



Published in final edited form as:

J Speech Lang Hear Res. 2012 February ; 55(1): 105–124. doi:10.1044/1092-4388(2011/10-0325).

Speech Recognition and Acoustic Features in Combined Electric and Acoustic Stimulation

Yang-soo Yoon¹, Yongxin Li², and Qian-Jie Fu¹

Yang-soo Yoon: yyoon@hei.org; Yongxin Li: lyxent@hotmail.com; Qian-Jie Fu: qfu@hei.org

¹Communication and Neuroscience Division, House Ear Institute, 2100 W. 3rd St., Los Angeles, CA 90057

²Department of Otolaryngology, Head and Neck Surgery, Beijing TongRen Hospital, Capital Medical University, Key Laboratory of Otolaryngology, Head and Neck Surgery, Ministry of Education of China, Beijing, People's Republic of China 100730

Abstract

Purpose—This study aimed to identify speech information processed by a hearing aid (HA) that is additive to information processed by a cochlear implant (CI) as a function of signal-to-noise ratio (SNR).

Method—Speech recognition was measured with CI alone, HA alone, and CI+HA. Ten participants were separated into two groups; good (aided pure-tone average (PTA) < 55 dB) and poor (aided PTA ≥ 55 dB) at audiometric frequencies ≤ 1 kHz in HA.

Results—Results showed that the good aided PTA group derived a clear bimodal benefit (performance difference between CI+HA and CI alone) for vowel and sentence recognition in noise while the poor aided PTA group received little benefit across speech tests and SNRs. Results also showed that a better aided PTA helped in processing cues embedded in both low and high frequencies; none of these cues were significantly perceived by the poor aided PTA group.

Conclusions—The aided PTA is an important indicator for bimodal advantage in speech perception. The lack of bimodal benefits in the poor group may be attributed to the non-optimal HA fitting. Bimodal listening provides a synergistic effect for cues in both low and high frequency components in speech.

Keywords

speech recognition; electric and acoustic stimulation; acoustic features; signal-to-noise ratio; bimodal benefit

Introduction

It is well known that combining electric speech processing with residual acoustic hearing leads to an improvement in speech understanding in noise. Some cochlear implant (CI) users wear a hearing aid (HA) in the nonimplanted ear (bimodal CI users; CI+HA), and other CI users wear HA in the implanted ear (hybrid CI users). In both cases, patients have the opportunity to benefit from the combination of electric and acoustic stimulation (EAS),

Send correspondence to: Yang-soo Yoon, House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057, yyoon@hei.org. Yongxin Li, Department of Otolaryngology, Head and Neck Surgery, Beijing TongRen Hospital, Capital Medical University, Key Laboratory of Otolaryngology, Head and Neck Surgery, Ministry of Education of China, Beijing People's Republic of China 100730, lyxent@hotmail.com.

leading to an EAS benefit, defined as the difference in speech performance between CI+HA and CI alone. The idea behind EAS is that the loss of sensory cells can be compensated by means of electric stimulation in the mid- to high-frequency range in combination with acoustic stimulation of the remaining low-frequency areas of the cochlear receptors.

Previous bimodal studies showed a clear trend in the improvement of speech intelligibility when acoustic hearing was added to electric hearing. Even though the degree of the bimodal benefit varies with experimental conditions such as test material, subject age, configuration of hearing thresholds for the acoustic side, and presence of a noise masker, generally the bimodal benefit ranges from 8% to 25% relative to the performance for CI alone (Hamzavi et al., 2004; Kong, Stickey & Zeng, 2005; Kiefer et al., 2005; Ching et al., 2006; Gifford et al., 2007; Dorman et al., 2008). The bimodal benefit is observed not only in speech perception but also in the perception of four-syllable numbers (Hamzavi et al., 2004). However, it should be noted that the majority of bimodal users received significant benefit in speech perception, but some portion of the subjects received no significant benefit (Beijen et al., 2008; Kong and Braida, 2010; Gstoettner et al., 2006; James et al., 2006; Kiefer et al., 2005).

Previous EAS studies showed similar amounts of an EAS benefit in both bimodal and hybrid CI users. Kiefer et al. (2005) measured sentence recognition in quiet and noise for hybrid, bimodal, and hybrid plus contralateral HA users. The difference in performance among these combinations was less than a few percentage points. Gantz et al. (2005) and Gifford et al. (2007) also showed similar patterns of results in bimodal and hybrid plus contralateral HA users. These findings suggest that an EAS benefit is not based on the mechanism of binaural hearing but the integration of speech information independently processed by CI and HA processors.

Another interesting finding in EAS studies is that there is a larger EAS benefit in noise than in quiet for both hybrid and bimodal users. Gstoettner et al. (2004) reported a significant EAS benefit (30% to 50%) for sentence recognition in noise for hybrid users and a few percentage point benefit in quiet. Turner et al. (2004) compared the speech reception threshold (SRT), defined as the signal-to-noise ratio (SNR) required for 50% correct performance in sentence recognition, between hybrid and traditional unilateral CI users. A significant EAS advantage (about 9dB SNR) was observed even though speech scores in quiet were similar between two groups. Kiefer et al. (2005) also reported an EAS benefit of 23% for sentence recognition in noise but only 8% in quiet. Even though the amount of residual hearing varied widely across studies, the EAS benefit in quiet for CNC word recognition is only a few percentage points, but the improvement in sentence recognition in noise was more than 10 percentage points (Gifford et al., 2007, Mok et al., 2006). Turner, Gantz, & Reiss (2008) also reported that a significant EAS advantage (4.2 dB SNR in SRT measures) for sentence perception between hybrid and traditional unilateral CI users was observed even though consonant recognition performance in quiet was equivalent. Dorman et al. (2008) also demonstrated that the EAS benefit in sentence recognition was 10 percentage points higher in noise than in quiet for 15 bimodal users.

When the benefit of EAS in speech recognition is addressed, the audiometric threshold of the acoustic ear should be considered as a covariate. The range of pure tone sensitivity that is known to provide a reliable bimodal benefit has been proposed. However, this is not an absolute requirement for a bimodal benefit because of poor correlation between EAS benefit and pure tone sensitivity for the acoustic side. Ching, Incerti, & Hill (2004) found no significant correlations between hearing thresholds at .25, .5, and 1 kHz for the acoustic ear and bimodal performance and the difference between CI+HA and CI alone performance. Gifford et al. (2007) also reported no significant relationship between bimodal benefit and

the slope of hearing thresholds for frequencies of .25 kHz and 1 kHz or .5 kHz and 1 kHz. In addition, no significant correlation was reported between a bimodal advantage for any CNC measure at 10 dB SNR from 13 school-aged children and aided thresholds at .25 and .5 kHz or 1 and 2 kHz (Mok et al., 2010). These results suggest that better sensitivity to low frequency pure tones does not necessarily lead to a greater bimodal advantage in speech perception.

In recent studies of EAS, the focus has been on finding the source of the EAS benefit. Some studies were conducted with real EAS patients (e.g., Kong, Stickney, & Zeng, 2005; Mok et al., 2010; Turner et al., 2004), and other studies were acoustic simulations of CI processing with normal hearing subjects (Chang, Bai, & Zeng, 2006; Brown and Bacon, 2009a; Kong and Carlyon, 2007; Turner et al., 2004). The effect of the fundamental frequency, F0, on speech perception has been primarily evaluated as one of the possible sources of an EAS benefit. Turner et al. (2004) reported that there is a significant EAS benefit in word recognition in a simulation study for a talker background masker but not for a white noise masker; this result was confirmed with three hybrid users (Turner et al., 2004). This result suggests that voice pitch perception, which aids in separating voices in a background of other talkers, contributes to an EAS advantage. Similarly for sentence recognition in noise, the bimodal benefit is largest when the difference in F0 between the target and talker masker was largest (Kong, Stickney, & Zeng, 2005). Two other studies (Chang, Bai, & Zeng, 2006; Qin and Oxenham, 2006) also showed significant improvement in sentence recognition when low-pass filtered speech containing F0 was presented to the acoustic ear. Using acoustic simulations of CI processing, Brown and Bacon (2009a) measured the EAS benefit with a tone modulated both in frequency with the dynamic F0 changes and in amplitude with the envelope of the 500Hz low-pass filtered speech. A significant EAS benefit was observed (ranging from 25% to 57%) with a tone carrying F0 and envelope cues. The EAS benefit was even greater when both cues were combined compared to either cue alone. This benefit was unaffected by the presence of a tone carrying these cues from a background talker. For both real hybrid and bimodal users, nearly identical patterns of results were observed with the same stimuli carrying F0 and envelopes cues (Brown and Bacon, 2009b). These results strongly suggest the importance of F0 and amplitude-envelope cues in target speech to be sources of an EAS benefit.

The EAS benefit was evaluated in terms of consonant feature perception with children at +10 dB SNR (four-talker babble noise) and in quiet (Ching et al., 2001). It was reported that manner information was a primary contributor to the EAS benefit in quiet. Two other studies measured speech information transmitted using CNC word recognition for adults (Mok et al., 2006) and school-aged children (Mok et al., 2010). The results showed that an EAS benefit in quiet arises from improved transmission of the low frequency components in speech such as nasals, semivowels, diphthongs and first formant (F1). In contrast, an EAS benefit in noise (+10 dB SNR in four-talker babble noise) arises from improved perception of both low and high frequency speech components. Recently, Kong and Braida (2010) investigated CI users' ability to integrate consonant and vowel acoustic cues across frequencies in quiet. They reported that no sizable bimodal benefits were measured for all consonant features, but that significant bimodal benefits were facilitated by better transmission of three vowel features (height in F1, backness in F2, and tense) only for those who demonstrated a significant bimodal benefit.

Previous research shows the existence of an EAS synergistic effect on speech understanding both in noise and quiet. However, the well-defined acoustic cues such as place and manner of articulation for consonants and F1 and F2 for vowels have not been evaluated with adult bimodal users as a function of SNR as a source of a possible EAS benefit. The relation between the aided PTA and the utilization of such acoustic cues in EAS mode for speech

perception is also not clear. To ascertain the source of an EAS benefit, it is important to characterize exactly what speech information is affected by EAS. It is also essential to quantify how much information and whether the use of such information is determined by the aided pure-tone average (PTA). The purpose of the present study was to characterize the source of an EAS benefit in consonant and vowel recognition in terms of acoustic features in patients with a conventional CI in one ear in conjunction with an HA in the other ear. Specifically, the issues were which acoustic features for consonant and vowel recognition contributed to EAS benefit, whether such EAS benefits are related to the aided PTA of the acoustic side, whether such an EAS benefit was a function of SNR, and whether sentence performance scores could be predicted from phoneme performance scores.

Methods

Subjects

Ten bimodal users (all post-lingually deafened except S2) participated in the study. Subjects were native speakers of American English and were between the ages of 25 and 77 years old. Subjects were recruited with no particular consideration of age at implantation, duration of profound deafness, period of bimodal experience, etiology, CI/HA type and configuration, and processor strategy. Subject demographics are shown in Table 1. A subject's unaided and aided audiometric thresholds along with mean threshold (thicker dotted line) in the nonimplanted ear are given in Fig. 1 along with the range of hearing threshold for the candidacy criteria in bimodal fitting suggested by MED-EL (dashed-dot line). The left panel of Fig. 1 shows audiometric thresholds for each of five listeners (average age of 69) with PTA > 55 dB hearing level (HL), and the right panel shows hearing thresholds for the other five listeners (average age of 45) with PTA ≤ 55 dB HL over frequencies of 0.25, 0.5, 0.75, and 1 kHz. The cutoff of 55 dB HL for grouping was chosen because it is the upper cutoff for the moderate hearing loss range (41 – 55 dB HL). However, this standard audiological definition was not strictly applied to avoid having the S4 with a PTA of 55 dB in the good group and the S7 with a PTA of 56 dB in the poor group. All subjects provided informed consent and all procedures were approved by the local Institutional Review Board.

Stimuli and Testing

Consonant recognition was measured in quiet and in noise at +5 dB and +10 dB SNR. Medial consonants included /b/, /d/, /g/, /p/, /t/, /k/, /m/, /n/, /f/, /s/, /ʃ/, /v/, /ð/, /z/, /dʒ/, and /tʃ/, presented in an /a/-consonant-/a/ context produced by five male and five female talkers (Shannon et al. 1999). Consonant recognition was measured using a 16-alternative forced choice (16AFC) paradigm. During testing, a stimulus was randomly selected (without replacement) from the stimulus set. The subject responded by pressing on one of the 16 response buttons labeled in a /a/-consonant-/a/ context. Each consonant syllable was presented 20 times (10 talkers x 2 repetitions) for each noise condition. No training or trial-by-trial feedback was provided during testing.

Vowel recognition was measured in quiet and in noise at +5 dB and +10 dB SNR. Medial vowels included 10 monophthongs (/i/, /I/, /e/, /æ/, /u/, /ʊ/, /ɑ/, /ʌ/, /ɔ/, /ɜ/) and 2 diphthongs (/əʊ/, /eI/), presented in an /h/-vowel-/d/ context ('heed', 'hid', 'head', 'had', 'who'd', 'hood', 'hod', 'hud', 'haved', 'heard', 'hoed', 'hayed') produced by five male and five female talkers (Hillenbrand et al., 1995). Vowel recognition was measured using a 12AFC paradigm. During testing, a stimulus was randomly selected (without replacement) from the stimulus set. The subject responded by pressing on one of the 12 response buttons labeled in a/h/-vowel-/d/ context. Each vowel syllable was presented 20 times (10 talkers x 2 repetitions) at each SNR. No training or trial-by-trial feedback was provided during testing.

Sentence recognition was measured in quiet and in noise at 0 dB, +5 dB, +10 dB SNR using hearing-in-noise test (HINT) sentences (Nilsson et al., 1994). The HINT sentences consist of 26 lists of ten sentences each produced by a male talker; the HINT sentences are of easy difficulty. During testing, a sentence list was selected (without replacement), and a sentence was randomly selected from within the list (without replacement) and presented to the subject who then repeated the sentence as accurately as possible. Since there was a total of 12 conditions for HINT test (3 listening modes x 4 SNRs) and each condition was tested with 3 sentence lists (i.e., a total of 36 lists are required), HINT lists that were presented at 0 dB SNR were repeated to reduce the effect of familiarization. The experimenter calculated the percentage of words correctly identified in the sentences. Sentence recognition for each set (10 sentences) was measured three times (total 30 different sentences) for each noise condition. No training or trial-by-trial feedback was provided during testing.

Consonant, vowel, and sentence recognition was measured under three listening conditions: CI alone, HA alone, and CI+HA. For testing in noise, speech-weighted steady noise (1000-Hz low-pass cutoff frequency, -12 dB/oct) was used. The SNR was calculated in terms of the long-term root-mean square of the speech signal and noise. Speech and noise were mixed at the target SNR. The combined speech and noise signal was delivered via an audio interface (Edirol UA 25) and amplifier (Crown D75A) to a single loudspeaker (Tannoy Reveal). Subjects were seated in a double-walled sound treated booth (IAC) directly facing the loudspeaker (0° azimuth) one meter away. Subjects were tested using their clinical speech processors and settings. The output level of the amplifier was set either 65 dB or 70 dB sound pressure level (SPL) for each listening condition (CI alone, HA alone, and CI +HA) according to the Cox loudness rating scale (Cox 1995) in response to 10 sentences in quiet. The output levels for each listening condition are listed for each subject in Table 1.

The subject's CI was turned off in the HA alone condition, and his or her HA was turned off and the non-implanted ear was plugged with an ear plug for the CI alone condition. These three listening conditions were evaluated in random order for each subject. Throughout this article, "bimodal or EAS benefit" refers to the difference between performance with CI+HA and performance with better ear (CI alone). For the present study, all of listeners showed higher performance with CI alone than that with HA alone.

Data Analysis for Speech Information Transmission

Information transmission analyses were performed (Miller & Nicely, 1955). For consonant perception, three acoustic categories (nine speech cues) were analyzed: voicing (voiced and unvoiced), manner of articulation (stop, nasal, fricative, and affricate), and place of articulation (front, middle, and back). For vowel perception, three acoustic categories (eight speech cues) were also analyzed: height of the first formant, F1, (high, middle, and low), place of the second formant, F2, (front, central, and back), and duration of sound (short and long). These speech information groups were chosen so that the frequencies important for their perception covered the range of frequencies in speech.

Vowel height refers to the vertical position of the tongue relative to either the roof of the mouth or the aperture of the jaw and is defined as the relative frequency of the F1. The F1 value is inversely related to the vowel height. Vowel backness is named for the position of the tongue during the articulation of a vowel relative to the back of the mouth and is defined as back or front according to the relative frequency of the F2. The F2 value is inversely correlated to vowel backness. Classification of these acoustic features is given in Table 2.

Results

Individual Speech Recognition Performance

Figure 2 shows individual performance scores along with mean scores (right panel) for the poor group on consonant (top row), vowel (middle row), and sentence (bottom row) recognition with CI+HA, CI alone, and HA alone. To make direct comparisons and be consistent with the presentation of consonant and vowel recognition scores, sentence recognition scores were plotted over three common SNRs used for all test materials. Note that the bimodal benefit for consonant recognition was less than 5 percentage points for all subjects across SNRs except S1 at +5 dB SNR (11.4% benefit). Similarly for vowel recognition less than 5 percentage points of the bimodal benefit was observed for the majority of the cases except S5 (7% at two lower SNRs; 13% in quiet) and S10 at +5 dB SNR (8% benefit). The bimodal advantage for sentence was also less than 5 percentage points except S1 at +10 dB SNR (8% benefit), S5 at +5 dB SNR (16% benefit), and S7 at +10 dB SNR (19% benefit). There was a noticeable ceiling effect in sentence recognition in quiet for all subjects. As expected, there was significant performance variability among listeners. S4 and S10 are relatively high performers across test materials and SNRs, while S1 is a low performer for consonant and S7 is a low performer for vowel and sentence. Variability is also substantial for each listening condition. For example, for consonant recognition with CI alone the difference between the lowest (S1 at +5 dB SNR) and the highest (S10 at +5 dB SNR) performance is approximately 45%.

Individual and mean (right panel) data for the good group were shown in Fig. 3 for consonant (top row), vowel (middle row), and sentence (bottom row) recognition with CI +HA, CI alone, and HA alone. Note that the bimodal benefit for consonant recognition was larger than 5 percentage points for all subjects across SNRs except S3 and S9 at +10 dB SNR and in quiet. The largest benefit was observed for S6 and S8. For vowel recognition, more than 10 percentage points of the bimodal benefit was observed in 4 out of 5 subjects across SNRs. The bimodal benefit for sentence recognition was larger 10 percentage points for all subjects except S9. Similar to the poor group, there was a clear ceiling effect in sentence recognition in quiet for all subjects. The variability for the good group is also noticeable mainly due to two poor performers (S2 and S8). For example, for consonant recognition with CI alone, the score of S2 at +5 dB SNR was only 10%, but that of the best performer (S3) was 57%.

Speech Recognition Comparisons between Two Groups

Group mean recognition scores with standard error were presented in Fig. 2 (right panel) for the poor aided PTA group and in Fig. 3 (right panel) for the good aided PTA group as a function of SNR. In general, bimodal benefit was more evident for listeners with good aided PTA than for listeners with poor aided PTA. Two-way repeated-measures analyses of variance (ANOVA) were performed with SNR and listening condition as factors. For the poor aided group, there was a significant effect of listening mode on consonant [$F(2,16)=67.3, p<0.001$], vowel [$F(2,16)=17.5, p<0.001$], and sentence recognition [$F(2,16)=88.3, p<0.001$]. The main effect of SNR was also significant for consonant [$F(2,16)=46.5, p<0.001$], vowel [$F(2,16)=23, p<0.001$], and sentence recognition [$F(2,16)=18.5, p<0.001$]. For the good aided group, the main effect of listening mode was significant for consonant [$F(2,16)=10, p<0.005$], vowel [$F(2,16)=29.7, p<0.001$], and sentence recognition [$F(2,16)=27.5, p<0.001$]. There was also a significant main effect of SNR for consonant [$F(2,16)=54.7, p<0.001$], vowel [$F(2,16)=21.4, p<0.001$], and sentence recognition [$F(2,16)=25.9, p<0.001$]. However, post-hoc pair-wise comparisons (Sidak method) showed that the bimodal advantage is significant for the good aided group for vowel and sentence perception at the lower two SNRs, indicated by asterisks in Fig. 3

($p < 0.05$). For the poor aided group, any pair-wise comparisons were not significant for each test material ($p > 0.05$).

Bimodal Benefit and Aided PTA

To evaluate the influence of aided thresholds on speech perception, comparisons were made between the two groups of listeners' mean performance scores (Fig. 2, right panel for the poor group and Fig. 3, right panel for the good group) for the HA alone condition. Two-way ANOVA were performed with aided PTA group and SNR as factors. The difference in group mean was significant for vowel [$F(1,24)=5.44, p < 0.05$] and sentence recognition [$F(1,24)=13.5, p < 0.001$] but not for consonant recognition [$F(1, 24)=3.9, p > 0.05$]. The main effect of SNR was also significant for consonant [$F(2,24)=4.93, p < 0.05$] and sentence recognition [$F(2,24)=6, p < 0.05$] but not for vowel recognition [$F(2,24)=1.98, p > 0.05$]. Post-hoc pair-wise comparisons (Sidak method) showed significant differences in performance for the HA alone condition between the two groups at higher SNRs for vowel and sentence recognition and in quiet for consonant recognition, denoted by plus symbols in Fig. 3 (right panel) ($p < 0.05$).

To evaluate the effect of aided PTA on the bimodal benefit, a correlation analysis was performed between the magnitude of the bimodal benefit and aided PTA at audiometric frequencies of 0.25, 0.5, 0.75, and 1 kHz for each speech test; these results are shown in Fig. 4. The correlation analysis showed no significant association in the bimodal benefit for consonant recognition averaged across SNR ($r(10) = -0.47, p > 0.05$) or at a specific SNR ($r(10) = -0.25, p > 0.05$ at +5 dB; $r(10) = -0.485, p > 0.05$ at +10 dB; $r(10) = -0.43, p > 0.05$ in quiet). A correlation between aided PTA and the bimodal benefit for sentence recognition was not significant over SNRs ($r = -0.43, p > 0.05$) or at each SNR ($r(10) = -0.25, p > 0.05$ at +5 dB; $r(10) = -0.54, p > 0.05$ at +10 dB; $r(10) = -0.06, p > 0.05$ in quiet). Correlation between aided PTA and the bimodal benefit in vowel perception (middle panel) was significant at the lower two SNRs ($r(10) = -0.68, p < 0.05$ at +5 dB; $r(10) = -0.64, p < 0.05$ at +10 dB), and the correlation was also significant between aided PTA and the bimodal benefit averaged across SNR ($r(10) = -0.65, p < 0.05$). Further correlation measures revealed that such a bimodal benefit for vowel perception is significantly attributed to the aided threshold at 0.5 kHz ($r(10) = -0.72, p < 0.05$).

Bimodal Benefit and Acoustic Features for Consonant

Figure 5 presents percent information transmitted for consonant recognition in terms of voicing, manner, and place of articulation for each group. Note that the good group transmitted most of the consonant features better with CI+HA than with CI alone, particularly for affricate and back in place. The difference in information transmitted between CI+HA and CI alone ranged from 15% to 18% for affricate and from 10% to 18% for back in place. For other features, the transmission was enhanced approximately from 5% to 10% across SNRs in the good group. The range of the feature transmission in HA alone was 20% to 40% depending on SNR.

In contrast, for the poor group, the difference in percent information transmitted between CI +HA and CI alone was less than 5 percentage points for all features analyzed across SNRs. The amount of information transmitted by HA alone in the poor group was consistently lower than that by HA alone in the good group. On the contrary, the amount of information transmitted by either CI+HA or CI alone in the poor group was consistently higher than that in the good group. For both groups, there was no ceiling or floor effect.

Figure 6 presents a summary of the bimodal benefit in percentage points for consonant recognition in terms of voicing, manner, and place of articulation for each group. In general,

the bimodal benefit is always greater for the listeners with better aided thresholds than the listeners with poorer aided thresholds for most of the features. For the poor group, a two-way repeated measure of ANOVA shows that none of the acoustic features is significantly different between CI+HA and CI alone; for the good group, the bimodal benefit was significantly attributed to low (voicing) and high frequency (stop and affricate) components in speech along with all three places which requires a wide range of spectral information. Statistical details are given in Table 3. The effect of SNR for both groups is significant for each of the acoustic features, but the pattern of the bimodal benefit for each feature is highly SNR-specific. For example, the magnitude of the bimodal benefit enhanced by a voicing cue is similar at +10 dB SNR and in quiet, but the bimodal benefit aided by an affricate manner of articulation is greater in quiet and noise at +5 dB SNR than that at +10 dB SNR.

To evaluate the effect of aided PTA on bimodal benefit, comparisons of the bimodal benefit between the two groups were made. The only significant benefit for the listeners with better aided hearing thresholds stemmed from acoustic cues: voiced, affricate, and back in place (See Table 3, last column). Other cues did not show a significant difference between groups. An asterisk symbol in Fig. 6 indicates a significant difference in bimodal benefit between the two groups by post-hoc pair-wise comparisons.

Bimodal Benefit and Acoustic Features for Vowel

Figure 7 shows the results of vowel feature analyses in terms of F1 and F2 along with duration for each group. For the good group, overall, relatively good enhancement was made for all features analyzed in CI+HA compared to CI alone. The improvement ranged from 8% to 20% depending on SNRs. Particularly central information in F2 was consistently transmitted 15% higher in CI+HA than in CI alone.

For the poor group, the transmission of the features improved less than 5 percentage points for all features across SNRs except for low in F1 at +5 dB SNR. The amount of information transmitted by CI+HA was similar between the two groups, but in CI alone, the poor group transmitted higher percentage of vowel information than the good group. The amount of information transmitted in HA alone was consistently lower in the poor group than in the good group. There were no ceiling or floor effects for either group.

A summary of the bimodal advantage for vowel recognition was presented in Fig. 8 in terms of F1 (first row) and F2 (second row) along with duration (last row) between the two groups. For the poor group, the transmission of speech information was not significantly improved with bimodal mode compared to CI alone; for the good group, all of the information features known to be embedded both in low and high spectral components in speech, except for short-duration cues, were significantly received with bimodal setting as opposed to CI alone. Statistical details are given in Table 4. In terms of a SNR effect, the bimodal benefit for the poor group was not significantly enhanced with increasing SNR for low in F1 height, central and back in F2 place, and short duration cues. For the good group, the bimodal benefit was significantly dependent on SNR for all features.

Comparisons in bimodal benefit between the two groups showed that four acoustic features (middle height in F1, central and back place in F2, and long duration) significantly contributed to the bimodal benefit for the good group compared to the poor group (see Table 4 for statistics). Post-hoc pair-wise comparisons showed a significant difference in the bimodal benefit between the two groups for all features, at least at one of the SNRs, as shown with an asterisk symbol.

Relationship between Contextual and Non-contextual Recognition

Figure 9 shows scatter plots in terms of bimodal benefit for consonant (left panel), vowel (middle panel), and phoneme (right panel) recognition relative to the bimodal benefit for sentence recognition. The bimodal benefits for phonemes were computed with

$\sqrt{C_{CIHA} \times V_{CIHA}} - \sqrt{C_{CI} \times V_{CI}}$ (Boothroyd & Nitttrouer, 1988), where C_{CIHA} and V_{CIHA} are the percent scores for consonant and vowel recognition in the bimodal mode, and C_{CI} and V_{CI} are the percent scores for consonant and vowel recognition in the CI alone mode. Each open circle represents an individual data point at each SNR. Linear regression analysis demonstrates that the magnitude of the bimodal benefit in sentence perception was significantly associated with the bimodal benefit for consonants ($r^2=.56$, $p<0.05$) and for phonemes ($r^2=.60$, $p<0.05$). However, the linear relationship between vowel and sentence recognition was not significant ($r^2=.47$, $p>0.05$).

Discussion

In this study, we characterized acoustic cues in speech recognition enhanced in a bimodal configuration. The bimodal benefit was greater for listeners with better aided PTA than that for the listeners with poorer aided PTA across SNR for all speech materials. The poorer aided PTA group received little bimodal benefits probably due to HA input being at or below threshold in some or all frequencies. For consonant and vowel recognition, the specific acoustic features were responsible for the bimodal benefit regardless of low- and high-frequency components in speech. Such an enhancement in the acoustic features is significantly influenced by aided hearing threshold as well. The bimodal benefit for sentence recognition is significantly correlated with that for consonant recognition, and the correlation becomes stronger when relating the magnitude of the advantages between sentence and phoneme scores.

Mean Bimodal Benefit in Quiet and in Noise

In the present study, the magnitude of mean bimodal advantage for consonant perception for the good group in quiet and in noise is 6 and 8 percentage points respectively (Fig. 3, right panel) which is comparable with the results of Dorman et al. (2008) who reported a benefit of 8% in quiet and of Mok et al. (2006, 2010) who reported a similar range at 10 dB SNR. For vowel perception (Fig. 3, right panel), the amount of the bimodal benefit in the present study, 7% in quiet, is lower than the 18% benefit reported in Dorman et al. (2008). The benefit for sentence recognition in the present study is 13% in noise and only 2% in quiet (Fig. 3, right panel). This is somewhat lower than values reported in the literature. Multiple studies reported a bimodal benefit of approximately 20% in noise at +5 dB SNR and/or +10 dB SNR (Ching, Bai, & Zeng, 2006; Dorman et al. 2008; Gifford et al., 2007; James et al., 2006; Kong et al., 2005). Dorman et al. (2008) reported that the bimodal benefit was 25% both at +5 dB and +10 dB SNR.

The reasons for the somewhat lower values of the bimodal benefit in the present study might be due to the different use of noise maskers and/or audiometric configurations for the acoustic side. Since the hearing thresholds are widely varied across these studies, a different type of noise masker might be a source for such a difference. Speech-weighted noise was used in the present study, but single or multiple talker babble noise was used in the other studies mentioned above. Brown and Bacon (2009a) demonstrated using a hybrid configuration simulation that the performance score for a speech-shaped masker was approximately 10 percentage points lower than that for the talker-babble masker.

For the poor group in the present study, both the bimodal and CI alone performance curves are almost identical across SNR and test materials (Fig. 2, right panel). This suggests that no

new speech information was added via HA or that speech cues provided by HA are not utilized by the auditory system. Little bimodal benefit for the poor group may be the result of impaired frequency difference limens in the region of elevated thresholds (Tyler et al., 1983; Peters & Moore, 1992) and inaudible F2 or higher formants. Thus no difference in the performance between CI+HA and CI alone suggests that HA does not add any further information regarding pitch and the first two formants when aided threshold is high enough (i.e., ≥ 55 dB HL at frequencies ≤ 1 kHz) or when the HA does not substantially improve thresholds.

Performance Comparisons for CI alone condition

One interesting observation from the group data (Fig. 2 and Fig. 3) is that the CI-alone condition was consistently lower (about 20 percentage points) in the good group (Fig. 3, right panel) compared to the poor group (Fig. 2, right panel). Such discrepancy might be simply attributed to performance variability across CI users. As shown in Fig. 2 and Fig. 3, two top performers (S4 and S10) in the poor group outperformed (more than 15% higher) any of subjects in the good group for consonant recognition. In addition, performances of the two poor performers (S2 and S8) in the good group were far below than group mean scores. Consequently two good performers in the poor group and two poor performers in the good group are major contributors to higher scores for CI-alone condition in the poor group than in the good group. As mentioned in the Method, subject S2 in the good group is only one with prelingual deafness. It is interesting to see how scores from S2 influence the difference in CI alone between two groups. Two-way ANOVA was performed for CI alone between two groups with and without data from S2 with group and SNR as factors. With data of S2 significant difference between two groups was observed for consonant ($F(1,24) = 10.3, p < 0.004$) and for sentence ($F(1,24) = 4.2, p < 0.05$), but not for vowel ($F(1,24) = 1.4, p > 0.05$). In contrast, without data of S2, difference in CI alone between two groups was not significant for consonant ($F(1,21) = 5.8, p > 0.05$), vowel ($F(1,21) = 1.3, p > 0.05$), and sentence ($F(1,21) = 0.95, p > 0.05$).

Correlation between Aided PTA and Bimodal Benefit

The present study shows that bimodal benefit is greater for the listeners with better aided thresholds than for the listeners with poorer aided thresholds (Figs. 2 and 3). This result is consistent with the result of Mok et al. (2006) but not of Mok et al. (2010). Mok et al. (2006) showed significant correlation between bimodal benefit and aided threshold at 1 kHz for consonants and vowels and at 2 kHz for consonants. In contrast, Mok et al. (2010) showed nonsignificant association between the bimodal benefit and low frequency aided thresholds (.25 and .5 kHz) for all CNC measures. This inconsistency might be due to the difference in hearing thresholds for subjects across studies. In addition, the study of Mok et al. (2006) was conducted in quiet with adults, and speech perception in Mok et al. (2010) was measured at +10 dB SNR with children. Ching, Incerti, & Hill (2004) reported no significant correlation between aided threshold and bimodal benefit in sentence recognition. Chang, Bai, & Zeng (2006) also reported in an acoustic simulation study nonsignificant correlations between aided PTA over frequencies of .5, 1, and 2 kHz and bimodal benefit even though the bimodal benefit decreased with an increase in hearing loss.

The correlation analysis in the present study (Fig. 4) shows a significant relationship between aided threshold and bimodal benefit in vowel recognition. This can be explained by Miller's vowel perceptual space theory: pitch plays a significant role in vowel perception by interacting with the frequency of F1 to form one dimension of a three-dimensional perceptual space (Miller, 1989). So, we would expect that vowel recognition would be significantly improved when HA is added. F1 could provide a frequency-appropriate reference against which higher frequency information provided by the CI would be

integrated. F1 could aid in perception of high- and mid-height vowels and formant transition (Strange et al., 1983).

The results of the present and Mok et al. (2006) studies suggest that individual variability in bimodal outcome could be partially accounted for by the differences in aided thresholds, and a better aided threshold does not necessarily result in a greater bimodal advantage for all test materials.

Information Transmission Analyses for Consonant Recognition

The results of the consonant feature analysis for the good group in the present study (Fig. 5) are consistent with those of Kong and Braida (2010) who measured similar quantities with 11 bimodal users in quiet. In both studies the HA provided information about nasal cues the most. Voicing information was similarly transmitted in CI alone between the two studies (55% in the good group of the present study and 56% in Kong and Braida (2010)). CI+HA provided a similar amount of benefit in voicing information transmitted in the two studies (58% in Kong and Braida (2010) and 62% in the good group). For other consonant features (stop and fricative), similar percent information transmitted was observed in the two studies as well.

The result of the present study in consonant recognition also shows that speech information required for the perception of higher frequency components (i.e., affricate and back in place) was also enhanced when HA was added, particularly for the listeners with good aided thresholds (Figs. 5 and 6). This result is consistent with those of Mok et al. (2010). This study characterized the bimodal benefits measured with 13 school-aged children at +10 dB SNR (four-talker babble masker). Data was analyzed in nine phonetic categories: fricatives, sibilants, plosives, F2, bursts, semivowels, nasals, diphthongs, and F1. Mok et al. (2010) reported a relatively large bimodal advantage (5% to 10%) for some of the other phoneme groups including those that require perception of higher-frequency information (e.g., fricatives, sibilants, F2). The results of the present study and Mok et al. (2010) suggest that the perception for cues embedded in higher frequencies is more important when listening in noise.

In contrast, our results are inconsistent with those of Mok et al. (2006) and of Kong et al. (2007). Mok et al. (2006) tested 14 adult bimodal users for CNC word recognition in quiet and analyzed the data in terms of the same features analyzed in Mok et al. (2010). They showed that only lower frequency features such as semivowel, nasal, diphthongs, and F1 contributed to the bimodal benefit. Kong et al. (2007) also reported a similar result for sentence recognition in noise, which they demonstrated by removing the phonetic cues in HA side (preserving the F0, voicing, and temporal envelope cues) and not observing a bimodal benefit at +10 dB and +15 dB SNR. This indicates that low-frequency phonetic cues are important for bimodal benefit, particularly at higher SNRs.

There could be a few factors contributing to the unexpected results between Mok et al. (2006) and Mok et al. (2010). First, the bimodal adults in Mok et al. (2006) were tested in quiet while bimodal children in Mok et al. (2010) were tested at +10 dB SNR. Combined with the result of the present study high-frequency speech cues were more utilized when listening in noise. Second, the bimodal children had been using the bimodal input since a young age while the bimodal adults only received CI in adulthood. Hence, it is possible that the children have learned to better integrate the signals from the two ears and that the mismatch between the ears at higher frequencies is less significant for children.

Relationship between Voiced/unvoiced Cue and F0

It is shown in the present study that in the good group both voiced and unvoiced information for consonant perception (Fig. 6) was significantly enhanced in CI+HA condition in noise compared to CI alone. Previous studies that examined speech information transmission in consonant perception also revealed significant transmission of voicing information with bimodal setting compared to with CI alone (Brown and Bacon, 2009a, 2009b). One reasonable thought related to voicing is that if F0 is one of the primary cues used by low-frequency acoustic hearing to improve electric hearing, then improvement for voiceless consonants should be minimal. The present results, however, indicate that this is not the case. These results suggest that voicing cues are not necessarily enhanced proportionally with F0 information. The results of the two previous studies support this idea as well (Brown and Bacon, 2009a and 2009b).

Brown and Bacon (2009a) showed 9 percentage points of bimodal benefit migrated by voicing cue alone under speech-shaped noise for sentence recognition using acoustic simulation. The same trend was observed with two bimodal CI users under a male and babble backgrounds (Brown and Bacon, 2009b). These results are in a good agreement with that of the present result for voicing cue. Brown and Bacon (2009b) also showed that additional 17 percentage points of the benefit were observed beyond voicing alone when a tone was modulated in frequency with the target talker's F0. These results suggest that F0 cue provides more cues, not accordingly with voicing cue. However, the sample size is too small to interpret this result in a meaning way. In addition, these two subjects in a study of Brown and Bacon (2009b) were tested in unaided condition under a male and babble background, and more importantly no background was presented in the acoustic ear. It might be interesting to test if F0 and temporal envelope cues are robust under speech-shaped noise environment in HA side.

F0 has also been shown to be important for the manner of articulation (Faulkner and Rosen, 1999). Ching et al. (2001) showed that significantly more manner information was received when listeners used bimodal hearing devices compared to using CI alone in quiet. In the present study, two manner cues, stop and affricate, were significantly better perceived only for the listeners with better aided threshold while information in nasal and fricative cues was not significantly enhanced for either group.

Information Transmission Analyses for Vowel Recognition

The results of the vowel feature perception in the present study (Fig. 7) are consistent with those of Kong and Braida (2010) who evaluated bimodal benefit in terms of vowel features in quiet. CI alone provided similar information transmission in height of F1 (61% in the good group of the present study and 53% in Kong and Braida (2010)) and place of F2 (69% in Kong and Braida (2010) and 63% in the good group) between the two studies. In CI+HA, height in F1 and place in F2 were similarly transmitted between the two studies. However, the HA provided height information of F1 about 46% for the good group and 27% in Kong and Braida (2010) and place in F2 about 43% for the good group and only 4% in Kong and Braida (2010) (mainly due to the greater variability ranging from 1% to 44%). Kong and Braida (2010) also reported that only those who demonstrated a substantial bimodal benefit showed significant improvement in all vowel features tested which is consistent with our findings.

All vowel features, except short duration, analyzed in the present study were significantly better perceived in a bimodal setting for the listeners with better aided PTA while none of the features, including short duration, were significantly perceived for the listeners with poorer aided PTA (Fig. 8). For the good group, the aided threshold at the range of F2

frequencies between 1 and 3 kHz is from 40dB HL to 80dB HL. This finer spectral information provided by HA can aid perception of some of vowels required for higher spectral components.

It is obvious that bimodal listening enhances acoustic cues in low height of F1 and high place of F2 spectral elements in vowels. Dellattre et al. (1952) showed a significant drop in vowel perception when any portion of the spectrum of F1 and F2 was removed. Dorman et al. (2005) showed that if all of the F2 information that fell under 2.1 kHz was eliminated, then there was no decrease in vowel perception with larger gaps above 2.1 kHz in frequency between CI in one ear and HA in the opposite ear using simulations. Multiple previous studies focused on F0 as a primary source for bimodal benefit (Brown and Bacon, 2009a, 2009b; Kong et al, 2005, 2007; Qin & Oxenham, 2006), but the result of the present study shows that F1 and F2 are also important components for vowel perception in bimodal setting.

Contextual and Non-contextual Comparisons in Bimodal Benefit

It is frequently stated, maybe widely believed, that the perception of continuous speech is a complex process since the integration of perceptual cues is governed by high-level language such as lexical, morphological, syntactic, and semantic constraints (Chomsky & Halle, 1968). Figure 9 shows that the advantage of a bimodal arrangement for sentence perception is significantly related with nonsense consonant and phoneme perception. This result is consistent with one of the findings that Rabinowitz et al. (1992) reported: consonant recognition is highly correlated with sentence recognition. It is known that consonants can be more affected by removing low-freq information, ≤ 0.8 kHz (Lippmann, 1996; Warren et al., 2004). Therefore, it is possible that sentence recognition was affected as well by better spectral resolution in the low frequency range for the better PTA group.

Cues in continuous speech can be classified into two types: acoustic and context cues. The context cues refer to redundancy, recoverable once given sufficient acoustic cues, along with the associated context information (Boothroyd & Nitttrouer, 1988; Bronkhorst *et al.*, 1993). Based on significant correlations in speech perception between contextual and non-contextual speech materials (Fig. 9), it can be stated that sentence understanding in noise can be accounted for by the acoustic cues extracted from random syllables (nonsense consonants and vowels). This result suggests the existence of such a connection between acoustic and context cues in a bimodal setting.

The Bimodal Benefit and Audibility/HA fitting

All the participants in the present study have been fitted by the same protocols and same audiologists to make sure that HA provides audible sounds. However, aided thresholds for the poor group were much poorer even though unaided thresholds were similar between groups (Fig. 1). It indicates a possibility that spectral gain prescribed to HAs might not be optimal for the poor group. This possibility is supported by the results of Ching et al. (2001) study. They showed that 15 of the 16 children required 6 dB more gain than prescribed (National Acoustic Laboratories-Revised, Profound) to balance the loudness of the implanted ear for speech signal presented at 65 dB SPL. It suggests that adjustment of HA gain to match loudness in the implanted ear for the poor group can facilitate integration of speech information across ears, leading to bimodal advantage.

To test this idea, a subject (S4) in the poor group was retested with a fixed 80 dB SPL presentation level for sentence recognition in noise and quiet for HA alone, CI alone and CI +HA conditions. This level provides the subject about 30 dB sensation level at frequencies between .25 and 1 kHz and about 15 dB sensation level at frequencies between 1 and 3

kHz (See aided audiogram for S4 in Fig. 1). Figure 10 shows performance with a presentation level of 80 dB SPL (right panel) along with original data (left panel) measured at 70 dB SPL for HA alone and at 65 dB SPL for CI alone and CI+HA conditions. No bimodal benefit was observed with 80 dB SPL presentation level even though performance improved for HA alone at +10 dB SNR and in quiet, but not at +5 dB SNR. In contrast, the performance dramatically dropped approximately 30% due to clipping for both CI alone and CI+HA conditions in noise, leading to bimodal interference at 80 dB SPL presentation level. This result suggests that optimal bimodal loudness balancing is needed to facilitate bimodal benefit; it is likely that some degree of bimodal benefit could be observed when 80 dB and 65 dB SPL levels for the HA side and CI side are presented.

Age Effects on Bimodal Benefit

In the good group (average age is 45), the three youngest subjects (S2: 44 years old, S6: 25 years old, and S9: 47 years old) showed the greatest EAS benefit, particularly for vowel and/or sentence recognition. S6 showed the benefit regardless of test materials, but S9 demonstrated the benefit only for vowel recognition. Two elderly subjects (S3 and S8) in the good group also showed a benefit for vowel and sentence tests. In the poor group, the youngest subject (S7: 56 years old) did not show a benefit. This inconsistent age effect is also reported in the study of Kong and Braida (2010). The four youngest bimodal subjects (<27 years old) out of 12 did not show a benefit in consonant and vowel recognition. However, substantial benefit was observed from two other subjects (64 and 57 years old) in vowel recognition. Therefore, it is unlikely that the bimodal benefit is accounted for by age only.

Limitations

In the present study, as subjects were tested in their clinical setting for each device, an optimization of a HA to complement a CI was not administrated in a systematic way. The frequency response of the HA should be optimized for speech understanding for each subject so that HAs have enough compression capability in dynamic range. Another issue is that loudness balancing should be performed and that sounds must be maintained at a comfortable listening level for input levels when the HA is used with a CI. As described in the Method section of this article, loudness was roughly balanced by adjusting the output level of the amplifier as the most comfortable level for each listening condition (CI alone, HA alone, and CI+HA). If serious consideration for loudness balance is necessary, then a loudness balancing procedure should be used to adjust the gain of the HA for different input levels so that loudness of speech is similar between ears. It is possible that bimodal benefit for the poor group might be enhanced with adjustment of gain for better loudness balance with CI.

Another limitation related to HA device is unknown interactions between the bimodal benefits and HA setting parameters. Each device and model has some of unique features such as different number of bands, channels, degree of compression, signal processing algorithm, and noise reduction technology along with different listening programs. All these factors may affect their aided PTA thresholds and HA performance. Even though it is difficult to expect how these features influence the outcome in bimodal settings, it is possible that one or more aspects of HA fitting will affect their bimodal benefits. Particularly considering a situation in which a noise reduction algorithm is not used for the CI side, but is included in the HA side, the CI-alone performance could drop significantly, which could explain greater bimodal benefit in noise, but lack of the benefit in quiet condition.

The lack of training prior to testing might be problematic, particularly for the CI alone and HA alone conditions. Given that performance for unilateral listening conditions was commonly measured by having bimodal CI users switch off one of their devices. This arrangement generates immediate disadvantage as this configuration does not characterize their everyday listening experience and in essence, changes how their auditory system process cues for speech recognition. Consequently, CI alone performance may be higher for listeners with poorer acoustic hearing threshold due to insufficient amplification in HA side.

Conclusions

In the present study, the source of the bimodal advantage in speech perception was characterized in terms of acoustic features for consonant, vowel, and sentence recognition in noise and in quiet. Speech recognition was measured in three conditions: CI alone, HA alone, and combined CI and HA conditions; speech and noise were presented at 0° azimuth. Results showed that bimodal advantage for vowel and sentence recognition depends on aided thresholds at audiometric frequency below 1 kHz, especially in noise. The poorer aided PTA group received little bimodal benefits probably due to HA input being at or below threshold in some or all frequencies. The results of information transmission analyses revealed that bimodal listening helps enhance the acoustic features both in low and high frequency components. For consonant recognition, low frequency voicing and high frequency stop and affricate cues along with places cues are the main sources of the bimodal benefit for the good aided group while none of these features contribute to the bimodal benefit for the poor aided group. Three acoustic cues, voiced, affricate, and back in place, are primary components used to differentiate the magnitude of the bimodal benefit in consonant recognition between the two aided PTA groups. For vowel recognition, low frequency cues (F1) and high frequency cues (F2) along with long duration cues are significant contributors to the bimodal benefit for the good aided group while none of these features are significant sources for the poor aided group. The magnitude of the bimodal advantage between the two aided PTA groups is differentiated by middle height in F1, central and back place in F2, and long duration cues. The enhancement of sentence recognition in bimodal setting is significantly associated with the bimodal benefit for consonant recognition, and correlation is stronger when relating the magnitude of the advantage between sentence and phoneme recognition (scores combined for consonants and vowels). The results suggest that aided hearing threshold should be carefully considered in order to maximize the advantage of the bimodal use in speech perception. The results also suggest that speech information processed by CI in conjunction with HA not only aided in the perception of low frequency components in speech but also helped process some of high frequency components.

Acknowledgments

We thank our participants for their time and effort. We thank Jaesook Gho, Tianhao Li, Aiguo Liu and Joseph Crew for their editorial assistance and valuable comments and Dr. Hou-Yong Kang for his help in data collection. We would also like to thank Dr. Eric Healy, Dr. Christopher Brown and one anonymous reviewer for their useful comments and suggestions. This work was supported by NIH grant 5R01DC004993.

References

- Beijen J, Mylanus E, Leeuw A, Snik A. Should a hearing aid in the contralateral ear be recommended for children with a unilateral cochlear implant? *Annals of Otology, Rhinology & Laryngology*. 2008; 117(6):397–403.
- Blamey P, Arndt P, Bergeron F, et al. Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants. *Audiology and Neurotology*. 1996; 1:293–306. [PubMed: 9390810]

- Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. *Journal of the Acoustical Society of America*. 1988; 84(1):101–114. [PubMed: 3411038]
- Bronkhorst AW, Bosman AJ, Smoorenburg GF. A model for context effects in speech recognition. *Journal of the Acoustical Society of America*. 1993; 93(1):499–509. [PubMed: 8423265]
- Brown CA, Bacon SP. Low-frequency speech cues and simulated electric-acoustic hearing. *Journal of the Acoustical Society of America*. 2009a; 125(3):1658–1665. [PubMed: 19275323]
- Brown CA, Bacon SP. Achieving electric-acoustic benefit with a modulated tone. *Ear and Hearing*. 2009b; 30(5):489–493. [PubMed: 19546806]
- Chang J, Bai J, Zeng FG. Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise. *IEEE Transactions on Biomedical Engineering*. 2006; 53:2598–2601. [PubMed: 17152439]
- Ching TY, Psarros C, Hill M, Dillon H, Incerti P. Should children who use cochlear implants wear hearing aids in the opposite ear? *Ear and Hearing*. 2001; 22:365–380. [PubMed: 11605945]
- Ching TY, Incerti P, Hill M. Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear and Hearing*. 2004; 25(1):9–21. [PubMed: 14770014]
- Ching TY, Incerti P, Hill M, van Wanrooy E. An overview of binaural advantages for children and adults who use binaural/bimodal hearing devices. *Audiology and Neurotology*. 2006; 11(suppl 1): 6–11. [PubMed: 17063004]
- Chomsky, N.; Halle, M. *The sound pattern of English*. Harper and Row Publishers; New York, NY: 1968.
- Cox RM. Using loudness data for hearing aid selection: the IHAF approach. *Hearing Journal*. 1995; 48(2):39–44.
- Dellatre P, Libermann AM, Cooper FS, Gerstman L. An experimental study of the acoustic determinants of vowel color. *Word*. 1952; 8:195–210.
- Dorman MF, Spahr AJ, Loizou PC, Dana CJ, Schmidt JS. Acoustic simulations of combined electric and acoustic hearing (EAS). *Ear and Hearing*. 2005; 26(4):371–380. [PubMed: 16079632]
- Dorman MF, Gifford RH, Spahr AJ, McKarns SA. The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiology and Neurotology*. 2008; 13(2):105–112. [PubMed: 18057874]
- Faulkner A, Rosen S. Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception. *Journal of the Acoustical Society of America*. 1999; 106:2063–2073. [PubMed: 10530029]
- Gantz BJ, Turner CW, Gfeller KE, Lowder MW. Preservation of hearing in cochlear implant surgery: advantages of combined electrical and acoustical speech processing. *Laryngoscope*. 2005; 115(5): 796–802. [PubMed: 15867642]
- Gifford RH, Dorman MF, McKarns SA, Spahr AJ. Combined electric and contralateral acoustic hearing: word and sentence recognition with bimodal hearing. *Journal of Speech Hearing Research*. 2007; 50(4):835–843.
- Gstoettner W, Helbig S, Maier N, Kiefer J, Radeloff A, Adunka OF. Ipsilateral electric acoustic stimulation of the auditory system: results of long-term hearing preservation. *Audiology and Neurotology*. 2006; 11(Suppl 1):49–56. [PubMed: 17063011]
- Gstoettner W, Kiefer J, Baumgartner WD, Pok S, Peters S, Adunka OF. Hearing preservation in cochlear implantation for electric acoustic stimulation. *Acta Oto-Laryngol*. 2004; 124(4):348–352.
- Hamzavi J, Pok SM, Gstoettner W, Baumgartner W. Speech perception with a cochlear implant used in conjunction with a hearing aid in the opposite ear. *International Journal of Audiology*. 2004; 43:61–65. [PubMed: 15035557]
- Hillenbrand J, Getty L, Clark M, Wheeler K. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*. 1995; 97:3099–3111. [PubMed: 7759650]
- James CJ, Fraysse B, Deguine O, Lenarz T, Mawman D, Ramos A, Ramsden R, Sterkers O. Combined electroacoustic stimulation in conventional candidates for cochlear implantation. *Audiology and Neurotology, Suppl*. 2006; 1:57–62.

- Kiefer J, Pok M, Adunka O, Stürzebeche RE, Baumgartner W, Schmidt M, Tillein J, Ye Q, Gstoettner W. Combined electric and acoustic stimulation of the auditory system: results of a clinical study. *Audiology and Neurotology*. 2005; 10(3):134–144. [PubMed: 15724084]
- Kong YY, Braida LD. Cross-frequency integration for consonant and vowel identification in bimodal hearing. *Journal of Speech Language Hearing Research*. 2010 in press.
- Kong YY, Stickney GS, Zeng FG. Speech and melody recognition in binaurally combined acoustic and electric hearing. *Journal of the Acoustical Society of America*. 2005; 117(3 Pt 1):1351–1361. [PubMed: 15807023]
- Kong YY, Carlyon RP. Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. *Journal of the Acoustical Society of America*. 2007; 121(6): 3717–3727. [PubMed: 17552722]
- Lippmann RP. Accurate consonant perception without mid-frequency speech energy. *IEEE Transaction on Speech and Audio Processing*. 1996; 4:66–99.
- Miller J. Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*. 1989; 85:2114–2134. [PubMed: 2659639]
- Miller GA, Nicely PE. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*. 1955; 27:338–352.
- Mok M, Grayden D, Dowell RC, Lawrence D. Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears. *Journal of Speech Hearing Research*. 2006; 49(2):338–351.
- Mok M, Galvin KL, Dowell RC, McKay CM. 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*. 2010; 15(1):44–56. [PubMed: 19468210]
- Moore BCJ. Dead regions in the cochlea: diagnosis, perceptual consequences, and implications for the fitting of hearing aids. *Trends In Amplification*. 2001; 5:1–34.
- Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*. 1994; 95 (2):1085–1099. [PubMed: 8132902]
- Peters RW, Moore B. Auditory filter shapes at low center frequencies in young and elderly hearing-impaired subjects. *Journal of the Acoustical Society of America*. 1992; 91:256–266. [PubMed: 1737876]
- Rabinowitz W, Eddington D, Delhorne L, Cuneo P. Relations among different measures of speech reception in subjects using a cochlear implant. *Journal of the Acoustical Society of America*. 1992; 92:1869–1881. [PubMed: 1401531]
- Shalloo JK, Arndt P, Turnacliffe KA. Expanded indications for cochlear implantation; Perceptual results in seven adults with residual hearing. *Journal of Speech Language Pathology Association*. 1992; 16(2):141–148.
- Shannon RV, Jansvold A, Padilla M, Robert M, Wang X. Consonant recordings for speech testing. *Journal of the Acoustical Society of America, (ARLO)*. 1999; 106 (6):L71–L74.
- Strange W, Jenkins JJ, Johnson T. Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*. 1983; 74:695–705. [PubMed: 6630725]
- Turner CW, Gantz BJ, Vidal C, Behrens A, Henry BA. Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing. *Journal of the Acoustical Society of America*. 2004; 115(4):1729–1735. [PubMed: 15101651]
- Turner CW, Gantz BJ, Reiss L. Integration of acoustic and electrical hearing. *Journal of Rehabilitation Research & Development*. 2008; 45(5):769–778. [PubMed: 18816425]
- Tyler R, Wood E, Fernandes M. Frequency resolution and discrimination of consonant and dynamic tones in normal and hearing-impaired listeners. *Journal of the Acoustical Society of America*. 1983; 74:1190–1199. [PubMed: 6643841]
- Warren RM, Bashford JA, Lenz PW. Intelligibility of bandpass filtered speech: Steepness of slopes required to eliminate transition band contributions. *Journal of the Acoustical Society of America*. 2004; 115:1292–1295. [PubMed: 15058351]

Qin MK, Oxenham AJ. Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech. *Journal of the Acoustical Society of America*. 2006; 119(4):2417–2426. [PubMed: 16642854]

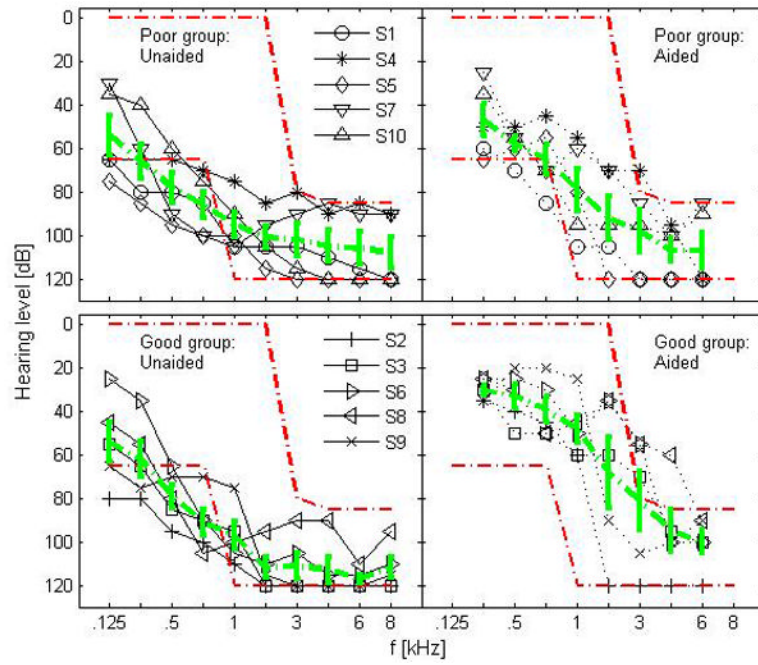


Figure 1. Individual and mean unaided and aided hearing threshold with standard error. The left panel presents hearing threshold for 5 listeners with aided PTA ≥ 55 dB HL over frequencies of 0.25, 0.5, 0.75, and 1 kHz while the right panel shows the threshold for another 5 listeners with aided PTA < 55 dB HL. Mean hearing threshold with standard error is indicated by thicker dotted line. The range of hearing threshold for the candidacy criteria in bimodal fitting set by MED-EL is indicated by dashed-dot line.

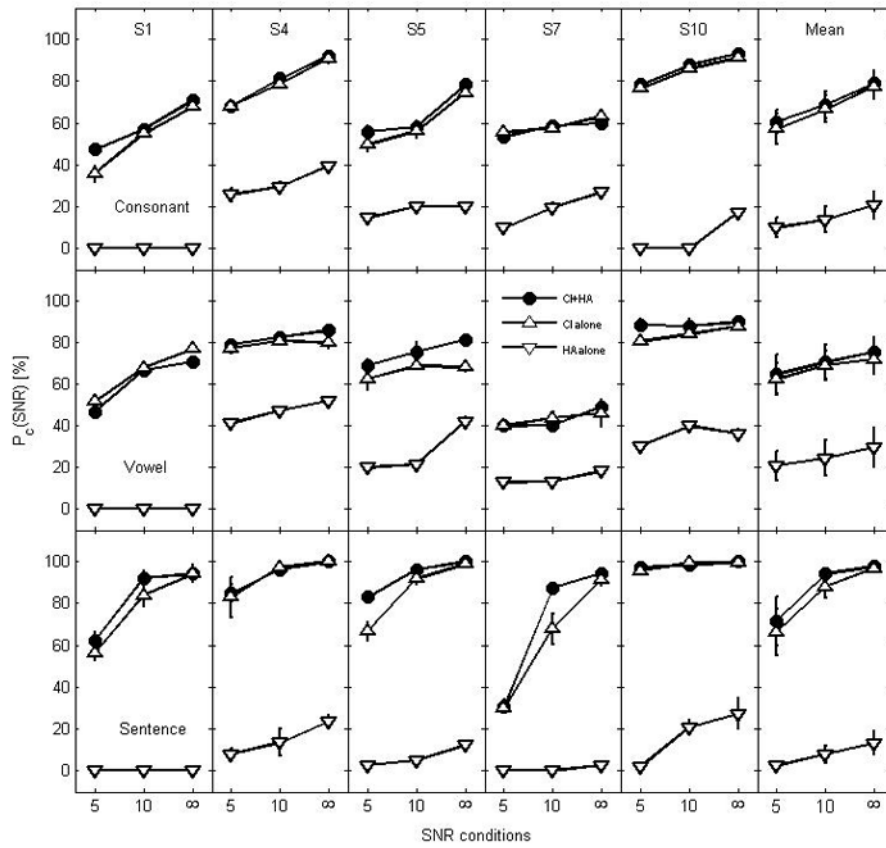


Figure 2. Individual and group mean recognition performance with standard error for the poor group as a function of SNR with CI+HA, CI alone and HA alone. Each row represents speech materials. Each column represents subject. Group mean scores are presented in the most right column. Infinity represents quiet condition.

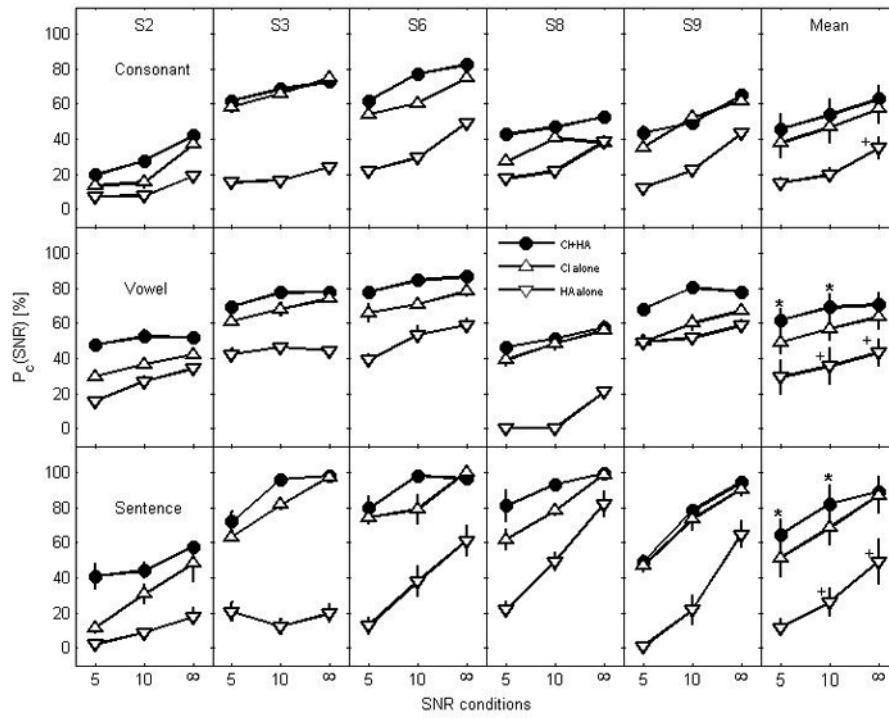


Figure 3. Individual and group mean recognition performance with standard error for the good group as a function of SNR with CI+HA, CI alone and HA alone. Each row represents speech materials. Each column represents subject. Group mean scores are presented in the most right column.

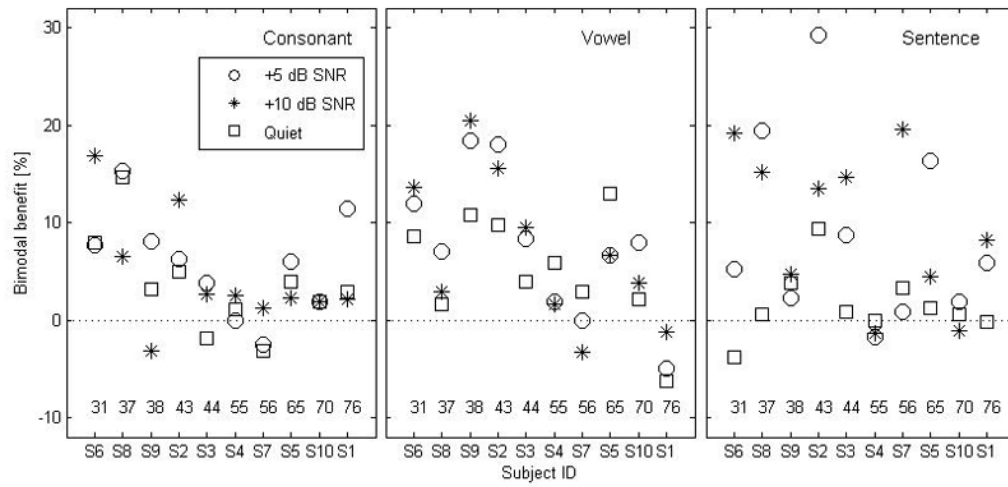


Figure 4. Bimodal advantage for individual listener in terms of aided PTA and SNR. Data point measured at each SNR is indicated with different symbols. The values of the aided PTA over frequencies of 0.25, 0.5, 0.75, and 1 kHz are also given at the bottom of each panel.

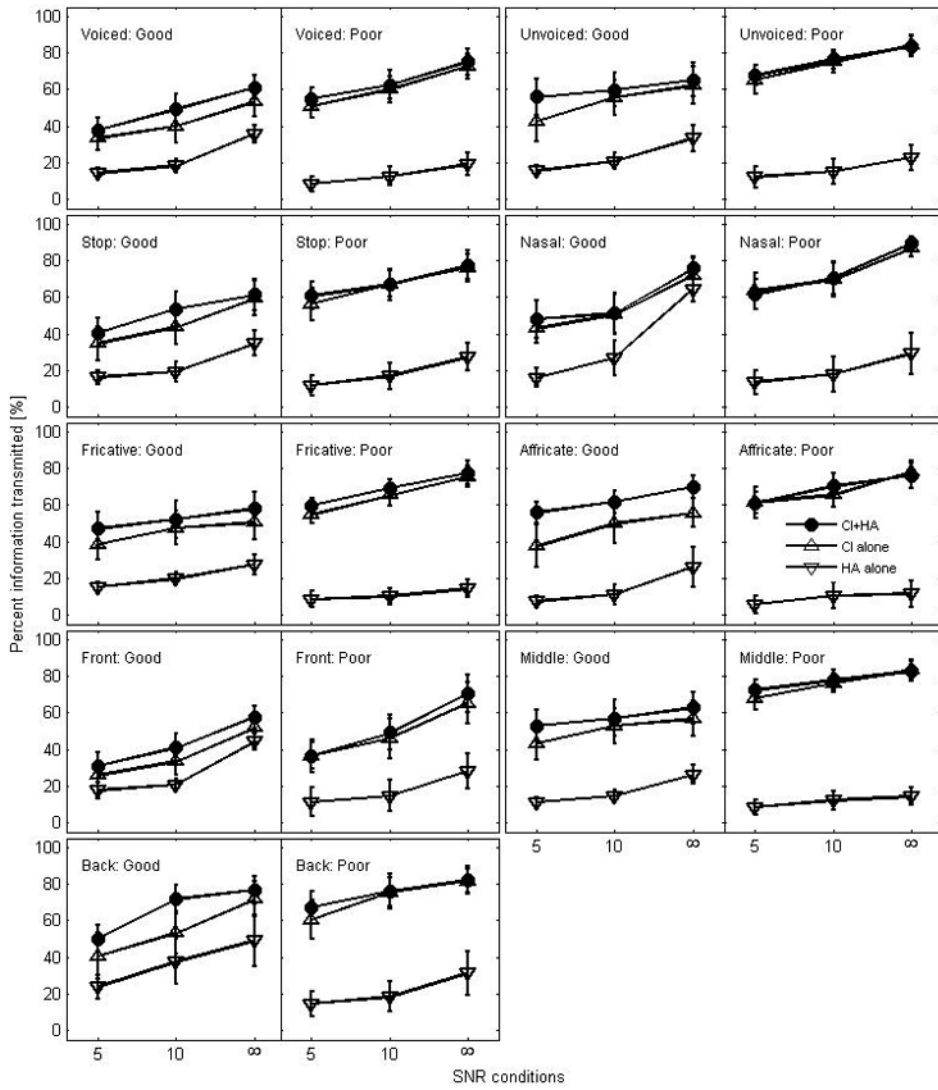


Figure 5. Percent information transmitted in consonant recognition for each group in terms of voicing, manner, and place cues.

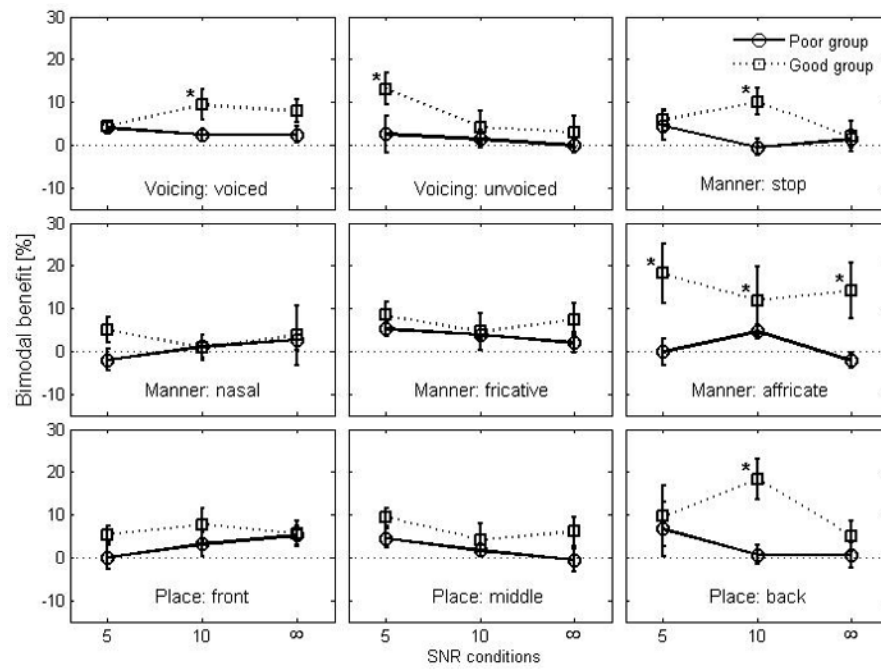


Figure 6. Mean bimodal benefit in consonant recognition for each group in terms of voicing, manner, and place cues. The error bar represents standard error.

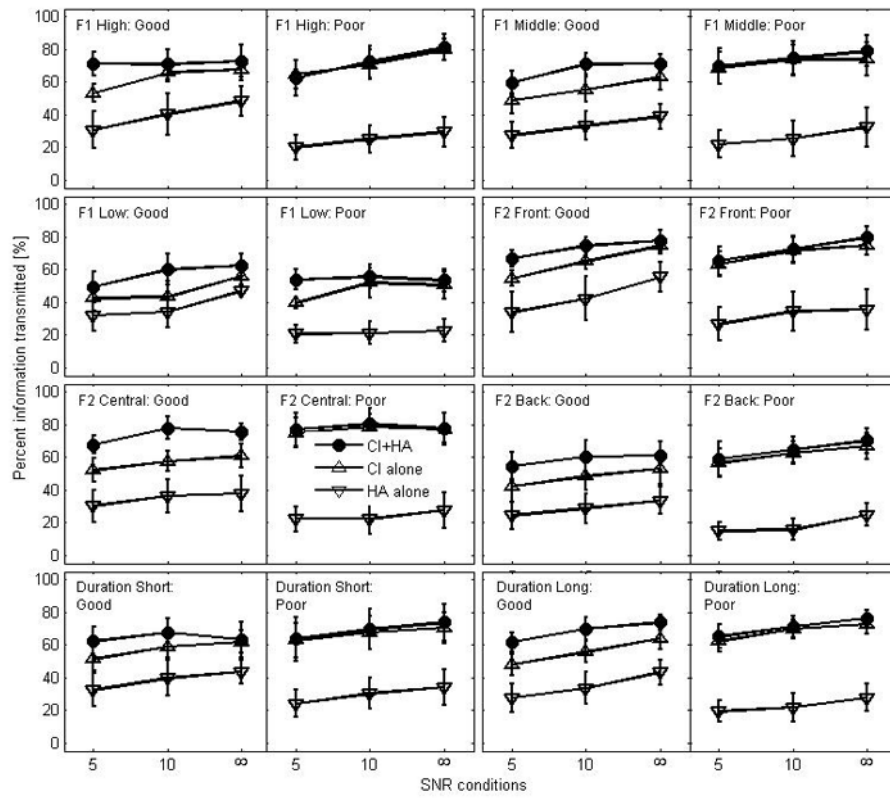


Figure 7. Percent information transmitted in vowel recognition for each group in terms of height of F1, place of F2, and duration cues.

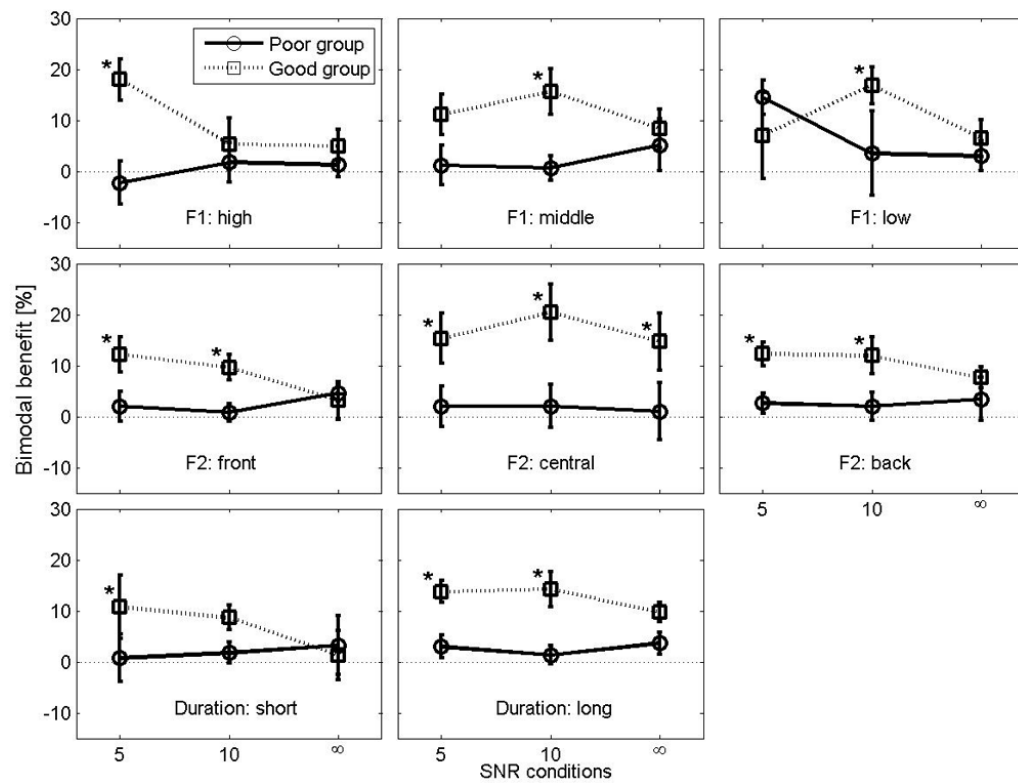


Figure 8. Mean bimodal benefit in vowel recognition for each group in terms of height of F1, place of F2, and duration of sounds. The error bar represents standard error.

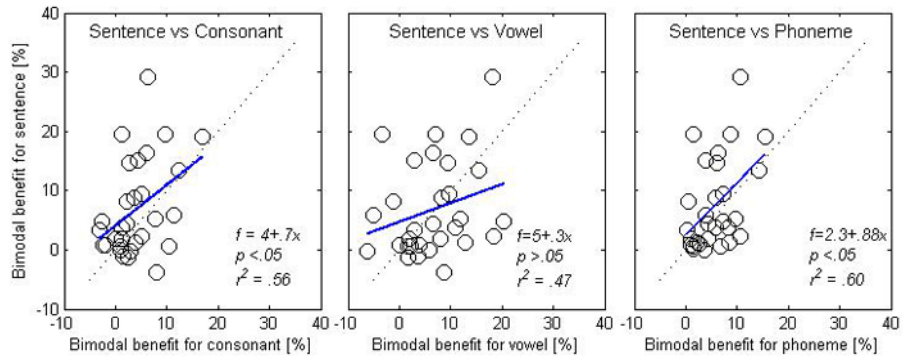


Figure 9. Relationship in bimodal advantage among speech test. The correlation in bimodal benefit between sentence and consonant recognition is given in the left panel, the correlation between sentence and vowel is in the middle panel, and the correlation between sentence and phoneme is in the right panel. The dotted line indicates iso benefit between speech tests, and the solid line indicates the linear fitting.

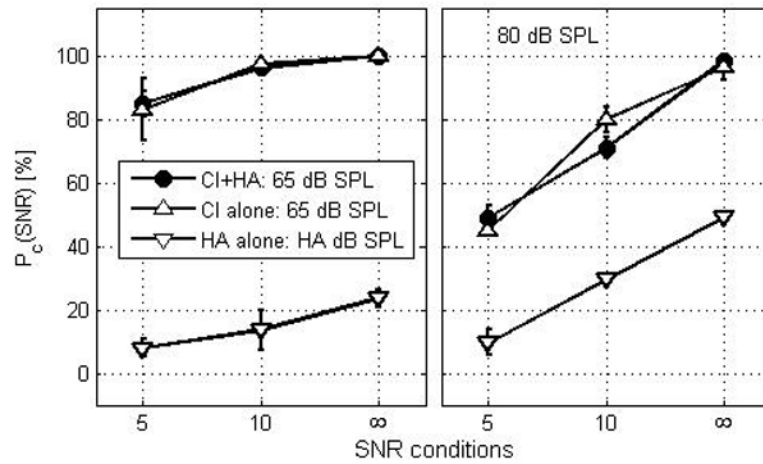


Figure 10. Comparisons in performance with single subject (S4) at two presentation level: 65 dB SPL for CI alone and CI+HA conditions and 70 dB SPL for HA alone (left panel) and 80 dB SPL (right panel) for all listening conditions.

Table 1

Subject demographics. R is right ear and L is left ear. Aided PTA = Pure tone average for frequencies of 0.25, 0.5, 0.75, and 1 kHz in the acoustic side, CI = cochlear implant, HA = hearing aid, SPEAK=Spectral PEAK, ACE=Advanced Combination Encoder, and HiRes=High Resolution.

ID (age, gender) (PTA in dB HL)	Device	Years of experience	Etiology	Device, Processor, strategy	Presentation level dB (A) SPL
S1 (77, F) (76)	R: CI	11	Unknown	Nucleus 24, Freedom, ACE Phonak, Savia	70
	L: HA	34	Unknown		70
	Both	11			65
S2 (44, M) (43)	R: CI	13	Rubella	Nucleus 24, Freedom, SPEAK Phonak, 313	70
	L: HA	43	Rubella		70
	Both	13			65
S3 (77, M) (44)	L: CI	11	Unknown	120, Harmony, HiRes Widex, Senso	65
	R: HA	11	Unknown		70
	Both	11			65
S4 (73, M) (55)	R: CI	1	Unknown	Nucleus 24, Freedom, ACE Phonex, CROS	65
	L: HA	36	Unknown		70
	Both	1			65
S5 (71, M) (65)	L: CI	3	Otosclerosis	Nucleus 24, Freedom, ACE Simens ETE	65
	R: HA	35	Otosclerosis		70
	Both	3			65
S6 (25, M) (31)	L: CI	1.5	Meningitis	Nucleus 5, Freedom, ACE Phonak, Supero	65
	R: HA	6	Meningitis		65
	Both	1.5			65
S7 (56, F) (56)	R: CI	2	Ototoxicity	C2, Harmony, HiRes Widex Body	70
	L: HA	8	Ototoxicity		70
	Both	2			70
S8 (69, M) (37)	L: CI	3	Unknown	Nucleus 24, Freedom, ACE Unitron, 360+	65
	R: HA	50	Unknown		70
	Both	3			65

ID (age, gender) (PTA in dB HL)	Device	Years of experience	Etiology	Device, Processor, strategy	Presentation level dB (A) SPL
S9 (47, M) (38)	L: CI	2	Unknown	Nucleus 24, Freedom, ACE Simens, Centra	65
	R: HA	42	Unknown		70
	Both	2			65
S10 (67, F) (70)	L: CI	7	Hereditary	3G, Freedom, ACE Resound, Metrix	65
	R: HA	37	Hereditary		70
	Both	7			65

Classification for acoustic features: For consonants, 0-voiced, 1-unvoiced, 2-stop, 3-nasal, 4-fricative, 5-affricate, 6-place, front, 7-place, middle, and 8-place, back; for vowels, 0-high F1, 1-middle F1, 2-low F1, 3-front F2, 4-middle F2, 5-back F2, 6-short duration, and 7-long duration.

Table 2

Consonants																
	/b/	/d/	/g/	/p/	/t/	/k/	/m/	/n/	/ŋ/	/s/	/ʃ/	/v/	/d/	/z/	/l/	/r/
Voicing	0	0	0	1	1	1	0	0	1	1	1	0	0	0	1	0
Manner	2	2	2	2	2	2	3	3	4	4	4	4	4	4	5	5
Place	6	7	8	6	7	8	6	7	6	7	7	6	7	7	7	7

Vowels														
	/æ/	/ɑ/	/ɔ/	/ɛ/	/e/	/i/	/ɪ/	/ə/	/o/	/ɒ/	/ʌ/	/ʊ/	/ɯ/	/u/
F1, Height	2	2	1	1	1	1	0	0	1	0	1	0	1	0
F2, Place	3	5	5	3	3	4	3	3	5	5	4	5	4	5
Duration	7	7	7	6	7	7	6	7	7	6	6	7	6	7

Table 3

The results of two-ways ANOVA for consonant bimodal advantage. The analysis was performed with two factors, listening Mode and SNR for each group. The same analysis was performed with two factors, Group and SNR between groups. Significant effect was indicated with bold face.

Consonant feature		PTA ≥ 55 dB HL	PTA < 55 dB HL	PTA ≥ 55 dB HL vs. PTA < 55 dB HL
Voicing	Mode	$F(1,8)=6.6, p>.05$	$F(1,8)=18.4, p<.01$	Group $F(1,16)=5, p<.05$
	SNR	$F(2,8)=23.3, p<.001$	$F(2,8)=39, p<.001$	SNR $F(2,16)=3, p>.05$
Unvoiced	Mode	$F(1,8)=2.8, p>.05$	$F(1,8)=5.4, p>.05$	Group $F(1,16)=2.1, p>.05$
	SNR	$F(2,8)=14.8, p<.005$	$F(2,8)=11, p<.005$	SNR $F(2,16)=4, p<.05$
Stop	Mode	$F(1,8)=8, p>.05$	$F(1,8)=5, p<.05$	Group $F(1,16)=1.4, p>.05$
	SNR	$F(2,8)=40, p<.001$	$F(2,8)=53, p<.001$	SNR $F(2,16)=1.5, p>.05$
Nasal	Mode	$F(1,8)=1, p>.05$	$F(1,8)=3, p>.05$	Group $F(1,16)=1, p>.05$
	SNR	$F(2,8)=7.3, p<.01$	$F(2,8)=30, p<.001$	SNR $F(2,16)=1.7, p>.05$
Manner	Mode	$F(1,8)=5.2, p>.05$	$F(1,8)=6.8, p>.05$	Group $F(1,16)=1, p>.05$
	SNR	$F(2,8)=12.6, p<.005$	$F(2,8)=7.5, p<.01$	SNR $F(2,16)=6, p>.05$
Affricate	Mode	$F(1,8)=.5, p>.05$	$F(1,8)=8.6, p<.05$	Group $F(1,16)=4.6, p<.05$
	SNR	$F(2,8)=9, p<.005$	$F(2,8)=9, p<.005$	SNR $F(2,16)=4, p>.05$
Front	Mode	$F(1,8)=1.4, p>.05$	$F(1,8)=14, p<.05$	Group $F(1,16)=1.5, p>.05$
	SNR	$F(2,8)=17, p<.001$	$F(2,8)=28, p<.001$	SNR $F(2,16)=8, p>.05$
Middle	Mode	$F(1,8)=1.7, p>.05$	$F(1,8)=9.4, p<.05$	Group $F(1,16)=3, p>.05$
	SNR	$F(2,8)=23, p<.001$	$F(2,8)=11, p<.005$	SNR $F(2,16)=2, p>.05$
Back	Mode	$F(1,8)=1.1, p>.05$	$F(1,8)=6.3, p<.05$	Group $F(1,16)=7.2, p<.05$
	SNR	$F(2,8)=10.2, p<.005$	$F(2,8)=38, p<.001$	SNR $F(2,16)=2, p>.05$

Table 4

The results of two-ways ANOVA for vowel bimodal advantage. The detail description is the same as those in Table 3.

Vowel feature	PTA \geq 55 dB HL	PTA < 55 dB HL	PTA \geq 55 dB HL vs. PTA < 55 dB HL
High	Mode	$F(1,8)=1, p>.05$	$F(1,16)=4.2, p>.05$
	SNR	$F(2,8)=10.4, p<.05$	$F(2,16)=1.8, p>.05$
Middle	Mode	$F(1,8)=4, p>.05$	$F(1,16)=6, p<.01$
	SNR	$F(2,8)=4.9, p>.05$	$F(1,16)=6, p<.05$
Low	Mode	$F(1,8)=2.4, p>.05$	$F(1,16)=2.4, p>.05$
	SNR	$F(2,8)=1.6, p>.05$	$F(2,16)=1.4, p>.05$
Front	Mode	$F(1,8)=1.8, p>.05$	$F(1,16)=3, p>.05$
	SNR	$F(2,8)=14, p<.005$	$F(2,16)=2.3, p>.05$
Central	Mode	$F(1,8)=14, p>.05$	$F(1,16)=5, p<.05$
	SNR	$F(2,8)=6, p>.05$	$F(2,16)=1.6, p>.05$
Back	Mode	$F(1,8)=1.6, p>.05$	$F(1,16)=9, p<.05$
	SNR	$F(2,8)=3, p>.05$	$F(2,16)=25, p>.05$
Short	Mode	$F(1,8)=2, p>.05$	$F(1,16)=1, p>.05$
	SNR	$F(2,8)=2.6, p>.05$	$F(2,16)=.97, p>.05$
Long	Mode	$F(1,8)=3, p>.05$	$F(1,16)=15, p<.05$
	SNR	$F(2,8)=8.2, p<.05$	$F(2,16)=.5, p>.05$