The role of continuous low-frequency harmonicity cues for interrupted speech perception in bimodal hearing

Soo Hee Oh^{a)} and Gail S. Donaldson^{b)}

*Department of Communication Sciences and Disorders, University of South Florida, PCD 1017,
4202 East Fowler Avenue, Tampa, Florida 33620, USA*

Ying-Yee Kong

*Department of Communication Sciences and Disorders, Northeastern University, 226 Forsyth Building,
360 Huntington Avenue, Boston, Massachusetts 02115, USA*

(Received 13 August 2015; revised 25 January 2016; accepted 29 March 2016; published online 11 April 2016)

Low-frequency acoustic cues have been shown to enhance speech perception by cochlear-implant users, particularly when target speech occurs in a competing background. The present study examined the extent to which a continuous representation of low-frequency harmonicity cues contributes to bimodal benefit in simulated bimodal listeners. Experiment 1 examined the benefit of restoring a continuous temporal envelope to the low-frequency ear while the vocoder ear received a temporally interrupted stimulus. Experiment 2 examined the effect of providing continuous harmonicity cues in the low-frequency ear as compared to restoring a continuous temporal envelope in the vocoder ear. Findings indicate that bimodal benefit for temporally interrupted speech increases when continuity is restored to either or both ears. The primary benefit appears to stem from the continuous temporal envelope in the low-frequency region providing additional phonetic cues related to manner and *F1* frequency; a secondary contribution is provided by low-frequency harmonicity cues when a continuous representation of the temporal envelope is present in the low-frequency, or both ears. The continuous temporal envelope and harmonicity cues of low-frequency speech are thought to support bimodal benefit by facilitating identification of word and syllable boundaries, and by restoring partial phonetic cues that occur during gaps in the temporally interrupted stimulus.

© 2016 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4945747>]

[TCB]

Pages: 1747–1755

I. INTRODUCTION

It is now widely accepted that electric-acoustic stimulation (EAS) in the form of bimodal hearing [cochlear implant (CI) supplemented by low-frequency acoustic hearing in the contralateral ear] or hybrid hearing (CI supplemented by low-frequency acoustic hearing preserved postoperatively in the same ear) has the potential to enhance speech understanding relative to a CI alone (see reviews by Dorman and Gifford, 2008, 2010). This is especially true when the target speech signal occurs in the presence of competing maskers or background noise (e.g., real EAS users: Kong *et al.*, 2005; Zhang *et al.*, 2010; Carroll *et al.*, 2011; Visram *et al.*, 2012a,b; Neuman and Svirsky, 2013; simulated EAS listeners: Qin and Oxenham, 2006; Li and Loizou, 2008; Tillery *et al.*, 2012).

In general, EAS benefits have been attributed to the fact that periodicity and harmonicity cues are represented more robustly in the low-frequency acoustic signal than in the electrically coded CI signal. This favorable representation of low-frequency cues in the acoustic signal is thought to support several mechanisms of enhancement. First, the low-frequency signal may provide the listener with segmental

speech cues (voicing, manner of articulation, and partial *F1* frequency cues) that are either complementary to, or redundant with, segmental cues available through the CI. By integrating the available speech cues across ears, the listener may be able to improve performance relative to performance with the CI alone (Kong and Braida, 2011; Sheffield and Zeng, 2012; Visram *et al.*, 2012a; Yang and Zeng, 2013). Second, harmonicity cues contained in the low-frequency acoustic signal may improve listeners' ability to segment syllable, word, and phrase boundaries, thereby helping them to accurately decode spectrally degraded signals from the CI ear (Spitzer *et al.*, 2009; Zhang *et al.*, 2010; Kong *et al.*, 2015). As discussed by Li and Loizou (2008) and Dorman and Gifford (2008), low-frequency fine-structure cues improve listeners' access to robust acoustic landmarks (Stevens, 2002), such as the onset of voicing, that mark syllable structure and word boundaries. Third, a process known as "glimpsing" may contribute to EAS benefit when speech occurs in competing backgrounds (Cooke, 2006; Kong and Carlyon, 2007; Li and Loizou, 2008; Brown and Bacon, 2009a,b). When a competing signal is present, portions of the target speech are masked by the interfering sound, causing temporal and spectral interruptions in the audible speech stream. The remaining small fragments or "glimpses" of the speech signal must then be decoded and integrated to reconstruct the target message. Cooke (2006) suggested that the local signal-to-noise ratio (SNR) required to generate usable

^{a)}Current address: Audiology Institute, Hallym University of Graduate Studies, 405, Yeoksam-no, Gangnam-gu, Seoul 135-841, Korea.

^{b)}Electronic mail: gdonalds@usf.edu

glimpses of the speech stimulus is ~ -5 dB; however, CI listeners may require more favorable SNRs due to their reduced spectral resolution.

Li and Loizou (2008) demonstrated that the SNR of noisy speech is generally higher in the low-frequency region than in mid- and high-frequency regions, allowing normal-hearing (NH) listeners to more easily extract low-frequency cues from the speech-plus-masker mixture. CI users are less able to take advantage of this factor due to poor spectral resolution in the tonotopic domain (which effectively reduces SNRs) and because temporal fine-structure cues are absent in the electrical stimulus. However, EAS has the potential to restore CI users' ability to make use of favorable SNRs in the low-frequency region. Li and Loizou (2008) argued that low-frequency residual hearing can give EAS listeners access to a continuous representation of the low-frequency speech signal, not available in the CI signal, that supports EAS benefit. Evidence supporting this view has been reported in studies by Kong and Carlyon (2007) and Brown and Bacon (2009b), which demonstrated that EAS benefit can occur even when the glimpsed low-frequency signal is limited to periodicity or harmonicity cues. In general, these studies suggest that voicing and F_0 contour cues contained in the low-frequency harmonic complex (HC) facilitate EAS listeners' ability to identify the voiced portion of the target speech when speech is embedded in noise.

Although glimpsing has traditionally been associated with speech perception in the presence of competing signals, temporally interrupted speech has been used to examine the potential benefit of bimodal hearing for a listening condition that is similar to speech perception in noise, i.e., one that requires the listener to fill-in brief, missing elements of the ongoing speech stream. The "on" portions of the interrupted speech are analogous to clearly perceived glimpses of the target speech while the "off" portions are analogous to the regions of low SNR where the target speech is inaudible due to masking. As implemented in the present study (5-Hz gating with a 50% duty cycle), temporal interruption results in 100-ms segments of speech alternated with 100-ms segments of silence. The silent segments tend to remove phonemes or syllables from the sentences, but not full words. Thus, the listener receives partial phonetic information for most words in each sentence, similar to the situation in which the syllabic energy peaks of one talker's speech interfere with the continuity of speech information from another. Although the glimpses of speech created by temporal interruptions are different in character from those occurring in a speech-noise mixture, the use of temporal interruptions (in lieu of a spectrally and temporally fluctuating masker) allows for systematic control of the low-frequency cues presented during gaps in the speech stream.

To investigate the role of low-frequency harmonicity cues related to the glimpsing mechanism for EAS benefit, the present study makes use of temporally interrupted sentences to examine the extent to which a continuous representation of the low-frequency speech stream facilitates listeners' ability to reconstruct the original sentence. Two experiments were conducted using NH listeners to simulate bimodal hearing. In both experiments, listeners heard 12-channel noise-vocoded speech in the simulated CI ear and low-pass (LP) filtered

signals in the opposite ear. Experiment 1 examined the benefit of restoring a continuous temporal envelope to the low-frequency ear while the vocoder ear received a temporally interrupted stimulus. We hypothesized that continuous low-frequency temporal envelope cues would enhance performance in the bimodal condition as compared to the unilateral (vocoder-alone) condition. Experiment 2 examined the effect of providing continuous harmonicity cues in the low-frequency ear, as compared to restoring a continuous temporal envelope (without periodicity cues) in the vocoder ear. In this case, we hypothesized that the bimodal condition would produce better performance than the unilateral condition owing to the contribution of low-frequency harmonicity cues in the low-frequency ear.

We tested NH listeners, as opposed to real bimodal listeners, because we were primarily interested in fundamental questions regarding the role of harmonicity cues to processes, such as glimpsing, that contribute to bimodal benefit. By testing young, NH listeners under simulated bimodal conditions, we were able to insure that the nature of sensory input (e.g., spectral resolution in the vocoder ear, low-frequency bandwidth in the LP ear) as well as top-down processing ability was relatively constant across individuals, thereby eliminating a source of variability expected among real bimodal listeners. This approach allowed us to evaluate the contribution of the harmonicity cue to bimodal benefit in the most favorable situation, i.e., when all acoustic cues provided were audible and processed efficiently by young listeners.

II. METHODS

A. Subjects

Twenty young normal-hearing (YNH) subjects (18–30 yr of age) were divided into 2 groups, with 12 subjects tested in experiment 1 and 8 subjects tested in experiment 2. All subjects were native speakers of American English. Subjects provided informed consent and were compensated for their participation. Study procedures were approved by the University of South Florida Institutional Review Board.

B. Stimuli

Stimuli were City University of New York (CUNY) sentences (Boothroyd *et al.*, 1985) recorded by an adult female speaker of standard American English in a conversational speaking style, as described by Kong *et al.* (2015). The CUNY sentence corpus consists of 60 lists sentences, and each list includes 12 sentences. Within each list, 1 sentence covers each of the following 12 topics: food, family, work, clothes, homes, animals, sports and hobbies, weather, health, seasons and holidays, money, and music. Sentence length varies from 3 to 14 words and each list includes 4 statements, 4 questions, and 4 commands. There are 2–9 key words per sentence, and 52–60 key words per list. Two example sentences from CUNY List 22, with key words underlined, are: *Take the dog for a walk. Do you think you will go to the mountains for your vacation this year?*

The recorded CUNY sentences were scaled to have a constant root-mean-square (RMS) amplitude, and then

processed using MATLAB R2012b (MathWorks, Inc., 1984, Natick, MA) and Praat (Boersma and Weenick, 2009) to generate five types of modified stimuli, described below, that were used in experiments 1 and 2 of the study. Figure 1 shows spectrograms for each of the five stimulus types for the initial portion of a single sentence.

1. Gated vocoded (gV) sentences

Sentences were square-wave gated with silence at a rate of 5 Hz (50% duty cycle; alternating 100 ms segments of speech and silence) with 5-ms raised-cosine ramps applied to onsets/offsets. Gated sentences were then subjected to 12-channel noise-band vocoding (Shannon *et al.*, 1995). Vocoder processing consisted of passing the signal through a Butterworth high-frequency emphasis filter; bandpass filtering the signal through a series of third-order, logarithmically spaced elliptical filters (spanning 80–8800 Hz); extracting the temporal envelope in each band using the Hilbert transform and using the envelope to modulate the amplitude of a white noise source; bandpass filtering the envelope-modulated noise using the same bandpass filters used in the analysis step; summing the noise bands across channels; and, finally, scaling the RMS of the vocoded sentence to match the intensity of the original gated sentence. gV sentences provided 100-ms glimpses of the speech signal that contained degraded spectral cues but lacked periodicity and temporal fine-structure information.

2. Noise-filled vocoded (nfV) sentences

Sentences were gated as described for the gV sentences, and then silent intervals were filled with speech-shaped noise that was amplitude modulated with the temporal envelope of

the unprocessed broadband sentence. The noise-filled sentences were subjected to 12-channel noise-band vocoding as described for the gV stimuli. nfV sentences provided a continuous representation of the temporal envelope of the broadband sentence (which was disrupted in the gV sentences), but eliminated spectral information during the noise-filled gaps.

3. Continuous low-pass harmonic complexes (cLPHCs)

Equal-amplitude HCs, representing the periodic components of voiced speech segments, were extracted from the original (unprocessed) sentences. The wideband HCs were then LP-filtered at 500 Hz (60 dB/octave) and amplitude modulated with the 500 Hz LP-filtered sentence envelope. cLPHC stimuli generally preserved the first three harmonics, providing a continuous representation of the temporal envelope of the low-frequency voiced segments of the sentence, as well as the F_0 frequency contour.

4. Gated low-pass harmonic complexes (gLPHCs)

Sentences were gated as described for the gV sentences. Equal-amplitude HCs were extracted from the gated sentences, LP-filtered at 500 Hz (60 dB/octave) and amplitude modulated with the envelope of the 500 Hz LP-filtered sentence. gLPHC stimuli were identical to the cLPHC stimuli except that representations of the temporal envelope and F_0 frequency were interrupted by intervals of silence imposed by the gating.

5. Noise-filled low-pass harmonic complexes (nfLPHCs)

Silent gaps in the gLPHC stimuli were filled with speech-shaped noise that was LP-filtered at 500 Hz and

FIG. 1. Spectrograms illustrating the five stimulus types used in experiments 1 and 2, for the first 1.5 s (underlined) of the sentence “Do you want to have a barbeque this evening.” The original (unprocessed) stimulus is also shown (upper left panel).

modulated with the temporal envelope of the 500-Hz LP-filtered sentence. nLPHC stimuli preserved the continuous temporal envelope of the LP-filtered sentence, but provided discontinuous (gated) information about voicing and the F_0 frequency contour.

Stimuli were played out from a personal computer through a Lynx L22 sound card (Lynx Studio Technology, Inc., Costa Mesa, CA), attenuated by a Tucker-Davis PA-5 attenuator (Tucker Davis Technology, Alachua, FL) and transmitted through Sennheiser HD 600 headphones (Sennheiser Electronic GmbH & Co., Wedemark, Germany) to the listener, who was seated inside a double-walled sound room.

Subjects heard vocoded stimuli (gV or nV) in one ear, with or without a low-frequency stimulus (gLPHC, nLPHC, or cLPHC) in the opposite ear. In total, six different listening conditions were used in experiments 1 and 2. These conditions are summarized in Table I.

Vocoded sentences and LPHC stimuli were presented at 70 dB sound pressure level (SPL) and 80 dB SPL, respectively. During pilot testing, these stimulus levels were found to produce comfortable loudness in both ears, and to produce balanced loudness across ears for the bimodal conditions.

Half of the subjects in each experiment received vocoder stimuli in the left ear and LP stimuli in the right ear; the other half received stimuli in the opposite configuration. During sentence recognition testing, the subject heard each sentence one time, and repeated as many words from the sentence as possible with guessing encouraged. Subjects received visual correct-answer feedback on training trials; no feedback was given on test trials. During test trials, subjects' verbal responses were recorded for later scoring. Two individuals scored each sentence independently for key words correct; a third scorer served as a tie-breaker when scoring differed between the first two scorers.

C. Training and testing

Prior to data collection, each subject underwent a period of familiarization and training. During the familiarization phase, the subject listened to one list of sentences, first in the unprocessed condition and, subsequently, in the gated-only, vocoded-only, and gated-vocoded conditions. This process allowed the subject to adjust incrementally to the gated-vocoded processing used in test sentences. The subject then completed a series of training trials consisting of 2–4 lists for each of the listening conditions to be included in the

formal testing. For experiment 1, training began with two lists of sentences in the gV condition, followed by two lists each in the gV + gLPHC, gV + nLPHC, gV + cLPHC, and nV + cLPHC conditions (ten lists total). For experiment 2, training began with 4 lists of sentences in the gV condition, followed by 4 lists each in the nV and nV + cLPHC conditions (12 lists total). Following training, each subject underwent baseline testing consisting of three sentence lists presented in the gV listening condition. Subjects who achieved an average score of 25% or higher on baseline testing proceeded to the main part of the experiment; subjects who failed to meet this criterion were dismissed from further testing. The baseline performance criterion was imposed to insure that estimates of bimodal benefit were not influenced by floor effects. A total of 25 subjects were recruited for the study in order to identify 20 who met the baseline criterion.

During formal testing, sentence stimuli were presented in blocks of four lists of sentences where each block included one practice list followed by three test lists for a given listening condition. To minimize the potential influence of learning effects (beyond the training period) on group outcomes, the order of listening conditions, and the assignment of lists to conditions, were randomized independently across subjects.

D. Calculation of bimodal benefit

Bimodal benefit was assessed using the *percentage-point gain* metric, which computes gain as the arithmetic difference in percent-correct score for the bimodal listening condition relative to the vocoder-alone condition. Percentage-point gain provides an intuitive measure of bimodal benefit, and has been used in a number of previous studies (e.g., Kong *et al.*, 2005; Kong and Carlyon, 2007; Başkent and Chatterjee, 2010; Başkent, 2012; Kong *et al.*, 2015). It was appropriate for the present study because a single measure of baseline performance (i.e., the gV listening condition) was used for all bimodal comparisons (Kong *et al.*, 2015).

III. EXPERIMENT 1

Experiment 1 examined the benefit of restoring a continuous temporal envelope to the low-frequency ear, while the vocoder ear received temporally interrupted sentences. Twelve YNH subjects completed the experiment. Each subject was tested in the following five listening conditions: gV, gV + gLPHC, gV + nLPHC, gV + cLPHC, nV + cLPHC. The gV and gV + gLPHC conditions provided baseline performance levels for temporally interrupted stimuli in the vocoder-alone and bimodal configurations. The remaining bimodal conditions added continuous representations of the temporal envelope and/or harmonicity cues in the LP ear (gV + nLPHC, gV + cLPHC), or added temporal envelope and harmonicity cues in the LP ear together with continuous temporal envelope cues in the vocoder ear (nV + cLPHC).

The left panel of Fig. 2 shows mean performance across the five listening conditions. It is apparent that the nV + cLPHC condition (far right), which provided continuous signals to both ears, produced a substantially higher mean score than the other bimodal listening conditions; however, it is not immediately clear whether the other bimodal

TABLE I. Listening conditions tested in experiments 1 and 2.

Condition	Vocoder ear	Low-frequency ear
gV	Gated	None
nV	Gated; silent interval filled with noise	None
gV + gLPHC	Gated	Gated
gV + nLPHC	Gated	Gated; silent interval filled with noise
gV + cLPHC	Gated	Continuous
nV + cLPHC	Gated; silent interval filled with noise	Continuous

FIG. 2. (Left) Mean percent-correct word recognition scores across 5 listening conditions for 12 YNH subjects. (Right) Scores for the bimodal conditions expressed as percentage-point gain. Error bars indicate ± 1 standard error of the mean.

listening conditions provided a benefit over the baseline (gV) condition.

To assess the differences in performance across the five listening conditions, a one-way repeated measures analysis of variance (RM ANOVA) was completed with the listening condition as the within-subject factor. The main effect of the listening condition was significant ($F[4,44] = 45.89$, $p < 0.001$). Planned *post hoc* pairwise comparisons were performed to evaluate bimodal benefit (four comparisons: each of the bimodal conditions compared to the gV condition) and differences in performance among the four bimodal conditions (six comparisons). The α -level criterion for statistical significance was adjusted to 0.005 based on Bonferroni correction (0.05/number of comparisons). *Post hoc* pairwise comparisons indicated that no bimodal benefit was achieved in the gV + gLPHC condition (i.e., gV + gLPHC produced performance similar to gV), but each of the remaining bimodal conditions produced significant bimodal benefit (i.e., gV + nfLPHC, gV + cLPHC, and nfV + cLPHC all produced significantly higher performance than gV) ($p < 0.005$). Pairwise comparisons also indicated that scores for the nfV + cLPHC condition were significantly greater than scores for the three other bimodal conditions (gV + gLPHC, gV + nfLPHC, and gV + cLPHC) ($p < 0.001$) and that scores for the gV + cLPHC condition were greater than scores for gV + gLPHC condition ($p < 0.005$). None of the remaining comparisons were significant ($p > 0.005$). The right panel of Fig. 2 replots the bimodal data in units of percentage-point gain. The nfV + cLPHC condition produced a percentage-point gain of 25.8 points, which was considerably higher than the gain for the next highest condition (gV + cLPHC, 9.5 points).

Overall, the data shown in Fig. 2 indicate that performance was improved across the bimodal listening conditions by increasing the continuity of signals provided to the low-frequency ear, or both ears. Several specific observations can be made: First, the bilaterally gated condition (gV + gLPHC) provided no bimodal benefit, suggesting that any useful information contained in the LPHC is neutralized by temporal interruptions. Second, the two conditions that restored continuity to the LP ear but not the vocoder ear (gV + nfLPHC; gV + cLPHC) provided modest amounts of bimodal benefit. However, bimodal benefit was no greater for the gV + cLPHC

condition than for the gV + nfLPHC condition. This finding is noteworthy because it suggests that the continuity of low-frequency temporal envelope cues is the primary factor underlying benefit in these conditions, regardless of whether the low-frequency temporal envelope contains continuous harmonics, or a combination of noise and harmonics. Third, restoring continuity of the temporal envelope in the vocoder ear (nfV + cLPHC condition) provided a clear increase in performance as compared to a similar condition in which the vocoder ear received a gated stimulus (gV + cLPHC). What is not clear from the data in Fig. 2 is whether improved performance in the nfV + cLPHC condition, relative to the gV + cLPHC condition, represents a bimodal effect. In other words, is it possible that a similar improvement could be achieved by restoring the temporal envelope in the vocoder ear alone, in the absence of low-frequency stimulation? This question is addressed in experiment 2.

IV. EXPERIMENT 2

Experiment 2 examined the benefit of providing continuous harmonicity cues in the low-frequency ear, as contrasted with the benefit of restoring a continuous temporal envelope (without fine-structure cues) in the vocoder ear. Eight YNH subjects completed the experiment. Each subject was tested in the following three listening conditions: gV, nfV, nfV + cLPHC.

Figure 3 shows the mean percent-correct data for the three listening conditions (left panel) and the corresponding percentage-point gains for the nfV and nfV + cLPHC conditions, relative to the gV condition (right panel). The nfV condition, which provided continuous temporal envelope cues in the vocoder ear, yielded mean performance that was intermediate to performance for the gV (baseline) and nfV + cLPHC conditions. However, statistical analysis failed to confirm a significant difference between scores for the gV and nfV conditions. In this analysis, a one-way RM ANOVA yielded a significant main effect of listening condition ($F[2,14] = 13.66$, $p < 0.001$). Three *post hoc* pairwise comparisons (α -level adjusted to 0.016 based on Bonferroni correction) indicated that performance for the bimodal condition (nfV + cLPHC) was significantly higher than performance for

FIG. 3. (Left) Mean percent-correct word recognition scores across three listening conditions for eight YNH subjects. (Right) Comparison of benefit across two listening conditions that provided continuity information in the vocoded ear, or both ears, expressed as percentage-point gain. Error bars indicate ± 1 standard error of the mean.

either unilateral condition (gV or nfV) ($p < 0.005$); however, performance was not significantly different for the nfV and gV conditions ($p > 0.016$). As shown in the right panel of Fig. 3, percentage-point gain increased from 12 points for the nfV condition to 25 points for the nfV + cLPHC condition.

Overall, the data from experiment 2 (Fig. 3) indicate that experiment 1 subjects' improved performance for the nfV + cLPHC condition, as compared to the gV + cLPHC condition (shown in Fig. 2), is a bimodal effect that cannot be explained solely by the restoration of temporal envelope cues in the vocoder ear. Instead, it suggests an interaction whereby the bimodal benefit associated with continuous temporal envelope cues in the LP ear is enhanced when the vocoder ear receives a continuous, rather than interrupted, signal.

Figure 4 compares performance for the nfV condition, from experiment 2, with performance for the gV + nfLPHC and gV + cLPHC conditions from experiment 1. It can be seen that performance was relatively similar across these

FIG. 4. Comparison of performance for one condition assessed in experiment 2 (nfV) and two conditions assessed in experiment 1 (gV + nfLPHC, gV + cLPHC). Error bars indicate ± 1 standard error of the mean.

conditions. Separate student's t -tests showed that differences were not significant between the nfV and gV + nfLPHC conditions ($p > 0.05$) or between the nfV and gV + cLPHC conditions ($p > 0.05$), further supporting the importance of continuous temporal envelope cues to bimodal benefit, as observed in experiment 1. This finding also indicates that temporal envelope cues provided in the low-frequency ear can effectively substitute for the loss of continuous temporal envelope cues in the vocoder ear.

V. DISCUSSION

A. Role of LF continuity in bimodal benefit

In experiment 1, we examined the extent to which a continuous signal in the low-frequency ear could aid the listener's ability to reconstruct temporally interrupted sentences presented to the vocoder ear. Findings showed that providing a continuous temporal envelope to the low-frequency ear improved performance as compared to interrupted vocoder speech alone. There are two possible explanations for this result. First, the temporal envelope may have provided an additional channel of information within the low-frequency region during silent gaps in the vocoder stimulus, conveying speech cues related to manner of articulation and $F1$ frequency (lower $F1$ s are associated with higher amplitudes in the low-frequency temporal envelope); these cues were then combined across ears and over time with speech cues extracted from glimpses of the vocoder speech. Second, the continuous low-frequency speech envelope may have facilitated integration of the glimpses of vocoder speech over time (i.e., streaming) to create a coherent auditory signal.

In experiment 2, we found that a bimodal condition that provided continuous temporal envelope and harmonicity cues in the low-frequency ear (nfV + cLPHC) improved performance over a noise-filled vocoder stimulus alone (nfV). Considered in isolation, this finding suggests that continuous periodicity/harmonicity information carried in the cLPHC can support improved identification of word and syllable boundaries (i.e., enhanced segmentation), as suggested

previously by Kong *et al.* (2015). However, the data obtained in experiment 1 with gated vocoder speech showed that similar bimodal benefit was achieved when the low-frequency ear received noise-filled LPHCs (gV + nLPHC condition) as when it received continuous LPHCs (gV + cLPHC condition). This finding suggests that harmonicity cues may be secondary in importance to continuous temporal envelope cues in mediating bimodal benefit, at least when the CI (or vocoder) ear is presented with temporally interrupted stimuli. In this case, the temporal envelope of low-frequency speech (with or without continuous harmonicity cues) likely provided segmental cues (voicing, manner of articulation, or *F1* frequency) during silent gaps in the vocoder stimulus. Any additional benefits of providing continuous harmonicity cues appear to be relatively small (or nonexistent) in experiment 1. The situation in experiment 2 is different. Here, the nFV stimulus, unlike the gV stimulus in experiment 1, provided a continuous speech envelope. Thus, the additional improvement observed in the bimodal condition (nFV + cLPHC), compared to the nFV condition, may be attributable to the saliency of harmonicity cues for speech segmentation.

Although the temporal envelope of continuous vocoder speech (or noise-filled vocoder speech) carries information concerning syllable and word boundaries, spectral degradation may hinder listeners' temporal processing abilities (Oxenham and Kreft, 2014) and negatively impact the perception of prosodic cues such as vowel length and word stress (Morris *et al.*, 2013). Kong *et al.* (2015) showed that simulated bimodal listeners were able to achieve significant amounts of bimodal benefit when receiving continuous vocoded sentences in one ear and a continuous LPHC in the other. Given that all speech materials were presented in quiet, they attributed the observed bimodal benefit to improved speech segmentation (i.e., enhanced definition of syllable and word boundaries) when robust harmonicity cues were present in the bimodal listening condition.

Improved lexical segmentation and improved perception of prosodic cues have been demonstrated in previous studies of EAS benefit (e.g., Spitzer *et al.*, 2009; Most *et al.*, 2011). Spitzer *et al.* (2009) investigated the possibility that low-frequency fine-structure cues contribute to bimodal benefit by supporting lexical segmentation. They solicited lexical boundary judgments for phrases with artificial syllabic stress patterns when low-frequency fine-structure cues were either preserved or absent, and contrasted performance between conventional unilateral CI (or vocoder-alone in simulation) and bimodal users (or simulated bimodal hearing). Their results showed that the presence of the low-frequency fine-structure cue that contains *F0* contour information influenced listeners' lexical boundary judgments to a greater extent than temporal envelope cues alone. These findings support the role of low-frequency fine-structure cues for the perception of speech prosody.

Most *et al.* (2011) examined the effect of low-frequency hearing on three tasks of prosody perception (intonation discrimination, identification of syllable stress patterns, and identification of word emphasis) in CI users with varied amounts of low-frequency residual hearing in the contralateral ear. As

a group, they reported significant bimodal benefit for all tasks. Syllable stress patterns and word emphasis are important cues for speech segmentation (Cutler and Carter, 1987; Cutler and Butterfield, 1992), and this finding further supports our argument that continuous harmonicity cues in the low-frequency region contribute to syllable/word boundary marking at the sentence level.

Taken together, the findings of Spitzer *et al.* (2009), Most *et al.* (2011), Kong *et al.* (2015), and the present study suggest that both harmonicity cues and the temporal envelope in the low-frequency region may contribute to bimodal benefit for speech recognition. When only glimpses of vocoder speech are available to the listener, the continuous signal in the LF ear (which arises from improved perception of the target speech at low frequencies) may contribute to speech perception (1) by transmitting additional segmental cues related to manner of articulation and/or *F1* frequency that are absent or weakly represented in the vocoded signal, and (2) by facilitating speech segmentation, thereby enhancing identification of syllable and word boundaries.

To summarize, when speech presented to the simulated CI (vocoder) ear is temporally interrupted, the loss of continuous temporal envelope cues and associated phonetic information during silent gaps in the speech stream contribute to reductions in speech recognition performance. The present data indicate that restoration of the temporal envelope in the low-frequency region is primarily responsible for bimodal benefit in this situation, presumably because the temporal envelope can facilitate speech segmentation as well as restore partial phonetic cues that occur during gaps in the temporally interrupted stimulus. Low-frequency harmonicity cues appear to provide additional benefit when a continuous representation of the temporal envelope is present in the low-frequency ear, or both ears. Although the present data were obtained using temporally interrupted speech, we would expect a similar contribution of low-frequency temporal envelope and harmonicity cues to be conferred when speech is presented in a background of fluctuating noise. There is preliminary evidence supporting this expectation in real bimodal listeners from a recent study by Visram *et al.* (2012a). In their second experiment, bimodal benefit was examined for sentence recognition in modulated noise masking. Four of six CI users showed improved performance for each of two conditions [low-frequency vocoded speech; low-frequency amplitude modulation and frequency modulation (AM-FM) cues provided in a pure-tone carrier] that provided temporal envelope cues, with or without *F0* cues, in the residual hearing ear. The bimodal benefits observed for these four subjects ranged from approximately 1 to 4 dB SNR (their Fig. 4).

Placed within a broader context, the low-frequency temporal envelope that contains harmonicity cues (i.e., the fine structure) is one of several low-frequency cues that may contribute to bimodal benefit in CI users. In a typical listening situation, the bimodal listener receives wide-band, spectrally degraded speech in the cochlear implant ear, and low-frequency speech in the residual hearing ear. Bimodal benefit reflects contributions from both segmental and suprasegmental cues, including *phonetic cues* related to voicing, manner,

and *F1* frequency and *periodicity/harmonicity cues* that facilitate lexical segmentation, as well as encoding prosody and emotional content. When the target speech occurs in noise or talker babble, low-frequency temporal envelope and fine-structure cues appear to take on greater importance due to the relatively greater SNR that exists in the low-frequency region. In this situation, a continuous representation of temporal envelope and fine-structure cues, preserved in the low-frequency ear, facilitates the streaming and reconstruction of the glimpses of target speech.

B. Applicability of findings to real bimodal listeners

The present findings were obtained by testing NH listeners with stimuli designed to simulate bimodal hearing. This approach allowed us to evaluate the role of the low-frequency temporal envelope and fine-structure cues to bimodal benefit in the most favorable situation when all acoustic cues provided were audible and processed efficiently by young adult listeners. The actual amount of benefit could be reduced to a variable extent among real bimodal listeners due to individual differences in neural interface in the implanted ear (e.g., [Bierer et al., 2011](#); [Pfungst et al., 2011](#)), degree of hearing loss or basic auditory function in the residual hearing ear (e.g., [Gifford and Dorman, 2012](#); [Neuman and Svirsky, 2013](#)), integration abilities across electric and acoustic stimulation ([Kong and Braida, 2011](#); [Yang and Zeng, 2013](#)), or a combination of these factors. However, bimodal listeners' perception of the low-frequency temporal envelope, which appears to underlie the bimodal benefit observed here, should be relatively robust to such differences, as it would require audibility only in the lowest frequency range (e.g., <250 Hz) ([Brown and Bacon, 2009a](#); [Zhang et al., 2010](#)). Thus, we would expect real bimodal listeners to benefit from this cue so long as the temporal envelope cue is audible.

ACKNOWLEDGMENTS

This work was completed as part of the doctoral dissertation of S.H.O., and was supported by the National Institutes of Health-National Institute on Deafness and Other Communication Disorders (NIH-NIDCD) Grant No. DC012300 (Y.-Y.K.). The authors thank Jean Krause and Catherine Rogers for useful discussions of the research, Ala Somarowthu for technical assistance, and Courtney Matthews for assistance with data collection and analysis. The authors also thank Tessa Bent and three anonymous reviewers for helpful comments on an earlier version of the manuscript.

Başkent, D. (2012). "Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric-acoustic stimulation," *J. Assoc. Res. Otolaryngol.* **13**, 683–692.

Başkent, D., and Chatterjee, M. (2010). "Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech," *Hear. Res.* **270**, 127–133.

Bierer, J. A., Faulkner, K. F., and Tremblay, K. L. (2011). "Identifying cochlear implant channels with poor electrode-neuron interfaces: Electrically evoked auditory brain stem responses measured with the partial tripolar configuration," *Ear Hear.* **32**, 436–444.

Boersma, P., and Weenink, D. (2009). Praat: Doing phonetics by computer (version 5.1.05) [computer program], <http://www.praat.org> (Last viewed 4/5/2016).

Boothroyd, A., Nanin, L., and Hnath, T. (1985). "A sentence test of speech perception: Reliability, set equivalence, and short term learning," Research report (Speech and Hearing Sciences Research Center, City University of New York, New York).

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

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

Carroll, J., Tiafen, 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**, 2054–2062.

Cooke, M. (2006). "A glimpsing model of speech perception in noise," *J. Acoust. Soc. Am.* **119**, 1562–1573.

Cutler, A., and Butterfield, S. (1992). "Rhythmic cues to speech segmentation: Evidence from juncture misperception," *J. Mem. and Lang.* **31**, 218–236.

Cutler, A., and Carter, D. M. (1987). "The predominance of strong initial syllables in the English vocabulary," *Comput. Speech Lang.* **2**, 133–142.

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

Dorman, M. F., and Gifford, R. (2010). "Combining acoustic and electric stimulation in the service of speech recognition," *Int. J. Audiol.* **49**, 912–919.

Gifford, R. H., and Dorman, M. F. (2012). "The psychophysics of low-frequency acoustic hearing in electric and acoustic stimulation (EAS) and bimodal patients," *J. Hear. Sci.* **2**, 33–44.

Kong, Y.-Y., and Braida, L. D. (2011). "Cross-frequency integration for consonant and vowel identification in bimodal hearing," *J. Speech Lang. Hear. Res.* **54**, 959–980.

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**, 3717–3727.

Kong, Y.-Y., Donaldson, G., and Somarowthu, A. (2015). "Effects of contextual cues on speech recognition in simulated electric-acoustic stimulation," *J. Acoust. Soc. Am.* **137**, 2846–2857.

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**, 1351–1361.

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

Morris, D., Magnusson, L., Faulkner, A., Jönsson, R., and Juul, H. (2013). "Identification of vowel length, word stress, compound words and phrases by postlingually deafened cochlear implant listeners," *J. Am. Acad. Audiol.* **24**, 879–890.

Most, T., Harel, T., Shpak, T., and Luntz, M. (2011). "Perception of supra-segmental speech features via bimodal stimulation: Cochlear implant on one ear and hearing aid on the other," *J. Speech Lang. Hear. Res.* **54**, 668–678.

Neuman, A. C., and Svirsky, M. A. (2013). "Effect of hearing aid bandwidth on speech recognition performance of listeners using cochlear implant and contralateral hearing aid (bimodal hearing)," *Ear Hear.* **34**, 553–561.

Oxenham, A. J., and Kreft, H. A. (2014). "Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing," *Trends Hear.* **18**, 1–14.

Pfungst, B. E., Coles, D. J., Hembrador, S., Kang, S. Y., Middlebrooks, J. C., Raphael, Y., and Su, G. L. (2011). "Detection of pulse trains in the electrically stimulated cochlea: Effects of cochlear health," *J. Acoust. Soc. Am.* **130**, 3954–3968.

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**, 2417–2426.

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

Sheffield, B. M., and Zeng, F.-G. (2012). "The relative phonetic contributions of a cochlear implant and residual acoustic hearing to bimodal speech perception," *J. Acoust. Soc. Am.* **131**, 518–530.

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–EL241.

- Stevens, K. N. (2002). "Toward a model for lexical access based on acoustic landmarks and distinctive features," *J. Acoust. Soc. Am.* **111**, 1872–1891.
- Tillery, K. H., Brown, C. A., and Bacon, S. P. (2012). "Comparing the effects of reverberation and of noise on speech recognition in simulated electric-acoustic listening," *J. Acoust. Soc. Am.* **131**, 416–423.
- Visram, A. S., Azadpour, M., Kluk, K., and McKay, C. M. (2012a). "Beneficial acoustic speech cues for cochlear implant users with residual acoustic hearing," *J. Acoust. Soc. Am.* **131**, 4042–4050.
- Visram, A. S., Kluk, K., and McKay, C. M. (2012b). "Voice gender differences and separation of simultaneous talkers in cochlear implant users with residual hearing," *J. Acoust. Soc. Am.* **132**, EL135–EL141.
- Yang, H. I., and Zeng, F.-G. (2013). "Reduced acoustic and electric integration in concurrent-vowel recognition," *Sci. Rep.* **3**, 1419–1423.
- Zhang, T., Dorman, M. F., and Spahr, A. J. (2010). "Information from the voice fundamental frequency (F_0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation," *Ear Hear.* **31**, 63–69.