Published in final edited form as:

Ear Hear. 2016 ; 37(5): 582–592. doi:10.1097/AUD.0000000000000298.

Top-down Processes in Simulated Electric-Acoustic Hearing: The Effect of Linguistic Context on Bimodal Benefit for Temporally Interrupted Speech

Soo Hee Oh¹, Gail S. Donaldson¹, and Ying-Yee Kong²

¹Department of Communication Sciences and Disorders, University of South Florida, PCD 1017, 4202 East Fowler Avenue, Tampa, Florida 33620

²Department of Communication Sciences and Disorders, Northeastern University, 226 Forsyth Building, 360 Huntington Ave., Boston, MA 02115

Abstract

Objectives—Previous studies have documented the benefits of bimodal hearing as compared with a CI alone, but most have focused on the importance of bottom-up, low-frequency cues. The purpose of the present study was to evaluate the role of top-down processing in bimodal hearing by measuring the effect of sentence context on bimodal benefit for temporally interrupted sentences. It was hypothesized that low-frequency acoustic cues would facilitate the use of contextual information in the interrupted sentences, resulting in greater bimodal benefit for the higher context (CUNY) sentences than for the lower context (IEEE) sentences.

Design—Young normal-hearing listeners were tested in simulated bimodal listening conditions in which noise band vocoded sentences were presented to one ear with or without low-pass (LP) filtered speech or LP harmonic complexes (LPHCs) presented to the contralateral ear. Speech recognition scores were measured in three listening conditions: vocoder-alone, vocoder combined with LP speech, and vocoder combined with LPHCs. Temporally interrupted versions of the CUNY and IEEE sentences were used to assess listeners' ability to fill in missing segments of speech by using top-down linguistic processing. Sentences were square-wave gated at a rate of 5 Hz with a 50 percent duty cycle. Three vocoder channel conditions were tested for each type of sentence (8, 12, and 16 channels for CUNY; 12, 16, and 32 channels for IEEE) and bimodal benefit was compared for similar amounts of spectral degradation (matched-channel comparisons) and similar ranges of baseline performance. Two gain measures, percentage-point gain and normalized gain, were examined.

Results—Significant effects of context on bimodal benefit were observed when LP speech was presented to the residual-hearing ear. For the matched-channel comparisons, CUNY sentences showed significantly higher normalized gains than IEEE sentences for both the 12-channel (20 points higher) and 16-channel (18 points higher) conditions. For the individual gain comparisons that used a similar range of baseline performance, CUNY sentences showed bimodal benefits that

Address correspondence to Gail S. Donaldson, Department of Communication Sciences and Disorders, University of South Florida, PCD 1017, 4202 East Fowler Avenue, Tampa, Florida 33620, USA. gdonalds@usf.edu.

Financial Disclosures/Conflicts of Interest: There are no conflicts of interest to report. This research was funded by NIH-NIDCD DC012300 (Y.-Y. Kong, P.I.)

were significantly higher (7 percentage points, or 15 points normalized gain) than those for IEEE sentences. The bimodal benefits observed here for temporally interrupted speech were considerably smaller than those observed in an earlier study that used continuous speech (Kong et al., 2015). Further, unlike previous findings for continuous speech, no bimodal benefit was observed when LPHCs were presented to the LP ear.

Conclusions—Findings indicate that linguistic context has a significant influence on bimodal benefit for temporally interrupted speech and support the hypothesis that low-frequency acoustic information presented to the residual-hearing ear facilitates the use of top-down linguistic processing in bimodal hearing. However, bimodal benefit is reduced for temporally interrupted speech as compared to continuous speech, suggesting that listeners' ability to restore missing speech information depends not only on top-down linguistic knowledge, but also on the quality of the bottom-up sensory input.

INTRODUCTION

A listener's ability to comprehend a spoken message depends on the reception of sensory inputs (bottom-up processes) as well as the application of world and linguistic knowledge to interpret the cues received (top-down processes). In bimodal hearing, also known as electric-acoustic stimulation (EAS), the listener receives sensory inputs through two sources: the electrically coded inputs supplied by a cochlear implant (CI) and the acoustic inputs supplied by low-frequency residual hearing in the contralateral ear. Compared to electric-alone stimulation, the addition of low-frequency acoustic cues improves the quality of the sensory information and facilitates the top-down processes that allow the listener to fill-in missing parts of the message. These advantages of EAS are particularly important when speech recognition occurs in challenging listening conditions, such as in the presence of background noise.

Several previous studies have explored the specific low-frequency cues that underlie *bimodal benefit*, i.e., the improved performance observed when low-frequency residual hearing is used to supplement implant-processed speech. These studies suggest that low-frequency cues related to voicing, pitch contour, the temporal envelope, and first formant (F1) frequency contribute to bimodal benefit for speech recognition in both quiet and noise (Kong and Carlyon 2007; Brown and Bacon 2009a,b; Zhang et al. 2010; Carroll et al. 2011; Kong and Braida 2011; Sheffield and Zeng 2012).

While it is generally accepted that low-frequency cues can enhance the sensory inputs received by a cochlear implant listener, the potential influence of low-frequency cues on top-down processes in bimodal hearing is less well understood. To date, two general approaches have been used to investigate the role of top-down processing in bimodal hearing. The first approach involves the use of temporally interrupted speech (Bakken and Chatterjee 2010; Chatterjee et al. 2010) and the assessment of phonemic restoration (PR) effects that occur when temporal interruptions are filled with noise (Bakken 2012). The second approach involves the use of speech materials that contain varying amounts of linguistic context (Brown and Bacon 2009a; Kong et al. 2015). Each of these approaches is described further in the sections that follow.

Temporally interrupted speech

In noisy environments, listeners can better extract spectrotemporal fragments, or “glimpses” of the speech signal in low-frequency regions where the signal-to-noise ratio (SNR) is generally more favorable than in higher frequency regions (Li and Loizou 2008).¹ This so-called glimpsing mechanism has been shown to contribute to bimodal benefit when speech is presented in a competing background (Kong and Carlyon 2007; Brown and Bacon 2009a). The perception of temporally interrupted speech represents a special condition of glimpsing. In most studies of temporally interrupted speech, speech stimuli (words or sentences) are multiplied with a square-wave gating function having a 50% duty-cycle. This results in a stimulus consisting of alternating, equal-duration segments of speech and silence. As with spectrotemporal glimpsing, “glimpses” of the preserved segments of speech may be extracted by the listener and integrated to construct a coherent speech stream. Because portions of the speech signal are missing, the reconstruction process requires the listener to make use of top-down processing to fill-in the missing sensory information.

Temporal interruptions cause proportionally larger reductions in sentence intelligibility for CI users as compared to normal-hearing (NH) listeners (Nelson et al. 2003; Nelson and Jin 2004; Chatterjee et al. 2010). This is presumed to occur because CI users receive spectrally degraded representations of the preserved segments of the speech signal, and additionally may have reduced temporal resolution (Busby and Clark, 1998; Fu et al., 2001) that leads to reduced glimpsing proficiency. However, CI users’ perception of temporally interrupted sentences may be substantially improved when low-frequency acoustic cues are added in the bimodal listening condition. In simulated bimodal listeners, Ba kent and Chatterjee (2010) found that combining acoustic low-pass (LP) filtered speech (< 500 Hz) with 4- or 8-channel vocoder processing improved intelligibility scores for temporally interrupted sentences by about 20 percentage points over the vocoder-alone condition. This led them to speculate that low-frequency acoustic cues enhance CI users’ ability to integrate preserved segments of the temporally interrupted signal.

Additional evidence of bimodal benefit has been observed in PR effects. PR refers to the improvement in speech recognition that occurs when the silent gaps in temporally interrupted speech are filled with noise (Warren 1970; Warren and Obusek 1971; Bashford et al. 1992). The fill-in noise is thought to improve speech intelligibility by restoring the listeners’ perception of stimulus continuity. Because the noise contains no actual speech cues, PR effects are attributed to the listener’s use of top-down processing to fill-in the missing bottom-up information. PR effects have not been studied extensively in EAS; however, a study by Ba kent (2012) examined PR in simulated CI and hybrid EAS listening conditions and showed small amounts of bimodal benefit. Hybrid stimulation refers to the variant of bimodal stimulation in which electric and acoustic signals are presented to the same ear. In Ba kent’s study, vocoded speech (4, 8, 16, or 32 channels) was presented either across the full range of speech frequencies, or limited to higher frequencies (> 500 Hz) in combination with acoustic stimulation at lower frequencies (< 500 Hz). Significant effects of

¹Previous studies by Kong and Carlyon (2007) and Brown and Bacon (2009a) have shown that the low-frequency pitch contour cue does not seem to facilitate sound source segregation in simulated EAS listening. Thus, it appears that the glimpsing mechanism for EAS benefit is more likely to arise from the more favorable SNR at low frequencies, as suggested by Li and Loizou (2008).

PR were observed for the highest channel conditions for both the vocoder-alone (6 percentage points for 32 channels) and hybrid listening conditions (7 and 9 percentage points, respectively, for 16 and 32 channels). Based on these results, Ba Kent suggested that hybrid stimulation may provide a small bimodal benefit for PR effects, attributable to an enhancement of top-down processing.

Speech stimuli varying in linguistic context

Additional evidence for a role of top-down processing in bimodal hearing has come from studies that investigated the effects of linguistic context on bimodal benefit. According to Needleman (1998, p. 306): “Linguistic context effects refer to the enhancement in speech recognition ability resulting from the contribution of word context to phoneme recognition, and the contribution of sentence context to word recognition.” Many factors are known to contribute to such context effects, including knowledge of the topic of the sentence; lexical factors such as word frequency, familiarity and meaning (i.e., real versus nonsense words); and prosody and intonation (Needleman, 1998).

To date, two studies have demonstrated evidence of context effects in bimodal benefit by testing simulated bimodal listeners with sentences having high versus low levels of sentence context (Brown and Bacon 2009a; Kong et al. 2015). Both studies used higher-predictability City University of New York (CUNY) sentences (Boothroyd et al. 1985) and lower-predictability IEEE sentences (IEEE 1969). Brown and Bacon (2009a) observed that bimodal benefit for CUNY sentences (57 percentage points) was larger than bimodal benefit for IEEE sentences (24–38 percentage points). However, the IEEE and CUNY sentences were tested in different groups of subjects, and factors such as speaking style were not matched across the two sets of sentence materials. Kong et al. (2015) measured context effects using a within-subjects repeated measures design and sentences that were recorded from a single female talker who was trained to use a constant speaking rate (words/minute) and range of pitch excursion (Hz) for both IEEE and CUNY sentences. Listeners heard noise-band vocoded speech (2 to 6 channels) in one ear paired with either 1) no stimulus, 2) LP speech (< 500 Hz) or 3) low-frequency harmonic complexes (LPHCs; < 500 Hz) in the opposite ear. LPHCs retain voicing, F0 contour, and low-frequency temporal-envelope cues, but lack other low-frequency phonetic cues (e.g., partial F1 frequency information) contained in LP speech. The effects of sentence context were evaluated by comparing percentage-point gain or normalized gain between the CUNY and IEEE sentences. Percentage-point gain was calculated as the difference in percent-correct score for the bimodal condition relative to the vocoder-alone condition. Normalized gain compared the bimodal benefit achieved by the listener to the total possible bimodal benefit, as expressed in Equation (1),

$$G = (B - V) / (100 - V), \quad \text{Eq. (1)}$$

where G is the normalized bimodal benefit, B is the percent-correct score measured in the bimodal listening condition and V is the percent-correct score measured with the vocoder alone.

Kong et al. (2015) found that the speech materials with higher levels of linguistic context (CUNY sentences) supported larger amounts of bimodal benefit than speech materials with lower levels of context (IEEE sentences). The largest bimodal benefits were observed when LP speech was presented to the simulated residual hearing ear. In this case, normalized gains were 30–35 points larger for CUNY sentences than for IEEE sentences in matched-channel comparisons (i.e., either 3 or 4 vocoder channels). The context effect was also observed when data were combined across channel conditions that yielded baseline (vocoder-alone) performance between 50 and 80%. In this case, CUNY sentences demonstrated greater gains than IEEE sentences by 9 percentage points and 33 points of normalized gain. Thus, clear effects of context were observed for bimodal benefit in the bimodal condition that paired vocoded speech with LP speech in the opposite ear. Context effects were also observed for bimodal benefit measured when vocoded speech was paired with LPHCs; however, both the magnitude of bimodal benefit and the effects of context on bimodal benefit were smaller for LPHCs than for LP speech.

Rationale for the present study

Kong et al.'s (2015) findings provide relatively strong evidence that residual acoustic hearing can enhance the use of contextual information in continuous speech, implicating an interaction between improved sensory input associated with the addition of low-frequency acoustic cues and top-down linguistic processing. On the other hand, Ba kent (2012) showed that bimodal hearing provides only weak benefits for restoring perceptual continuity in temporally interrupted sentences, suggesting that the top-down benefits of low-frequency residual hearing may be weakened when a spectrally degraded speech stream is further degraded by temporal interruptions.

The primary goal of the present study was to investigate the potential effects of context on bimodal benefit for temporally interrupted speech as a means of further investigating the role of top-down processing in EAS. Temporally interrupted signals were used because they provide an additional challenge to the listener beyond the spectral degradation imposed by vocoder processing. That is, they force the listener to rely even more heavily on top-down processing to fill in the missing information corresponding to temporal gaps in the signal. In addition, the process by which listeners extract information from the 'on' portions of speech and fill in missing information from the 'off' portions of speech is similar to the glimpsing process that has been implicated as contributing to bimodal benefit when speech occurs in the presence of background noise. Unlike the dips that occur in a speech-noise mixture, however, the gaps in temporally interrupted speech are systematic and can be precisely controlled.

A secondary goal of the study was to determine whether bimodal benefit for temporally interrupted speech was supported to the same extent by LPHCs as by LP speech. As noted earlier, Kong et al. (2015) found that LPHCs supported bimodal benefit when tested with continuous sentences, but that the magnitude of benefit was less for LPHCs than for LP speech. This finding is consistent with the notion that LPHCs contain some, but not all, of the relevant low-frequency speech cues that support bimodal benefit with LP speech, including cues to voicing, pitch contour, and the temporal envelope. It has been speculated

that an important role of low-frequency harmonicity cues is to support lexical segmentation – i.e., to assist the listener with identification of word and phrase boundaries [e.g., Tyler and Cutler (2009) in normal hearing; Spitzer et al. (2009) and Hu and Loizou (2010) in bimodal hearing]). However, imposing temporal interruptions on LPHCs could reduce their effectiveness in this regard; thus, it might be expected that LPHCs would provide reduced bimodal benefit for temporally interrupted speech as compared to continuous speech. The present study examined whether LPHCs support significant bimodal benefit for temporally interrupted sentences, or if the gains attributable to LPHCs for continuous stimuli are negated by temporal interruptions.

To address the study goals, simulated bimodal listeners were tested in three listening conditions (vocoder-alone, vocoder combined with LP speech, and vocoder combined with LPHCs) using temporally interrupted CUNY and IEEE sentences. Bimodal benefit was compared across the two sentence types, using both matched-channel comparisons and for individual gain data with similar ranges of baseline (vocoder-alone) performance. It was hypothesized that, similar to continuous speech, low-frequency acoustic cues would facilitate the use of linguistic context in the interrupted sentences, resulting in greater bimodal benefit for the higher context CUNY sentences than for the lower context IEEE sentences. However, the context effect was expected to be smaller for interrupted speech than for continuous speech because temporal degradation of the bottom-up speech cues was expected to limit listeners' use of top-down linguistic processing.

Simulated (normal-hearing) listeners were used rather than real bimodal listeners so that low-frequency audibility and degree of spectral degradation would be similar across listeners. It was important to maintain a constant degree of spectral degradation because the effects of spectral degradation and temporal interruption are likely to interact with the effects of top-down processing, the main focus of the present study. From a broader perspective, the use of simulations allowed us to examine the effect of linguistic context while controlling for *both* the degree of spectral degradation and different levels of baseline performance (Kong et al., 2015). In addition, the potential effects of age could be more easily controlled when using normal hearing listeners, given that bimodal listeners represent a relatively small portion of the population and vary in age.

MATERIALS AND METHODS

Subjects

Subjects were young adults, 18 to 30 years of age, who were native speakers of American English and who had NH bilaterally as characterized by pure tone thresholds ≤ 20 dB HL at octave frequencies between 250 and 8000 Hz. Thirteen subjects were recruited from the student population at the University of South Florida in order to identify twelve subjects who met the baseline performance criterion described later. Subjects underwent informed consent prior to testing and were compensated on an hourly basis for their participation. All procedures were approved by the USF Institutional Review Board.

Stimuli

Speech stimuli consisted of unprocessed and modified versions of the CUNY and IEEE sentences recorded by an adult female speaker of standard American English in a conversational speaking style. They were the same sentence recordings used by Kong et al. (2015). As noted earlier, the talker maintained a consistent speaking rate (words/minute) and pitch excursion (Hz) across all sentences in the CUNY and IEEE corpora. The original recorded sentences were processed as described below. All processing was performed using custom scripts written for MATLAB R2012b (MathWorks, Inc., 1984, Natick, MS) except where otherwise noted.

The recorded CUNY and IEEE sentences were first scaled to have a constant root-mean-square (RMS) amplitude. The RMS-equated sentences were then square-wave gated with silence at a rate of 5 Hz (50% duty cycle i.e., 100 ms segments of speech alternating with 100 ms segments of silence) with 5-ms raised cosine ramps applied to the onsets and offsets of each speech segment. This produced temporally interrupted sentences similar to those used in Chatterjee et al. (2010). Sentences always began with a full segment (100 ms) of speech.

The gated sentences were further modified to create three types of stimuli, as described in the following sections:

Gated-vocoded (gV) stimuli—The gated sentences were processed through a noise band vocoder (NBV) to preserve temporal-envelope cues, similar to the signal processing that occurs in CI systems (Dudley 1939; Shannon et al. 1995). Each sentence was first passed through a high frequency Butterworth pre-emphasis filter. It was then bandpass filtered through a series of third order elliptical filters (30 dB/octave rolloff) into the desired number of logarithmically-spaced frequency bands (Greenwood 1990) with a combined frequency range of 80 to 8800 Hz. The temporal envelope in each band was extracted using the Hilbert transform and the envelope was used to modulate the amplitude of a white noise source. The envelope-modulated noise was bandpass filtered using the same filter parameters used in the analysis stage of processing, after which the modulated noise bands were summed across frequency channels to produce a noise-band vocoded stimulus with the same frequency range as the original stimulus. Finally, the RMS intensity of the vocoded sentence was scaled to match the intensity of the original (unprocessed) gated sentence.

CUNY sentences were processed into 8, 12 and 16 vocoder channels while IEEE sentences were processed into 12, 16 and 32 vocoder channels. These channel conditions were selected on the basis of pilot data with three goals in mind: 1) to avoid floor effects in the vocoder-alone listening condition and to avoid ceiling effects in the bimodal listening condition, 2) to allow for comparison of bimodal benefit for the CUNY and IEEE sentences over similar ranges of vocoder-alone (baseline) performance, and 3) to insure that two vocoder-channel conditions (i.e., 12 and 16 channels) were used for both the CUNY and IEEE sentences so that the effects of sentence context could be evaluated for equal numbers of channels (i.e., equal amounts of spectral information) as well as for similar ranges of baseline performance.

Gated low-pass speech (gLPsp) stimuli—Gated sentences were LP filtered at 500 Hz using Butterworth filters with a rolloff slope of 60 dB/octave to simulate residual low-frequency hearing.

Gated low-pass harmonic complex (gLPHC) stimuli—Equal-amplitude harmonic complexes (HCs) representing voiced speech segments were extracted from the unprocessed sentences using the Praat speech analysis software (Boersma and Weenink 2009). The HCs extracted in this way had flat amplitude envelopes but preserved the F0 frequency contours of the original speech. Next, the HCs were LP filtered at 500 Hz (60 dB/octave, generally preserving the first three harmonics) and amplitude modulated with the envelope of the corresponding continuous LP sentences.

General Procedures

Stimuli were played 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 routed to Sennheiser HD 600 headphones (Sennheiser electric GmbH & CO.KG, 1945, Germany) to the listener who was seated inside a double-walled sound booth. During speech recognition testing, subjects' verbal responses were captured by a MXL Pro Con AC 404 room microphone (Marshall Electronics, El Segundo, CA) and stored as digital files (Cool Edit Pro 2.0, Syntrillium Software Corporation, Phoenix, AZ) for later scoring.

Half of the subjects listened to vocoder stimuli with the left ear and the other half listened to vocoder stimuli with the right ear. In the bimodal listening conditions, low-frequency stimuli (LP speech or LPHCs) were presented to the opposite ear. Vcoded sentences were presented at 70 dB SPL; LP speech and LPHC stimuli were presented at 75 and 80 dB SPL, respectively. In pilot testing, these presentation levels were consistently found to produce comfortable loudness percepts, and to produce balanced loudness across ears for the bimodal listening conditions (Dorman et al. 2014). Because small changes in presentation level had little effect on pilot subjects' perception of balanced loudness, loudness balancing was not performed with individual test subjects.

During sentence recognition testing, the subject was seated in front of a computer monitor and keyboard inside the sound booth. On training trials, each sentence was presented once and the subject pressed a key to view the correct sentence displayed in written form on the computer monitor. On test trials, each sentence was presented once and the subject responded verbally, without correct-answer feedback. Subjects were instructed to repeat as many words as possible from each sentence, and were strongly encouraged to guess the missing parts of sentences. They were given as much time as desired to respond and were given short breaks as needed to insure that they maintained a high level of concentration and attention to the task. A mandatory break of at least 5 minutes duration was imposed approximately every 30 minutes.

Study procedures were completed over seven testing sessions (Figure 1), with each session lasting 1.5 – 2 hours. During **session 1**, the subject completed initial administrative procedures (informed consent, hearing history form) and underwent hearing screening to

confirm NH bilaterally (i.e., thresholds ≤ 20 dB HL at octave frequencies between 250 Hz and 8 kHz). Thereafter, the subject completed an initial period of familiarization and training for the sentence recognition tasks followed by baseline testing to determine whether he or she met the performance criteria (see below and Fig. 1) required to continue to the main study. If the performance criteria were met, the subject was scheduled for **sessions 2 – 7**, which were devoted to sentence recognition testing. All sessions were completed within a four week time period. Details of the procedures used for familiarization, training and testing are provided in the next section.

A repeated measures design was used in which all subjects were tested in all stimulus conditions. Three listening conditions were tested (gV , $gV+gLPsp$ and $gV+gLPHC$) for each of two sentence types (CUNY, IEEE) and 3 vocoder channel conditions (8, 12 and 16 channels for CUNY; 12, 16 and 32 channels for IEEE), leading to a total of 18 stimulus conditions.

Neither the $gLPHC$ nor $gLPsp$ stimuli were tested alone (i.e., in the absence of contralateral stimulation) because previous evidence indicated that both would support near-zero intelligibility. First, Kong and Carlyon (2007) reported that continuous LPHCs, presented alone, produced zero intelligibility; therefore, it was assumed that temporally interrupted versions of these stimuli ($gLPHC$ stimuli) would support zero intelligibility. Second, during pilot testing, it was determined that gated LP speech ($gLPsp$) produced negligible sentence intelligibility for both the CUNY and IEEE sentences.

Training and Testing

The overall plan for training and testing was designed after completing a series of small pilot studies that evaluated several possible training and testing paradigms. The final plan was intended to provide adequate training and testing to ensure that stable estimates of performance were obtained for the 9 test conditions (3 listening condition $\times$ 3 vocoder channel conditions) per sentence type while requiring a feasible amount of test time per subject (10 – 12 hours across 7 test sessions). In addition, it avoided the need to use any sentence list more than once.

Session 1 began with a familiarization phase in which the experimenter played one list each of CUNY and IEEE sentences to the subject, beginning with unprocessed sentences, and proceeding to gated-only, vocoded-only (12-channel CUNY and 16-channel IEEE) and, finally, gated-vocoded versions of the same sentences. In pilot testing, this progressive familiarization approach was found to help subjects acclimate to the difficult listening task of identifying sentences that were both gated and vocoded. The familiarization phase was completed in approximately 15 minutes.

Following familiarization, the training phase began, during which the subject completed the sentence recognition task for 12 lists of sentences with correct-answer feedback. These consisted of 6 lists each of 12-channel CUNY and 16-channel IEEE sentences, divided between the gV (2 lists), $gV+gLPsp$ (2 lists), and $gV+gLPHC$ (2 lists) listening conditions. After each sentence was presented, the subject had the opportunity to practice repeating the sentence verbally before pressing a key to receive correct-answer feedback in the form of the

written sentence displayed on the computer monitor. The training conditions were selected to include both sentence types (CUNY, IEEE) and all three of the listening conditions (gV , $gV+gLpSp$, and $gV+gLPHC$). However, due to limitations in the number of sentence lists available, only the middle of three vocoder conditions was trained for each sentence type (12 channel for CUNY, 16 channel for IEEE). The training phase was completed in 30 – 40 minutes.

Finally, after the training phase was completed, baseline testing was performed to determine whether the subject met the minimum performance criteria needed to qualify for participation in the main study. Testing included 3 lists each of CUNY and IEEE sentences in the vocoder-alone listening condition (gV stimuli) without feedback. The subject was required achieve mean scores of 25% correct for the 12-channel CUNY sentences and 15% correct for the 16-channel IEEE sentences in order to qualify for further testing. As indicated earlier, twelve of thirteen potential participants met the minimum performance criteria and were tested in the main study. Baseline testing was completed in approximately 20 minutes.

During formal testing (sessions 2–7), sentence stimuli were presented in blocks of 9 lists of sentences, where each block was devoted to a single sentence type (CUNY or IEEE) and listening condition (gV , $gV+gLpSp$, or $gV+gLPHC$), but included all three vocoder channel conditions (8, 12 and 16 channels for CUNY, or 12, 16 and 32 channels for IEEE). The first three lists in each block were practice lists (1 list per channel condition) which included correct-answer feedback; the remaining 6 lists were actual test lists (2 consecutive lists per channel condition, with random ordering of channel conditions). Each test session included one block of CUNY sentences and one block of IEEE sentences with the order of listening and channel conditions assigned randomly (without replacement) across blocks. Half of the subjects always completed CUNY sentences first in each test session followed by IEEE sentences; the other half always completed IEEE sentences first followed by CUNY sentences. Due to the limited number of CUNY sentence lists, 8 CUNY lists from the first 8 testing lists with 8 and 12 channels were used twice for the last 8 practice lists; however, only novel lists were used during actual testing. None of the IEEE lists were repeated during training, practice or actual testing.

Subjects' recorded responses were scored in terms of the percentage of key words correctly identified, using scoring sheets developed for this purpose. Errors related to plurality, verb tense, or adjective suffix (-s, -ed, d, ing, or -y) were disregarded so long as the word stem was correct. Responses were scored independently by two examiners, one of whom was the first author (a non-native English speaker) and the other of whom was a native English speaker. On average, differences in scores were always less than 2 percentage points for a given sentence list. Nonetheless, a third scorer, who was a native English speaker, acted as a tie-breaker for any sentences where scoring differed between the first two scorers.

Statistical analyses

In the sections that follow, raw scores were transformed to rationalized arcsine units (RAUs; Studebaker, 1985) prior to statistical analysis. The transformed data sets were subjected to parametric analysis if relevant assumptions were met. However, when RAU transformation did not allow the data to meet required assumptions for parametric analysis, non-parametric

tests were applied to the raw (untransformed) data. For individual gain comparisons, all gain scores were transformed to z-scores to minimize the subject factor.

RESULTS

Mean performance across sentence, channel, and listening conditions

Figure 2 shows mean word recognition scores obtained from the 12 NH listeners across sentence, channel, and listening conditions. Overall, mean scores varied from 29 to 76% correct (30 to 75 RAUs) across conditions. Of the 216 scores measured (12 listeners x 18 conditions), the lowest score was 18% (18 RAUs) and the highest score was 87.8% (89 RAUs), and only four scores were greater than 85% correct (85 RAUs). Thus, it is unlikely that the pattern of results was influenced by either floor or ceiling effects.

Several expected trends can be observed in the mean data. First, performance for CUNY sentences was higher than performance for IEEE sentences when compared for the same channel and listening conditions. Second, performance improved with each increase in the number of spectral channels, for both IEEE and CUNY sentences. Third, performance varied systematically with listening condition across the six combinations of sentence type and number of channels: The *gV+LPsp* condition produced the highest mean scores, the *gV* condition produced the lowest mean scores, and the *gV+LPHC* condition produced intermediate mean scores that were only slightly higher than the *gV* scores.

The above trends were further evaluated by statistical analysis:

1. The effect of spectral degradation (number of channels) was evaluated separately for IEEE and CUNY sentences and for each listening condition. Friedman's test confirmed significant differences in performance across channels for both the IEEE and CUNY sentences ($p < 0.001$) and Wilcoxon Signed Ranks comparisons indicated that performance improved with each increase in the number of channels for each sentence type ($p < 0.01$).
2. The 12-channel data and the 16-channel data were analyzed separately to confirm the effects of sentence type and listening condition. In each case, a 2-way repeated measures analyses of variance (RM ANOVA) was performed with sentence type and listening condition as the within-subject factors. The main effect of sentence type was significant for both the 12- and 16-channel data (12 channel: $F[1,11] = 417.53$, $p < 0.001$, $\eta_p^2 = 0.974$; 16-channel: $F[1,11] = 354.08$, $p < 0.001$, $\eta_p^2 = 0.970$). Collapsed across the three listening conditions, the mean percent-correct scores for IEEE and CUNY sentences were 32% and 59.7%, respectively for the 12-channel condition, and 42.6% and 69.9%, respectively for the 16-channel condition. The main effect of listening condition was also significant for both the 12-channel and 16-channel analyses (12 channel: $F[2,22] = 9.73$, $p = 0.001$, $\eta_p^2 = 0.469$; 16-channel: $F[2,22] = 11.43$, $p < 0.001$, $\eta_p^2 = 0.510$). Subsequent pairwise comparisons indicated, for both the 12- and 16-

channel data, that scores for the $gV+gLpSp$ condition were significantly higher than those for the gV condition ($p = 0.001$) but scores for the gV and $gV+gLPHC$ conditions were not significantly different ($p > 0.05$). In summary, both analyses (12-channel data and 16-channel data) indicated a significant bimodal benefit for gated LP speech, but no bimodal benefit for gated LPHCs.

3. To determine whether a benefit attributable to LPHCs could be observed for any of the individual channel conditions for either IEEE or CUNY sentences, scores for the $gV+gLPHC$ and gV conditions were compared separately for each channel and sentence condition (i.e., IEEE 12, 16 and 32 channels and CUNY 8, 12 and 16 channels). These comparisons indicated that scores for the $gV+gLPHC$ condition were not significantly different than scores for the gV condition for any of the six conditions (Wilcoxon Signed Ranks test, $p > 0.1$). To determine whether performance for the $gV+gLpSp$ condition produced higher performance than the $gV+gLPHC$ condition, additional paired comparisons were performed. Findings indicated that the $gV+gLpSp$ condition produced significantly higher scores than the $gV+gLPHC$ condition for CUNY sentences with 8 and 16 vocoder channels [Wilcoxon Signed Ranks with $p=0.025$ criterion for significance after Bonferroni correction: 8-ch ($p=0.003$), 16-ch ($p=0.023$)]. A similar trend was observed for the remaining four conditions, however, these comparisons failed to reach significance [Wilcoxon Signed Ranks with $p=0.025$ criterion: CUNY 12-ch ($p=0.041$), IEEE 12-channel ($p=0.099$), IEEE 16-ch ($p=0.108$), IEEE 32-ch ($p=0.136$)]. Because the gated LPHC stimuli failed to produce a significant bimodal benefit, subsequent analyses considered only the bimodal benefit associated with LP speech.

Effects of context on bimodal benefit

The key question to be answered by this experiment was whether high-predictability (CUNY) sentences support larger amounts of bimodal benefit than low-predictability (IEEE) sentences. This question was addressed using two types of comparisons: 1) matched-channel comparisons, which evaluated differences in bimodal benefit for CUNY versus IEEE sentences separately for the 12- and 16-channel conditions, and 2) comparisons of individual subjects' gain scores for CUNY and IEEE sentences within the same range of baseline (i.e., vocoder-alone) performance. Matched-channel comparisons are valuable because they resemble the situation that exists for real CI users who have constant spectral resolution regardless of the type of speech material used to evaluate bimodal benefit. However, the interpretation of findings from matched-channel comparisons can be confounded by differences in baseline performance across sentence materials. This potential confound is alleviated when the individual gain scores obtained with different speech materials are compared over a similar range of baseline (vocoder-alone) scores; however, as described below, such comparisons are valid only if bimodal benefit does not vary systematically with baseline performance.

Two complementary measures of benefit were used for both the matched-channel and individual gain comparisons:

percentage-point gain and normalized gain—Percentage-point gain is calculated as the arithmetic difference in percent-correct scores obtained with, and without, low-frequency acoustic information. It provides a straightforward measure of bimodal benefit that has been used in several previous studies (Kong and Carlyon 2007; Brown and Bacon 2009a, b; Ba kent and Chatterjee 2010; Ba kent 2012). The primary limitation of the percentage-point gain measure is that it does not account for differences in baseline performance; for example, an increase in performance from 20 to 30 percent correct, and an increase in performance from 80 to 90 percent both yield a percentage-point gain of 10 points. The normalized gain measure (described previously; Eq. 1), considers bimodal benefit relative to the maximum possible benefit that can be obtained and, in this way, provides an index of bimodal benefit that compensates for baseline performance. An improvement in either percentage-point gain or normalized gain indicates improved speech understanding and, thus, is relevant to real-world listening.

Matched-channel comparisons—Figure 3 shows the $gV+gLPsp$ data from Figure 2 for the 12-channel and 16-channel conditions, replotted in terms of bimodal benefit. Bimodal benefit is expressed as percentage-point gain [$gV+gLPsp - gV$] in the left panel, and as normalized gain [Eq. (1)] in the right panel.

When expressed as percentage-point gain, mean bimodal benefit fell within a narrow range of values (9–15 points) regardless of sentence type or channel condition. Percentage-point gains (left side of Fig. 3) were slightly larger for CUNY sentences than for IEEE sentences for both the 12-channel and 16-channel conditions and there was a trend for CUNY sentences to produce more bimodal benefit than IEEE sentences for the 12-channel comparison (Wilcoxon Signed Ranks test, $p = 0.08$). However, the effect of context did not reach statistical significance for either the 12- or 16-channel comparison (Wilcoxon Signed Ranks test, $p = 0.39$).

When expressed in terms of normalized gain (right side of Fig. 3), mean bimodal benefit ranged from 12 to 32%. The average normalized gain was significantly greater for CUNY sentences as compared to IEEE sentences for both the 12-channel (Wilcoxon Signed Ranks, $p = 0.023$) and 16-channel data (Wilcoxon Signed Ranks, $p = 0.015$).

To summarize, matched-channel comparisons revealed a significant effect of sentence context on bimodal benefit expressed as normalized gain, but showed no clear effect of sentence context on bimodal benefit expressed as percentage-point gain.

Individual gain comparisons—Individual subjects' bimodal benefit scores, expressed as percentage-point gains or normalized gains, are plotted as a function of baseline performance in Figure 4. In these figures, individual subjects' gain scores for IEEE and CUNY sentences are plotted as a function of baseline (gV) performance, regardless of the channel condition in which they were obtained.

The left panel of Figure 4 shows the bimodal benefit achieved with LP speech ($gV+gLPsp$ condition) expressed as a percentage-point gain. In order to eliminate a relationship observed between percentage-point gain and baseline performance for CUNY scores (Spearman correlation, $r = 0.472$, $p = 0.004$), and allow comparison of CUNY scores to IEEE scores, it was necessary to restrict baseline performance for both sentence types to a range of 25–65%. A Kruskal-Wallis test performed on the restricted data set revealed that CUNY sentences produced an average of 7 percentage points more gain than IEEE sentences, and that this difference was significant ($p < 0.005$).

The right panel of Figure 4 shows normalized gain scores for the the full set of CUNY and IEEE data. For IEEE sentences, normalized gain varied from –12.7 to 42.4% over a range of baseline performance from 18 to 65%, while for CUNY sentences, normalized gain varied from –15.2 to 66.2% over a slightly higher range (27.4 to 77%). There was no significant relationship between normalized gain and baseline performance for either IEEE or CUNY sentences (Spearman correlation, $p > 0.1$); therefore, no restriction of the data was required. A Kruskal-Wallis test applied to the data showed that average normalized gains were 15 points greater for CUNY sentences than for IEEE sentences and that this difference was significant ($p < 0.001$).

To summarize, the individual gain comparisons for both percentage-point gain and normalized gain indicated that CUNY sentences produced significantly more bimodal benefit than IEEE sentences. When combined with the significant effects of context described for normalized-gain measure in the matched-channel comparisons, these findings support the notion that linguistic context contributes to bimodal benefit for temporally interrupted speech.

DISCUSSION

Bimodal benefits for temporally interrupted speech

As described earlier, bimodal benefit has only been evaluated for temporally interrupted speech in one previous study (Ba kent and Chatterjee 2010); thus, it is useful to compare the findings of that study with the present results. Ba kent and Chatterjee (2010), examined bimodal benefit for speech interrupted at a 1.5 Hz interruption rate (50% duty cycle) and four conditions of noise-band vocoding (4, 8, 16, and 32 channels) using high-context Dutch sentences. Unlike the present study, bimodal benefit was evaluated using a hybrid stimulation condition in which 500 Hz LP speech was combined with vocoded sentences in the same ear. The simulated CI-alone condition yielded scores ranging from 5% (4 channels) to 56% (32 channels) correct, while the hybrid listening condition produced scores ranging from 25% (4 channels) to 68% (32 channels) correct. On average, the bimodal benefit associated with hybrid stimulation was about 14 percentage points across the 4- to 32-channel conditions.

A comparison of the 8- and 16-channel results from Ba kent and Chatterjee's (2010) study with the 8- and 16-channel CUNY data from the present study shows that fairly similar levels of baseline performance, and nearly equal amounts of bimodal benefit, were achieved in the two studies. For the vocoder-alone condition, Ba kent and Chatterjee reported average

scores of 30% for the 8-channel condition and 50% for the 16-channel condition; in the present study, average scores were 34% (8 channels) and 65% (16 channels), respectively. Average bimodal benefits were essentially equal across studies at about 17 percentage points for the 8-channel condition and 10 percentage points for the 16-channel condition. The slightly higher baseline performance in the present study may be due to differences in temporal interruption rates and/or test materials, as well as the training subjects' received prior to testing. The slower interruption rate used by Ba Kent and Chatterjee (1.5 Hz) produces 333-ms silent gaps, thereby preserving or removing relatively large segments of the speech stream (syllables or words), while the faster rate used in our study (5 Hz) produces 100-ms silent gaps, thereby preserving or removing smaller segments of the speech stream (phonemes or syllables). In addition, although the Dutch sentences used by Ba Kent and Chatterjee were described as being high in context, there may be differences in the amount of context available in their sentences as compared to the CUNY sentences used in the present study. Overall, however, the two studies produced similar amounts of bimodal benefit for high-context, temporally interrupted sentences, despite differences in temporal interruption rate and bimodal configuration.

Effects of context on bimodal benefit in interrupted speech

As indicated earlier, the effect of context on bimodal benefit has been systematically evaluated in only one previous study (Kong et al. 2015), although preliminary evidence of context effects was reported in a prior study by Brown and Bacon (2009a). As noted previously, Kong et al. used the same recordings of IEEE and CUNY sentences used in the present study; thus, it is possible to directly compare findings for continuous and interrupted conditions across the two studies. Kong et al. reported that CUNY sentences produced significantly greater bimodal benefit than IEEE sentences for continuous speech; differences attributable to context were approximately 18 percentage points or 30 points of normalized gain. The present study showed smaller context effects for temporally interrupted speech. Specifically, the context effect in the present study was 7 percentage points, or 14 points of normalized gain, which is less than half the magnitude of the context effect reported by Kong et al. Closer examination of the bimodal gains achieved for CUNY and IEEE sentences in the two studies reveals that bimodal benefits are greater for continuous sentences than for interrupted sentences regardless of the context level of the sentences. Specifically, for CUNY sentences, bimodal benefit was 39 percentage points or 68% normalized gain for continuous sentences and 15 percentage points or 29% normalized gain for interrupted sentences. For IEEE sentences, bimodal benefit was 20 percentage points or 37% normalized gain for continuous sentences and 8 percentage points or 15% normalized gain for interrupted sentences.

It is important to note that Kong et al.'s findings for continuous speech were achieved by using speech that was severely spectrally degraded (2 to 6 channels of noise-band vocoding) while the present study applied more moderate amounts of spectral degradation (8 to 16 channels for CUNY; 12 to 32 channels for IEEE) to sentences that were temporally interrupted at a rate of 5 Hz. That is, the two studies used different combinations of spectral and temporal degradation to achieve similar ranges of performance in the vocoder-alone condition. It appears that the two types of sensory (bottom-up) disruptions interact

differently with top-down linguistic processing to produce larger context effects for continuous sentences that are severely spectrally degraded than for temporally interrupted sentences that have undergone more moderate amounts of spectral degradation. In this regard, it is possible that top-down linguistic processing is less able to compensate for reduced sensory information when portions of the speech signal are completely missing (during silent gaps) than when the speech signal is continuously present, but with poorer spectral resolution.

As anticipated by our hypotheses, the present study demonstrated evidence of interactions between bottom-up and top-down processes for bimodal perception of temporally interrupted speech. When speech is temporally interrupted, bottom-up speech cues are substantially reduced, increasing the listener's need to apply top-down processing to fill-in the missing information. However, top-down repair mechanisms can only be effective when at least some bottom-up cues exist to provide the scaffolding to which top-down repair can be applied. The present data suggest that the addition of low-frequency acoustic cues that may exist in the residual hearing ear facilitates the top-down repair process, either by augmenting the amount of sensory (bottom-up) cues available to the listener, or by facilitating the listener's use of top-down processing. The fact that bimodal benefit occurred even though word recognition was absent (near 0%) for LP speech alone argues against an additive effect of percent-correct scores across ears. Thus, it appears that interactions between bottom-up and top-down processing were primarily responsible for the bimodal benefits observed in this case.

Compared to the Kong et al. (2015) study that demonstrated significant (5–15 percentage points) bimodal gains for both CUNY and IEEE sentences when continuous LPHCs were paired with continuous vocoded sentences, the present study did not show significant bimodal benefits when the LPHCs and sentences were temporally interrupted. There were no obvious differences between subjects or procedures between the two studies. Thus, the present findings suggest that the low-frequency cues contained in LPHCs are less useful when degraded by silent gaps. As suggested by several previous authors (Kong and Carlyon, 2007; Spitzer et al., 2009; Hu and Loizou, 2010), LPHCs provide voicing cues that mark syllable boundaries to facilitate the glimpsing of target speech in noise. However, it appears that this mechanism is no longer operable for gated speech perception. An interesting question is whether LPHCs presented continuously in the low-frequency ear can enhance bimodal benefit when temporally interrupted sentences are presented to the vocoder ear. Such a benefit, if observed, would imply a role of low-frequency harmonicity cues in the listener's ability to integrate information from available speech segments and reconstruct a coherent message with the aid of top-down processing. However, if low-frequency harmonicity cues are primarily helping the listener to identify word boundaries within the vocoded signal, then bimodal benefit for LPHCs may occur only when both ears receive continuous signals. Further investigation of the role of low-frequency continuity in bimodal benefit is underway in our laboratory (Oh, 2015).

In contrast to the finding of lack of significant bimodal benefit with LPHCs produced for temporally interrupted speech, some benefit was preserved in the present study when temporally interrupted LP speech was presented to the residual hearing ear. This finding

suggests that at least some cues present in LP speech are resistant to the effects of temporal interruption.

Context effects for real cochlear implant users

The present study tested NH listeners in simulated listening conditions; thus, it is important to consider whether the findings reported here would apply to real CI users who have residual acoustic hearing in the contralateral ear. Performance for the vocoder-alone (*gV*) condition may be compared to data obtained by Chatterjee et al. (2010) in six postlingually-deafened CI users with Cochlear Freedom or N24 devices. Chatterjee et al. used Hearing in Noise Test (HINT) sentences, and used gating parameters identical to those in the present study (5 Hz, 50% duty cycle). Mean performance was approximately 98% correct for continuous sentences and 17% correct for the temporally interrupted sentences. In comparison, the most similar gated condition in the present study (8-channel CUNY sentences, *gV* condition) yielded slightly higher mean performance of 33% correct. It is difficult to directly compare these findings, given differences in sentence materials and potential differences in effective spectral resolution (8-channel vocoding vs. the resolution available to individual CI users). Nonetheless, the broad similarity of mean performance levels (17% versus 33%) suggests that our NH simulation provided a reasonable approximation of the outcomes expected in real CI users.

With respect to bimodal benefit, two general factors could contribute to different outcomes among real CI users as compared to our simulated listeners. First, the degree of low-frequency audibility can impact bimodal benefit (Sheffield and Zeng 2012; Yang and Zeng 2013; Zhang et al., 2013). In the present study, all subjects received an equal range of low-frequency audibility in the simulated residual ear (i.e., 500 Hz LP filtered speech). However, real CI users have varying degrees of residual hearing available, typically aided by amplification. Extended hearing in the LP ear may enhance the listener's access to spectral cues (i.e., F1), whereas more restricted low-frequency audibility, or reduced spectral resolution (Spahr et al. 2007; Zhang et al., 2013), may reduce their access to such cues. Second, CI users may have a reduced ability to integrate speech cues across ears in the bimodal listening condition as compared to simulated NH listeners; such reduced integration ability has been speculated to underlie reduced bimodal benefit observed for some real CI users in previous studies (Kong and Braida 2011; Yang and Zeng 2013), possibly related to reduced audibility in the residual hearing ear (Yang and Zeng 2013).

Clinical application

Currently, the decision whether to implant the second ear of a unilateral CI user does not make use of systematic assessment procedures; thus, clinicians are limited in their ability to make well-reasoned recommendations for individual patients. Evaluation may rely on phoneme or word recognition scores, or on conventional sentence materials without regard to the role of linguistic context on bimodal benefit. Findings from Kong et al. (2015) and the present study indicate that larger bimodal benefits are produced for high-context, as compared to low-context, sentences. These findings suggest that bimodal benefit could be evaluated either using high-context materials alone, to optimize the possibility of observing bimodal benefit in real-life listening situations, or using both high and low-context materials

to estimate bimodal benefit under listening conditions that require differing amounts of perceptual ability and linguistic processing.

SUMMARY AND CONCLUSIONS

The role of sentence context for bimodal benefit in temporally interrupted sentences was investigated in NH listeners, using high and low context sentences. Temporally interrupted IEEE (low-context) and CUNY (high-context) sentences were processed with noise band vocoding to simulate the CI ear, and were LP filtered at 500 Hz to simulate the residual hearing ear. Three different listening conditions (gV , $gV+gLPsp$, and $gV+gLPHC$) were tested across different channel conditions that yielded baseline performance ranging from approximately 30 to 60 percent correct. Bimodal benefits were expressed as percentage-point gains and normalized gains.

Bimodal benefits obtained with gated low-pass speech ($gV+gLPsp$ condition) were significantly larger for CUNY sentences than for IEEE sentences. For individual gain comparisons, both percentage-point and normalized gains were larger for high context sentences than for low context sentences. For matched-channel comparisons, normalized gains were larger for high-context sentences than for low-context sentences. These findings indicate that linguistic context has a significant effect on bimodal benefit for temporally interrupted speech. Although the magnitude of context effect on bimodal benefit for temporally interrupted sentences was less than that for continuous sentences described in an earlier study (Kong et al. 2015), findings from both studies support the hypothesis that low-frequency acoustic information presented to the residual hearing ear facilitates the use of top-down linguistic processing in bimodal hearing.

Bimodal benefits were not observed when gated LPHCs were presented to the simulated residual hearing ear ($gV+gLPHC$ condition). This finding suggests that temporal interruptions negate the usefulness of low-frequency harmonicity cues that support bimodal benefit for continuous stimuli.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was completed as part of the doctoral dissertation of the first author, and was supported by NIH-NIDCD DC012300 (Y.-Y. Kong, P.I.). The authors thank Jean Krause and Catherine Rogers for useful discussions of the research, Ala Somarowthu for software development, and Katherine Austin and Shaina Maldonado for assistance with data analysis. We also thank Mario Svirsky and three anonymous reviewers for helpful comments on an earlier version of the manuscript.

References

- Bashford JA, Riener KR, Warren RM. Increasing the intelligibility of speech through multiple phonemic restorations. *Perception Psychophysics*. 1992; 51:211–217. [PubMed: 1561046]
- Ba kent D, Chatterjee M. Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech. *Hearing Research*. 2010; 270:127–133. [PubMed: 20817081]

- Ba kent D. Effect of speech degradation on top-down repair: phonemic restoration with simulations of cochlear implants and combined electric-acoustic stimulation. *Journal of the Association for Research in Otolaryngology*. 2012; 13:683–692. [PubMed: 22569838]
- Boothroyd, A.; Hanin, L.; Hnath, T. A sentence test of speech perception: reliability, set equivalence, and short term learning. New York: Speech and Hearing Sciences Research Center, City University of New York; 1985.
- Boersma P, Weenink D. PRAAT: doing phonetics by computer [computer program]. 2009 Version 5.1.05.
- Brown CA, Bacon S. Low-frequency speech cues and simulated electric-acoustic hearing. *Journal of the Acoustical Society of America*. 2009a; 125:1658–1665. [PubMed: 19275323]
- Brown CA, Bacon S. Achieving electric-acoustic benefit with a modulated tone. *Ear and Hearing*. 2009b; 30:489–493. [PubMed: 19546806]
- Busby PA, Clark GM. Gap detection by early-deafened cochlear implant subjects. *Journal of the Acoustical Society of America*. 1998; 105:1841–1852. [PubMed: 10089607]
- Carroll J, Tiaden S, Zeng FG. Fundamental frequency is critical to speech perception in noise in combined acoustic and electric hearing. *Journal of the Acoustical Society of America*. 2011; 130:2054–2062. [PubMed: 21973360]
- Chatterjee M, Peredo F, Nelson D, Ba kent D. Recognition of interrupted sentences under conditions of spectral degradation. *Journal of the Acoustical Society of America*. 2010; 127:EL37–41. [PubMed: 20136176]
- Dorman MF, Loizou P, Wang S, Zhang T, Spahr T, Loisselle L, Cook S. Bimodal cochlear implants: the role of acoustic signal level in determining speech perception benefit. *Audiology and Neuro-Otology*. 2014; 19:234–238. [PubMed: 24992987]
- Dudley H. The automatic synthesis of speech. *Proceedings of the National Academy of Sciences USA*. 1939; 25:377–83.
- Fu QJ, Galvin JJ, Wang X. Recognition of time-distorted sentences by normal-hearing and cochlear-implant listeners. *Journal of the Acoustical Society of America*. 2001; 109:379–384. [PubMed: 11206166]
- Greenwood DD. A cochlear frequency-position function for several species- 29 years later. *Journal of the Acoustical Society of America*. 1990; 87:2592–2605. [PubMed: 2373794]
- Hu Y, Loizou PC. Effects of introducing low-frequency harmonics in the perception of vocoded telephone speech. *Journal of the Acoustical Society of America*. 2010; 128:1280–1289. [PubMed: 20815463]
- IEEE. IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*. 1969; 17:225–246.
- Kong YY, Carlyon RP. Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. *Journal of the Acoustical Society of America*. 2007; 121:3717–3727. [PubMed: 17552722]
- Kong YY, Braida LD. Cross-frequency integration for consonant and vowel identification in bimodal hearing. *Journal of Speech Language Hearing Research*. 2011; 54:959–980.
- Kong YY, Donaldson G, Somarowthu A. Effects of contextual cues on speech recognition in simulated electric-acoustic stimulation. *Journal of the Acoustical Society of America*. 2015; 137:2846–2857. [PubMed: 25994712]
- Li N, Loizou PC. A glimpsing account for the benefit of simulated combined acoustic and electric hearing. *Journal of the Acoustical Society of America*. 2008; 123:2287–2294. [PubMed: 18397033]
- Needleman AR. Quantification of context effects in speech perception: influence of prosody. *Clinical Linguistics and Phonetics*. 1998; 12:305–327.
- Nelson PB, Jin SH, Carney AE, Nelson DA. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America*. 2003; 113:961–968. [PubMed: 12597189]
- Nelson PB, Jin SH. Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America*. 2004; 115:2286–2294. [PubMed: 15139640]

- Oh, SH. Doctoral dissertation. 2015. Top-down processes in simulated combined electric-acoustic hearing: The effect of context and the role of low-frequency cues in the perception of temporally-interrupted speech. Retrieved from ProQuest Dissertations and Theses. (Accession Order No. AAI 3666210)
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995; 270:303–304. [PubMed: 7569981]
- Sheffield BM, Zeng FG. The relative phonetic contributions of a cochlear implant and residual acoustic hearing to bimodal speech perception. *Journal of the Acoustical Society of America*. 2012; 131:518–530. [PubMed: 22280613]
- Spahr AJ, Dorman MF, Loisel LH. Performance of patients using different cochlear implant systems: effects of input dynamic range. *Ear and Hearing*. 2007; 28:260–275. [PubMed: 17496675]
- Spitzer S, Liss J, Spahr T, Dorman M, Lansford K. The use of fundamental frequency for lexical segmentation in listeners with cochlear implants. *Journal of the Acoustical Society of America*. 2009; EL125:236–241.
- Studebaker GA. A “Rationalized” Arcsine Transform. *Journal of Speech and Hearing Research*. 1985; 28:455–462. [PubMed: 4046587]
- Tyler MD, Cutler A. Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America*. 2009; 126:367–376. [PubMed: 19603893]
- Yang HI, Zeng FG. Reduced acoustic and electric integration in concurrent-vowel recognition. *Scientific Reports*. 2013; 3:1419. [PubMed: 23474462]
- Warren RM. Perceptual restoration of missing speech sounds. *Science*. 1970; 167:392–393. [PubMed: 5409744]
- Warren RM, Obusek CJ. Speech perception and phonemic restorations. *Perception and Psychophysics*. 1971; 9:358–362.
- Zhang T, Dorman MF, Spahr AJ. Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. *Ear and Hearing*. 2010; 31:63–69. [PubMed: 20050394]
- Zhang T, Spahr AJ, Dorman MF, Saoji A. Relationship between auditory function of nonimplanted ears and bimodal benefit. *Ear and Hearing*. 2013; 34:133–141. [PubMed: 23075632]

Short Summary

The effect of sentence-level context on simulated bimodal hearing was investigated for temporally interrupted speech. A bimodal condition using low-pass speech demonstrated a significant effect of context on bimodal benefit, supporting the hypothesis that low-frequency acoustic cues facilitate top-down processing in bimodal hearing. However, bimodal benefits for temporally interrupted sentences were smaller than those reported previously for continuous sentences, suggesting that listeners' ability to restore missing speech information depends strongly on the quality of the sensory input. That is, listeners can successfully make use of top-down knowledge only when provided with a critical amount and quality of bottom-up speech information.

Figure 1.
Summary of activities completed during each of seven test sessions.

Figure 2.

Mean percent-correct word recognition scores across 6 stimulus conditions and 3 listening conditions for 12 subjects. Error bars indicate ± 1 standard error of the mean.

Figure 3.

Comparison of bimodal benefit in the $gV+gLpSp$ listening condition for CUNY and IEEE sentences, for the 12- and 16-channel conditions. Benefit is shown as normalized gain (right) and percentage-point gain (left).

Figure 4.

Individual subjects' percentage-point gain scores (left panel) and normalized gain scores (right panel) as a function of baseline performance for CUNY and IEEE sentences. Scores in the left panel were restricted in baseline value (25–63%) to satisfy statistical requirements (see text).