Effects of contextual cues on speech recognition in simulated electric-acoustic stimulation

Ying-Yee Kong^{a)}

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

Gail Donaldson

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

Ala Somarowthu

Department of Bioengineering, Northeastern University, 360 Huntington Avenue, Boston,
Massachusetts 02115, USA

(Received 5 December 2013; revised 10 April 2015; accepted 11 April 2015)

Low-frequency acoustic cues have shown to improve speech perception in cochlear-implant listeners. However, the mechanisms underlying this benefit are still not well understood. This study investigated the extent to which low-frequency cues can facilitate listeners' use of linguistic knowledge in simulated electric-acoustic stimulation (EAS). Experiment 1 examined differences in the magnitude of EAS benefit at the phoneme, word, and sentence levels. Speech materials were processed via noise-channel vocoding and lowpass (LP) filtering. The amount of spectral degradation in the vocoder speech was varied by applying different numbers of vocoder channels. Normal-hearing listeners were tested on vocoder-alone, LP-alone, and vocoder + LP conditions. Experiment 2 further examined factors that underlie the context effect on EAS benefit at the sentence level by limiting the low-frequency cues to temporal envelope and periodicity (AM + FM). Results showed that EAS benefit was greater for higher-context than for lower-context speech materials even when the LP ear received only low-frequency AM + FM cues. Possible explanations for the greater EAS benefit observed with higher-context materials may lie in the interplay between perceptual and expectation-driven processes for EAS speech recognition, and/or the band-importance functions for different types of speech materials. © 2015 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4919337>]

[BRM]

Pages: 2846–2857

I. INTRODUCTION

It is well established that low-frequency acoustic cues, accessed via residual acoustic hearing in the non-implanted ear, can improve speech perception for many cochlear-implant (CI) users, even when the low frequency cues, by themselves, support no measurable speech intelligibility. This phenomenon, referred to as *bimodal benefit* or *electric-acoustic stimulation (EAS) benefit*, has received considerable attention in recent years as a possible alternative to traditional cochlear implantation. Bimodal hearing is most often configured with a cochlear implant in one ear and a hearing aid on the opposite ear; however, recent improvements in surgical techniques also allow for the use of residual hearing in the implanted ear in some patients (e.g., [Lenarz et al., 2013](#); [Kopelovich et al., 2014](#)).

Most previous studies of bimodal hearing (e.g., [Kong and Carlyon, 2007](#); [Li and Loizou, 2008](#); [Brown and Bacon, 2009a,b](#); [Spitzer et al., 2009](#); [Zhang et al., 2010](#)) have focused on the role of specific speech information provided by the residual hearing ear including fundamental frequency (F0) contour, temporal envelope, voicing, and first formant

(F1) frequency cues. The periodicity cues that contain voicing and/or F0 contour information have received considerable attention because such cues are unavailable in the CI ear due to the removal of temporal fine structure from the signal during CI speech processing. F0 contour and voicing information convey several types of suprasegmental cues: First, the F0 contour, in conjunction with stress patterns, codes intonation patterns in speech, allowing the listener to extract the emotional intent of the speaker and distinguish questions from statements. Second, these cues code the gender and resonance patterns of the speaker, allowing the listener to adjust their listening expectations to the individual characteristics of the speaker. Finally, F0 contour cues may contribute to lexical segmentation, i.e., the listener's ability to segment the continuous speech stream into individual words prior to further processing. The role of F0 contour cues in lexical segmentation has been demonstrated in normal-hearing listeners (e.g., [Tyler and Cutler, 2009](#)) and preliminary evidence suggests a similar role for F0 contour cues in bimodal hearing (e.g., [Spitzer et al., 2009](#); [Hu and Loizou, 2010](#)).

While there is broad consensus that low-frequency cues can support an improved representation of acoustic speech cues, their possible contributions to other processes

^{a)}Electronic mail: yykong@neu.edu

(interactive or feedforward) that influence the recognition of words in continuous speech streams are less well understood. The speech perception literature includes reports of several phenomena that implicate an influence of prior knowledge of spoken language on word recognition. For example, word recognition performance can be affected by the phonemic restoration (PhR) effect (e.g., Warren, 1970; Warren and Obusek, 1971; Bashford and Warren, 1979), semantic and/or syntactic effects (e.g., Boothroyd and Nitttrouer, 1988), linguistic priming (e.g., Garrett and Saint-Pierre, 1980) and lexical frequency effects (Luce and Pisoni, 1998). Most researchers agree that listeners' prior knowledge of the spoken language interacts with perceptual processes during speech recognition, and some attribute these expectation-driven effects to a top-down process that influences perceptual analysis (e.g., TRACE model of speech perception: McClelland and Elman, 1986).

The potential role of expectation-driven processes in bimodal hearing is still poorly understood, and only a few studies have systematically examined these effects. Başkent and colleagues investigated listeners' ability to perceptually fill in missing speech information when receiving spectrally degraded speech, using both simulations of CI and EAS (Başkent and Chatterjee, 2010; Başkent, 2012; Bhargava *et al.*, 2014) and real CI users (Bhargava *et al.*, 2014). Başkent (2012) studied the effect of PhR with simulated EAS. She reported that severe degradation of the speech signal via vocoder processing reduced the restoration of missing speech information. A significant effect of PhR was observed at the two highest spectral resolution conditions tested (16 and 32 vocoder channels). However, the PhR effect was only slightly greater for simulated EAS than for simulated CI. Thus, Başkent's findings provide only weak evidence for listeners' use of prior knowledge to fill in missing speech information in bimodal hearing.

If the interaction between prior knowledge of the spoken language and perceptual processing does, in fact, contribute to bimodal benefit, then it might be expected that the magnitude of bimodal benefit would vary in proportion to the amount of linguistic context contained in the stimulus materials. In this case, the effects of applying prior knowledge may be as important as acoustic cues in contributing to EAS benefit.

While the effects of linguistic context on EAS benefit have not been systematically assessed, Brown and Bacon (2009b) provided preliminary evidence that simulated bimodal benefit may be greater for higher-context speech materials [City University of New York (CUNY) sentences; Boothroyd *et al.*, 1988] than for lower-context materials [Institute of Electrical and Electronics Engineers (IEEE) sentences; IEEE, 1969]. Their study was not specifically intended to evaluate context effects, however, and as acknowledged by the authors, differences in the sentence materials other than context level could potentially account for the differences in benefit they observed. Specifically, the two sets of sentence materials they used were produced by different talkers; and the IEEE sentences were produced with a conversational style whereas CUNY sentences were produced at a slower rate with a highly articulated speaking style.

The goal of the present study was to further assess the potential effects of linguistic context on EAS benefit in simulated EAS listening as a means of elucidating the possible role of prior knowledge of spoken language in bimodal hearing. To address this goal, we measured EAS benefit with different types of speech materials varying in linguistic context, from nonsense syllables to high-context sentences, while controlling for potentially confounding acoustic factors such as differences in speaking rate (Brown and Bacon, 2009b). In the first experiment, we systematically varied the extent of spectral degradation in the vocoder speech while presenting a fixed amount of acoustic information [500-Hz low-pass (LP) filtered speech] to the LP ear. On the basis of previous findings that demonstrated significant PhR with greater spectral resolution (Başkent and Chatterjee, 2010; Başkent, 2012) and possible effects of context on EAS benefit (Brown and Bacon, 2009b), we hypothesized that the supplemental sensory input provided in the EAS condition, relative to the vocoder-alone condition, would interact with non-perceptual processing for spoken word recognition. More specifically, we predicted that EAS benefit would increase as a function of the amount of context available in the speech signal, from isolated words to words in sentences, and from words in low-context sentences to words in high-context sentences.

A second experiment was completed in which the simulated residual hearing ear received only F0 contour and temporal envelope cues extracted from the low-frequency speech. This allowed us to examine the effect of further degradation of the LP signal on the use of linguistic context, and investigate the mechanism that underlies the context effect on EAS benefit when equating the amount of spectral content and baseline performance across the low and high context sentence materials.

II. EXPERIMENT 1: QUANTIFYING EAS BENEFIT AND CONTEXT EFFECT

A. Method

1. Subjects

Forty normally hearing young adults between 18 and 30 years of age participated in this experiment. All were native speakers of American English and had no known history of speech or hearing problems.

2. Stimuli

Test stimuli included four different sets of speech materials with different amounts of contextual constraints—CVC nonsense syllables, CVC words, IEEE sentences, and CUNY sentences.

a. CVC nonsense syllables and CVC words. Phonetically balanced lists of CVC syllables were used, identical to those used in Boothroyd and Nitttrouer (1988). Each list consisted of ten syllables constructed from the same pool of ten initial consonants, ten medial vowels, and ten final consonants. Twelve of the lists consisted of nonsense syllables; the other 12 lists consisted of meaningful words (see Boothroyd and Nitttrouer,

1988, for details of the construction of these lists). The 24 lists were recorded by an adult female speaker of American English who spoke each syllable in isolation. The recorded syllables were then scaled to have equal root-mean-square (rms) amplitudes.

b. CUNY and IEEE sentences. CUNY and IEEE sentences were used to measure word recognition in sentences. Both CUNY and IEEE corpuses have a large number of sentences organized into lists (60 lists with 12 sentences per list for CUNY and 72 lists with 10 sentences per list for IEEE) that allow for a within-subject test design and for practice before data collection. The CUNY sentences are relatively easy and contain high levels of context (e.g., “I have a sore throat and a very bad cough”), whereas IEEE sentences are more difficult with lower levels of context (e.g., “The birch canoe slid on the smooth planks”). The high levels of context in the CUNY sentences allow for greater predictability of individual words in the sentences. The CUNY and IEEE sentences were recorded by a different adult female speaker of American English than the speaker who produced the CVC words and syllables. To minimize potential confounding factors related to speaking styles, the speaker held the speaking rate and articulation effort constant across sentence lists and sentence sets. Statistical analysis using a Student *t* test confirmed that speaking rates, calculated as number of words per minute (WPM), were not statistically different [$t(1426) = 0.23$, $p > 0.05$] for the CUNY (mean WPM = 198.8) and IEEE (mean WPM = 198.5) sentences. This speaking rate is consistent with a conversational speaking style (Picheny *et al.*, 1986). Variations of the average speaking rate for each list were within 10% of the mean of the sentence set (i.e., 179–219). In addition, a *t* test was performed to confirm that pitch excursion within a sentence, calculated as the range of F0 in semitones over the duration of a sentence (i.e., semitones per second; de Pijper, 1983), was not statistically different [$t(1426) = 1.59$, $p > 0.05$] between the CUNY (mean = 4.3) and IEEE (mean = 4.2) sentences. Variations in the average pitch excursion for each list were within half a semitone of the mean pitch excursion across the entire sentence set (i.e., 3.7–4.7). Each recorded sentence was scaled to have equal rms amplitude.

Recorded stimuli were subjected to two types of signal processing: LP filtering and noise channel vocoding. LP filtering was performed using Butterworth filters with a rolloff slope of 60 dB/octave and a cutoff frequency of 500 Hz. These LP parameters mimicked a sloping hearing loss above 500 Hz, an audiometric configuration commonly seen in real

EAS users. Kong and Braidia (2011) used this LP filtering with a group of normal-hearing listeners, and they reported that percent information transmission for consonant and vowel recognition in quiet was similar to that obtained from a group of real EAS users. Noise channel vocoding was intended to simulate aspects of CI processing related to (1) reduction of spectral resolution and temporal fine structure cues, and (2) preservation of temporal envelope cues. The system used for channel-vocoding processing was similar to that described by Shannon *et al.* (1995). In this system, broadband speech was first processed through a pre-emphasis filter and then band-pass filtered into a number of logarithmically spaced frequency bands (see cutoff frequencies for each channel in Table I). The amplitude envelope of the signal was extracted from each band by full-wave rectification and LP filtering with a 400-Hz cutoff frequency. The envelope extracted from each frequency band was used to modulate white noise, which was then filtered by the same band-pass filter used to generate the frequency band in the analysis stage. All bands were then summed to produce the final vocoded stimulus.

The glimpsing mechanism that is thought to underlie EAS benefits for speech in noise is likely to be more effective for sentences than for isolated words (Kong and Carlyon, 2007; Brown and Bacon, 2009b). Using this mechanism, listeners extract small spectrotemporal regions of the speech stream (so called “glimpses”) with favorable signal-to-noise ratios and integrate those segments into a coherent speech stream (Li and Loizou, 2008). To minimize the contribution of this factor and to better isolate the effect of context on speech recognition in EAS, speech materials (isolated CVC syllables and words, and sentences) were presented in quiet.

3. Procedures

All stimuli were presented from a LynxTWO sound card using 16-bit resolution at a 44.1-kHz sampling rate and routed to Sennheiser HD 600 headphones at an rms level of 70 dBA. Each type of speech material was presented in three listening conditions: LP-alone, vocoder-alone, and EAS (i.e., LP + vocoder). There were five vocoder conditions differing in the number of vocoder channels, from two channels to six channels (i.e., 2ch, 3ch, 4ch, 5ch, and 6ch vocoder). This resulted in a total of 11 conditions (one LP, five vocoder, and five EAS). Subjects were divided into five groups of eight, with each group was presented with one of the vocoder-channel conditions. Thus, each subject was tested

TABLE I. Cutoff frequencies for different vocoder-channel conditions.

Channel Condition	Cutoff frequencies for each analysis/synthesis filter					
	1	2	3	4	5	6
2	80–1250	1250–8000				
3	80–624	624–2373	2373–8000			
4	80–424	424–1250	1250–3234	3234–8000		
5	80–329	329–832	832–1844	1844–3886	3886–8000	
6	80–275	275–624	264–1250	1250–2373	2373–4388	4388–8000

with the LP, one vocoder, and one EAS condition for each type of speech material. For example, subject 1 (S1) was tested with LP, 6ch-vocoder alone, and 6ch-vocoder + LP for each of the four speech materials (CVC nonsense syllables, CVC words, IEEE sentences, and CUNY sentences). Each subject was tested four times on different days with each session lasting 2–3 h. Within each group, half of the subjects were presented with vocoder stimuli to the left ear and LP stimuli to the right ear; the remaining half received the stimuli on the opposite sides. Also within each group, the order of presentation of the vocoder and EAS conditions was counter-balanced across subjects. Each subject was tested on sentence recognition first followed by word/syllable recognition. For sentence recognition within each group, half of the subjects were tested with CUNY first followed by IEEE. For word recognition, half of the subjects were tested with words first followed by nonsense syllables.

For each listening condition, subjects first received practice trials with visual feedback that included presentation of words or sentences shown on a computer screen. The purpose of the practice was to familiarize the subjects with the speech materials, degraded speech, and the test protocols. For sentence recognition, subjects practiced listening to four lists of CUNY or IEEE sentences, depending on the test condition, prior to testing. Each subject was then tested with four new lists of sentences per listening condition. The test lists were randomly selected without repetition.

For word recognition with isolated CVC words and CVC nonsense syllables, subjects practiced listening to two lists of 50 CNC words (Peterson and Lehiste, 1962), rather than the test stimuli. Different speech materials were used for practice because there are only 12 lists of CVC words and 12 lists of nonsense syllables available, not enough for both practice and testing. For testing, each subject was presented with five lists of CVC words and nonsense syllables per listening condition. The lists that were used for the vocoder-alone and EAS conditions were randomly selected without repetition. However, because of the limited number of lists for CVC words and nonsense syllables, the remaining two lists and three of the lists that were presented for either the vocoder or EAS condition were used for the LP-alone condition. The LP condition was tested before the vocoder and EAS condition for half of the subjects and after for the remaining half. There was a 1 to 2 week time lapse between the testing of the LP condition and the vocoder-alone and EAS conditions.

For the word and sentence stimuli, subjects listened to the stimulus and were instructed to type as much of the stimulus as they could into the computer. For the nonsense syllables, subjects were told that the stimuli presented to them were not real words, and they were instructed to repeat their responses verbally into a microphone placed directly in front of them. A verbal response was necessary to minimize the potential confound of variability in subjects' ability to transcribe the sounds they heard. For all tasks and listening conditions, subjects were encouraged to guess if necessary. All written and verbal responses were stored for offline scoring and analysis.

4. Data analyses

CVC words and syllables were scored according to the percentage of phonemes and whole syllables correctly recognized. The scoring of nonsense syllables was performed independently by two examiners who had formal training in phonetics. When the examiners disagreed on a particular response, the first author, who also had formal training in phonetics, made the final scoring decision. For sentence recognition, scoring was based on percentage of keywords correctly recognized in the sentences.

The EAS benefit attributable to adding low-frequency speech cues was calculated using two approaches.¹ The first approach was to calculate the percentage-point difference in scores achieved in the vocoder-alone and the EAS conditions [i.e., percentage-point gain (PPG)]. This approach may underestimate EAS benefit if the performance in the EAS condition approaches ceiling. The second approach was to calculate percent normalized gain (NG), a metric established in the audio-visual literature to quantify speechreading benefit where the visual cues are considered supplemental information to the auditory cues (e.g., Sumby and Pollack, 1954; Rabinowitz *et al.*, 1992; Grant and Seitz, 1998; Kirk *et al.*, 2007)

$$G = (\text{EAS} - \text{VOC}) / (100 - \text{VOC}), \quad (1)$$

where G is the percentage-point gain normalized by the potential benefit possible given an individual's baseline (vocoder-alone) score. This NG measure is intended to compensate for differences in baseline performance.

The effect of linguistic context on EAS benefit was revealed by comparing the amount of gain (PPG or NG) across different speech material types. Specifically, gain was compared in a way that took into consideration (1) vocoder-channel conditions (i.e., amount of spectral cues in the vocoder speech), as well as (2) differences in vocoder-alone baseline performance. For matched-channel comparisons, we compared the gain across different speech materials when the amount of *acoustic* information delivered to the listeners was the same. Because of the fact that the assumption of normality was not met in the data obtained for some measures (Shapiro-Wilk test, $p < 0.05$), the Wilcoxon Signed-Ranks test was performed on the within-subject data to determine whether speech recognition scores or gains were significantly different for lower versus higher context speech materials. To determine the effect of context on EAS benefit, Mann-Whitney (comparison between two groups) or Kruskal-Wallis (comparison across three or more groups) tests were performed on the between-subject data after taking into consideration differences in the baseline performance. Spearman correlations were used to determine potential relationships between pairs of variables.

B. Results

1. CVC nonsense syllables and words

a. Percent-correct scores. Figure 1 shows the mean percent-correct scores for each vocoder-channel and EAS

FIG. 1. Percent-correct scores for (top) phoneme and (bottom) syllable recognition for (left) nonsense syllables and (right) words for three listening conditions (vocoder, LP, and EAS).

condition. The left and right panels show the performance for CVC nonsense syllables and real words, respectively. The top panels show the percentage of phonemes correctly recognized, and the bottom panels show the percentage of whole syllables correctly recognized. The dashed lines represent the mean scores for the LP condition. Because the LP condition (i.e., 500-Hz LP) was the same for all vocoder and EAS conditions, the mean score was obtained by averaging the LP scores from all five groups of subjects.

For both phoneme and whole syllable recognition, there was a monotonic increase in performance as a function of the number of vocoder channels, except for a slight drop in EAS performance at six channels for the nonsense syllables. When comparing between the left and right panels, average performance across vocoder channels was slightly better ($p \leq 0.05$) for the real words than for the nonsense syllables for both vocoder and EAS conditions. For 500-Hz LP speech, average performance was similar ($p > 0.05$) for nonsense syllables and real words, both for phonemes (36%–38%) and syllables (6%–7%).

b. Benefit of additional low-frequency cues. As demonstrated in Fig. 1, performance with EAS was better than with the vocoder alone, regardless of the number of vocoder channels. On average (across all channel conditions), the percentage-point difference between the EAS and vocoder conditions was significant ($p \leq 0.001$) for phoneme recognition (nonsense, 13 points; real words, 14 points) and for whole syllable recognition (nonsense, 11 points; real words, 15 points). We calculated the PPG and NG to show the benefit of adding the low-frequency cues to the vocoder speech. Data points with very low ($<10\%$) and very high ($>90\%$) percent recognition in the vocoder-alone condition were excluded from these analyses to limit the influence of floor

FIG. 2. Average (left) PPG and (right) NG for (top) phoneme and (bottom) whole syllable recognition at different vocoder-channel conditions. Gains are plotted separately for nonsense syllables (black bars) and for real words (white bars). The asterisk indicates significant difference between nonsense syllables and real words. Data points for the 2ch condition for syllable recognition are excluded from this plot because the performance level was below 10% correct.

and ceiling effects. Figure 2 shows the average PPG (left) and NG (right) for phoneme (upper) and syllable (lower) recognition with nonsense syllables and real words in the different vocoder-channel conditions. Figure 3 shows the individual data for PPG (left) and NG (right) for phoneme

FIG. 3. Individual data for (left) PPG and (right) NG for (top) phoneme and (bottom) syllable recognition as a function of vocoder-alone baseline performance. Gains are plotted separately for nonsense syllables (black) and for real words (white). Data fit with a linear curve indicates significant correlation between the gain and baseline performance.

(upper) and syllable (lower) recognition as a function of baseline (vocoder-alone) performance with nonsense syllables and real words.

A Kruskal-Wallis test showed that neither the PPG nor NG measures varied significantly across the vocoder-channel conditions ($p > 0.05$). In addition, although PPG and NG varied among subjects, neither measure was systematically related to baseline performance ($p > 0.05$) with the exception of PPG for phoneme recognition. Spearman correlations showed that PPG was significantly negatively correlated with the baseline phoneme recognition performance for both nonsense syllables ($r = -0.666$, $p \leq 0.001$) and real words ($r = -0.614$, $p \leq 0.001$; see the linear curve fit in Fig. 3 upper left panel). However, the slopes of the correlation functions were similar between nonsense syllables and real words ($t = -0.210$, $p > 0.05$).² These statistical results confirmed the validity of examining the effects of lexical context (nonsense syllables versus real words) on PPG and NG using a Kruskal-Wallis test applied to the gain values.

When calculated as PPG, EAS benefit was not significantly different between nonsense syllables and real words for phoneme ($p > 0.05$) and syllable ($p > 0.05$) recognition. However, when calculated as NG, EAS benefit was slightly greater with real words than with nonsense syllables for both phoneme (31% vs 23%; $p \leq 0.05$) and whole syllable recognition (22% vs 15%; $p \leq 0.05$). Further pairwise comparisons using a Wilcoxon Signed-Ranks test showed that NG was significantly greater ($p \leq 0.05$) at the 4ch and 6ch conditions for phoneme recognition, and at the 6ch condition for syllable recognition.³ The increase in EAS benefit with stronger context cues cannot be explained by differences in LP-alone performance because, as mentioned earlier (see Fig. 1), the LP-alone performance was similar for the nonsense and real word stimuli for both phoneme and whole syllable recognition. In other words, whereas the LP ear received the same acoustic information for both nonsense syllables and real words, the addition of linguistic context in the LP ear did not improve phoneme and syllable percent-correct recognition scores.

2. IEEE and CUNY sentences

a. Percent-correct scores. Figure 4 shows the mean percent-correct scores for each vocoder channel and EAS condition for word recognition with IEEE (left) and CUNY

FIG. 4. Percent-correct word recognition for (left) IEEE and (right) CUNY sentences for three listening conditions (vocoder, LP, and EAS).

sentences (right). The dashed lines represent the mean scores, averaged across all subjects, for the LP condition. Similar to CVC syllables, there was a monotonic increase in performance as a function of the number of vocoder channels although the mean CUNY scores were limited by ceiling effects beginning at 4–6 channels. When comparing between the left and right panels, average performance across vocoder channels was considerably better ($p \leq 0.001$), by about 24 percentage points, for the higher-context CUNY sentences than for the lower-context IEEE sentences. This is true for both the vocoder-alone and EAS conditions. The LP-alone condition also showed significantly better word recognition with CUNY sentences than with IEEE sentences, with a difference of about 32 percentage points ($p \leq 0.001$).

b. Benefit of additional low-frequency cues. Percent-correct word recognition in the EAS listening condition was better than that in the vocoder-alone condition, regardless of the number of vocoder channels. Analyses were performed to compare PPG and NG between the IEEE and CUNY sentences, again taking into consideration vocoder-channel conditions as well as vocoder-alone baseline performance. As mentioned earlier, the vocoder-alone performance approached ceiling ($>90\%$) with 5 and 6 channels for CUNY sentences, while it was close to floor in the 2ch ($<10\%$) condition for IEEE sentences. Data points with very low ($<10\%$) and very high ($>90\%$) percent-correct recognition in the vocoder-alone condition were excluded from further analyses. Thus, when comparing benefit based on the number of vocoder channels, we considered only the data in the 3ch and 4ch conditions for both sentence types. The average vocoder-alone scores were 36% and 43% for IEEE sentences and 69% and 75% for CUNY sentences for the 3ch and 4ch vocoder condition, respectively.

Figure 5 shows the mean PPG and NG for IEEE and CUNY sentences in the 3ch (left) and 4ch (right) conditions. A Wilcoxon Signed-Ranks test showed no significant difference for PPG between the IEEE and CUNY sentences for either the 3ch ($p > 0.05$) or 4ch ($p > 0.05$) conditions. However, the difference in EAS benefit across sentence types was significant when calculated as NG for both channel conditions (3ch, $p \leq 0.05$; 4ch, $p \leq 0.001$).

FIG. 5. Average PPG and NG for IEEE (black) and CUNY (white) sentences in the (left) 3ch and (right) 4ch vocoder condition. The asterisk indicates a significant difference between IEEE and CUNY.

FIG. 6. Individual data for (left) PPG and (right) NG as a function of vocoder-alone baseline performance with IEEE (black) and CUNY (white) sentences. Data points with very low (<0.10) and very high (>0.90) recognition probabilities in the vocoder-alone condition are excluded. The lower panels replot the PPG and NG with a more restricted range of baseline performance (50%–85%). Data fit with a linear curve showed a significant correlation between gain and baseline performance.

The failure to observe a significant context effect for PPG for the matched channel comparisons is likely due to the baseline difference between IEEE and CUNY. This difference is shown in Fig. 6, which plots the individual data for PPG (upper left) and NG (upper right) as a function of vocoder-alone baseline performance with IEEE and CUNY sentences. Within the 10%–90% baseline performance range, Spearman correlations showed that PPG was significantly correlated with the baseline performance for both IEEE ($r = -0.764$, $p \leq 0.001$) and CUNY ($r = -0.963$, $p \leq 0.001$) sentences. The slopes of the correlation functions were different between IEEE and CUNY sentences.⁴ Similar to CVC syllables, however, NG was not systematically related to baseline performance for either IEEE ($r = -0.011$, $p > 0.05$) or CUNY ($r = -0.091$, $p > 0.05$) sentences. In order to overcome the problem of different regression slopes which would invalidate application of the Kruskal-Wallis test, we restricted the data to a range of baseline performance (50%–85%) where the IEEE and CUNY scores overlapped and the slopes of the correlation functions were similar between the two types of sentences (i.e., data mostly from 5ch and 6ch conditions for IEEE sentences, and 3ch and 4ch from CUNY sentences) [$t = 0.554$, $p > 0.05$]. As seen in the bottom panels in Fig. 6, which shows the restricted data sets, both PPG and NG were larger for CUNY sentences than for IEEE sentences in the majority of cases. A Kruskal-Wallis test showed significant differences in both PPG ($p \leq 0.05$) and NG ($p \leq 0.001$) between the IEEE and CUNY sentences. On average, PPG and NG values were greater for CUNY than for IEEE sentences by 9 points and 33 points, respectively.

Unlike isolated syllables, which demonstrated similar LP performance for the no-context and context conditions

(i.e., nonsense syllables and words, respectively), the LP condition produced significantly higher word recognition scores for the CUNY sentences than for the IEEE sentences. Although comparisons were made at similar vocoder-alone performance levels between IEEE and CUNY sentences, recognition performance was different in the LP ear between these two sentence types. Thus, it is unclear which of several possible factors may underlie the greater EAS benefit observed for the higher-context (CUNY) sentences. These factors include (1) the greater context effect for the LP stimuli than for the vocoder stimuli, (2) a mechanism in which the combined vocoder and LP information facilitates the use of contextual cues compared to the vocoder-alone condition, and (3) an additive effect such that the LP-alone scores simply add to the vocoder scores as described by the probability-summation rule, i.e., $P_c = P_{lp} + P_{voc} - P_{lp}P_{voc}$ (Kong and Carlyon, 2007; Micheyl and Oxenham, 2012).

Regardless of the underlying contributing factor(s), however, the fact that NG was found to be significantly greater for CUNY sentences than for IEEE sentences in the matched-channel comparisons, where subjects received the same amount of acoustic information for both sentence types, suggests that stronger linguistic context provides greater EAS benefit.

III. EXPERIMENT 2: FACILITATION OF USE OF PRIOR KNOWLEDGE WITH LIMITED LOW-FREQUENCY CUES

The first experiment evaluated the effect of linguistic context on EAS benefit by systematically varying the amount of spectral information in the vocoder ear, while fixing the spectral information in the LP ear. This second experiment further evaluated the context effect on EAS benefit when the LP ear received more degraded low-frequency information that contained only F0 contour and temporal envelope (AM + FM) cues extracted from the low-frequency speech. This simulated a common situation in which the EAS user has very limited residual hearing in the non-implanted ear that contributes zero percent speech intelligibility. While it has been demonstrated that the limited low-frequency cues could improve speech intelligibility in noise when combined with the CI/vocoder speech (Kong and Carlyon, 2007; Brown and Bacon, 2009a,b) via the glimpsing mechanism, it is unclear if the very limited acoustic information in the form of AM + FM cue provided by the LP ear could also facilitate the use of prior knowledge. In addition, by restricting the cues to F0 and temporal envelope, which would equate the baseline performance level across the low and high context sentence materials in the LP ear, it allowed us to infer whether context effects could be attributed to probability summation across ears. At the same time, it avoided the ceiling effect in the EAS condition for the high-context sentence materials encountered in the first experiment. Stimuli for this experiment were the same CUNY and IEEE sentences used in experiment 1; however, the LP ear received more limited speech information, in the form of an amplitude-modulated LP filtered harmonic complex (HC).

A. Method

1. Subjects

Eight normally hearing young adults between 18 and 25 years of age participated in this experiment. All were native speakers of American English and had no known history of speech or hearing problems.

2. Stimuli

Test stimuli were derived from the CUNY and IEEE sentences used in experiment 1. As noted earlier, these stimuli were matched for speaking rate and pitch excursion, unlike the stimuli used in the study of [Brown and Bacon \(2009b\)](#). The same channel vocoder as in experiment 1 was used to process the sentence materials into 2 or 4 channels. Equal-amplitude HCs were created with F0s that followed the F0 contours of the original sentences. This processing was performed in PRAAT ([Boersma and Weenink, 2009](#)). The HCs were then LP filtered at 500 Hz, using the same filtering parameters as experiment 1. Finally, the LP-filtered HCs were amplitude modulated with the temporal envelope of the 500-Hz LP speech. These processing steps preserved the temporal envelope of the LP speech and the F0 contours of the original speech, but eliminated other possible low-frequency cues such as F1 frequency. As reported by [Kong and Carlyon \(2007\)](#), speech intelligibility was zero with only AM + FM cues for high-context sentences, and preliminary testing with two subjects also showed zero percent word recognition with the CUNY sentences.

3. Procedures

CUNY and IEEE sentences were presented in two listening conditions: Vocoder alone and EAS (i.e., HC + vocoder). When combined with the two vocoder-channel conditions (2ch-vocoder and 4ch-vocoder), this resulted in a total of eight test conditions (2 vocoder channels $\times$ 2 listening conditions $\times$ 2 sentence materials). Each subject was tested in all eight conditions. Similar to experiment 1, sentence materials and listening conditions were counterbalanced across subjects. Half of the subjects were tested with two vocoder channels first, for both the vocoder-alone and EAS conditions, followed by four vocoder channels; the remaining half were tested in the reverse order. Other methods, including the method of stimulus presentation and procedures for practice and testing, were the same as those used in experiment 1.

4. Data analyses

Scoring procedures were the same as in experiment 1 and EAS benefit was again calculated as PPG and NG. The effects of context on EAS benefit were evaluated by comparing PPG and NG between CUNY and IEEE sentences using (1) the same number of vocoder channels, and (2) a similar range of baseline performance. Shapiro-Wilk tests were used to confirm a normal distribution ($p > 0.05$) of the data for each test condition. Subsequently, repeated-measures analysis of variance (ANOVA) and paired t -tests were used to determine if either percent-correct scores or EAS benefit was

significantly different between pairs of test conditions. Correlational analyses were performed using Pearson's correlation.

B. Results

1. Percent-correct scores

The top panels in Fig. 7 show the mean percent-correct scores for each vocoder channel and EAS condition for word recognition with IEEE (left) and CUNY sentences (right). As expected, speech recognition performance was higher for the 4ch condition as compared to the 2ch condition. Averaged percent-correct scores across vocoder channels were greater for CUNY than for IEEE sentences by 22 and 18 percentage points for the vocoder-alone [$F(1,7) = 85.91$, $p \leq 0.001$] and EAS [$F(1,7) = 183.27$, $p \leq 0.001$] conditions, respectively. Also, average percent-correct scores across

FIG. 7. Percent-correct scores for (left) IEEE and (right) CUNY sentences for vocoder alone and EAS conditions are shown in the top panels. The middle panels compare the (left) PPG and (right) NG between IEEE (black bars) and CUNY (white bars) sentences in the 2ch and 4ch conditions. The lower panels show individual PPG and NG for both sentence types as a function of vocoder-alone baseline performance. Note that the PPGs and NGs were converted into z-scores to account for the subject factor. In the bottom two panels, all four data series demonstrated a correlation between gain and baseline performance.

vocoder channels were greater for the EAS condition than for the vocoder-alone condition for both IEEE [$F(1,7) = 8.91$, $p \leq 0.05$] and CUNY [$F(1,7) = 32.15$, $p \leq 0.001$] sentences.

2. Benefit of additional low-frequency cues

The middle panels of Fig. 7 show PPG (left) and NG (right) for both IEEE and CUNY sentences with 2 and 4 vocoder channels. The lower panels show individual PPG (left), and NG (right) as a function of baseline performance. Note that both PPG and NG values were converted into z-scores to account for the subject factor.

When comparing gain differences between IEEE and CUNY sentences, two-way repeated-measures ANOVAs (2 sentence types $\times$ 2 channel conditions) were performed. For PPG, significant main effects of sentence type [$F(1,7) = 18.14$, $p \leq 0.005$] and channel condition [$F(1,7) = 6.85$, $p \leq 0.05$] were found. The interaction between the two factors was not significant [$F(1,7) = 0.33$, $p > 0.05$]. Averaged across channel conditions, PPG was seven points greater for CUNY than for IEEE sentences. Paired t -tests confirmed that PPG was greater for CUNY sentences than for IEEE sentences at each channel condition [2ch, 6 points $t(7) = 2.73$, $p \leq 0.05$; 4ch, 8 points $t(7) = 2.97$, $p \leq 0.05$]. For NG, the main effect of sentence type was significant [$F(1,7) = 39.51$, $p \leq 0.001$], but the main effect of channel condition was not [$F(1,7) = 0.938$, $p > 0.05$]. However, there was a significant interaction between these two factors [$F(1,7) = 9.11$, $p \leq 0.05$]. Paired t -tests showed that NG was greater for CUNY than for IEEE sentences by 9 points for the 2ch condition [$t(7) = 3.81$, $p \leq 0.01$] and by 26 points for the 4ch condition [$t(7) = 5.04$, $p \leq 0.001$].

Both PPG and NG varied across subjects and across baseline performance. PPG decreased as a function of baseline performance for both IEEE ($r = -0.658$, $p \leq 0.01$) and CUNY ($r = -0.490$, $p \leq 0.05$) sentences. NG decreased with baseline performance for IEEE sentences ($r = -0.536$, $p \leq 0.05$), but increased for CUNY sentences ($r = 0.648$, $p \leq 0.01$). As seen in Fig. 7, PPG and NG were greater for CUNY than for IEEE sentences for the majority of cases. Paired- t tests were performed to compare PPG and NG between IEEE and CUNY sentences at a baseline range where the two sentence types overlapped (i.e., 10%–50%, 4ch IEEE vs 2ch CUNY). Results showed significant greater PPG [$t(7) = 7.35$, $p \leq 0.001$] and NG [$t(7) = 5.02$, $p \leq 0.005$] for CUNY sentences than for IEEE sentences by 10 and 14 points, respectively. Taken together, the results in this experiment confirm the findings from experiment 1 that EAS benefit is augmented by contextual cues.

IV. DISCUSSION

The current study investigated the effect of linguistic context on EAS benefit in simulated EAS listening. We tested speech recognition with materials varying in the amount of context from nonsense syllables to isolated words, low-context sentences and high-context sentences. To minimize potential confounds, we controlled for differences in speaking styles, such as speaking rate and pitch excursion for the sentences materials. We showed that speech

recognition performance was generally greater in the simulated EAS condition than in the vocoder-alone condition across all types of speech materials. The effect of linguistic context on EAS benefit was revealed by comparing the gain (PPG or NG) between two types of speech materials that contained greater or lesser amounts of contextual cues.

Overall, our findings suggest that (1) EAS benefit is greater for speech materials with higher context than those with lower context, even when spectral content and baseline performance are matched in both the vocoder and LP ears. (2) Context effects persist across conditions that provide varied amounts of spectral information in both the vocoder and LP signals. Given that significant effects of context were observed consistently for the NG measure, but not the PPG gain measure in experiment 1, however, it is important to observe a degree of caution in interpreting these findings.

A. Influence of outcome measures

The choice of an appropriate metric for measuring bimodal benefit takes on particular importance when comparisons of benefit are made across materials that produce different levels of baseline performance (e.g., different levels of spectral resolution, or different levels of linguistic context) as done in the present study. PPG represents a straightforward metric that has been used in most previous EAS studies; however, a possible limitation of PPG is that it implicitly assumes that a given percentage-point improvement in performance is equally meaningful whether applied to a low or high baseline performance. In our view, NG represents a more practically relevant measure of bimodal benefit because it reflects the proportion of available gain (i.e., between baseline performance and 100%) attributable to bimodal benefit. A potential limitation of NG is that it would appear to favor listening conditions that generate higher levels of baseline performance because a fixed percentage-point increase in performance translates to a higher NG when baseline performance is higher. Interestingly, however, in our data the NG is relatively consistent across baseline levels. On the other hand, our data reveal a trend for PPG to decrease as a function of baseline performance (phonemes in Fig. 3, IEEE and CUNY sentences in Figs. 6 and 7), suggesting that the PPG measure may be inherently biased against conditions that generate higher levels of baseline performance. This would result in a reduced likelihood of observing context effects for PPG measures as compared to NG measures, consistent with our findings.

B. Different levels of contextual cues

Sources of linguistic context include (1) phonotactic constraints, which govern the acceptable sequences in which phonemes can occur in a given language; (2) lexical context, which distinguishes real words (that exist in the lexicon) from other phonotactically correct words; and (3) sentential context, which refers to the increase in predictability of words that occur in sentence context, due to the listener's world and syntactic knowledge. In experiment 1, we showed that the average NGs for the 3ch and 4ch vocoder conditions were considerably greater for sentences (IEEE, 38%; CUNY, 70%)

than for isolated syllables (nonsense syllables, 11%; real words, 18%). At the sentence level, the difference in NG between IEEE and CUNY sentences was as large as 33 points. This finding is consistent with the notion that sentence meaning represents the most effective source of contextual constraint, as discussed by [Boothroyd and Nitttrouer \(1988\)](#).

The use of AM + FM stimuli for the LP ear in experiment 2 minimized the possibility that the increased EAS benefit observed for the higher context CUNY sentences in experiment 1 could be attributed to probability summation. The average NG difference between IEEE and CUNY sentences was 23 points, again greater than what we observed for isolated syllables. Taken together, then, the results of experiments 1 and 2 support the notion that EAS benefit increases as the listener is given greater opportunity to apply prior knowledge of the spoken language.

C. Contribution of additional low-frequency cues to the use of prior knowledge for spoken-word recognition

The question remains as to why additional low-frequency cues contribute more strongly to the understanding of higher context materials as compared to lower context materials. Two possible explanations are discussed below.

1. Interaction between sensory input and context information

Our findings suggest a combined effect of improved sensory input and the use of prior linguistic knowledge on EAS benefit for spoken-word recognition. In the case of the LP AM + FM signal in experiment 2, the LP signal does not contribute to speech intelligibility for percent-correct word recognition (consistent with [Kong and Carlyon, 2007](#)), but it enhances word recognition performance when combined with vocoded speech in a quiet background. The surprising finding in our study is that an LP AM + FM signal enhanced speech intelligibility to a greater extent for higher context CUNY sentences than for lower context IEEE sentences, even when the spectral content and the baseline percent-correct word recognition levels were matched across the two types of speech materials in *both* ears. This EAS benefit, observed with the AM + FM signal, may stem from an improvement in the listener's identification of syllable and word boundaries ([Spitzer et al., 2009](#)).

Among contemporary models of speech perception, the interactive models (e.g., TRACE, [McClelland and Elman, 1986](#)) consider linguistic context effects to reflect top-down processes that interact strongly with bottom-up processes, not only helping the listener to identify the most likely target word, but also influencing the listener's interpretation of the bottom-up (acoustic-phonetic) cues. According to this view of speech processing, our current findings suggest that improved sensory input (e.g., due to better segmentation) in the EAS condition facilitates top-down repair mechanisms. Other researchers have also argued for the existence of a role of top-down processing or interaction between sensory and top-down processing in EAS benefit. For example, using a

PhR paradigm [Başkent et al. \(2009\)](#) showed that top-down restoration of spectrally degraded speech was reduced when the speech signal was interrupted temporally by silent intervals, possibly due to interference with auditory object formation. In another study, [Başkent \(2012\)](#) showed that a significant restoration benefit was found for vocoder speech only when the vocoder signal provided relatively good spectral resolution (i.e., 32 channels). In addition, the benefit of adding low-frequency information in a simulated EAS condition provided slightly more benefit than it provided in a vocoder-alone condition. The observation of PhR effects only in conditions that provide relatively high levels of spectral resolution supports an interaction between bottom-up and top-down processing. Although we did not observe a systematic effect of spectral resolution on the EAS context effect in the present research, there was a trend for the context benefit (i.e., the gain difference between lower and higher context speech materials) to be greater for higher channel vocoder conditions as compared to lower channel conditions (cf. Figs. 2 and 7). This suggests that the less degraded bottom-up speech signals were better able to facilitate the top-down process, resulting in greater EAS benefit.

In contrast to interactive models of speech perception, autonomous models of speech perception (e.g., Shortlist-B; [Norris and McQueen, 2008](#)) postulate a strictly feedforward process of word recognition. The recently revised Shortlist B model argues that information related to linguistic context and word frequency provides additional evidence to “revise the listener's prior beliefs” in the system. Under this view, the context effect for EAS benefit observed in the present study could be modeled as the combined effects of improved sensory inputs and prior linguistic knowledge (i.e., the interaction between perceptual process and prior knowledge) on the alternation of the prior probability of encountering a particular word.

The present study was not designed to test a specific model of speech perception, and further discussion of how the expectation-driven effects are being realized (i.e., top-down vs feedforward) is beyond the scope of this paper. In general, it appears that our findings of an interaction between perceptual and contextual information could potentially be supported by both types of models.

2. Differential weighting of frequency bands

Another possible explanation for the context effect in EAS benefit is that band-importance functions derived using the Speech Intelligibility Index [SII; [American National Standards Institute \(ANSI\), 1997](#)] are different across different speech materials. The SII describes the relative contribution of each spectral band to the total speech information available in a stimulus. Regardless of bandwidth condition, it has been shown that isolated words (e.g., CID-22 or NU6 words) have greater band-importance values than nonsense syllables in the low-frequency bands (<1000 Hz; [ANSI, 1997](#)). It has also been shown that band-important functions differ for lower- versus higher-context sentences. Using the “compound” technique, [Healy et al. \(2013\)](#) created band-importance functions for the final word in the high-

predictability and low-predictability SPIN sentences spoken by a male speaker (Kalikow *et al.*, 1977; Bilger *et al.*, 1984). They showed that while the two functions were generally similar in shape, there were differences in the importance of individual bands between the two functions. When summed across the low-frequency bands up to 840 Hz (a frequency range that is relevant to the LP stimuli used in the current study), the importance value was greater for the high-predictability than for the low-predictability sentences by about 0.04 (i.e., about 13% greater for high-predictability sentences). A similar magnitude of difference was also observed between the high-predictability SPIN sentences and the isolated CID-W22 words.

If differences in the importance functions are similar for the CVC words and nonsense syllables, and for the CUNY sentences and IEEE sentences used here, then these differences could explain, or at least contribute to, the differences in EAS benefit observed in experiments 1 and 2. The idea is that low-frequency cues in the LP ear enhance the combined perception of low-frequency speech cues across the two ears. If the low-frequency information carries greater importance for the higher-context speech materials, then EAS benefit should be greater for higher-context speech materials relative to lower-context speech materials. Although this explanation is plausible, it should be noted that band-importance functions are talker and speech material specific (Healy *et al.*, 2013). Thus, without directly generating the band-importance functions for our isolated syllable and sentence materials, we cannot be certain that they differ in the same way as the isolated words and nonsense syllables reported in ANSI (1997), and as the low- and high-predictability SPIN sentences reported by Healy *et al.* (2013).

D. Future research

The role of the listener's prior knowledge of spoken language in word recognition, and bimodal hearing in particular, is complex; thus, there is a need for additional research in this area. Most standardized speech materials that are currently available were developed when speech perception research focused primarily on perceptual factors. These materials are not particularly well suited to investigations of listeners' use of contextual information, or the interactions between sensory information and prior knowledge. Thus, future investigations related to EAS benefit would benefit from the development of materials more ecologically relevant to speech and language comprehension. For example, it would be useful to have standardized tests of listening comprehension (e.g., Best *et al.*, 2013) that are stratified in terms of word- and sentence-level context and/or factors that may influence lexical segmentation (e.g., cues that encode word boundaries).

Related to the discussion of band-importance functions above, it may be fruitful to investigate differences in the importance functions of different types of speech materials presented with and without supplemental low-frequency cues. EAS involves the integration of two sources of speech information across ears or across different frequency regions, a process similar to the integration of speech cues across auditory and visual

inputs (Ronan *et al.*, 2004; Kong and Braida, 2011). Thus, the ANSI (1997) method for estimating band-importance functions for combined audio-visual inputs (p. 20) could likely be used to evaluate importance functions for EAS.

E. Clinical implications

The current findings indicate that EAS benefit is greatest when listeners are tested with speech materials that contain high levels of linguistic context. This suggests that EAS benefit may be greater in everyday communication than when measured in the typical laboratory or clinical setting. Current assessment tools for speech perception in CI users include isolated words and simple sentences, but the selection of speech materials to evaluate EAS benefit varies across clinics. In some cases, the decision to implant the second ear in a unilateral CI patient may be based on the patient's apparent lack of EAS benefit, as reflected in tests of phoneme or isolated word recognition. Our results suggest that higher context sentence materials may be more appropriate for evaluating EAS benefit, so long as ceiling effects are avoided. It is currently not known how the use of prior knowledge of spoken language (e.g., the ability to make use of sentence-level context) or lexical segmentation is affected if the residual low-frequency hearing ear is replaced by a second implant. This issue has potentially important clinical consequences and should be investigated in future research.

The present work tested speech recognition in quiet in order to examine the effects of context while minimizing potential confounds related to the glimpsing mechanism. The reliance on prior knowledge during speech perception is almost certainly greater when portions of the speech signal are masked by noise than when speech occurs in quiet. Thus, in realistic listening situations, which often include background noise, the magnitude of context effects on EAS benefit could be even larger than those observed here.

ACKNOWLEDGMENTS

The authors thank Dr. Joshua Bernstein for his helpful comments and suggestions on the relationship between EAS benefit and band-importance functions for different types of speech materials. We thank Dr. Ben Munson for his helpful comments on earlier versions of the manuscript. Finally, we thank Soo Hee Oh and the Biostatistic Service Center at Northeastern University for assistance with data analysis. This work was supported by NIH-NIDCD DC012300.

¹We view the low-frequency cues as supplemental cues to the vocoder speech. Hence, in this paper, EAS benefit is considered as the increase in speech recognition performance in the EAS condition compared with the vocoder-alone condition.

²The normality assumption was met for this comparison. Thus, linear regression analysis was performed to compare the slopes of the two functions.

³Pairwise comparisons between nonsense and real words for syllable recognition were not performed for the 2ch, 3ch, and 4ch conditions due to limited data after we excluded data <10% and >90%.

⁴A regression analysis to compare the slopes of the two functions was not performed due the fact that normality was not met for the CUNY PPG data.

- American National Standards Institute (ANSI) (1997). S3.5 R2007, *American National Standard Methods for the Calculation of the Speech Intelligibility Index* (Acoustical Society of America, New York).
- Bashford, J. A., Jr., and Warren, R. M. (1979). "Perceptual synthesis of deleted phonemes," *J. Acoust. Soc. Am.* **65**, S112.
- 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.
- Başkent, D., Eiler, C., and Edwards, B. (2009). "Effects of envelope discontinuities on perceptual restoration of amplitude-compressed speech," *J. Acoust. Soc. Am.* **125**, 3995–4005.
- Best, V., Keidser, G., Buchholz, J., and Freeston, K. (2013). "Psychometric effects of adding realism to a speech-in-noise test," *J. Acoust. Soc. Am.* **133**, 3379.
- Bhargava, P., Gaudrain, E., and Başkent, D. (2014). "Top-down restoration of speech in cochlear-implant users," *Hear. Res.* **309**, 113–123.
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeccowski, C. (1984). "Standardization of a test of speech perception in noise," *J. Speech Hear. Res.* **27**, 32–48.
- Boersma, P., and Weenink, D. (2009). "PRAAT: Doing phonetics by computer (version 5.1.05) [computer program]."
- Boothroyd, A., Hnath-Chisolm, T., Hanin, L., and Kishon-Rabin, L. (1988). "Voice fundamental frequency as an aid to the speechreading of sentences," *Ear Hear.* **9**, 306–312.
- Boothroyd, A., and Nittrouer, S. (1988). "Mathematical treatment of context effects in phoneme and word recognition," *J. Acoust. Soc. Am.* **84**, 101–114.
- 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.
- de Pijper, J. R. (1983). *Modeling British English Intonation* (Foris Publications, Dordrecht, Netherlands), pp. 1–150.
- Garrett, R. L., and Saint-Pierre, L. (1980). "Message comprehension as a function of intelligibility and advanced knowledge of the topic," *J. Aud. Res.* **20**, 71–80.
- Grant, K. W., and Seitz, P. F. (1998). "Measures of auditory-visual integration in nonsense syllables and sentences," *J. Acoust. Soc. Am.* **104**, 2438–2450.
- Healy, E. W., Yoho, S. E., and Apoux, F. (2013). "Band importance for sentences and words reexamined," *J. Acoust. Soc. Am.* **133**, 463–473.
- Hu, Y., and Loizou, P. C. (2010). "Effects of introducing low-frequency harmonics in the perception of vocoded telephone speech," *J. Acoust. Soc. Am.* **128**, 1280–1289.
- Institute of Electrical and Electronics Engineers (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoustics* **17**, 227–246.
- Kalikow, D. N., Stevens, K. N., and Elliot, L. L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* **61**, 1337–1351.
- Kirk, K. I., Hay-McCutcheon, M. J., Holt, R. F., Gao, S., Qi, R., and Gehrlein, B. L. (2007). "Audiovisual spoken word recognition by children with cochlear implants," *Audiol. Med.* **5**, 250–261.
- Kong, Y.-Y., and Braid, 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.
- Kopelovich, J. C., Reiss, L. A., Oleson, J. J., Lundt, E. S., Gantz, B. J., and Hansen, M. R. (2014). "Risk factor for loss of ipsilateral residual hearing after hybrid cochlear implantation," *Otol. Neurotol.* **35**, 1403–1408.
- Lenarz, T., James, C., Cuda, D., Fitzgerald O'Connor, A., Frachet, B., Frijns, J. H., Klenzner, T., Laszig, R., Manrique, M., Marx, M., Merkus, P., Mylanus, E. A., Offeciers, E., Pesch, J., Ramos-Macias, A., Robier, A., Sterkers, O., and Uziel, A. (2013). "European multi-centre study of the Nucleus Hybrid L24 cochlear implant," *Int. J. Audiol.* **52**, 838–848.
- 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.
- Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model," *Ear Hear.* **19**, 1–36.
- McClelland, J. L., and Elman, J. L. (1986). "The TRACE model of speech perception," *Cog. Psychol.* **18**, 1–86.
- Michéyl, C., and Oxenham, A. J. (2012). "Comparing models of the combined-stimulation advantage for speech recognition," *J. Acoust. Soc. Am.* **131**, 3970–3980.
- Norris, D., and McQueen, J. M. (2008). "Shortlist B: A Bayesian model of continuous speech recognition," *Psychol. Rev.* **115**, 357–395.
- Peterson, G. E., and Lehiste, I. (1962). "Revised CNC lists for auditory tests," *J. Speech Hear. Disord.* **27**, 62–70.
- Picheny, M. A., Durlach, N. I., and Braid, L. D. (1986). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," *J. Speech Hear. Res.* **29**, 434–446.
- Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (1992). "Relations among different measures of speech reception in subjects using a cochlear implant," *J. Acoust. Soc. Am.* **92**, 1869–1881.
- Ronan, D., Dix, A. K., Shah, P., and Braid, L. D. (2004). "Integration across frequency bands for consonant identification," *J. Acoust. Soc. Am.* **116**, 1749–1762.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- 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.
- Sumbly, W. H., and Pollack, I. (1954). "Visual contribution to speech intelligibility in noise," *J. Acoust. Soc. Am.* **26**, 212–215.
- Tyler, M. D., and Cutler, A. (2009). "Cross-language differences in cue use for speech segmentation," *J. Acoust. Soc. Am.* **126**, 367–376.
- Warren, R. M. (1970). "Perceptual restoration of missing speech sounds," *Science* **167**, 392–393.
- Warren, R. M., and Obusek, C. J. (1971). "Speech perception and phonemic restorations," *Percept. Psychophys.* **9**, 358–362.
- Zhang, T., Dorman, M. F., and Spahr, A. J. (2010). "Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation," *Ear Hear.* **31**, 63–69.