Kang et al.) could be due to differences in the stimuli and experimental design. Previous studies used discrimination tasks while this study assessed the dissimilarities between musical instrument sounds. However, the patterns of results in timbre perception with EAS were similar to those observed for speech recognition. First, stronger correlations between perceptual dimensions and acoustic features were found in many subjects and in the group data with CI-alone compared to HA-alone. This is consistent with the phoneme recognition results, in which almost all subjects (except for C7) achieved better phoneme recognition scores with CI-alone compared to HA-alone. Second, unlike pitch perception in which there was a lack of combined benefit, almost half of the EAS listeners in this study showed a considerably enhanced correlation between spectral centroid and Dim 2 in the CI+HA condition compared to the better single device condition. The relationship between speech and timbre was further supported by the consistent patterns of results with regard to the combined (bimodal or bilateral) benefit for vowel and timbre perception in almost all subjects (with only one exception – C2 in the bimodal group). Spectral envelope cues are important for both vowel recognition and timbre perception (McAdams et al. 1995; Marozeau et al. 2003; Caclin et al. 2005; Kong et al. 2011). Listeners from both the bimodal and bilateral CI groups who demonstrated significant combined benefit for vowel recognition (including F1 and F2 perception) also showed a considerably stronger correlation between spectral centroid and Dim 2 in the combined condition. At the same time, listeners who did not demonstrate combined benefit for vowel recognition also did not show a stronger correlation in the combined condition.

The current dominant view for vowel perception is that the frequencies of the first two or three lowest formants are the primary cues. Spectral centroid measure is one of the measures of spectral envelope. The majority of the stimuli used in this investigation have a single peak in the spectrum and low energy for high-frequency harmonics. As shown in Fig. 2, the two spectra that represent the general spectral shape of the stimulus set resemble the formant peak (spectral envelope feature) for vowels. The values of spectral centroid in the stimulus set are between 800 and 2000 Hz, coinciding with the F1 and F2 frequencies in many static vowel stimuli used in this study (F1 range: 300 to 1000 Hz; F2 range: 900 to 3000 Hz). The finding of a close relationship between the spectral centroid cue for timbre perception and vowel recognition in the context of combined benefit in this study is also supported by an alternative model of static vowel perception, known as the spectral shape model (e.g., Bladon & Lindblom 1981; Ito et al. 2001). This model considers that vowels can be correctly identified based on global spectral characteristics rather than individual formants; thus, correct vowel recognition relies on the integration of broadband spectral information. A recent study by Fox et al. (2011) demonstrated the important role of spectral centroid in both F1 and F2 regions for vowel perception. Fox et al. replaced one of the formants in a two-formant stimulus with two pairs of sine waves. The intensities of these sine waves were modified to cause variation in the spectral centroid of the sine waves. They reported that the identification of the vowel was highly influenced by the change of the spectral centroid.

The current findings on the similar patterns of results between timbre and speech perception with regard to the combined benefit have significant scientific and clinical implications,

including: (1) listeners can combine spectral envelope cues from different sources (i.e., across ears in bimodal listeners) to enhance perception for speech and non-speech signals; (2) the reduced benefit for phoneme recognition observed in some bimodal listeners might not be attributable to the imperfect internal representation of the phonemes, particularly for individuals with a long duration of hearing loss, as suggested by Kong & Braida (2011); and (3) benefits from perceptual training regimens that utilize non-speech stimuli could generalize to speech recognition.

CONCLUSIONS

The present study investigated the contributions of temporal and spectral envelope cues for timbre perception and the relationship between timbre and speech perception in bimodal and bilateral CI users. The main findings are as follows:

1. Bimodal and bilateral CI users relied primarily on the temporal envelope cue (impulsiveness) to perceive differences in timbre, and the spectral envelope cue (spectral centroid) was a less reliable or salient cue for timbre perception.
2. Combined benefit for timbre perception was found in some bimodal CI users and one bilateral CI user, in which the correlation between spectral centroid and one of the perceptual dimensions was considerably strengthened. This suggests that, unlike pitch perception, the combined use of devices could potentially improve the ability of these listeners to use the spectral envelope cues to perceive timbre.
3. There was a close relationship between timbre perception and vowel recognition with regard to combined (bimodal or bilateral) benefit. That is, individuals who demonstrated a significant combined benefit for vowel recognition also showed an enhanced correlation value between spectral centroid and Dim 2.

Acknowledgments

We are grateful to all the listeners for their participation in the experiment. We would like to thank Professor Stephen McAdams for providing the stimuli. This work was supported by NIH-NIDCD (R03 DC009684-03) and Northeastern University Provost Faculty Development Research Grant to the first author (YYK).

References

ANSI. American National Standard – Psychoacoustical Terminology S3.20. New York: American National Standards Institute; 1973.

Bladon RAW, Lindblom B. Modeling the judgment of vowel quality differences. J Acoust Soc Am. 1981; 69:1414–1422. [PubMed: 7240572]

Borg, I.; Groenen, PJF. Modern Multidimensional Scaling: Theory and Applications. New York: Springer, New York; 1997.

Caclin A, McAdams S, Smith BK, et al. Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. J Acoust Soc Am. 2005; 118:471–482. [PubMed: 16119366]

Carroll JD, Chang JJ. Analysis of indifferences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. Psychometrika. 1970; 35:283–319.

Dorman MF, Gifford RH. Combining acoustic and electric stimulation in the service of speech recognition. Int J Audiol. 2010; 49:912–919. [PubMed: 20874053]

Dorman MF, Gifford RH, Spahr AJ, et al. The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. Audiol Neurootol. 2008; 13:105–112. [PubMed: 18057874]

Fox RA, Jacewicz E, Chang C-Y. Auditory spectral integration in the perception of static vowels. J Speech Lang Hear Res. 2011 Epub ahead of print.

Gfeller K, Lansing CR. Melodic, rhythmic, and timbral perception of adult cochlear implant users. J Speech Hear Res. 1991; 34:916–920. [PubMed: 1956198]

Gfeller K, Christ A, Knutson JF, et al. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. J Am Acad Audiol. 2000; 11:390–406. [PubMed: 10976500]

Gfeller K, Knutson JF, Woodworth G, et al. Timbral recognition and appraisal by adult cochlear implant users and normal-hearing adults. J Am Acad Audiol. 1998; 9:1–19. [PubMed: 9493937]

Gfeller K, Oleson J, Knutson JF, et al. Multivariate predictors of music perception and appraisal by adult cochlear implant users. J Am Acad Audiol. 2008; 19:120–134. [PubMed: 18669126]

Gfeller KE, Olszewski C, Turner C, et al. Music perception with cochlear implants and residual hearing. Audiol Neurotol. 2006; 11:12–15.

Gfeller K, Witt S, Adamek M, et al. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. J Am Acad Audiol. 2002; 13:132–145. [PubMed: 11936169]

Grey JM. Multidimensional perceptual scaling of musical timbres. J Acoust Soc Am. 1977; 61:1270–1277. [PubMed: 560400]

Ito M, Tsuchida J, Yano M. On the effectiveness of whole spectral shape for vowel perception. J Acoust Soc Am. 2001; 110:1141–1149. [PubMed: 11519581]

Kang R, Nimmons GL, Drennan W, et al. Development and validation of the University of Washington Clinical Assessment of Music Perception test. Ear Hear. 2009; 30:411–418. [PubMed: 19474735]

Kendall RA, Carterette EC. Verbal attributes of simultaneous wind instruments timbres: I. von Bismarck's adjectives. Mus Percept. 1993; 10:445–468.

Kong Y-Y, Braida LD. Cross-frequency integration for consonant and vowel identification in bimodal hearing. J Speech Lang Hear Res. 2011; 54:959–980. [PubMed: 21060139]

Kong Y-Y, Stickney GS, Zeng F-G. Speech and melody recognition in binaurally combined acoustic and electric hearing. J Acoust Soc Am. 2005; 117:1351–1361. [PubMed: 15807023]

Kong Y-Y, Mullangi A, Marozeau J, et al. Temporal and spectral cues for musical timbre perception in electric hearing. J Speech Lang Hear Res. 2011; 54:981–994. [PubMed: 21060140]

Kong Y-Y, Zeng F-G. Temporal and spectral cues in Mandarin tone recognition. J Acoust Soc Am. 2006; 120:2830–2840. [PubMed: 17139741]

Krimphoff J, McAdams S, Winsberg S, et al. Caracterisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique. Journal de Physique. 1994; 4:625–628.

Krumhansl, CL. Why is musical timbre so hard to understand?. In: Nielzen, S.; Olsson, O., editors. Structure and perception of electroacoustic sound and music. Amsterdam, the Netherlands: Elsevier Scientific; 1989. p. 43-53.

Looi V, McDermott H, McKay C, et al. Comparisons of quality ratings for music by cochlear implant and hearing aid users. Ear Hear. 2007; 28:59S–61S. [PubMed: 17496649]

Looi V, McDermott H, McKay C, et al. The effect of cochlear implantation on music perception by adults with usable pre-operative acoustic hearing. Int J Audiol. 2008; 47:257–268. [PubMed: 18465410]

McAdams S, Winsberg S, Donnadieu S, et al. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychol Res. 1995; 58:177–192. [PubMed: 8570786]

McCallum R. Effects of conditionality on indscal and alscal weights. Psychometrika. 1977; 42:297–305.

McDermott HJ. Music perception with cochlear implant: a review. Trends Amplif. 2004; 8:49–82. [PubMed: 15497033]

Marozeau J, de Cheveigne A, McAdams S, et al. The dependency of timbre on fundamental frequency. J Acoust Soc Am. 2003; 114:2946–2957. [PubMed: 14650028]

Mok M, Grayden D, Dowell RC, et al. Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears. J Speech Lang Hear Res. 2006; 49:338–351. [PubMed: 16671848]

Mok M, Galvin KL, Dowell RC, et al. Speech perception benefit for children with a cochlear implant and a hearing aid in opposite ears and children with bilateral cochlear implants. Audiol Neurotol. 2010; 15:44–56.

Nimmons GL, Kang RS, Drennan WR, et al. Clinical assessment of music perception in cochlear implant listeners. Otol Neurotol. 2008; 29:149–155. [PubMed: 18309572]

Peeters, G. CUIDADO I S T Project Report. Paris, France: Institut de Recherche et Coordination Acoustique/Musique (Ircam); 2004. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Retrieved from Ircam website: http://recherche.ircam.fr/equipes/analyse-synthese/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf

Pratt RL, Doak PE. A subjective rating scale for timbre. J Sound Vibration. 1976; 43:317–328.

Schulz, E.; Kerber, M. Music Perception with the Med-El Implants. In: Hochmair-Desoyer, IJ.; Hochmair, ES., editors. Advances in Cochlear Implants. Vienna: Manz; 1994.

Sucher CM, McDermott HJ. Bimodal stimulation: benefits for music perception and sound quality. Cochlear Implants Int. 2009; 10(S1):96–99. [PubMed: 19230032]

Shannon RV, Jensvold A, Padilla M, et al. Consonant recordings for speech testing. J Acoust Soc Am. 1999; 106:L71–74. [PubMed: 10615713]

Shannon RV, Zeng F-G, Kamath V, et al. Speech recognition with primarily temporal cues. Science. 1995; 270:303–304. [PubMed: 7569981]

Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. Nature. 2002; 416:87–90. [PubMed: 11882898]

von Bismarck G. Sharpness as an attribute of the timbre of steady sounds. Acustica. 1974; 30:159–172.

van Tasell DJ, Greenfield DG, Logemann JJ, et al. Temporal cues for consonant recognition: training, talker generalization, and use in evaluation of cochlea implants. J Acoust Soc Am. 1992; 92:1247–1257. [PubMed: 1401513]

van Tasell DJ, Soli SD, Kirby VM, et al. Speech waveform envelope cues for consonant recognition. J Acoust Soc Am. 1987; 82:1152–1161. [PubMed: 3680774]

Wessel D. Timbre space as a musical control structure. Comput Mus J. 1979; 3:45–52.

Won JH, Drennan WR, Kang RS, et al. Psychoacoustic abilities associated with music perception in cochlear implant users. Ear Hear. 2010; 31:796–805. [PubMed: 20595901]

Yoon Y-S, Liu A, Fu QJ. Binaural benefit for speech recognition with spectral mismatch across ears in simulated electric hearing. J Acoust Soc Am. 2011a; 130:EL94–100. [PubMed: 21877777]

Yoon Y-S, Li Y, Kang H-Y, et al. The relationship between binaural benefit and difference in unilateral speech recognition performance for bilateral cochlear implant users. Int J Audiol. 2011b; 50:554–565. [PubMed: 21696329]

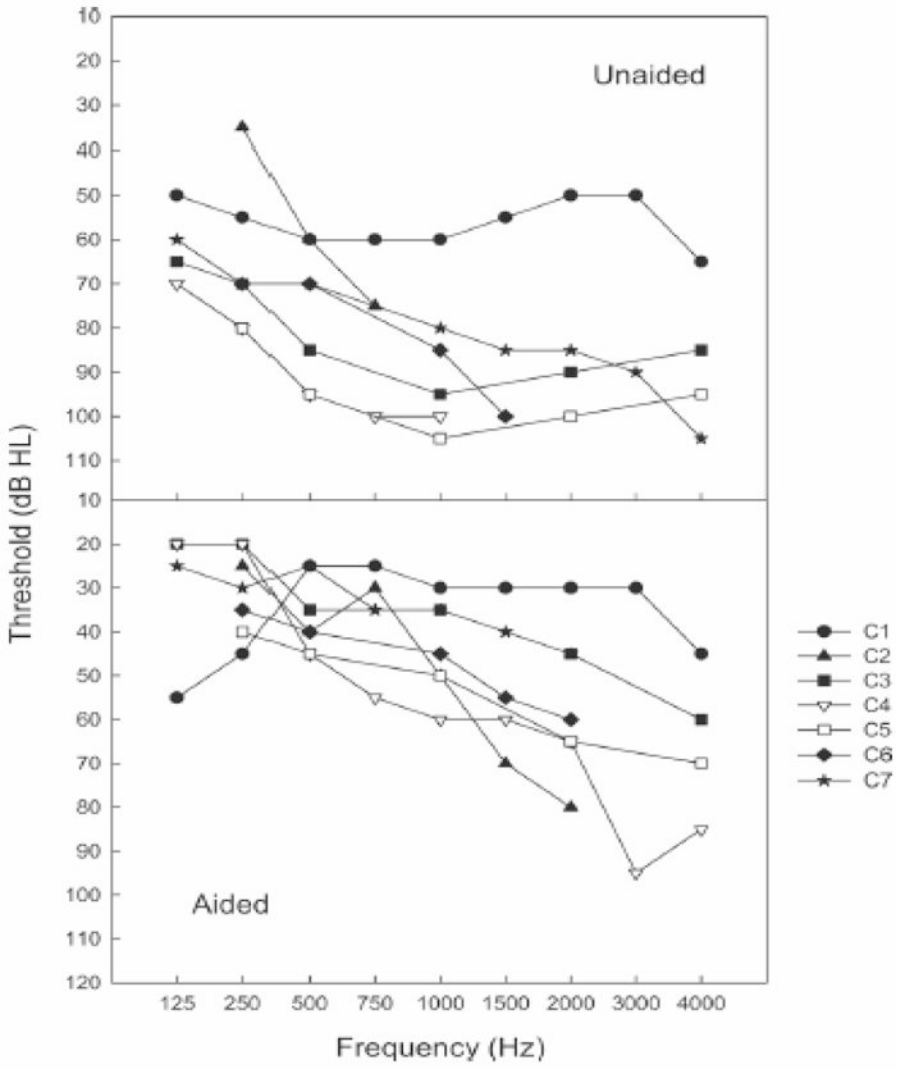

Figure 1.
Unaided (top) and aided (bottom) thresholds of the seven bimodal CI users.

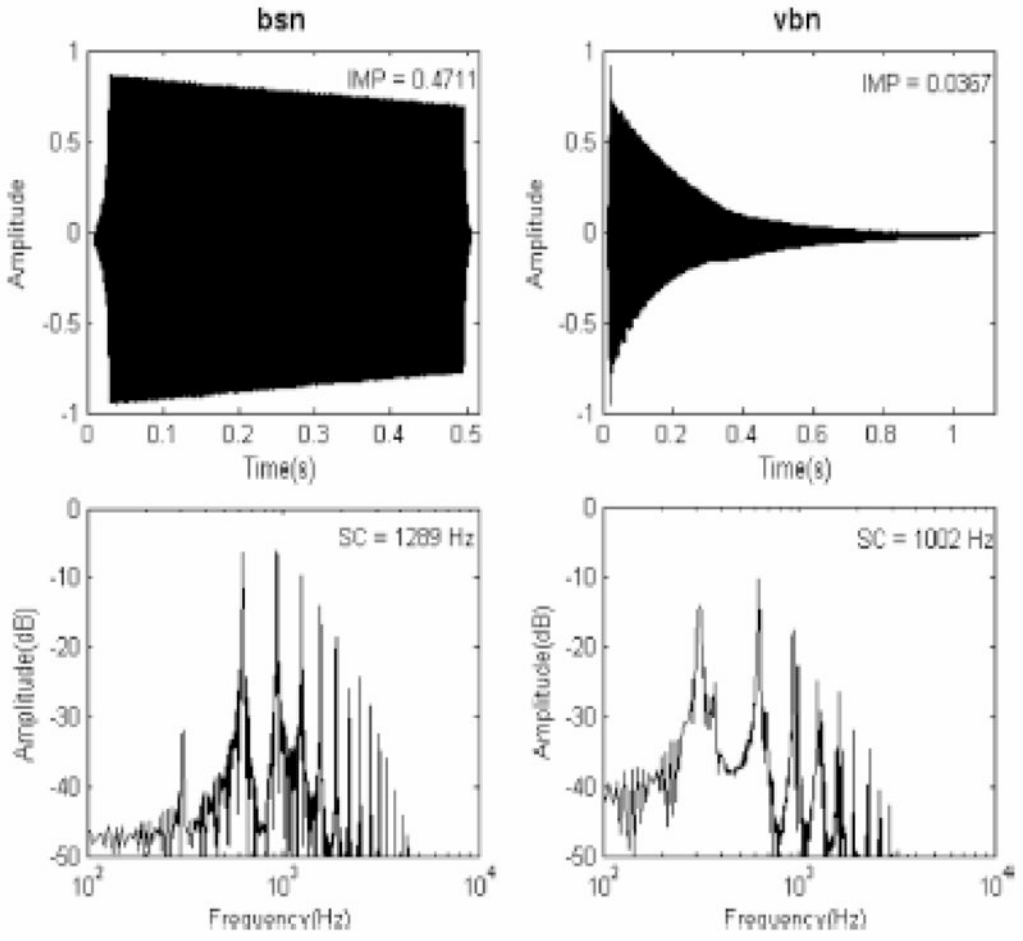

Figure 2.
Time waveforms (top) and spectra (bottom) of two musical stimuli – bsn (left) and vbn (right). The impulsiveness (IMP) and spectral centroid (SC) values of each stimulus are shown in the upper right corner in each panel.

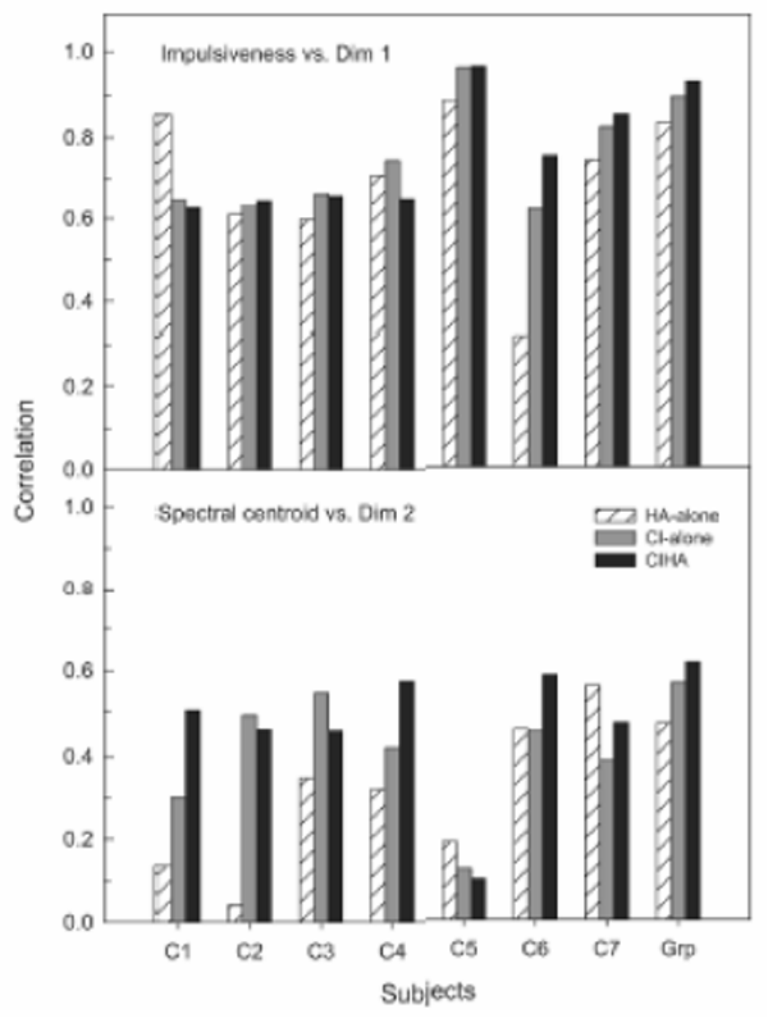

Figure 3.
Correlations between impulsiveness and Dim1 (top) and between spectral centroid and Dim 2 (bottom) for individual subject and group data for each listening condition in the bimodal CI group.

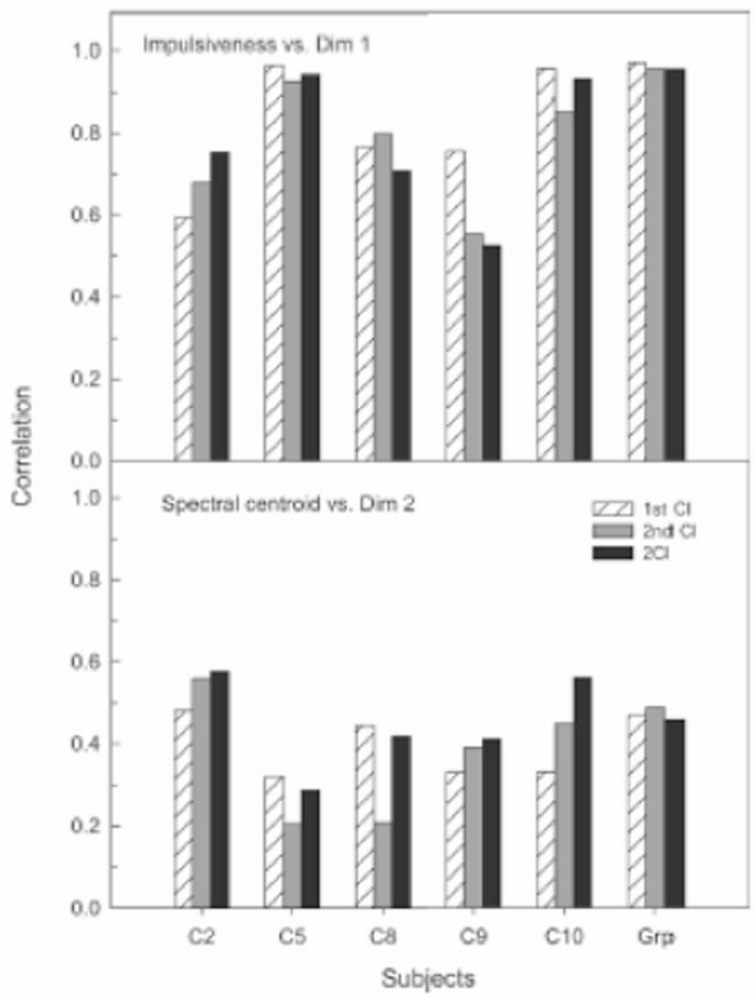

Figure 4.
Correlations between impulsiveness and Dim1 (top) and between spectral centroid and Dim 2 (bottom) for individual subject and group data for each listening condition in the bilateral CI group.

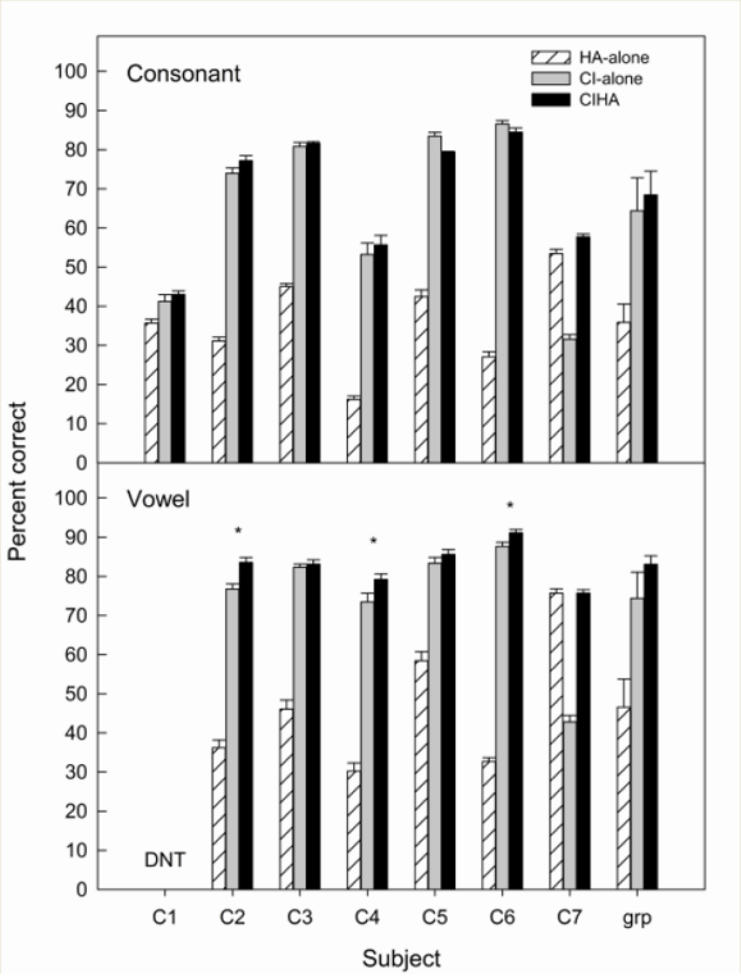

Figure 5.

Overall percent correct consonant (top) and vowel (bottom) recognition scores for individual bimodal subject and group data for each listening condition. Error bars represent standard error of the mean. Significant combined benefit is marked with an asterisk (*). Subject C1 was not tested on vowel recognition. (data extracted from Kong and Braida, 2011 with permission)

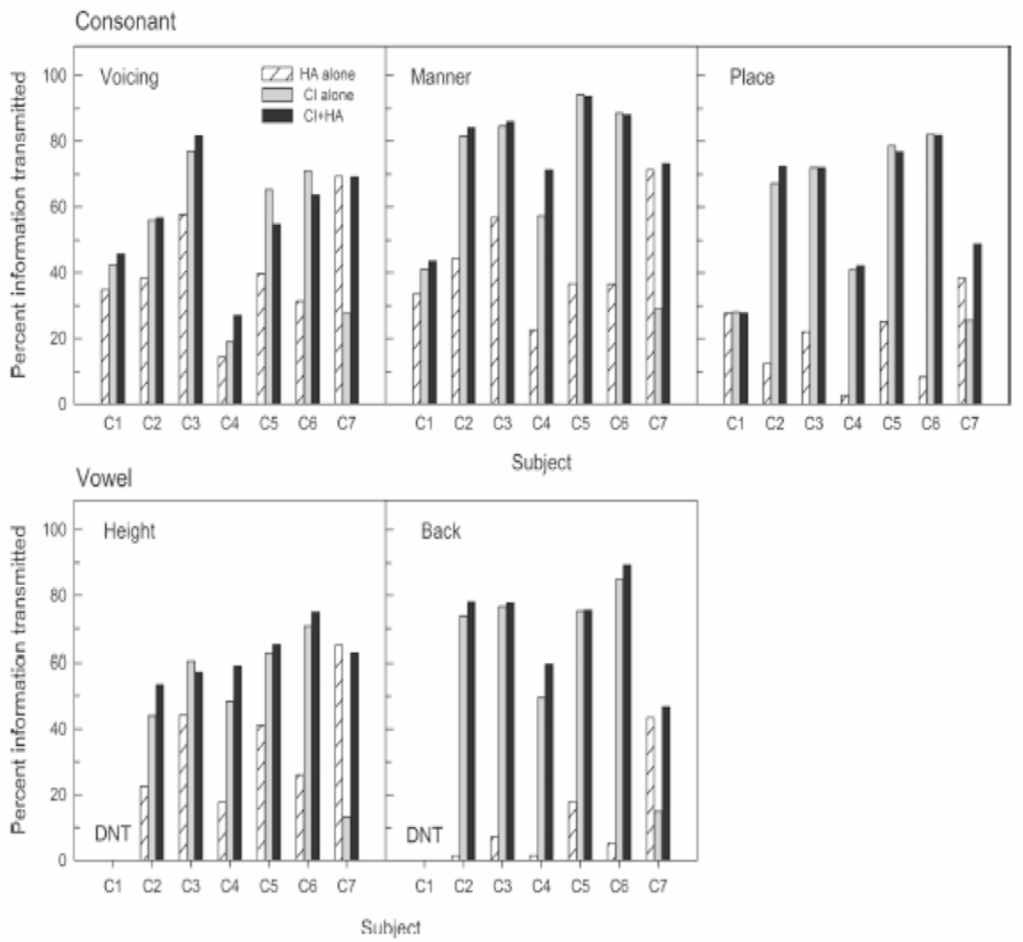

Figure 6.

Percent information transmission for consonant features – voicing, manner, place (top), and vowel features – height and back (bottom) for individual bimodal subject for each listening condition.

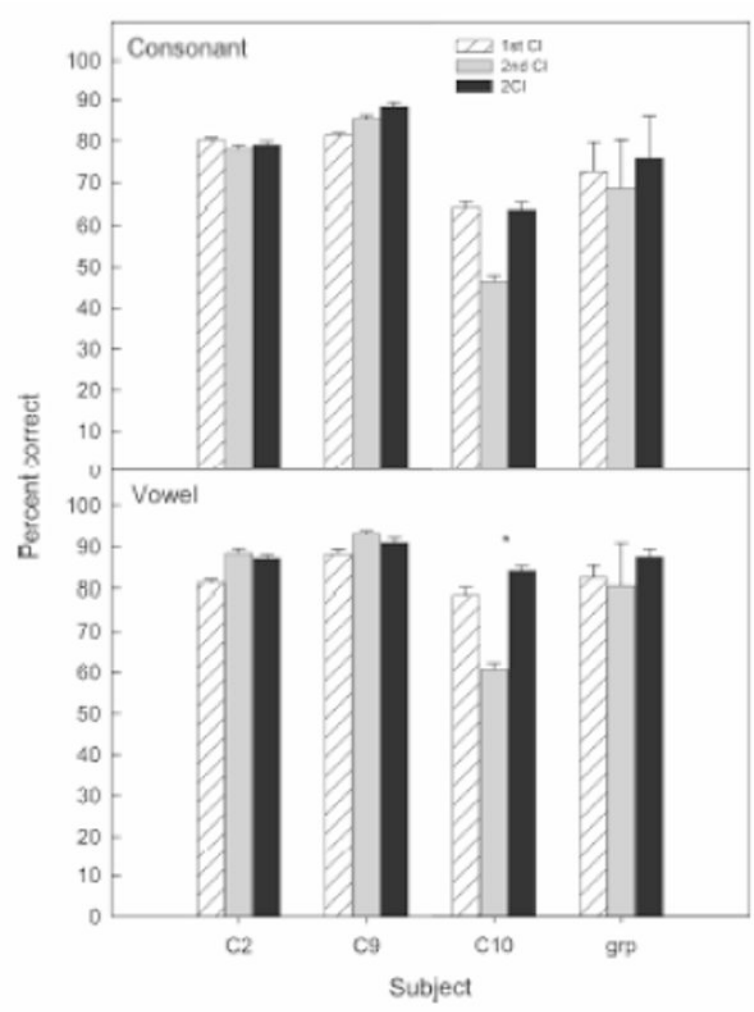

Figure 7.
Overall percent correct consonant (top) and vowel (bottom) recognition scores for each of the three bilateral subjects tested and the group data for each listening condition. Error bars represent standard error of the mean. Significant combined benefit is marked with an asterisk (*).

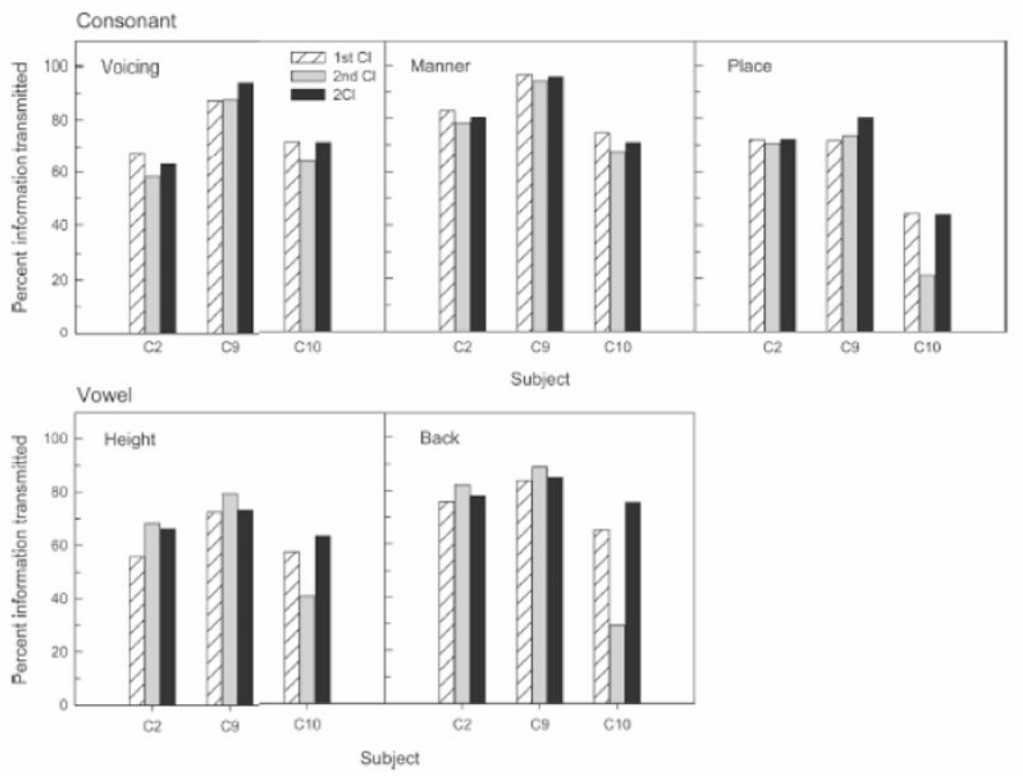

Figure 8.
Percent information transmission for consonant features – voicing, manner, place (top), and vowel features – height and back (bottom) for each of the three bilateral subjects for each listening condition.

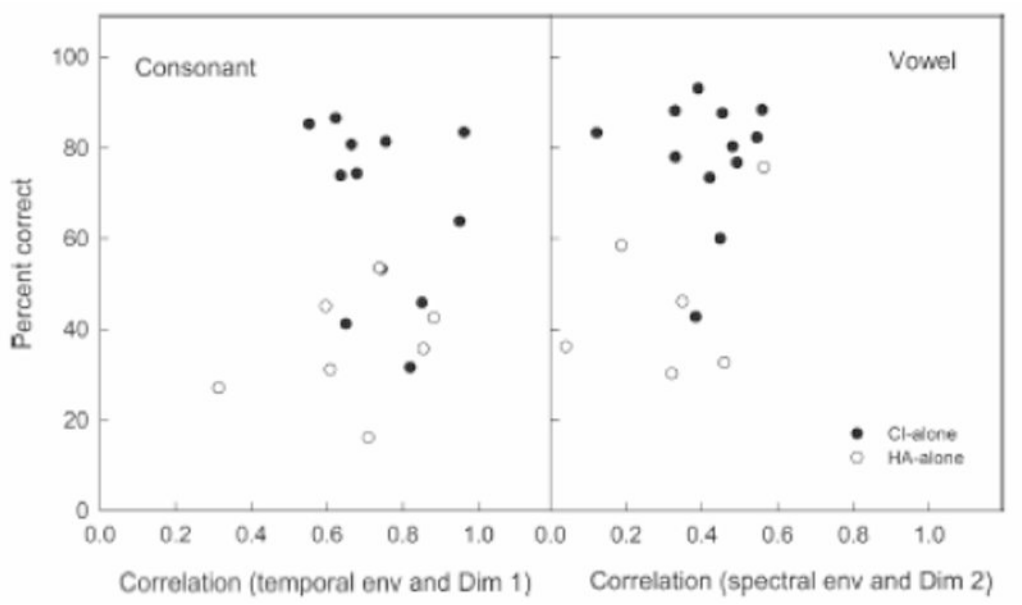

Figure 9.
Relationship between overall percent correct consonant (left) and vowel (right) recognition scores and correlation values between the perceptual dimensions and their corresponding acoustic features (impulsiveness vs. Dim 1 and spectral centroid vs. Dim 2).

Table I

Bimodal CI subjects' detailed demographic information

Subject	Age	Gender	Etiology (CI ear)	CI Ear	Onset HL (CI ear)	Yrs severe HL prior CI	CI Processor	Yrs of CI use	PTA*
C1	46	M	Unknown	L	mid-20s	5	Harmony	1.5	56.25
C2	64	F	Unknown	L	37	10	Freedom	2	83.75
C3	26	M	Genetic	R	Birth	21	Freedom	4.5	85
C4	57	M	Premature birth	L	5	6	Harmony	5	98.75
C5	16	F	Mondini	R	Birth	3	Freedom	1	95
C6	19	F	Unknown	L	Birth	9	ESPrit 3G	10	86.25
C7	65	F	Meningitis	R	2	58	Harmony	5.5	76.25

*Pure-tone average of unaided thresholds across 250, 500, 1000, and 2000 Hz.

Table II

Bilateral CI subjects' detailed demographic information

Subject	Age*	Gender	Etiology	Ear	Onset HL	1st CI Yrs of severe HL	Processor	Yrs of CI use**	Ear	Onset HL	2nd CI Yrs of severe HL	Processor	Yrs of CI use
C2	66	F	Unknown	L	37	10	Nucleus 5	4	R	37	13	Nucleus 5	0.5
C5	17	F	Mondini	R	Birth	3	Freedom	2	L	Birth	5	Freedom	0.7
C8	48	F	Unknown	R	5	15	Freedom	2	L	5	16	Freedom	1
C9	15	F	Cogan	L	Birth	10	Freedom	5	R	7	5	Freedom	1
C10	63	M	Unknown	L	5	3	Geneva	23	R	5	26	Harmony	3

*Age at the time of testing.

**Years of CI use at the time of testing