The relative phonetic contributions of a cochlear implant and residual acoustic hearing to bimodal speech perception[a)]

Benjamin M. Sheffield[b)]
Hearing and Speech Research Laboratory, Department of Otolaryngology–Head and Neck Surgery, 110 Med Sci E, University of California, Irvine, California 92697-1275

Fan-Gang Zeng
Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology–Head and Neck Surgery, 110 Med Sci E, University of California, Irvine, California 92697-1275

(Received 27 December 2010; revised 18 October 2011; accepted 20 October 2011)

The addition of low-passed (LP) speech or even a tone following the fundamental frequency (F0) of speech has been shown to benefit speech recognition for cochlear implant (CI) users with residual acoustic hearing. The mechanisms underlying this benefit are still unclear. In this study, eight bimodal subjects (CI users with acoustic hearing in the non-implanted ear) and eight simulated bimodal subjects (using vocoded and LP speech) were tested on vowel and consonant recognition to determine the relative contributions of acoustic and phonetic cues, including F0, to the bimodal benefit. Several listening conditions were tested (CI/Vocoder, LP, $T_{F0\text{-env}}$, CI/Vocoder + LP, CI/Vocoder + $T_{F0\text{-env}}$). Compared with CI/Vocoder performance, LP significantly enhanced both consonant and vowel perception, whereas a tone following the F0 contour of target speech and modulated with an amplitude envelope of the maximum frequency of the F0 contour ($T_{F0\text{-env}}$) enhanced only consonant perception. Information transfer analysis revealed a dual mechanism in the bimodal benefit: The tone representing F0 provided voicing and manner information, whereas LP provided additional manner, place, and vowel formant information. The data in actual bimodal subjects also showed that the degree of the bimodal benefit depended on the cutoff and slope of residual acoustic hearing. [DOI: 10.1121/1.3662074]

PACS number(s): 43.71.Ky, 43.66.Hg, 43.71.Es [AA] Pages: 518–530

I. INTRODUCTION

With growing acceptance of the cochlear implant (CI) and relaxation of implant candidacy criteria, there is an increasing CI population with residual functional acoustic hearing, allowing combined electric and acoustic stimulation (EAS). Bimodal listeners (CI in one ear and hearing aid in the other), who comprise the overwhelming majority of EAS users, have shown improved speech recognition over CI-only performance (e.g., Shallop *et al.*, 1992; Armstrong *et al.*, 1997; Tyler *et al.*, 2002; Ching *et al.*, 2004; Kong *et al.*, 2005; Mok *et al.*, 2006; Dorman *et al.*, 2008). The bimodal benefit has been observed for speech presented in both quiet and steady-state noise, but it is most pronounced in the presence of competing speech (Turner *et al.*, 2004; Dorman *et al.*, 2005; Kong *et al.*, 2005). This benefit is generally attributed to the addition of low-frequency information that is provided by the hearing aid but not well transmitted by the CI. However, it remains a matter of debate which specific cues contained within the low-frequency acoustic signal are responsible for the benefits observed.

Much of the discussion regarding the bimodal benefit to CI speech perception has focused on the role of fundamental frequency (F0). The perceptual pitch of a talker's voice, relayed by the F0 of the speech signal, is known to be poorly conveyed by the CIs of today (e.g., Qin and Oxenham, 2005; Vongphoe and Zeng, 2005) but is available within the low frequency acoustic signal for those with residual acoustic hearing. Several hypotheses have been put forth as to how F0 information provided to the acoustic-hearing ear might benefit CI speech perception (see Brown and Bacon, 2010, for a summary). One possibility is that the more salient pitch cue from the acoustic signal allows listeners to segregate concurrent sound sources in a competing-talker environment. The idea is that listeners might use the pitch contour of the target talker's voice as a cue to perceptually segregate it from the interfering talkers (Darwin and Hukin, 2000; Turner *et al.*, 2004; Kong *et al.*, 2005; Qin and Oxenham, 2006; Chang *et al.*, 2006). For example, Kong *et al.* (2005) demonstrated that, for a male target, the bimodal benefit was greater with a female masker than with a male masker (i.e., when the mean F0s of the target and masker were further separated). The authors suggested that listeners may have correlated the improved pitch cue from the non-implanted ear with the relatively weak pitch cue from the implant to enhance source segregation. This theory follows naturally from studies with normal hearing listeners that show F0 to

[a)]Portions of this work were presented at the Conference on Implantable Auditory Prostheses, Lake Tahoe, CA, July 2009 and at the Association for Research in Otolaryngology MidWinter Meeting, Baltimore, MD, Feb. 2009.

[b)]Current address: Audiology and Speech Center, Walter Reed National Military Medical Center, 8901 Rockville Pike, Bethesda, MD 20889. Author to whom correspondence should be addressed. Electronic mail: benjamin.m.sheffield@us.army.mil

be a strong cue for sound source segregation (e.g., Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990). However, the mechanism has not proven effective for hearing impaired listeners (Summers and Leek, 1998) and several studies have since questioned whether F0 information in the non-implanted ear contributes to source segregation in bimodal listening situations (e.g., Kong and Carlyon, 2007; Brown and Bacon, 2009a,b).

Another possibility is that the amplitude envelope of the F0 component provides a cue for "glimpsing" the target during the short periods when the instantaneous level of a fluctuating masker is low. Kong and Carlyon (2007) showed that the bimodal benefit does not necessarily depend on pitch information as a partial benefit remained at low SNRs even when the signal presented to the acoustic ear was limited to a tone fixed at 150 Hz and modulated by the voiced portions of target speech. The authors suggested that this voicing cue may allow proper glimpsing of the target to improve the functional SNR. The glimpsing hypothesis received further support from Li and Loizou (2008a,b) who showed that increasing spectral contrast in the low-frequency region enhances the ability to glimpse the target and contributes to the overall EAS advantage, because SNRs are generally more favorable in this region.

A third possibility is that the low-frequency acoustic information available to the residual-hearing ear might improve performance simply by relaying information about the speech itself, rather than or in addition to providing cues for glimpsing the target within the masker or for segregating simultaneous sources. For example, Ching (2005) demonstrated that residual low frequency hearing increases consonant voicing and manner of articulation information in bimodal listeners. Similarly, Mok *et al.* (2006) showed increased transfer of low frequency phonemes (dipthongs, semivowels, nasals and the first formant frequency) with bimodal hearing relative to the CI alone. In fact, it has been suggested that low-frequency phonetic information (e.g., vowel formant or consonant cues) contained within the acoustic signal is primarily responsible for the EAS benefit (Kong and Carlyon, 2007). It is also possible that a substantial portion of this phonetic information is contained within the F0 component of the speech signal. Brown and Bacon, using simulated EAS subjects (Brown and Bacon, 2009a) and actual EAS subjects (Brown and Bacon, 2009b), demonstrated that a tone frequency-modulated by the F0 and amplitude-modulated by the envelope of the target talker provided as much benefit as low-passed (LP) target speech. They suggested that, in addition to glimpsing cues, the F0 component may provide information to more directly improve speech perception, including phonetic cues (e.g., consonant voicing or manner of articulation), lexical boundaries, contextual emphasis, or prosodic information. If so, then the F0 information relayed by the residual acoustic hearing for bimodal CI listeners would be more important to speech perception broadly, and not only in situations requiring glimpsing or concurrent source segregation. Furthermore, this would support the notion that CI users with even very low-frequency residual hearing (<125 Hz) may be able to achieve significant EAS benefits (Cullington *et al.*, 2010; Zhang *et al.*, 2010; Brown and Bacon, 2010).

It is clear from the body of literature regarding EAS that low-frequency phonetic cues present in the acoustic signal are at least partially responsible for the benefits observed in speech perception. Furthermore, it seems plausible that the F0 cues contained within the acoustic signal might be responsible for providing at least a portion of the phonetic information available to the residual-hearing ear. Yet, few studies have examined information transfer of the various acoustic and phonetic cues in bimodal hearing (Ching *et al.*, 2004; Mok *et al.*, 2006) and none have explicitly examined the role of F0 in this regard. Therefore, the overall objective of the present study was to increase understanding of the particular acoustic and phonetic cues that contribute to the bimodal benefit in speech perception. Our first aim was to test the hypothesis that residual low-frequency acoustic hearing provides phonetic information which, when combined with the CI, produces a boost in speech understanding beyond that of the CI alone. A second aim was to test the hypothesis that the F0 cues available within the low-frequency signal provide some of this phonetic information to boost speech understanding with a CI.

To achieve these goals, the present study systematically measured vowel and consonant perception in quiet and in speech-spectrum-shaped noise (0 dB SNR) for eight actual bimodal listeners as well as eight simulated bimodal listeners. Data from speech presented to the implant alone, the acoustic-hearing ear alone (using LP speech or an acoustic representation of F0 extracted from speech), and both ears simultaneously were collected and analyzed to determine the relative transfer of phonetic information, including first and second formant frequency, voicing, manner, and place of articulation. This method of information transfer analysis (Miller and Nicely, 1955) will not only directly determine the relative phonetic contributions of each hearing modality in isolation, but also which particular phonetic cues contribute to the benefit in speech perception when these modalities are combined. Furthermore, comparing the results from LP speech and a tone that follows the contour and amplitude envelope of the F0 extracted from speech may provide insight into the components of the acoustic signal that are important for achieving this benefit.

Our hope is that these findings may in turn influence CI and/or hearing aid processing strategies for individuals with bimodal or hybrid EAS hearing (i.e., with residual acoustic hearing in the implanted ear; see Buchner *et al.*, 2009; von Illberg *et al.*, 1999; Dorman *et al.*, 2005; Gantz *et al.*, 2006) by highlighting the parts of speech that may be most beneficial if targeted by each device. Secondly, these findings may help to inform surgical decisions regarding cochlear implantation for individuals with residual acoustic hearing. For example, testing a patient to determine the phonetic cues that their acoustic hearing ear or ears provide(s) may influence decisions about whether to implant bilaterally or unilaterally, as well as *which* ear to implant if implanting unilaterally. Furthermore, if F0 indeed plays an important role in EAS speech perception, the indication is that CI users with even very low-frequency residual hearing may be able to achieve significant benefits, which may further discourage patients with acoustic hearing in this range from undergoing bilateral implantation or to perhaps consider hybrid implantation. Lastly, it should be noted that the present study focuses on the bimodal benefit with regard to segmental features of speech (i.e., vowels and consonants). It does not attempt to examine the effect of residual acoustic hearing on the perception of other phonetic features such as word stress, intonation, or lexical tones with a CI, for which F0 is likely to be an important cue.

II. METHODS

A. Subjects

1. Actual bimodal subjects

Eight CI users between the ages of 46 and 83 years (mean 63.0 years; 3 female, 5 male) participated in the study and are referred to as the actual bimodal group. Each subject also wore a hearing aid in his or her non-implanted ear on a daily basis. Figure 1 shows group mean and individual-subject audiograms. Table I displays demographic information for each individual subject. Four subjects used a Nucleus 24 device (Cochlear Corporation, Lane Cove, Australia), two used a HiRes Harmony device (Advanced Bionics Corp., Valencia, CA, USA), one used a Clarion CII device (Advanced Bionics Corp.), and another used a Combi 40+ device (MED-EL, Innsbruck, Austria). All subjects were post-lingually deafened, native American English speakers, and had acquired at least 12 months of listening experience post implantation prior to testing.

2. Simulated bimodal subjects

Eight normal-hearing, native American English speakers, between the ages of 17 and 33 years old (mean 22.9 years, 5 female, 3 male), also participated in this study and are referred to as the simulated bimodal group. All subjects had audiometric thresholds below 20 dB HL for pure-tone frequencies between 125 and 8000 Hz. The experimental procedures were approved by the University of California, Irvine Institutional Review Board. Simulations of hearing loss and cochlear implants will be described below.

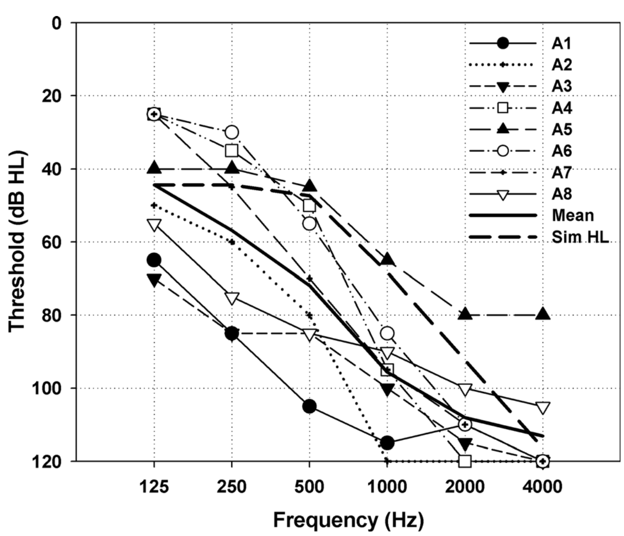

FIG. 1. Pure-tone thresholds for the actual bimodal subjects. Also plotted are the actual bimodal group mean and the simulated hearing loss (fourth order Butterworth filter with a lowpass cutoff at 500 Hz) provided to the simulated bimodal group. This demonstrates the differences in the effective slope of hearing loss for the two groups, particularly between 125 and 1000 Hz.

TABLE I. Biographical data of actual bimodal subjects.

Subject	Gender	Age (years)	CI Experience (years)	Etiology	Device
A1	F	71	7.6	Genetic	Clarion II
A2	M	74	2.9	SNHL	Nucleus 24
A3	F	55	3.4	Unknown	HiRes Harmony
A4	F	65	5.7	Unknown	Nucleus 24
A5	M	59	4.8	Unknown	Combi
A6	M	51	1.8	Unknown	Nucleus 24
A7	M	83	6.1	Unknown	Nucleus 24
A8	M	46	2.6	Ototoxicity	HiRes Harmony

B. Stimuli

Vowel stimuli consisted of 11 tokens in the /hVd/ context (Hillenbrand, 1995) spoken by a male talker with a mean F0 of 101 Hz (range, 97–106 Hz). Only 11 of the standard 12 vowels were presented, with the vowel stimulus /hod/ eliminated to avoid confusion with /hawed/, whose relative pronunciations are indistinguishable in the region of the United States where testing was conducted. Consonant stimuli consisted of 20 tokens in the /aCa/ context (Shannon, 1999) spoken by the same male talker with a mean F0 of 106 Hz (range, 99–120 Hz).

The original vowel and consonant stimuli were processed to create five conditions: (1) CI/Vocoder, (2) LP, (3) $T_{F0\text{-env}}$, (4) CI/Vocoder + LP, and (5) CI/Vocoder + $T_{F0\text{-env}}$. All five conditions were presented with the signal in quiet and four of the five were also presented with the signal in noise (0 dB SNR) as described below. All processing was conducted in MATLAB and stimuli were presented using a Creative Labs 24-bit external USB sound card at a 44.1 kHz sampling rate. For the conditions involving the CI, the actual bimodal group had the original stimuli played directly to their CI via direct-connect cable either in quiet or in combination with speech-spectrum-shaped noise at 0 dB SNR. For the simulated bimodal group, the original stimuli were processed to create a 4-channel CI simulation (vocoder) and presented via headphone to the left ear. To create the CI simulation, stimuli were processed using a 4-band noise excited vocoder (Shannon *et al.*, 1995) with cutoff frequencies spaced logarithmically between 250 and 8700 Hz (Greenwood, 1990). Stimuli were then passed through a bank of third order elliptical filters and full-wave rectified to obtain the envelope in each channel. The envelopes were each used to modulate a broadband signal. Finally, these modulated signals were passed through the original analysis filters again to eliminate sidebands induced outside the channel and summed to create the Vocoder stimulus. These methods follow closely those of Carroll *et al.* (2011). For conditions where the Vocoder signal was presented in noise, speech-spectrum-shaped noise was combined with the original signal at 0 dB SNR prior to processing.

The LP and $T_{F0\text{-env}}$ conditions were processed identically for both groups. The LP condition was created by low-pass filtering the original stimuli either alone or in combination with speech-spectrum-shaped noise at 0 dB SNR using a fourth order Butterworth filter with a cutoff at 500 Hz. Figure 1 shows the cutoff and slope of this simulated hearing loss characteristic (thick dashed line). The $T_{F0\text{-env}}$ condition was created by first extracting the time-varying F0 from the original stimuli using the STRAIGHT program (Kawahara *et al.*, 1999). The amplitude envelope of the F0 contour was then extracted by low-pass filtering the original stimuli using a fourth order Butterworth filter with a cutoff equal to the maximum frequency of the F0 contour for each stimulus, full-wave rectifying, and low-pass filtering again with a 20 Hz cutoff to minimize the introduction of sidebands. Finally, the extracted amplitude envelope and frequency contour were used to modulate a sinusoidal carrier, with the unvoiced speech segments represented by silence. These methods resulted in an acoustic representation of F0 that provides amplitude, duration, and frequency cues, as well as their derivatives, such as AM and FM. This condition is similar to the AM-FM condition of Kong and Carlyon (2007), the $T_{F0\text{-env}}$ condition of Brown and Bacon (2010), and the Target F0 condition of Carroll *et al.* (2011). The CI/Vocoder + LP and CI/Vocoder + $T_{F0\text{-env}}$ stimuli were dichotic combinations of CI or Vocoder stimuli presented to one ear and the LP or $T_{F0\text{-env}}$ stimuli presented to the other ear.

Note that for the $T_{F0\text{-env}}$ and CI/Vocoder + $T_{F0\text{-env}}$ conditions, the $T_{F0\text{-env}}$ signal was always presented in quiet (even in the conditions where noise was presented to the CI or Vocoder). This was to allow a more sensitive measure of its phonetic contributions (Brown and Bacon, 2009a). Similarly, an additional condition was tested for the actual bimodal group in noise, in which the LP signal was presented in quiet when combined with the CI signal, forming a sixth listening condition (CI + LP Q). Again, this manipulation was to allow a more sensitive measure of the acoustic and phonetic contributions from LP speech.

All stimuli were presented acoustically with Sennheiser HDA 200 Audiometric headphones except the CI signal for the actual bimodal group, which was presented via direct-connect cable to the subject's CI. Stimuli for the actual bimodal listeners were presented at the subject's most comfortable level (MCL) in both the CI and the non-implanted ear with the acoustic signal amplified linearly and presented to the subject's non-implanted ear unaided (hearing aid removed). In order to eliminate differences due to the individualized hearing aid programs of our bimodal subjects, they did not wear their hearing aids during testing. For the simulated bimodal group, the Vocoder signal was always presented to the left ear, whereas the LP and $T_{F0\text{-env}}$ signals were always presented to the right ear. All stimuli were presented at an average level of 70 dB SPL for the simulated bimodal group. Experiments were conducted in a double-walled sound booth.

C. Procedure

Computer interfaces were created using MATLAB to display a push-button grid representing the 11 vowel tokens or 20 consonant tokens used for each task. Subjects were presented with a token and asked to select which token they heard. For the actual bimodal group, the experimenter first determined the MCL for each ear independently by presenting a token and asking the subject whether the volume should be increased or decreased. Tokens were then presented binaurally and the levels were further adjusted to maintain comfort.

Prior to testing, all subjects were given training for the CI/Vocoder, LP, and $T_{F0\text{-env}}$ conditions in quiet using an additional computer interface that allowed subjects to click on the pushbutton grid to hear each corresponding token. Subjects were able to become familiar with the tokens under each condition and to compare and contrast the differences between tokens prior to testing. Each training session was limited to 2 min. Subjects were also given at least two practice rounds for the CI/Vocoder, LP, and $T_{F0\text{-env}}$ conditions in quiet prior to any formal data collection for the experimental tests. In the practice rounds, as well as in the experimental tests, stimuli were presented in random order such that each token was presented a total of three times per test (3×11 tokens $= 33$ tokens total for vowel recognition and 3×20 tokens $= 60$ tokens total for consonant recognition per test). Feedback was not provided after each selection, but subjects briefly viewed the percent correct score at the end of each test, as well as the confusions made, which were displayed in a matrix. The practice rounds were repeated for each of the CI/Vocoder, LP and $T_{F0\text{-env}}$ conditions in quiet, interleaved with training sessions, until performance stabilized. Performance was considered stable for each condition when the difference in scores between two consecutive tests was less than 13 percentage points for the vowel recognition task and less than 7 percentage points for consonant recognition. These relative percentages were chosen to allow for a maximum difference of four correctly identified tokens between consecutive tests. For all subjects, training/practice was provided for the CI/Vocoder condition first, followed by the LP condition and, finally, the $T_{F0\text{-env}}$ condition.

After the training/practice rounds, the experimental tests were conducted. The simulated bimodal subjects were tested in a total of nine conditions, comprised of all five listening conditions in quiet and four of the five listening conditions in noise ($T_{F0\text{-env}}$ not tested in noise) for each task. Actual bimodal subjects were tested in a total of 10 conditions, comprised of all five listening conditions in quiet and four of five listening conditions in noise ($T_{F0\text{-env}}$ not tested in noise) plus an additional condition (CI + LP Q). The testing order of conditions was randomized. A percentage-correct score and confusion matrix were saved after the completion of each test. Once all conditions were completed, the tests were administered a second time with a different randomization of the condition order. Percentage-correct scores from the two test iterations of each condition were averaged and the confusion matrices summed.

D. Data analysis

Percentage-correct phoneme recognition scores were analyzed using a repeated-measures analysis of variance

TABLE II. Classification of vowels by acoustic features.[a]

	had	hawed	hayed	head	heed	herd	hid	hoed	hood	hud	who'd
IPA Symbol	æ	ɔ	e	ɛ	i	ɝ	ɪ	o	ʊ	ʌ	u
Duration	2	3	2	2	2	2	1	2	1	1	2
F1	4	4	2	3	1	3	2	3	2	4	1
F2	3	1	4	3	4	2	4	1	2	2	2

[a]Duration coding: 1=short (<640 ms), 2=medium (640–680 ms), 3=long (>680 ms). F1 coding: 1=low (<420 Hz), 2=med-low (440–500 Hz), 3=med-high (500–600 Hz), 4=high (>680 Hz). F2 coding: 1=low (<1150 Hz), 2=med-low (1260–1390 Hz), 3=med-high (1550–1800 Hz), 4=high (>2000 Hz).

(ANOVA). Paired *t* tests with a Bonferroni adjustment were used for pairwise comparisons of the experimental conditions with a significance level of $p<0.05$.

In addition, information transfer (IT) was analyzed for the transmission of phonetic and acoustic features (Miller and Nicely, 1955). Vowel stimuli were classified according to the duration categories as well as first (F1) and second (F2) formant frequency categories (Table II). Consonant stimuli were classified according to voicing, manner and place of articulation (Table III), similar to the classification used by Xu *et al.* (2005). Confusion matrices were used to calculate the individual IT for each subject under each listening condition and then averaged across listeners to obtain the group-average IT for each condition. Paired *t* tests with a Bonferroni adjustment were used for pairwise comparisons of the relative group-average IT percentages. As demonstrated by Sagi and Svirsky (2008), pooling the subjects' matrices prior to IT analysis would have reduced small-sample estimation bias in the absolute IT values, but at the expense of being able to make valid statistical comparisons between conditions. Because we were much more interested in the relative differences between experimental conditions rather than in absolute IT values, the average IT method was chosen instead.

III. RESULTS

A. Vowels

1. *Vowel recognition*

Figure 2 shows vowel recognition for actual [Fig. 2(a)] and simulated [Fig. 2(b)] subjects in quiet (left panels) and in noise (right panels). The mean percent correct scores are plotted for the different listening conditions with error bars representing the standard error. Chance performance for this experiment is 9.1% (horizontal dashed lines).

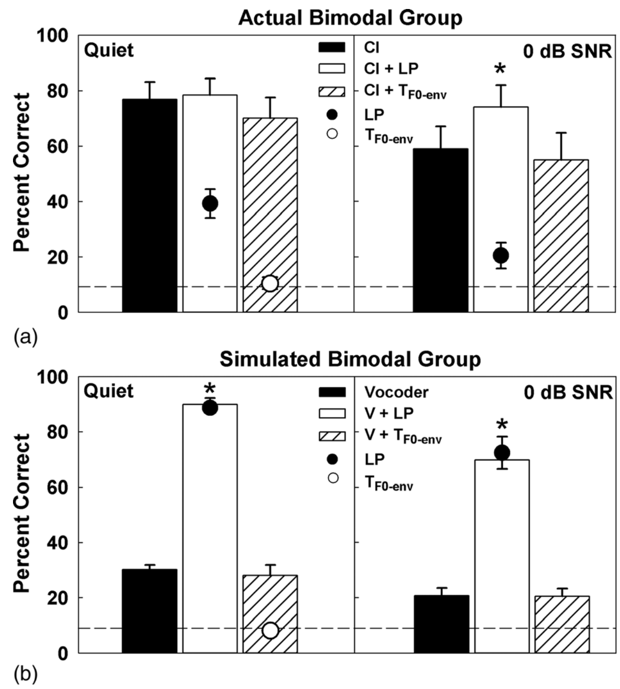

FIG. 2. Vowel recognition mean scores for (a) the actual bimodal group and (b) the simulated bimodal group for each listening condition in quiet and in noise. Stars indicate statistical significance ($p<0.05$) with respect to the CI condition. Error bars represent mean standard error. Note that $T_{F0\text{-env}}$ is not plotted in the right panels because this condition was not tested at 0 dB SNR; see the left panel for $T_{F0\text{-env}}$ scores in quiet.

For the actual bimodal group, the only acoustic information that significantly enhanced CI performance for vowel recognition was LP speech in the noise condition [Fig. 2(a)]. In quiet, the CI, CI + LP, and CI + $T_{F0\text{-env}}$ conditions yielded similar mean scores (76.9%, 78.4%, and 70.1%, respectively) [$F(2,14) = 3.4$, $p>0.05$], suggesting that the addition of acoustic speech information did not improve performance relative to the CI alone. However, at 0 dB SNR, introduction of the LP signal significantly increased CI performance from 58.9 to 74.1% correct ($p<0.01$) [$F(2,14) = 11.6$, $p<0.005$]. A mean score of 54.9% was observed for the CI + $T_{F0\text{-env}}$ condition at 0 dB SNR, which was not significantly different from the CI condition ($p = 1.0$). This suggests that the benefit observed from the LP signal was due to cues other than those present in the $T_{F0\text{-env}}$ signal.

Performance was near chance level with the $T_{F0\text{-env}}$ condition in quiet (open circles, 10.2%), but above chance for the LP condition in quiet and at 0 dB SNR (filled circles,

TABLE III. Classification of consonants by phonetic/acoustic features.[a]

	aBa	aDa	aGa	aPa	aTa	aKa	aFa	aVa	aTHa	aZa	aSa	aSHa	aJa	aCHa	aMa	aNa	aLa	aRa	aYa	aWa
IPA Symbol	b	d	g	p	t	k	f	v	ð	z	s	ʃ	ʤ	tʃ	m	n	l	r	y	w
Voicing	1	1	1	0	0	0	0	1	1	1	0	0	1	0	1	1	1	1	1	1
Manner	1	1	1	1	1	1	2	2	2	2	2	2	3	3	4	4	5	5	5	5
Place	1	3	5	1	3	5	1	1	2	3	3	4	4	4	1	3	3	3	4	5

[a]Voicing coding: 1=voiced, 0=unvoiced. Manner coding: 1=plosive, 2=fricative, 3=affricate, 4=nasal, 5=glide. Place coding: 1=labial, 2=dental, 3=alveolar, 4=palatal, 5=velar.

39.2% and 20.5%, respectively). Interestingly, when the LP speech signal was presented in quiet and combined with CI in noise (CI + LP Q, 74.1%), it did not further benefit the actual bimodal listeners as a group beyond what was observed for CI + LP at 0 dB SNR (74.1%) (not shown). Therefore, this condition was not plotted nor included in the IT analysis for vowel recognition.

Compared with the actual bimodal subjects, the simulated bimodal subjects performed differently [Fig. 2(b)]. Like the actual bimodal group, LP speech significantly enhanced Vocoder performance, but the $T_{F0\text{-env}}$ signal did not [Fig. 2(a)]. Also, simulated performance was near chance level with the $T_{F0\text{-env}}$ condition in quiet (open circles, 8.3%) but significantly above chance for the LP condition in quiet and at 0 dB SNR (filled circles, 88.6% and 72.3%, respectively). However, three significantly different trends were also noted. First, the Vocoder condition (30.1%) yielded lower scores than the mean performance seen in the CI condition for the actual bimodal group (76.9%) [$F(1,7) = 51.2$, $p < 0.001$]. Second, a reverse trend was observed for the LP condition, showing better simulated performance (88.6%) than actual performance (39.2%) [$F(1,7) = 63.4$, $p < 0.001$]. Third, note similarly high-level performance between the Vocoder + LP and LP conditions (90.0% and 88.6%, respectively). The differences in performance between the actual and simulated groups are addressed in the discussion section.

Similar performance trends for the simulated group extended to the noise condition where the results for Vocoder + LP (69.7%) were significantly better than for the Vocoder alone (20.6%), but not LP alone (72.3%). Scores for the Vocoder + $T_{F0\text{-env}}$ condition were not significantly different from the Vocoder alone in quiet or in noise, and subjects performed at chance level for the $T_{F0\text{-env}}$ condition in quiet, coinciding with results for the actual bimodal group.

2. *Vowel information transfer analysis*

Figure 3 shows information transfer (IT) for the actual bimodal group in noise during vowel recognition. The mean IT percentage for 3 acoustic features (duration, F1, F2) are plotted for the different listening conditions. A significant main effect of listening condition was observed for F1 [$F(2,14) = 13.6$, $p < 0.005$], as well as F2 [$F(2,14) = 7.8$, $p < 0.05$], but not for duration [$F(2,14) = 1.1$, $p > 0.1$]. Observe that the main effects for both formants were entirely due to the additional LP Speech, as the combination of CI + LP provided a significant IT increase relative to the CI alone for both F1 and F2 ($p = 0.008$ and $p = 0.03$, respectively). In comparison, no significant differences were observed between CI + $T_{F0\text{-env}}$ and CI. Furthermore, the absolute IT values were near zero for the $T_{F0\text{-env}}$ condition in all three features. The IT results indicate that the bimodal benefit observed in vowel recognition was, at least in part, due to improved formant discrimination from the acoustic side, but with little or no assistance from F0 or durational cues.

Although not shown, the vowel IT results for the simulated bimodal group were similar to the actual bimodal group in that the $T_{F0\text{-env}}$ signal did not provide a benefit for any of the three features in quiet or in noise when combined with the CI simulation. However, the two groups differed slightly in the benefits seen from the LP signal. For the simulated subjects, LP significantly improved IT of all 3 features when combined with the CI simulation (Vocoder + LP), including duration, which was not observed for the bimodal group. Furthermore, these trends were present in the quiet condition as well as in noise for the simulated subjects. It should be noted, however, that simulated CI + LP performance was not significantly better than the LP condition for any feature in quiet or in noise. This indicates that the simulated subjects were relying almost entirely on information from the LP signal during simulated bimodal vowel perception.

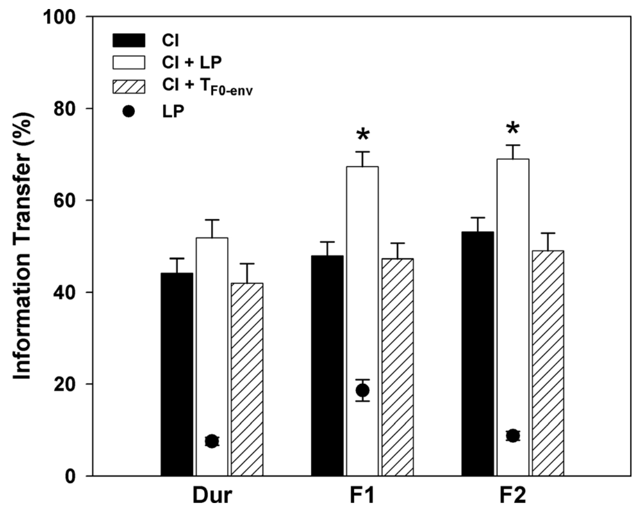

FIG. 3. Information transfer results of the actual bimodal group for vowel recognition in noise. The features of interest are vowel duration (Dur), first formant frequency (F1) and second formant frequency (F2). Stars indicate statistical significance ($p < 0.05$) with respect to the CI condition. Error bars represent mean standard error.

3. *Vowel results summary*

Compared to CI/Vocoder alone, the additional LP signal significantly improved vowel recognition at 0 dB SNR. The improvement was 15.2 percentage points for the actual bimodal group and 49.1 for the simulated bimodal group, with the simulated listeners appearing to rely primarily on the LP signal. IT analysis indicated that the residual acoustic hearing of the actual bimodal subjects provided access to information regarding both formant frequencies with little or no assistance from F0, amplitude envelope, or durational cues.

B. Consonants

1. *Consonant recognition*

Figure 4 shows consonant recognition for actual [Fig. 4(a)] and simulated [Fig. 4(b)] subjects in quiet (left panels) and in noise (right panels). The mean percent correct scores are plotted for the different listening conditions with error bars representing the standard error. Chance level performance for this experiment is 5% (horizontal dashed lines).

For the actual bimodal group [Fig. 4(a)], a significant main effect of listening condition was observed in quiet [$F(2,14) = 13.7$, $p < 0.001$] and in noise [$F(3,21) = 15.1$, $p < 0.001$]. In quiet, mean performance for the CI + LP

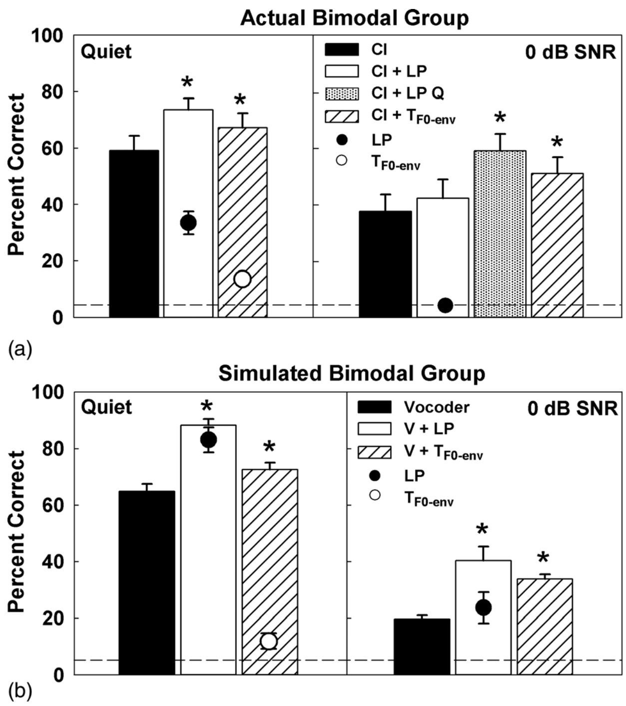

FIG. 4. Consonant recognition mean scores for (a) the actual bimodal group and (b) the simulated bimodal group for each listening condition in quiet and in noise. The CI + LP Q condition is also plotted for the actual bimodal group in noise. Stars indicate statistical significance (p<0.05) with respect to the CI condition. Error bars represent mean standard error. Note that the LP Q and $T_{F0\text{-env}}$ conditions are not included in the right panels because these conditions were only tested in quiet; refer to the left panels for these scores.

(mean ± s.e., 73.5 ± 4.0%) and CI + $T_{F0\text{-env}}$ (67.2 ± 5.1%) conditions were both significantly greater than performance for the CI alone (59.1 ± 5.2%). This indicates that both the LP and $T_{F0\text{-env}}$ signals contained speech information that benefitted consonant perception relative to the CI alone. In noise, on the other hand, while performance for the CI + $T_{F0\text{-env}}$ condition (50.9 ± 5.8%) was significantly better than for CI (37.5 ± 6.0%), the LP speech had little effect when both signals were at 0 dB SNR (CI + LP, 42.2 ± 6.7%). (Keep in mind that the $T_{F0\text{-env}}$ signal remained in quiet for the CI + $T_{F0\text{-env}}$ conditions at 0 dB SNR.) Note also that performance was at chance level for the LP condition in noise (4.7 ± 0.9%), indicating that the SNR may have been too low to observe a benefit from the LP signal for these subjects in consonant recognition. However, when the LP signal was held in *quiet* and combined with the CI signal at 0dB SNR (CI + LP Q, 59.0 ± 6.0%) there was a statistically significant 21.5 percentage point benefit over the CI condition (p < 0.001).

A clear difference between consonant and vowel recognition was the contribution from the low frequency tone. Recall that this signal had no effect on vowel recognition. For consonant recognition, the $T_{F0\text{-env}}$ condition yielded above chance performance on its own (13.9 ± 1.3%) and improved performance when combined with the CI. In fact, performance for the CI + $T_{F0\text{-env}}$ condition was not statistically different from either the CI + LP condition in quiet (p>0.1) or the CI + LP Q condition in noise (p>0.1), implying that much of the information which boosted performance when LP speech was combined with the CI was contained within this tone.

Comparing Figs. 4(a) and 4(b), the performance differences between the actual and simulated groups were much less pronounced for consonant recognition than those observed for vowel recognition. For instance, simulated performance for the Vocoder condition in quiet (64.9 ± 2.5%) was not significantly different from the actual CI performance (59.1 ± 5.2%), [$F(1,7) = 1.4$, p>0.1]. Also in line with the results from the actual bimodal group, a significant main effect of listening condition was observed both in quiet [$F(2,14) = 43.2$, p<0.001] and in noise [$F(2,14) = 25.4$, p<0.005]. In quiet, this effect was due to a significant improvement in both the Vocoder + LP (88.1 ± 2.2%) and Vocoder + $T_{F0\text{-env}}$ (72.6 ± 2.4%) conditions over the Vocoder alone. The same trends were observed at 0 dB SNR, with the Vocoder + LP (40.3 ± 4.9%) and Vocoder + $T_{F0\text{-env}}$ (33.9± 1.6%) conditions both providing a significant benefit over the Vocoder alone (19.7 ± 1.3%). In quiet, the effect of LP speech on Vocoder performance can be attributed to the fact that LP performance was significantly greater than Vocoder alone and was no different than performance for Vocoder + LP. But not in noise, where there was a real combination of Vocoder and LP information.

The two groups differed in two primary areas, which were also observed for vowel recognition. First, Vocoder performance was significantly worse than CI performance at 0 dB SNR [$F(1,7)$=10.6, p<0.05]. Secondly, the opposite trend was noted for the LP condition, which was significantly better for the simulated bimodal group than for the actual bimodal group, overall. Together, these trends were likely responsible for the statistical benefit observed for the Vocoder + LP condition at 0 dB SNR relative to the Vocoder alone, which was not observed for the actual bimodal subjects.

2. Consonant information transfer analysis

Figure 5 shows information transfer (IT) for actual (left) and simulated (right) subjects during consonant recognition at 0 dB SNR. The mean IT percentage for three acoustic/phonetic features (voicing [Fig. 5(a)], manner [Fig. 5(b)], and place [Fig. 5(c)]) is plotted for each listening condition. Significant main effects of listening condition were observed for both groups for all three features.

Prior to describing the rest of the IT results, it should first be noted that the LP condition provided minimal information transfer for the actual bimodal group. Furthermore, the CI + LP condition did not provide a significant increase in IT for any of the three features relative to the CI alone. However, the IT for all three consonant features increased significantly when the LP signal in quiet was added to the CI signal in noise (CI + LP Q). Therefore, we focused on this condition when comparing IT results for the actual and simulated subjects.

For the simulated bimodal group, which performed much better with the LP condition, the Vocoder + LP condition also provided a significant increase over the Vocoder alone for all three consonant features. For both groups, additional acoustic access to F0 and amplitude envelope (CI/Vocoder + $T_{F0\text{-env}}$) resulted in significant increases of voicing and manner information, but not place information,

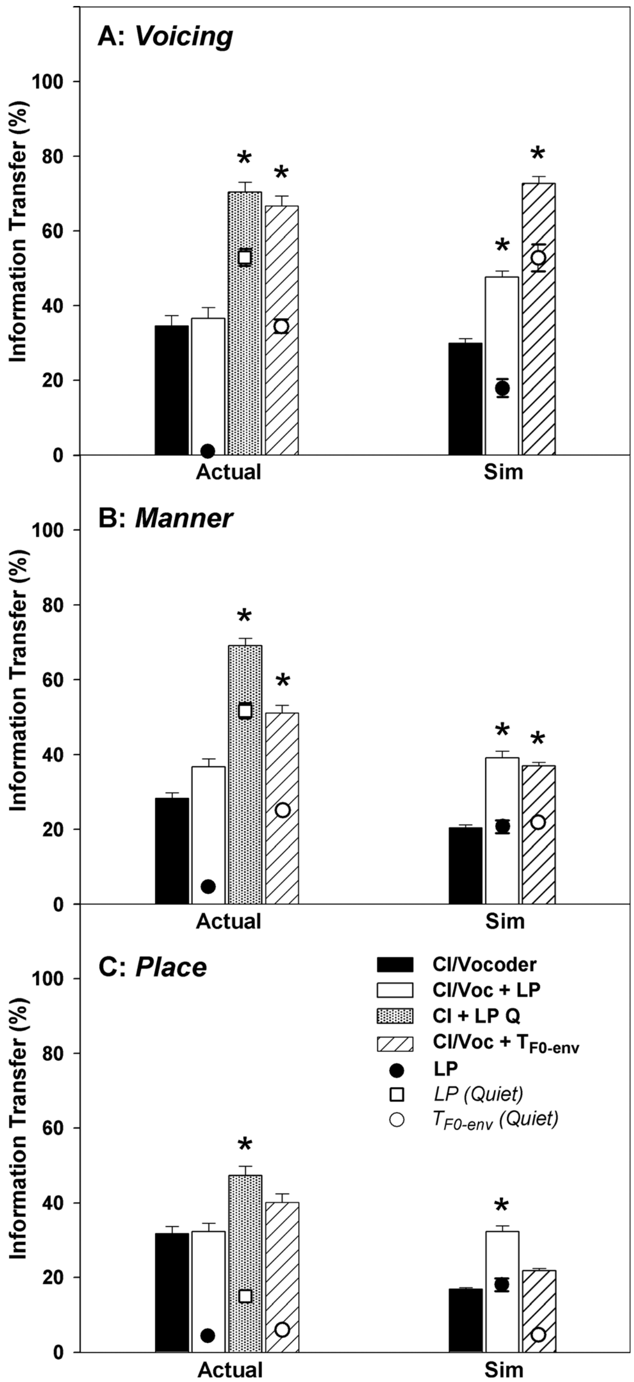

FIG. 5. Information transfer results of both groups for consonant recognition at 0 dB SNR. The features of interest are voicing, manner of articulation and place of articulation. Stars indicate statistical significance ($p<0.05$) with respect to the CI condition. Error bars represent mean standard error.

relative to the CI alone. Similar results were observed for the quiet conditions (not shown). Also of note is that there was no statistical difference in voicing IT between the CI + $T_{F0\text{-env}}$ and CI + LP Q conditions for the actual bimodal subjects (p=1.0), suggesting that the $T_{F0\text{-env}}$ signal provided most of the increase in voicing information over the CI alone.

3. Consonant results summary

Similar to vowel recognition results, the additional LP signal significantly improved CI and vocoder consonant recognition in quiet and noise for both groups. IT analysis indicated that much of the benefit in CI and vocoder consonant perception can be attributed to the $T_{F0\text{-env}}$ signal, which significantly increased voicing and manner information, but not place information. These findings contrast with the vowel recognition task for which the $T_{F0\text{-env}}$ signal had no effect and where information contained in the LP signal (i.e., formant frequencies) provided the most benefit to CI or Vocoder vowel perception.

IV. DISCUSSION

A. Effect of residual acoustic hearing on CI vowel perception

In the vowel IT analysis, an increase in F1 information was observed from adding LP acoustic speech to the CI signal. This result was not surprising, as many of the vowel tokens had F1 frequencies within the residual hearing range of these CI subjects. However, an overall increase in F2 information was also observed. *Post hoc* analysis of the individual results showed that the magnitude of the F2 increase was inversely correlated with the slope of the subject's hearing loss. Pearson product-moment correlations were used to analyze the relationship between the slope of each subject's audiogram between 125 and 500 Hz, expressed in dB/octave (M = 13.8, SD = 6.4), and the percentage-point difference in IT between the CI + LP and CI conditions for F2 (M = 15.9, SD = 14.1). This correlation was found to be statistically significant [$r^2 = 0.77$, $p<0.05$]. Generally, subjects with shallower sloping audiograms showed a greater increase in F2 information when the LP signal was combined with the CI, whereas F2 likely fell outside the audible frequency range of listeners with steeper sloping audiograms. Slope of hearing loss was not found to be a statistically significant predictor of overall vowel recognition performance. However an inverse correlation [$r^2 = 0.71$, $p<0.05$] was observed between hearing threshold at 1 kHz and the percentage-point benefit in vowel recognition for the CI + LP condition relative to the CI condition (M = 15.2, SD = 11.1). Looking at other studies exploring correlations between audiogram measures and the bimodal benefit in sentence perception, the evidence for predicting the bimodal benefit is not strong (Ching *et al.*, 2004; Gifford *et al.*, 2007). It is probable that because the acoustic signal in these experiments was amplified linearly, the effects on audibility of absolute threshold and the slope of hearing loss were exacerbated relative to what would be observed with a hearing aid fitting that includes dynamic range compression. This could explain why the degree of bimodal benefit in vowel perception was highly dependent on each individual's residual acoustic hearing profile when it has not been shown to be true for sentence perception in other studies. The acoustically provided $T_{F0\text{-env}}$ signal had no effect on vowel recognition for either group (bimodal or simulated). However, this result is not surprising, as the primary cues for vowel recognition are manifested by a well-defined formant structure (Stevens, 1998a). Furthermore, the vowel tokens were spoken by a normal talker and did not contain any superimposed suprasegmental cues, such as stress or intonation where the

$T_{F0\text{-}env}$ signal may have helped (Bunton, 2006; Ladefoged, 1982). It should be noted that the $T_{F0\text{-}env}$ signal extracted for the present study contains information not only about pitch variations, but also the onset/offset timing of the talker's voice and amplitude envelope in the F0 range. Although none of these cues could be expected to help with formant perception, it was hypothesized that the voicing onset/offset cue might provide some information regarding vowel duration at least with $T_{F0\text{-}env}$ alone. However, this was not observed in the results. This is likely due to the fact that the recorded vowel tokens used in the experiment had relatively little variation in duration compared to vowels observed in other studies (e.g., Hillenbrand *et al.*, 1995).

B. The role of F0 in bimodal consonant perception

The contributions from the acoustic $T_{F0\text{-}env}$ signal were very different for consonant recognition than for vowel recognition. For consonants, the $T_{F0\text{-}env}$ signal provided a significant boost to CI performance both in quiet and at 0 dB SNR (8.1 and 13.4 percentage points, respectively). In fact, performance for the $CI + T_{F0\text{-}env}$ condition was not statistically different from that of the CI + LP condition for the actual bimodal group (Quiet, $p = 0.11$; 0 dB SNR, $p = 0.18$). That the $T_{F0\text{-}env}$ signal boosted CI/Vocoder consonant perception in quiet and in steady-state noise implies that contributions of voicing and/or F0 to the EAS benefit are not limited to source segregation or glimpsing cues at low SNRs, because neither environment involved a competing talker or a fluctuating masker Instead, as hypothesized, the $T_{F0\text{-}env}$ signal provided phonetic information about the target speech itself. Specifically, the $T_{F0\text{-}env}$ signal contained information about consonant voicing and manner of articulation that enhanced the perception of these features relative to the CI condition. Because the $T_{F0\text{-}env}$ signal used in these experiments contained amplitude, duration, and frequency cues, as well as their derivatives such as AM and FM, it is important to examine which cues contained within this signal may be responsible for the benefits observed.

It is not surprising that the $T_{F0\text{-}env}$ signal increased consonant voicing information, as there are several obvious acoustic correlates to voicing contained within this signal. Assuming non-whispered speech, for nasals and glides the mere presence of F0 in the consonantal segment can be an indicator of whether it was voiced or voiceless (Stevens, 1998b). For plosives and fricatives, the duration of the preceding vowel can indicate whether the consonant is voiced (Umeda, 1975). An additional acoustic correlate to voicing for plosives, particularly in quiet, is the voice onset time (VOT), which is defined as "the period between the release of the plosive and the beginning of voicing in the [following] vowel" (Jiang *et al.*, 2005). This might not be a viable cue for $T_{F0\text{-}env}$ alone, because the burst has been filtered out of the signal, but when combined with the CI, where the burst is presumably preserved, it may provide an additional cue to voicing. What all of these acoustic features have in common is that they only rely on the timing of the $T_{F0\text{-}env}$ signal (i.e., when it is on and when it is off). In other words, the fundamental frequency is irrelevant, meaning that this information could be transmitted even with a stationary F0 signal (i.e., no F0 variation) as was suggested by Kong and Carlyon (2007). The increase in manner of articulation is less intuitive. Previous studies have shown that F0 carries manner information during normal acoustic listening scenarios (e.g., Breeuwer and Plomp, 1986; Boothroyd, 1988; Faulkner and Rosen, 1999), but to the authors' knowledge the present study is the first to explicitly demonstrate that an acoustic representation of F0 enhances manner perception for EAS users. To examine the effects of the $T_{F0\text{-}env}$ signal on manner in greater detail, IT analysis was conducted using the actual bimodal group consonant confusion matrices (0 dB SNR) with manner of articulation separated into its main binary features (plosive, fricative, affricate, nasal, and glide). Interestingly, $CI + T_{F0\text{-}env}$ appeared to enhance all the features of manner relative to the CI alone, although it was only significant for nasals, glides and affricates. To be sure that these benefits were independent of redundancies from the voicing feature, the analysis was repeated using the sequential information analysis technique (Wang and Bilger, 1973), in which the effect of the voicing feature was partialed out of the analysis. The results are included in Fig. 6, with the CI + LP Q and LP (Quiet) conditions also provided for comparison. Significant differences between CI and $CI + T_{F0\text{-}env}$ remained for the nasal, glide, and affricate features. In fact, for affricates the $T_{F0\text{-}env}$ benefit was not statistically different from that provided by LP ($p = 1.0$). It remains unclear, however, which component(s) of the $T_{F0\text{-}env}$ signal were responsible for the relative increases observed for each feature of manner. Removing effects of the voicing feature from the analysis only takes care of information about whether the consonant contained voicing at all (i.e., *if* the consonant is voiced). But *when* the voicing occurs relative to the other acoustic events in the token can still enhance manner cues. For example, if the voicing occurs during the consonantal segment this can indicate a nasal or glide. Again, this cue is

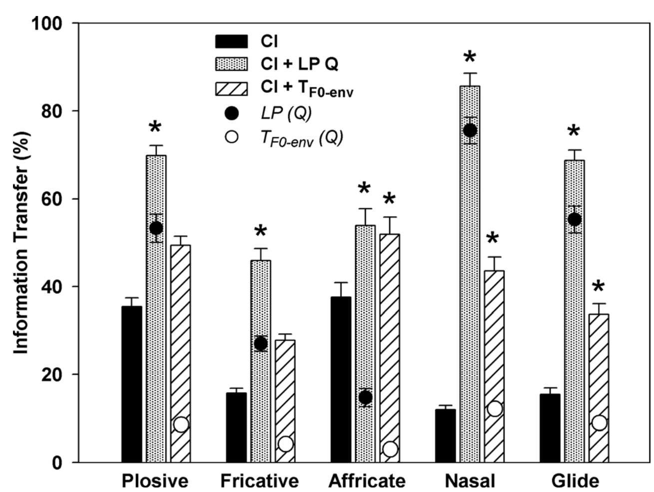

FIG. 6. Sequential information transfer results of the actual bimodal group for consonant recognition at 0 dB SNR, concentrating on the contributions of the $T_{F0\text{-}env}$ signal to manner of articulation cues. The features of interest are plosive, fricative, affricate, nasal, and glide. Results shown are second iteration of analysis with influences from the voicing feature removed. The CI + LP Q and LP (quiet) conditions are also plotted for comparison. Stars indicate statistical significance ($p<0.05$) with respect to the CI condition. Error bars represent mean standard error.

independent of the fundamental frequency and could be delivered with any audible tone gated on and off with voicing. However, for plosives, fricatives, and affricates, if the $T_{F0\text{-env}}$ signal enhances manner cues at all, it would likely be due to amplitude envelope information or to discontinuities in the F0 contour rather than the onset/offset timing of voicing in the consonantal segment. Slow time-varying envelope fluctuations from 2 to 50 Hz can provide information about manner (Rosen, 1992), so it is possible that the amplitude envelope of our $T_{F0\text{-env}}$ signal (smoothed with a 20 Hz filter) may have boosted manner cues. Other studies have shown that the amplitude envelope can be influential for voiceless affricate-fricative distinctions (e.g., Dorman *et al.*, 1980; Gerstman, 1957; Howell and Rosen, 1983; Repp *et al.*, 1978), so this could explain the boost in affricate cues in particular. As far as F0 contour, Faulkner and Rosen (1999) showed that variations in the F0 contour had a significant effect on audiovisual consonant recognition due to an improvement of nasal and plosive manner perception. They also showed a significant benefit in manner perception from F0 onset/offset timing. Here, too, it is likely that the improved manner perception for bimodal consonant recognition is from some combination of F0 variation, onset/offset timing, and amplitude envelope. However, each of these cues would have to be tested individually to more precisely quantify their relative influences. It would be interesting to repeat the same experiment comparing a stationary F0 signal (amplitude envelope and frequency variations eliminated to isolate F0 timing information) to F0 signals that also provide either amplitude envelope or pitch variation cues [similar to Brown and Bacon (2010) and Carroll *et al.* (2011) but with phonemes instead of sentences].

For completeness, it should be noted that neither group showed an increase in IT for place of articulation from the additional $T_{F0\text{-env}}$ signal, which is not surprising given that place of articulation cues are heavily dependent on high-frequency (>600 Hz) fine-structure information (Rosen, 1992) and formant trajectories in the adjacent vowels (Stevens, 1980), neither of which we would expect to the $T_{F0\text{-env}}$ signal to deliver. It should also be stated that the consonantal findings in this study are pertinent to non-whispered /aCa/ syllables. It is not necessarily clear that the results would generalize to other consonant tokens, as vowel context and syllable position can have a large effect on the acoustic and perceptual correlates of features (Blumstein and Stevens, 1980).

As it stands, the present study provides an additional explanation for the EAS benefits observed during word/sentence recognition when the acoustic input is limited to a tone representing the target F0 (e.g., Brown and Bacon, 2009a,b; Zhang *et al.*, 2010; Carroll *et al.*, 2011). Our findings confirm previous assumptions that acoustic features in the F0 region provide information about the speech itself, namely by enhancing voicing and manner of articulation cues.

C. Performance differences between the actual and simulated bimodal groups

Actual and simulated bimodal performance differed, particularly for the CI/Vocoder and LP conditions. The simulated subjects performed poorly with the vocoder in comparison to CI performance for the actual subjects for two possible reasons. First, the simulated group had little listening experience using the vocoder (minutes of practice) compared to the experienced CI users (years). Secondly, the vocoder was limited to 4 independent frequency channels, which might yield performance similar to the lowest-performing CI users.

The opposite trend was observed for the LP condition, whereby the simulated bimodal group outperformed the actual bimodal group. There were several possible contributing factors to this performance difference. The most obvious factor is the difference between the available bandwidth in the simulated hearing loss provided to the simulated bimodal subjects and that available in the mean residual acoustic hearing for the actual bimodal subjects (Fig. 1). The 500 Hz cutoff of the simulation is on the higher end of the actual bimodal group, and the slope is shallower in comparison. Also, despite measurable thresholds in the upper frequencies for the actual bimodal group, it is possible that these regions may not actually be receiving useful input, as there may be dead regions in the cochlea of these listeners (Summers *et al.*, 2003; Moore, 2004). These factors, combined with a compressed dynamic range and the fact that testing was unaided, likely reduced the effective acoustic frequency range of the actual bimodal listeners relative to the LP condition for the simulated listeners. In addition to differences in audibility, the actual bimodal listeners may have had a distorted representation of the phonemes even when the signal was audible (Plomp, 1986; Ching *et al.*, 2001). For example, it has been suggested that reduced frequency selectivity (Tyler, 1982; Leek and Summers, 1993, 1996) and/or a reduction in the ability to use temporal fine structure information (Buss *et al.*, 2004; Lorenzi *et al.*, 2006; Hopkins *et al.*, 2008) might reduce the salience of speech cues for hearing impaired listeners compared to listeners with normal hearing. Lastly, the two groups were from vastly disparate age ranges (17–33 years for the normal hearing listeners and 46–83 years for the actual bimodal listeners), so it is possible that cognitive factors may have contributed to differences in performance between groups.

The observed performance differences limit interpretation of the simulation data, particularly for the vowel recognition task where the LP signal appeared to dominate the Vocoder signal for the simulated listeners. However, performance differences between the two groups were much less pronounced for the consonant recognition task. Thus, the simulated group results provide a supportive comparison to the results observed for the actual bimodal listeners in consonant recognition.

D. Mechanisms and implications

The present results add to existing knowledge about the mechanisms and implications of bimodal hearing. The results clearly demonstrate a dual mechanism in the bimodal benefit: On one hand, a tone following the F0 contour of target speech and modulated with an amplitude envelope within the F0 range provides voicing and manner information, which improves consonant but not vowel recognition. On the other hand, LP speech provides additional manner, place, and formant information that improves both consonant and vowel recognition. Understanding these mechanisms has significant impact on processing and rehabilitation strategies of bimodal hearing.

First, because an acoustic representation of F0 improves general speech recognition via increased access to consonant cues, in addition to providing prosodic cues and/or lexical boundaries (Spitzer *et al.*, 2009), the implication is that this information is important for speech in all cases, not just in situations requiring higher-level mechanisms such as source segregation or glimpsing. These findings strengthen the argument for preserving even the small amount of low frequency residual hearing available to many CI users. With mounting evidence of the benefits of bilateral implantation for localization and speech perception in spatially separated noise (e.g., Gantz *et al.*, 2002; Schleich *et al.*, 2004; Litovsky *et al.*, 2009), many individuals are now receiving a second CI. Nevertheless, bilateral implantation comes at the cost of whatever residual acoustic hearing the patient may have in the second ear prior to implantation (not taking into consideration hybrid EAS). Therefore, the benefits of bilateral implantation must continue to be weighed against the bimodal benefits provided by residual acoustic hearing (e.g., Nittrouer and Chapman, 2009; Mok *et al.*, 2007; Litovsky *et al.*, 2009; Cullington and Zeng, 2011). Preservation of this hearing is especially beneficial because it can be provided at a cost that is a fraction of bilateral implantation. Also, while data were collected using bimodal listeners (residual hearing in the non-implanted ear), there is no reason to think that the findings would not apply to other EAS listeners (i.e., to those with acoustic hearing in the implanted ear). For example, Buchner *et al.* (2009) has already demonstrated that acoustic information below 300 Hz, while completely unintelligible on its own, significantly improved speech perception in hybrid EAS patients. Additionally, we acknowledge that there are many other potential benefits of EAS that the study does not tap into (e.g., tonal languages, binaural cues, music perception, etc.), for which perseveration of acoustic hearing is important.

Second, the present results show that LP speech improves phoneme recognition beyond what is provided from the $T_{F0\text{-env}}$ signal via increased performance in formant recognition and place of articulation cues. However, overall information transfer for place was weak relative to the other consonant features tested. If the present results were extended to the audiovisual case, one might find an even greater benefit from LP or $T_{F0\text{-env}}$, as visual cues are known to be highly complementary to low-frequency acoustic hearing, in part by providing this missing information regarding place of articulation (Grant and Walden, 1996). Being that much of communication happens face to face, EAS listeners often have access to this additional modality (i.e., the visual modality). By enhancing the cues which are not adequately provided by either the CI or low-frequency hearing (namely, place and higher-order formants), additional visual cues may help to complete the picture (Sheffield *et al.*, 2011).

Third, the findings also emphasize the importance of exploring alternative means to deliver F0 information, whether through improved fine-structure processing in CIs (Riss *et al.*, 2008), or by shifting F0 into regions of audibility, as suggested by Brown and Bacon (2010). Even for those without any residual hearing, recent research has shown that F0 can be provided via a tactile aid to improve speech and music perception in CI users (Huang *et al.*, 2010).

Finally, the present results showed significant correlations between hearing profile and vowel perception, highlighting the potential for not only predicting the bimodal benefit prior to implantation, but also optimizing this benefit after the implantation. Unfortunately, the evidence for predicting the EAS benefit based solely on the audiogram is not strong for bimodal listeners (Ching *et al.*, 2004; Gifford *et al.*, 2007) or for hybrid EAS listeners (Luetje *et al.*, 2007; Gifford *et al.*, 2008). It is possible that combining the audiogram with performance in other tasks that measure frequency selectivity or that rely on temporal fine structure processing ability may provide better predictive power of this benefit. However, much more work remains in this area before these possibilities can be realized.

IV. SUMMARY

The relative phonetic contributions of electric and acoustic hearing were evaluated in actual and simulated bimodal listeners. The additional low-frequency signal was either low-passed speech (LP) or a tone following the F0 contour of target speech and modulated with an amplitude envelope of the maximum frequency of the F0 contour ($T_{F0\text{-env}}$). Information transfer was analyzed for both vowel and consonant recognition and related to the audiometric configuration of the actual bimodal subjects. The first aim was to test the hypothesis that, when combined with the CI, the residual low-frequency acoustic hearing provides phonetic information which improves speech understanding beyond that of the CI alone. A second aim was to test the hypothesis that F0 cues available within the low-frequency signal contribute some of this phonetic information. Results are summarized as follows:

(1) LP improved both vowel and consonant recognition by increasing the transfer of vowel formant cues (F1 and F2) and all three consonant features (voicing, manner, and place);
(2) the $T_{F0\text{-env}}$ signal did not improve vowel recognition but significantly improved consonant recognition by increasing the transfer of information regarding the voicing and manner cues;
(3) the benefit provided by LP in vowel recognition was dependent on the cutoff and slope of the listener's hearing loss.

The results confirm previous suggestions that low-frequency acoustic hearing provides phonetic cues that boost CI speech perception, and the F0 component of the acoustic signal allows some access to this phonetic information. These findings emphasize the importance of preserving even very low frequency residual hearing for bimodal listeners.

ACKNOWLEDGMENTS

The authors thank Jeff Carroll and Stephanie Tiaden for their preliminary work, Jennifer Yuan and Benjamin Dong for their help collecting data from the simulated bimodal subjects, and Joshua Bernstein, Abeer Alwan, and three anonymous reviewers for their constructive comments and suggestions on an earlier draft of the manuscript. We would also like to thank all participants in the study for their time and dedication. This work was supported by NIH Grants 1RO1 DC008858 (F.G.Z.) and P30 DC008369 (Raju Metherate).

Armstrong, M., Pegg, P., James, C., and Blamey, P. (**1997**). "Speech perception in noise with implant and hearing aid," Am. J. Otolaryngol. **18**(6, Suppl.), S140–S141.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Blunstein, S. E., and Stevens, K. N. (**1980**). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **67**(2), 648–662.

Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (**1988**). "Voice fundamental frequency as an auditory supplement to the speechreading of sentences," Ear Hear. **9**, 306–312.

Breeuwer, M., and Plomp, R. (**1986**). "Speechreading supplemented with auditorily presented speech parameters," J. Acoust. Soc. Am. **79**, 481–499.

Brokx, J., and Nooteboom, S. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Brown, C. A., and Bacon, S. P. (**2009a**). "Low-frequency speech cues and simulated electric-acoustic hearing," J. Acoust. Soc. Am. **125**(3), 1658–1665.

Brown, C. A., and Bacon, S. P. (**2009b**). "Achieving electric-acoustic benefit with a modulated tone," Ear Hear. **30**, 489–493.

Brown, C. A., and Bacon, S. P. (**2010**). "Fundamental frequency and speech recognition in noise," Hear. Res. **266**, 52–59.

Buchner, A., Shussler, M., Battmer, R. D., Stover, T., Lesinski-Schiedat, A., and Lenarz, T. (**2009**). "Impact of low-frequency hearing," Audiol. Neurotol. **14**(Suppl. 1), 8–13.

Bunton, K. (**2006**). "Fundamental frequency as a perceptual cue for vowel identification in speakers with Parkinson's disease," Folia. Phoniatr. Logop. **58**, 323–339.

Buss, E., Hall, J. W., and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Carroll, J., Tiaden, S., and Zeng, F. G. (**2011**). "Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing," J. Acoust. Soc. Am. **130**(4), 2054–2062.

Chang, J. E., Bai, J. Y., and Zeng, F. G. (**2006**). "Unintelligible low-frequency sound enhances simulated cochlear-implant speech recognition in noise," IEEE Trans. Biomed. Eng. **53**(12), 2598–2601.

Ching, T. Y. C. (**2005**). "The evidence calls for making binaural-bimodal fittings routine," Hear. J. **58**(11), 32–41.

Ching, T. Y. C., Dillon, H., Katsch, R., and Byrne, D. (**2001**). "Maximizing effective audibility in hearing aid fitting," Ear Hear. **22**, 212–224.

Ching, T. Y. C., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**, 9–21.

Cullington, H. E., and Zeng, F. G. (**2010**). "Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing contralateral ear," Ear Hear. **31**(1), 70–73.

Cullington, H. E., and Zeng, F. G. (**2011**). "Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification," Ear Hear. **32**(1), 16–30.

Darwin, C. J., and Hukin, R. W. (**2000**). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," J. Acoust. Soc. Am. **107**(2), 970–977.

Dorman, M. F., Gifford, R., Spahr, A. J., and McKarns, S. (**2008**). "The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies," Audiol. Neurootol. **13**, 105–112.

Dorman, M. F., Raphael, L. J., and Isenberg, D. (**1980**). "Acoustic cues for a fricative-affricate contrast in word-final position," J. Phonet. **8**, 397–405.

Dorman, M. F., Spahr, A. J., Loizou, P. C., Dana, C. J., and Schmidt, J. S. (**2005**). "Acoustic simulations of combined electric and acoustic hearing (EAS)," Ear. Hear. **26**(4), 371–380.

Faulkner, A., and Rosen, S. (**1999**). "Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception," J. Acoust. Soc. Am. **106**, 2063–2073.

Gantz, B. J., Turner, C., and Gfeller, K. E. (**2006**). "Acoustic plus electric speech processing: Preliminary results of a multicenter clinical trial of the Iowa/Nucleus Hybrid implant," Audiol. Neurotol. **11**(Suppl. 1), 63–68.

Gantz, B., Tyler, R. S., Rubenstein, J., Wolaver, A., Lowder, M., Abbas, P., Brown, C., Hughes, M., and Preece, J. P. (**2002**). "Binaural cochlear implants placed during the same operation," Otol. Neurotol. **23**, 169–180.

Gerstman, L. J. (**1957**). "Perceptual dimensions for the friction portions of certain speech sounds," Ph. D. thesis, New York University, New York, NY.

Gifford, R. H., Dorman, M. F., McKarns, S. A., and Spahr, A. J. (**2007**). "Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing," J. Speech Lang. Hear. Res. **50**, 835–843.

Gifford, R. H., Dorman, M. F., Spahr, A. J., Bacon, S. P., Skarzynski, H., and Lorens, A. (**2008**). "Nonlinear cochlear function following cochlear implantation," J. Acoust. Soc. Am. **124**(4), 2164–2173.

Grant, K. W., and Walden, B. E. (**1996**). "Evaluating the articulation index for auditory-visual consonant recognition," J. Acoust. Soc. Am. **100**, 2415–2424.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species–29 years later," J. Acoust. Soc. Am. **87**(6), 2592–2605.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "The effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Howell, P., and Rosen, S. (**1983**). "Production and perception of rise time in the voiceless affricate/fricative distinction," J. Acoust. Soc. Am. **73**, 976–984.

Huang, J., Sheffield, B. M., and Zeng, F. G. (**2010**). "Auditory-tactile integration enhances cochlear implant speech perception in noise," *Proceedings of Assoc. Res. Otolaryngol. MidWinter Meeting*, Anaheim, CA, USA, February, Abstract 620.

Jiang, J., Chen, M., and Alwan, A. (**2005**). "On the perception of voicing in syllable-initial plosives in noise," J. Acoust. Soc. Am. **119**(2), 435–445.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207.

Kong, Y. Y., and Carlyon, R. P. (**2007**). "Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation," J. Acoust. Soc. Am. **121**(6), 3717–3727.

Kong, Y. Y., Stickney, G. S., and Zeng, F.-G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**(3), 1351–1361.

Ladefoged, P., (**1982**). *Syllables and Suprasegmental Features*, in *A Course in Phonetics*, edited by P. Ladefoged (New York, Harcourt Brace Jovanovich), pp. 219–237.

Leek, M. R., and Summers, V. (**1993**). "Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise," J. Acoust. Soc. Am. **94**(6), 3127–3137.

Leek, M. R., and Summers, V. (**1996**). "Reduced frequency selectivity and the preservation of spectral contrast in noise," J. Acoust. Soc. Am. **100**(3), 1796–1806.

Li, N., and Loizou, P. C. (**2008a**). "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**(3), 1673–1682.

Li, N., and Loizou, P. C. (**2008b**). "A glimpsing account for the benefit of simulated combined acoustic and electric hearing," J. Acoust. Soc. Am. **123**(4), 2287–2294.

Litovsky, R. Y., Parkinson, A., and Arcaroli, J. (**2009**). "Spatial hearing and speech intelligibility in bilateral cochlear implant users," Ear Hear. **30**, 419–431.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. **103**, 18866–18869.

Luetje, C. M., Thedinger, B. S., Buckler, L. R., Dawson, K. L., and Lisbona, K. L. (**2007**). "Hybrid cochlear implantation: Clinical results and critical review of 13 cases," Otol. Neurotol. **28**, 473–478.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Mok, M., Galvin, K. L., Dowell, R. C., and McKay, C. M. (**2007**). "Spatial unmasking and binaural advantage for children with normal hearing, a cochlear implant and a hearing aid, and bilateral implants," Audiol. Neurotol. **12**, 295–306.

Mok, M., Grayden, D., Dowell, R. C., and Lawrence, D. (**2006**). "Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears," J. Speech Lang. Hear. Res. **49**, 338–351.

Moore, B. C. J. (**2004**). "Dead regions in the cochlea: conceptual foundations, diagnosis, and clinical applications," Ear. Hear. **25**(2), 98–116.

Nittrouer, S., and Chapman, C. (**2009**). "The effects of bilateral electric and bimodal electric acoustic stimulation on language development," Trends Amplif. **13**(3), 190–205.

Plomp, R. (**1986**). "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," J. Speech Hear. Res. **29**, 146–154.

Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0. discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.

Qin, M. K., and Oxenham, A. J. (**2006**). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," J. Acoust. Soc. Am. **119**(4), 2417–2426.

Repp, B. H., Liberman, A. M., Eccardt, T., and Pesetsky, D. (**1978**). "Perceptual integration of acoustic cues for stop, fricative and affricate manner," J. Exp. Psychol. Hum. Percept. Perform. **4**, 621–637.

Riss, D., Arnolder, C., Baumgartner, W. D., Kaider, A., and Hamzavi, J. S. (**2008**). "A new fine structure speech coding strategy: Speech perception at a reduced number of channels," Otol. Neurotol. **29**, 784–788.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Sagi, E., and Svirsky, M. A. (**2008**). "Information transfer analysis: A first look at estimation bias," J. Acoust. Soc. Am. **123**(5), 2848–2857.

Schleich, P., Nopp, P., and D'Haese, P. (**2004**). "Head shadow, squelch and summation effects in bilateral users of the MED-EL Combi 40/40+ cochlear implant," Ear Hear. **25**, 197–204.

Shallop, J. K., Arndt, P. L., and Turnacliff, K. A. (**1992**). "Expanded indications for cochlear implantation: Perceptual results in seven adults with residual hearing." J. Spoken Lang. Path. Audiol. **16**, 141–148.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**(5234), 303–304.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.

Sheffield, B. M., Bernstein, J. G., and Schuchman, G. (**2011**). "Characterizing the benefit of residual acoustic hearing to auditory and auditory-visual cochlear implant consonant perception," *Proceedings of Assoc. Res. Otolaryngol. MidWinter Meeting*, Baltimore, MD, February, Abstract 959.

Spitzer, S., Liss, J., Spahr, T., Dorman, M., and Lansford, K. (**2009**). "The use of fundamental frequency for lexical segmentation in listeners with cochlear implants," J. Acoust. Soc. Am. **125**, EL236–EL240.

Stevens, K. N. (**1980**). "Acoustic correlates of phonetic categories," J. Acoust. Soc. Am. **68**(3), 836–842.

Stevens, K. N. (**1998a**). "Vowels: Acoustic events with a relatively open vocal tract," in *Acoustic Phonetics*, edited by K. N. Stevens (MIT Press, Cambridge, MA), pp. 257–322.

Stevens, K. N. (**1998b**). "Sonorant consonants," in *Acoustic Phonetics*, edited by K. N. Stevens (MIT Press, Cambridge, MA), pp. 487–556.

Summers, V., and Leek, M. R. (**1998**). "F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss," J. Speech Lang. Hear. Res. **41**, 1294–1306.

Summers, V., Molis, M. R., Musch, H., Walden, B. E., Surr, R. K., and Cord, M. T. (**2003**). "Identifying dead regions in the cochlea: Psychophysical tuning curves and tone detection in threshold-equalizing noise," Ear. Hear. **24**(2), 133–142.

Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**(4), 1729–1735.

Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (**2002**). "Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization," Ear Hear. **23**, 98–105.

Tyler, R. S., Wood, E. J., and Fernandes, M. (**1982**). "Frequency resolution and hearing loss," Br. J. Audiol. **16**(1), 45–63.

Umeda, N. (**1975**). "Vowel duration in American English," J. Acoust. Soc. Am. **58**(2), 435–445.

Vongphoe, M., and Zeng., F. G. (**2005**). "Speaker recognition with temporal cues in acoustic and electric hearing," J. Acoust. Soc. Am. **118**(2), 1055–1061.

von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Sturzebecker, E., and Klinke, R. (**1999**). "Electric-acoustic stimulation of the auditory system," ORL J. Otorhinolaryngol. Relat. Spec. **61**, 334–340.

Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.

Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**(5), 3255–3267.

Zhang, T., Dorman, M. F., and Spahr, A. J. (**2010**). "Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation," Ear Hear. **31**, 63–69.