Published in final edited form as:
J Speech Lang Hear Res. 2014 August 1; 57(4): 1512–1520. doi:10.1044/2014_JSLHR-H-13-0127.

The perception of consonants in reverberation and noise by adults fitted with bimodal devices

Michelle Mason and **Kostas Kokkinakis**
Department of Speech-Language-Hearing, University of Kansas, 1000 Sunnyside Avenue, Lawrence, KS 66045, USA

Abstract

Purpose—The purpose of this study was to evaluate the contribution of a contralateral hearing aid (HA) to the perception of consonants, in terms of voicing, manner, and place of articulation cues in reverberation and noise by adult cochlear implantees aided by bimodal fittings.

Method—Eight post-lingually deafened adult cochlear implant listeners with a fully inserted cochlear implant in one ear and low-frequency hearing in the other ear were tested on consonant perception. The subjects were presented with consonant stimuli processed in the following experimental conditions: one quiet condition, two different reverberation times (0.3 s and 1.0 s), and the combination of two reverberation times with a single signal-to-noise ratio (SNR = 5 dB).

Results—Consonant perception improved significantly when listening in combination with a contralateral hearing aid (HA) as opposed to listening with a cochlear implant (CI) alone in 0.3 s and 1.0 s of reverberation. Significantly higher scores were also noted when noise was added to 0.3 s of reverberation.

Conclusion—A considerable benefit was noted from the additional acoustic information in conditions of reverberation and reverberation plus noise. The bimodal benefit observed was more pronounced for voicing and manner of articulation than for place of articulation.

Keywords

hearing aid; cochlear implant; bimodal benefit; voicing; manner; place; reverberation; noise

1. Introduction

Currently, almost all cochlear implant devices perform well in quiet settings. The majority of congenitally deaf cochlear implant recipients can achieve high open-set speech recognition scores regardless of the device or speech coding strategy used (e.g., see Skinner et al., 2002; Spahr and Dorman, 2004). Hence, it is no surprise that cochlear implantation has become increasingly prevalent. Yet, optimizing speech perception by users of such devices in complex listening conditions to better approximate their performance to that of individuals with normal auditory function continues to be an active area of research.

Please address correspondence to: Kostas Kokkinakis, Ph.D., Department of Speech-Language-Hearing, University of Kansas, 1000 Sunnyside Avenue, Lawrence, KS 66045. Phone: (785) 864-4467; Fax: (785) 864-3974 kokkinak@ku.edu.

A prime example of these efforts is that the trend to opt for electrical stimulation in both ears (bilateral electrical hearing) has recently begun to shift. An increasing number of patients who have functional residual acoustic hearing (up to frequencies in the range of 500 or 750 Hz) are now offered the option of complementing the cochlear implant (CI) on one side with a contralateral hearing aid (HA) on the non-implanted side. This paradigm is known as bimodal stimulation (CI+HA). Bimodal stimulation combines electrically elicited with acoustically elicited information. This has been shown to provide additional low-frequency cues through the hearing aid normally not transmitted by the CI device (Tyler et al., 2002; Ching et al., 2004, 2007). Attempting to elicit an electro-acoustic advantage has also led efforts towards preserving existing natural hearing during implantation in order to complement the high-frequency cues conveyed by electrical hearing with acoustic information. To preserve low-frequency hearing, it has become a relatively common practice to implant a short electrode array to a depth of approximately 8 to 10 mm beyond the cochleostomy instead of the traditional long electrode that is normally inserted in depths approaching or even extending beyond 1.5 turns of the cochlea (e.g., see Gantz and Turner, 2003; Gifford et al., 2007, 2008).

The combination of electric hearing and either ipsilateral or contralateral acoustic hearing has been shown to enhance the perception of acoustic cues both in quiet and in noise (e.g., see Turner et al., 2004; Mok et al., 2006; Zhang et al., 2010; Incerti et al., 2011; Sheffield and Zeng, 2012). A benefit due to electro-acoustic stimulation has been observed even when the acoustic stimulation alone provides little to no speech understanding. The observed benefit stemming from the combined use of acoustic and electric stimulation is often attributed to a better delivery of the fundamental frequency (F0) cue, which is present in the low-frequency acoustic signal (Brown and Bacon, 2010; Cullington and Zeng, 2011). The F0 cue is delivered primarily from the added acoustic stimulation, since electrical stimulation provides only weak low-frequency acoustic information. Zhang et al. (2010) demonstrated a significant improvement in the perception of monosyllabic words in quiet when acoustic information was added. An average improvement of 22 percentage points over the CI-alone condition was observed in the CI+HA condition even when the complementary acoustic information only contained frequencies up to 125 Hz. Mok et al. (2006) observed that a significant bimodal advantage (i.e., the benefit arising from aided by both a HA and CI as opposed to relying on a CI alone) existed in +10 dB signal-to-noise ratio (SNR) noise as well as in quiet. The authors theorized that phonetic information in the low-frequency region (e.g., consonant cues) contained within the acoustic signal was primarily responsible for the benefit observed in noise.

Kong and Carlyon (2007) investigated the contribution of F0 and low-frequency cues in speech recognition using vocoded speech. The authors used noise-vocoder simulations (left ear) combined with contralateral simulations of profound high-frequency hearing loss (right ear) to assess performance in noise by normal-hearing listeners. They found that improved speech recognition in noise was due to low-frequency phonetic information rather than solely F0 cues. That is, the combined hearing advantage remained even when the F0 cue was removed. Therefore, the authors suggested that voicing or temporal glimpsing cues may also play a role in enhancing speech recognition in noise. Sheffield and Zeng (2012) tested hearing-impaired listeners with bimodal fittings on both vowel and consonant recognition in noise. The authors argued the presence of a dual mechanism relating to the bimodal benefit. Adding an acoustic tone modulated to represent the F0 and amplitude envelope contours of the target speech was shown to provide enough voicing and manner information to improve consonant recognition in noise. Yet, improved perception of both vowels and consonants was observed with the addition of acoustically presented low-pass filtered speech. This suggests that the addition of complementary low-frequency acoustic cues on the opposite side should improve overall phoneme recognition at a greater extent than by simply presenting information about the amplitude envelope of the F0 component (e.g, see Kong and Carlyon 2007; Sheffield and Zeng, 2012).

Despite a growing body of literature emphasizing the benefits of bimodal speech perception in noise, to-date there has been only a single study on the performance of electro-acoustic stimulation in reverberation. Tillery et al. (2012) used simulated electro-acoustic hearing to measure sentence recognition in reverberation. Normal-hearing listeners were tested on sentences corrupted with reverberation times equal to 0, 0.125, 0.25, 0.5, and 1.0 s using simulated electro-acoustic and electrical hearing. They found that speech intelligibility of vocoded speech alone (simulated electrical hearing) dropped by 30 percentage points when sentences were corrupted with a short reverberation time of 0.125 s. The addition of the acoustic component (500 Hz low-pass filtered speech) yielded an average advantage of almost 40 percentage points across different acoustic conditions. This addition of low-pass filtered speech to the vocoded signal restored a substantial degree of intelligibility in reverberation but only a negligible benefit was noted in reverberated noise. These findings confirm results from Kokkinakis et al. (2011) who showed that speech intelligibility by cochlear implant users in reverberation will degrade exponentially as reverberation increases. Tillery et al. (2012) and Kokkinakis et al. (2011) used slightly different reverberation times and two different subject populations (NH and CI listeners, respectively). Both studies, however, concluded that the intelligibility of electrically-processed IEEE sentences (IEEE, 1969) will drop by more than 60 percentage points on average when reverberation increases from 0 s to 1.0 s.

The effects of reverberation when combined with background noise on normal-hearing listeners (Neuman et al., 2010) and on adults with cochlear implants (Hazrati and Loizou, 2012) have also been recently evaluated. In the Neuman et al. (2010) study, adults were tested with Bamford-Kowal-Bench Speech in Noise (BKB-SIN) sentences corrupted with four-talker babble and reverberation times equal to 0.3, 0.6 and 0.8 s. A higher signal-to-noise ratio was required as reverberation time increased in order to achieve 50% speech intelligibility. The average speech reception threshold (SRT) for the adult participants was around −2.5 dB in the anechoic condition. Their mean threshold elevated to 0 dB in reverberation. The combined (adverse) effects of reverberation plus noise were greater than those introduced by either reverberation or noise alone for cochlear implant listeners (Hazrati and Loizou, 2012).

Speech intelligibility by cochlear implant listeners in reverberation and reverberation plus noise has been shown to decrease substantially. Yet, an insufficient amount of data exists to conclude on the effects of reverberation and reverberation plus noise on cochlear implant and bimodal devices at the phonemic (e.g., consonant) level. The objective of the present study was to conduct a comprehensive assessment of the effects of reverberation and reverberation plus noise on the linguistic cues of English consonants (e.g., manner and place of articulation as well as voicing). An information transfer analysis was performed in each listening condition to assess the availability of these linguistic cues. We tested whether the addition of a HA in the contralateral ear elicits a bimodal benefit in this regard and to which degree such a benefit can be observed in reverberation and reverberation plus noise.

2. Methods

2.1 Subjects

Eight post-lingually deafened adult cochlear implant users with a fully inserted (i.e., long electrode array) cochlear implant in one ear and aided low-frequency hearing in the other ear took part in the study. All subjects were native speakers of American English and had acquired at least 12 months of experience with their device post-implantation prior to testing. Unaided air conduction thresholds were obtained in the non-implanted ear using TDH-50 headphones and a GSI-16 audiometer at 250, 500, 1,000, 2,000, 4,000, and 8,000 Hz. The individual audiometric thresholds in the non-implanted ears are plotted in Figure 1.

Eight young normal-hearing (NH) listeners with no history of speech or hearing disorders were also recruited from the undergraduate student population of the University of Kansas. Extra credit was given for participation. This study was approved by the Human Subjects Committee of the University of Kansas in Lawrence (HSCL). All subjects gave informed consent prior to the beginning of testing and a case history interview was conducted with each subject to determine eligibility. The subjects were paid an hourly wage for their participation. Table 1 provides complete demographic information for the individuals tested.

2.2 Hearing Aid Fittings

All of the bimodal participants tested were experienced hearing aid users. Prior to enrollment in our study, all of the bimodal subjects were questioned on the use of their hearing aid. Only users who used the HA device at least 75% of the time relatively to CI use were included in the study. Although all of the subjects tested, used their hearing aid on a regular basis, the type of HA worn and the method of fitting the HA differed among subjects. Hence, prior to the first study visit, each subject’s audiologist was contacted to verify the hearing aid settings.

All of the participants tested wore behind-the-ear (BTE) hearing aid devices. Subjects were fitted with either semi-occluding or fully-occluding earmolds. Prior to testing, the hearing aids were programmed in basic mode. In basic mode, the hearing aid used only a single omnidirectional microphone and all advanced features, except for feedback management, were disabled, including directional processing and digital noise reduction (DNR). For all participants, the volume on the HA was adjusted to provide comfortable listening for soft speech in quiet conditions.

2.3 Stimuli

Twenty consonant sounds of American English (broad transcription) were organized into nonsense syllables using the /aCa/ context, where C = p, b, t, d, k, g, f, v, ð, s, z, ʃ, ʧ ʤ, m, n, r, l, w, and j. Four male American English speakers produced the syllables resulting in 80 speech tokens in total (20 consonants × 4 speakers). The stimuli were recorded inside a double-walled IAC sound-attenuating booth. All stimuli were recorded at the sampling rate of 44,100 Hz. The root-mean-square amplitude of all syllables was equalized to the same value corresponding to approximately 65 dB SPL.

MATLAB was used to generate the stimuli corrupted with reverberation and reverberation plus noise. Head related transfer functions (HRTFs) recorded by Rychtarikova et al. (2009) were used to simulate reverberant conditions. To obtain measurements of HRTFs, the authors used a CORTEX MKII manikin artificial head placed inside a rectangular reverberant room with dimensions 5.50 m × 4.50 m × 3.10 m (length × width × height) and a total volume of 76.80 m^3. The average reverberation time of the experimental room (average in one-third-octave bands with center frequencies between 125 Hz and 4,000 Hz) before any modification was equal to 1.0 s. The average reverberation time was reduced to 0.3 s by increasing the number of acoustic panels and by adding floor carpeting as well as highly absorbent rectangular acoustic boards to the room. This latter value corresponds to a well-dampened room and is typical of the lowest reverberation time that might be found in a small office room.

To obtain HRTF recordings for each reverberation condition, the artificial head was placed in the middle of a ring of 1.25 m inner diameter. A single-cone loudspeaker (Tannoy Reveal 501A) was placed at a 0° azimuth in the frontal plane. A two-channel sound card (VX Pocket 440 DIGIGRAM) and DIRAC 5.5 software type 7841 (Bruel and Kjaer Sound and Vibration Measurement Systems) were used to generate the acoustic impulse responses for the creation of the test materials. All recordings were conducted using identical microphones to those used in modern BTE speech processors. For the reverberation plus noise conditions, speech-shaped noise (1,000 Hz low-pass cut-off frequency, −12 dB per octave) was used. The reverberation plus noise corrupted stimuli were generated by first adding the speech-shaped noise to the anechoic signal at the target signal-to-noise ratio and then by convolving with each of the two room acoustic impulse responses corresponding to each reverberation condition as described in Tillery et al. (2012).

2.4 Procedure

Subjects participated in a total of five experimental conditions corresponding to: (1) one anechoic (no reverberation) and quiet (no noise) condition ($RT = 0$ s, $SNR = \infty$), (2) two different reverberation times ($RT = 0.3$ s and $RT = 1.0$ s), and (3) the combination of the two different reverberation times with a single SNR level (e.g., $RT = 0.3$ s, $SNR = +5$ dB). We denote the anechoic and quiet stimuli as Q, the reverberant stimuli as R (e.g., R03 and R10) and the reverberant plus noisy stimuli as R + N (e.g., R03N5 and R10N5).

In total, there were 2,000 stimuli presented (20 consonant syllables × 4 talkers × 5 listening conditions × 5 repetitions). For all the different listening conditions, each bimodal subject was tested with: cochlear implant (CI) alone, and cochlear implant plus hearing aid (CI+HA). In the CI-alone condition, the HA device was turned off and removed. A silicone rubber earplug was inserted in the non-implanted ear. The earplug was fitted with the maximum depth of insertion that could be safely achieved (80% in the canal). According to the manufacturer's specifications, the attenuation provided by the earplug was approximately 30 dB for frequencies between 125 Hz and 1,000 Hz. A USB audio capture device (M-Audio C 600) was connected to the computer and provided separate left and right audio output jacks and a volume control dial. The M-Audio interface was connected to the cochlear implant and the hearing aid devices via direct audio input (DAI). This type of connection ensured that the everyday programs on both devices were used.

A graphical user interface (GUI) was developed in MATLAB and was used to present the stimuli and collect responses from the participants. The GUI software enabled selected sound files on the computer to be presented to listeners at a controlled presentation level. The files were presented in a random order with the responses entered directly into the program. Each participant was instructed on how to use the GUI prior to testing. As described to the participants, the task was to identify the consonant in the syllable by selecting it from the 20 consonant candidates displayed on the screen. The GUI also included sample words that served as a reference for the target sound.

For all testing conditions, the participants were asked to use the settings (e.g., programs) that they would normally use for everyday listening. Participants were also asked to adjust the individual settings of the HA device to match the loudness of their CI device for the voice of the tester, which was presented at 65 dB SPL. Note that in the CI-alone conditions, the settings of the CI device (volume and sensitivity controls) were not adjusted again in order to ensure that the same settings were retained on the CI side for both the CI+HA and CI experimental conditions. In the normal-hearing group of listeners, the same stimuli (e.g., consonant syllables) were delivered through high quality circumaural headphones (Sony MDR–7509HD) at the same presentation level.

The participants were seated inside the double-walled IAC sound-attenuating booth. They listened to all conditions and selected the token that they perceived on the GUI. The presentation of conditions was randomized to decrease order effects. For each listening condition, each test block consisted of 400 syllables and all four talkers randomly produced a token. The next token was randomly selected and presented once the subject responded. The test was divided into five sessions (one session per listening condition) and each session took each subject approximately 1 hour to complete. The participants were given a 10-minute break every 60 minutes between the testing sessions to control fatigue. No feedback was provided.

3. Results

3.1 Consonant recognition

Figure 2 depicts the mean percent correct scores for consonant recognition obtained in the various listening and device conditions. The different device conditions are shown along the abscissa. For each listening condition, the scores obtained from the NH population are plotted for comparison. Two-way analysis of variance (ANOVA) (with repeated measures) using the five listening conditions (quiet, two reverberation times, two reverberation times plus noise) and two device conditions (CI and CI+HA) as within-subject factors indicated a significant effect of the device ($F[1, 7] = 52.77$, $p = 0.001$), a significant effect of the listening condition ($F[4, 28] = 133.62$, $p < 0.001$), and a significant interaction ($F[4, 28] = 2.87$, $p = 0.04$) between the listening condition and the device used. ANOVA was conducted on the rationalized arcsine unit-transformed values in order to normalize the variance of the scores. A critical value of alpha equal to 0.05 was set as the significance level.

Post-hoc comparisons using Tukey's honest significant difference tests indicated that consonant recognition scores obtained in the CI+HA conditions were not significantly higher than those obtained in the CI-alone condition in quiet ($p = 0.08$) and in the R10N5 condition ($p = 0.06$). However, the bimodal listeners tested scored significantly higher when listening in combination with the contralateral HA than when listening through their CI device only in the R03 condition ($p < 0.001$), R10 condition ($p = 0.003$) and R03N5 condition ($p = 0.001$). Significant improvements in consonant perception were observed when listeners relied on bimodal stimulation as opposed to when listening through the CI alone. This was the case in most listening conditions. The bimodal advantage (averaged across all subjects) was shown to be significant in both the R03 and R10 conditions (e.g., reverberation only), and was equal to 20 percentage points and 15 percentage points, respectively. A significant mean bimodal advantage of 17 percentage points was also observed in the R03N5 condition (e.g., reverberation plus noise). No significant bimodal benefit was observed in quiet (11 percentage points) or the R10N5 condition (10 percentage points). The average bimodal benefit for the subjects tested was calculated after scores obtained with CI-alone were subtracted from the scores obtained in the CI+HA device condition. The bimodal benefit is plotted in Figure 3 for all listening conditions tested.

3.2 Information transmission analysis

In order to determine whether specific consonant features differed across different listening conditions and devices, the 20 consonants were coded according to their articulatory features, which included voicing (i.e., voiced and unvoiced), manner of articulation (i.e., stop, fricative, affricate, nasal, liquid, glide) and place of articulation (i.e., bilabial, labiodental, interdental, alveolar, palatal, and velar). The consonants were coded as described in Table II. The confusion matrices for each participant and combination of listening condition and device condition (i.e., 8 subjects × 2 devices × 5 listening conditions) were re-arranged and analyzed based on the transmitted information of articulatory features as per Miller and Nicely (1955).

The information transmitted mean scores obtained for voicing, manner and place of articulation in all five different stimulus conditions are plotted in Figure 4. A three-way ANOVA (with repeated measures) using the listening condition (quiet, reverberation only, and reverberation plus noise), device (CI or CI+HA), and feature (voicing, manner, place) as within-subject factors indicated that there was a statistically significant three-way interaction. Therefore, three sets of two-way ANOVA (with repeated measures) were conducted to examine the effects of the listening condition as well as the effects of the device for all three different consonant features. For the feature of voicing, the ANOVA indicated a significant effect of the device ($F[1, 7] = 26.58$, $p = 0.001$), a significant effect of the listening condition ($F[4, 16] = 115.25$, $p < 0.001$), and a significant interaction ($F[4, 28] =8.48$, $p = 0.002$) between listening condition and device. For manner of articulation, the ANOVA indicated a significant effect of the device ($F[1, 7] = 11.49$, $p = 0.03$), a significant effect of the listening condition ($F[4, 28] = 120.98$, $p < 0.001$), and a significant interaction ($F[4, 28] = 4.08$, $p = 0.02$) between listening condition and device. For place of articulation, the ANOVA indicated a significant effect of the device ($F[1, 7] = 23.92$, $p = 0.009$), a significant effect of the listening condition ($F[4, 28] = 80.61$, $p < 0.001$), and a significant interaction ($F[4, 28] = 4.59$, $p = 0.04$) between listening condition and device.

Post-hoc comparisons using Tukey's honest significant difference tests indicated that more voicing information was transmitted when relying on CIs and HAs than when listening through CIs alone in the R03 condition ($p < 0.001$), R10 condition ($p = 0.006$), and R03N5 condition ($p = 0.002$). No significant differences were found for the transmission of voicing between the two different device conditions in quiet ($p = 0.607$) and the R10N5 condition ($p = 0.081$). Significant differences between devices were found in the R03 condition ($p < 0.001$) and R03N5 condition ($p = 0.025$) for manner of articulation. The information transmission scores for manner were not different between the two devices in the R10 condition ($p = 0.07$). The scores for manner of articulation were not significantly different in quiet ($p = 0.35$) and the R10N5 condition ($p = 0.187$). The transmission of place of articulation was not significantly different between the CI+HA and CI-alone in the R10 condition ($p > 0.05$), R03N5 condition ($p > 0.05$), and R10N5 condition ($p > 0.05$). However, the place of articulation features for the consonant stimuli were transmitted more efficiently through the bimodal fitting in quiet ($p = 0.006$) and the R03 condition ($p = 0.001$).

Post-hoc comparisons using Tukey's honest significant difference tests were conducted between different listening conditions separately for each feature. These indicated that bimodal listeners achieved better overall identification of consonant voicing cues in 0.3 s of reverberation than in 1.0 s of reverberation ($p = 0.01$). For the same device condition, the difference between 0.3 s and 1.0 s of reverberation was not significant ($p > 0.05$) in terms of manner of articulation. The perception of place was better in 0.3 s than in 1.0 s of reverberation ($p < 0.001$). The differences observed between the R03N5 and R10N5 conditions for the bimodal device were not significant for voicing ($p > 0.05$), manner ($p > 0.05$) or place of articulation ($p > 0.05$).

4. Discussion and Conclusions

The purpose of this study was to evaluate the specific contribution of the acoustic information to the perception of consonants in terms of voicing, manner, and place of articulation cues in adult cochlear implantees fitted with a contralateral hearing aid. We tested the hypothesis that the additional acoustic information provided in the low-frequency range due to the contribution of the HA device would yield better perception of these cues in reverberation and reverberation plus noise than that obtained with the CI alone.

The magnitude of the average bimodal advantage for consonant perception observed in quiet was not significant. In the quiet and anechoic condition (RT = 0 s, SNR = ∞), voicing and manner information was transmitted to a higher degree than information regarding place of articulation for both devices. Overall, the voicing feature was perceived better than the manner feature, which was also perceived better than the place feature. This has also been reported by other studies with CI and CI+HA devices (e.g., see Teoh et al., 2003; Incerti et al., 2011; Sheffield and Zeng, 2012), as well as NH listeners (Miller and Nicely, 1955).

The information transmission analysis in the quiet and anechoic condition for both devices (CI and CI+HA) suggested that the amount of voicing and manner information transmitted in the CI+HA condition was only marginally higher than the voicing and manner of articulation information transmitted in the CI-alone conditions (see Figure 4). In contrast, the transmitted place of articulation scores were considerably higher in quiet when listeners made use of the HA device (e.g., average 60%) than when consonants were perceived only through the CI device (e.g., mean value of 40%). The boost of information transfer with regards to place of articulation was not enough to yield a significant advantage in consonant perception in the optimal (quiet) listening condition.

The addition of the contralateral HA device significantly increased the transmission of voicing information over the CI alone in the reverberation-only conditions (R03 and R10). A significant difference was noted in the voicing transmission scores in the R03 and R10 conditions between the CI+HA and CI device conditions. Relative to electrical stimulation, the transmission of voicing features increased considerably in 0.3 s and 1.0 s of reverberation with bimodal stimulation. In contrast, a significant benefit due to the addition of the contralateral HA in the transmission of manner and place of articulation was observed only when reverberation was equal to 0.3 s. This benefit was around 25 percentage points for manner of articulation and 30 percentage points for place of articulation.

These findings suggest that low-frequency cues (e.g., voicing bars) as well as temporal voice onset time (VOT) cues were more resistant to the effects of reverberation (e.g., 0.3 s and 1.0 s) when acoustic hearing was available. This could be attributed to the fact that the contralateral HA provided more temporal speech cues than the cochlear implant alone. Because of the improved transmission of acoustic periodicity cues, the ability to distinguish between voiced and voiceless sounds in consonant phonemes was improved in reverberation. It seems that bimodal listeners can make use of this additional information across the two ears to improve consonant recognition even when confronted with a fairly long reverberation time (e.g., 1.0 s). Despite the fact that additive reverberation could blur the temporal onset and offset of the syllables, the overall temporal contour of the signal was at least partially preserved with the addition of the HA.

Additive acoustic reverberation affected the manner of articulation to a higher degree than voicing. Bimodal users scored high in terms of voicing and manner of articulation only in 0.3 s of reverberation. The amount of information transferred for both of these features decreased as reverberation time increased from 0.3 s to 1.0 s. The bimodal benefit due to the HA in the latter listening condition remained for voicing information but was not present for manner of articulation. Information about manner of consonant articulation relies mainly on the efficient transmission of amplitude envelope cues. The scores obtained for manner of articulation in the CI+HA condition indicated that such cues were not altered substantially in 0.3 s of reverberation. However, reverberant overlap-masking effects became dominant when the additive acoustic energy increased to 1.0 s (Kokkinakis and Loizou, 2011). It is possible that the reverberant energy originating from preceding phonemes leaked and filled the succeeding gaps present in low-energy consonants. This introduced a large number of manner of articulation errors.

The contribution of the bimodal stimulation to the perception of place was also found to be greater in 0.3 s than in 1.0 s of reverberation. High-frequency spectral (fine-structure) information and formant transitions are two main cues for the perception of place of articulation (Sheffield and Zeng, 2012). As the stimuli consists of an /aCa/ format, the HA enhanced the formant onset during the transition from consonant to vowel, while the CI enhanced the mid- to high-frequency spectral information (Incerti et al., 2011). Thus, bimodal listeners received a sufficient combination of cues to improve the perception of place of articulation when faced with a mildly distorted signal in 0.3 s of reverberation.

In the reverberation plus noise conditions, the CI+HA scores for voicing and manner of articulation were substantially better than the CI-only scores when reverberation was equal to 0.3 s. The benefit observed in voicing and manner for the reverberation plus noise condition (R03N5) was also significant. The scores for place of articulation were not different between the two devices. None of the consonant features were perceived well in 1.0 s of reverberation and +5 dB noise when listening through a CI+HA over a CI-alone.

The literature attributes the bimodal benefit observed in noise largely to the glimpsing of acoustic cues (e.g., see Kong and Carlyon, 2007). That is, the overall temporal fine structure is expected to improve by introducing acoustic cues in the low-frequency range (< 500 Hz). Glimpsing, also known as “dip-listening”, refers to the ability of detecting the target signal during the amplitude dips of a fluctuating masker (Li and Loizou, 2008). The improved temporal fine information due to the HA enhances the ability to listen in the dips and can thus facilitate better glimpsing. The glimpsing model can effectively explain why the perception of voicing and manner increased with the addition of a HA to the CI when noise (SNR = +5 dB) was added to the relatively short amount of reverberation. The perception of voicing and somewhat manner of articulation seemed to be more dependent on temporal information, which was relatively better preserved even in the presence of background noise.

Information about place of articulation was not preserved in the same listening condition. The interaction of noise and reverberation distorted all of the features examined but was most detrimental for place. This was due to the masking of high-frequency cues caused by noise. There is ample evidence to suggest that there is a predominance of place of articulation errors in reverberation and reverberation plus noise conditions (Helfer, 1994). In the Helfer (1994) study, consonant perception in NH listeners was measured for stimuli distorted by reverberation, noise, and reverberation plus noise in a binaural listening paradigm. It was noted that while voicing and manner of articulation were quite resistant to background noise, in general, place of articulation yielded the most errors.

Our findings point to a considerable benefit from the contribution of the HA device in listening situations whereby the stimuli were distorted by reverberation and reverberation plus noise. The bimodal benefit observed was more pronounced for voicing and manner of articulation than for place of articulation. This suggests that users of bimodal fittings will benefit from the added low-frequency acoustic information. It is expected that users will gain access to better voicing and manner cues that fall within this range. In contrast, place of articulation relied more on high-frequency cues transmitted predominantly through the CI. In the present study, voicing was found to be the most salient feature of consonants for the bimodal group not only in quiet, but also in reverberation as well as reverberation plus noise. The addition of the HA did not yield significantly higher scores for any of the features analyzed in severe reverberation plus noise. This implies that future speech enhancement strategies should focus on restoring both the temporal and spectral cues of speech.

Acknowledgments

This work was supported by grant R03 DC 00 8882 awarded from the NIDCD to Kostas Kokkinakis. We thank the individuals who participated in this study.

References

Brown CA, Bacon SP. Fundamental frequency and speech intelligibility in noise. Hearing Research. 2010; 266:52–59. [PubMed: 19748564]

Ching TYC, Incerti P, Hill M. Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. Ear and Hearing. 2004; 25:9–21. [PubMed: 14770014]

Ching TYC, Van Wanrooy E, Dillon H. Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review. Trends in Amplification. 2007; 11:161–192. [PubMed: 17709573]

Cullington HE, Zeng FG. Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification. Ear and Hearing. 2011; 32:16–30. [PubMed: 21178567]

Gantz BJ, Turner CW. Combining acoustic and electrical hearing. Laryngoscope. 2003; 113:1726–1730. [PubMed: 14520097]

Gifford RH, Dorman MF, Spahr AJ, McKarns SA. Combined electric and contralateral acoustic hearing: word and sentence intelligibility with bimodal hearing. Journal of Speech, Language and Hearing Research. 2007; 50:835–843.

Gifford RH, Dorman MF, Skarzynski H, Lorens A. Hearing preservation surgery: Psychophysical estimates of cochlear damage in recipients of a short electrode array. Journal of the Acoustical Society of America. 2008; 124:2164–2173. [PubMed: 19062856]

Hazrati O, Loizou PC. Tackling the combined effects of reverberation and masking noise using ideal channel selection. Journal of Speech, Language, and Hearing Research. 2012; 55:500–510.

Helfer KS. Binaural cues and consonant perception in reverberation and noise. Journal of Speech and Hearing Research. 1994; 37:429–438. [PubMed: 8028325]

IEEE. IEEE recommended practice speech quality measurements. IEEE Transactions on Audio and Electroacoustics. 1969; AU17:225–246.

Incerti P, Ching TYC, Hill A. Consonant perception by adults with bimodal fitting. Seminars in Hearing. 2011; 32:90–102.

Kokkinakis K, Hazrati O, Loizou PC. A channel-selection criterion for suppressing reverberation in cochlear implants. Journal of the Acoustical Society of America. 2011; 129:3221–3232. [PubMed: 21568424]

Kokkinakis K, Loizou PC. The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners. Journal of the Acoustical Society of America. 2011; 130:1099–1102. [PubMed: 21895052]

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; 121:3717–3727. [PubMed: 17552722]

Li N, Loizou PC. A glimpsing account for the benefit of simulated combined acoustic and electric hearing listeners. Journal of the Acoustical Society of America. 2008; 123:2287–2294. [PubMed: 18397033]

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, Language and Hearing Research. 2006; 49:338–351.

Neuman AC, Wroblewski M, Hajicek J, Rubinstein A. Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults. Ear and Hearing. 2010; 31:336–344. [PubMed: 20215967]

Rychtarikova M, van den Bogaert T, Vermeir G, Wouters J. Binaural sound source localization in real and virtual rooms. Journal of the Audio Engineering Society. 2009; 57:205–220.

Sheffield BM, Zeng FG. The relative phonetic contributions of a cochlear implant and residual acoustic hearing to bimodal speech perception. Journal of the Acoustical Society of America. 2012; 131:518–530. [PubMed: 22280613]

Skinner MW, Holden LK, Whitford LA, Plant KL, Psarros C, Holden TA. Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. Ear and Hearing. 2002; 23:207–223. [PubMed: 12072613]

Spahr AJ, Dorman MF. Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices. Archives Otolaryngology—Head and Neck Surgery. 2004; 130:624–628.

Teoh SW, Neuburger HS, Svirsky MA. Acoustic and electrical pattern analysis of consonant perceptual cues used by cochlear implant users. Audiology and Neuro-Otology. 2003; 8:269–285. [PubMed: 12904682]

Tillery KH, Brown CA, Bacon SP. Comparing the effects of reverberation and of noise on speech recognition in simulated electric-acoustic listening. Journal of the Acoustical Society of America. 2012; 131:416–423. [PubMed: 22280603]

Tyler RS, Parkinson AJ, Wilson BS, Witt S, Preece JP, Noble W. Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization. Ear and Hearing. 2002; 23:98–105. [PubMed: 11951854]

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; 115:1729–1735. [PubMed: 15101651]

Zhang T, Spahr AJ, Dorman MF. Frequency overlap between electric and acoustic stimulation and speech-perception benefit in patients with combined electric and acoustic stimulation. Ear and Hearing. 2010; 31:195–201. [PubMed: 19915474]

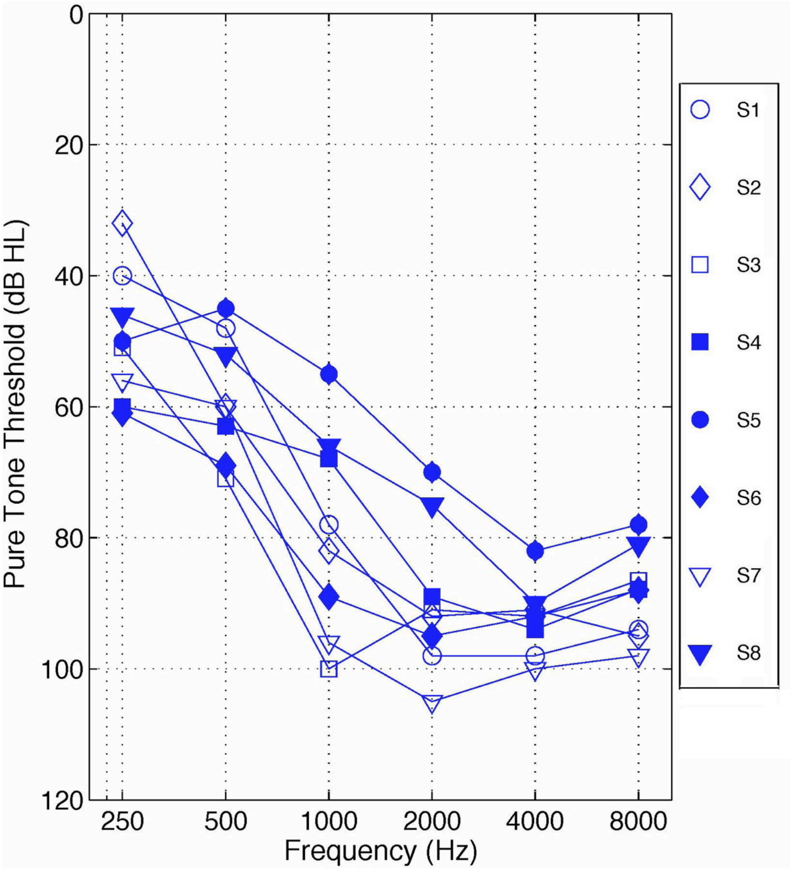

Figure 1.

Individual unaided audiometric thresholds for the eight bimodal (CI+HA) participants measured in the ear contralateral to the implant.

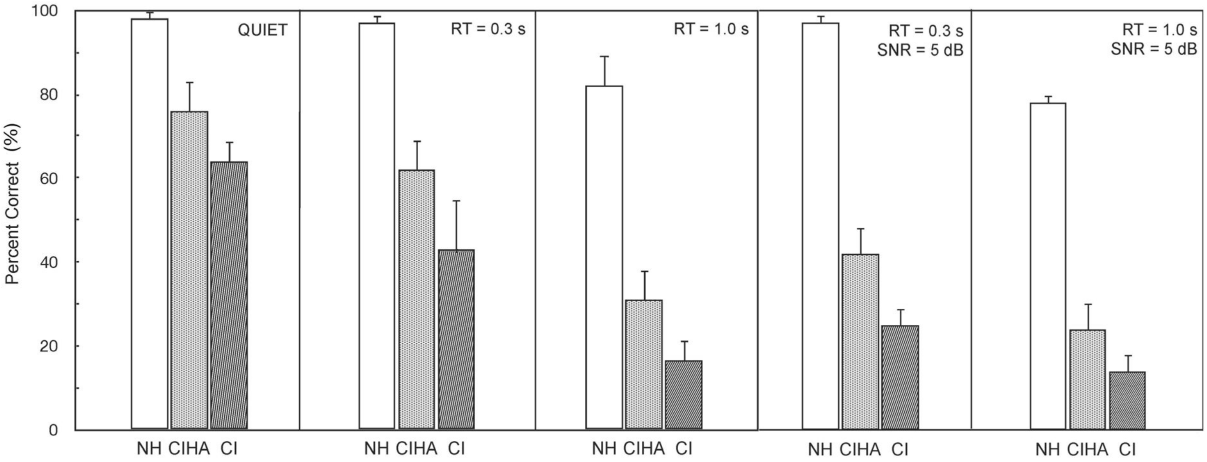

Figure 2.

Mean percent correct scores obtained from eight bimodal listeners tested on consonant syllables when aided with a cochlear implant (CI) alone and a cochlear implant and hearing aid (CIHA) in quiet, two reverberation conditions (R03 and R10) and two reverberation plus noise conditions (R03N5 and R10N5). Mean percent scores obtained from eight young normal-hearing (NH) listeners also plotted for comparison. Error bars represent standard deviations.

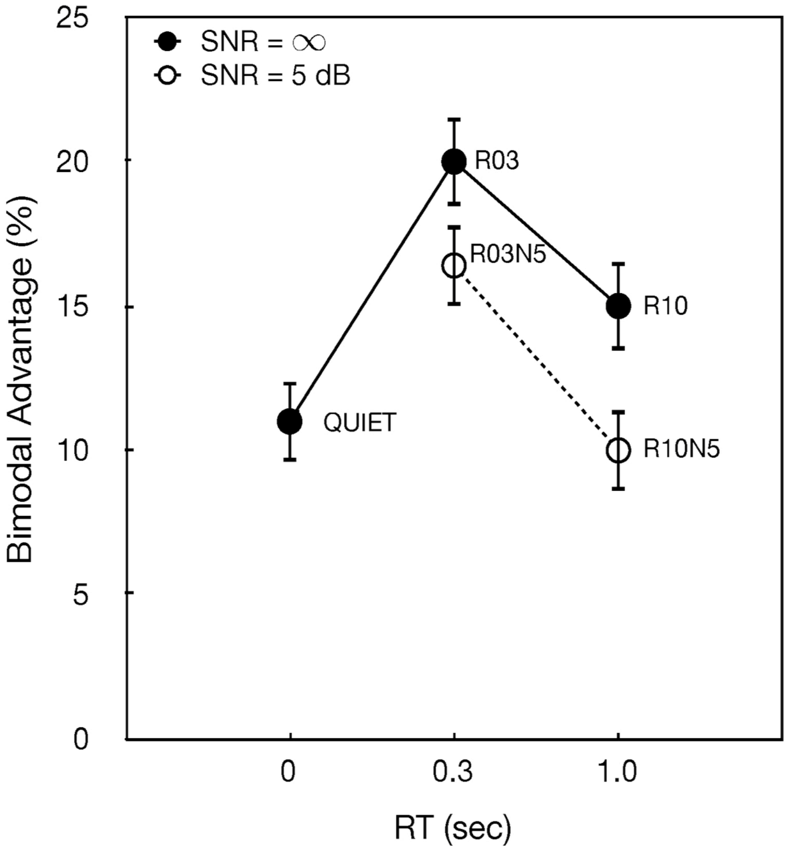

Figure 3.

Mean bimodal advantage for all eight bimodal listeners tested plotted against the different stimulus conditions. Error bars represent standard deviations.

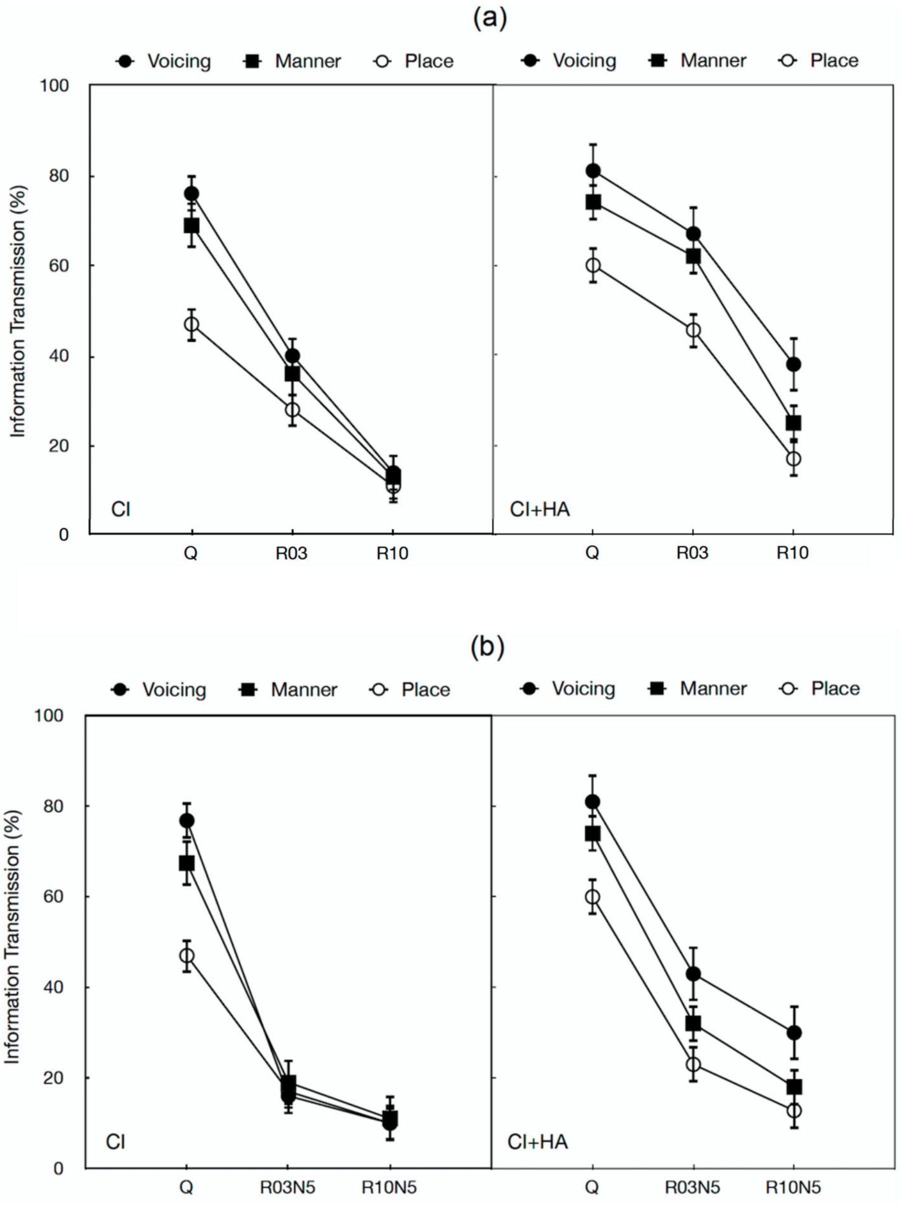

Figure 4.

Mean percent relative information transmitted for voicing, manner, and place of articulation in (a) anechoic and quiet (Q) and reverberation only (R03 and R10) and (b) anechoic and quiet (Q) and reverberation plus noise at +5 dB SNR (R03N5 and R10N5), is shown for the cochlear implant (CI) alone device condition and the cochlear implant and hearing aid device condition (CI+HA). Error bars represent standard deviations.

Table 1

Demographic details of participants.

Subject	Gender	Age at Testing	Months of Experience with Bimodal Stimulation	Etiology of Hearing Loss	Cochlear Implant Type	Contralateral Hearing Aid Type	Consonant-Nucleus-Consonant Word-Recognition Score in Quiet (% Correct)
S1	M	54	38	Hereditary	Nucleus 5	Phonak Naida	89%
S2	M	67	24	Noise exposure	Nucleus 5	Phonak Naida	94%
S3	F	57	27	Unknown	Nucleus 5	ReSound Pixel	86%
S4	F	58	21	Unknown	Nucleus 5	Phonak Naida	74%
S5	F	45	20	Unknown	Nucleus 5	Phonak Naida	86%
S6	F	46	78	Hereditary	Nucleus 5	Phonak Naida	79%
S7	M	88	32	Noise exposure	Nucleus 5	Phonak Exelia	70%
S8	F	71	78	Unknown	Nucleus 5	Oticon 380P	92%
MEAN		60.8	39.8				84%
SD		14.2	24.3				9%

Table 2

Articulatory features for the twenty American English consonants used in the study.

Consonant	Voicing	Manner	Place
p	voiceless	stop	bilabial
b	voiced	stop	bilabial
t	voiceless	stop	alveolar
d	voiced	stop	alveolar
k	voiceless	stop	velar
g	voiced	stop	velar
f	voiceless	fricative	labiodental
v	voiced	fricative	labiodental
ð	voiced	fricative	interdental
s	voiceless	fricative	alveolar
z	voiced	fricative	alveolar
ʃ	voiceless	fricative	palatal
ʧ	voiceless	affricate	palatal
ʤ	voiced	affricate	palatal
m	voiced	nasal	bilabial
n	voiced	nasal	alveolar
r	voiced	liquid	palatal
l	voiced	liquid	alveolar
w	voiced	glide	bilabial
j	voiced	Glide	palatal