# Influences of noise-interruption and information-bearing acoustic changes on understanding simulated electric-acoustic speech[a)]

Christian Stilp[b)]
*Department of Psychological and Brain Sciences, University of Louisville, Louisville, Kentucky 40292, USA*

Gail Donaldson and Soohee Oh[c)]
*Department of Communication Sciences and Disorders, University of South Florida, PCD 1017, 4202 East Fowler Avenue, Tampa, Florida 33620, USA*

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

(Received 6 January 2016; revised 18 August 2016; accepted 28 October 2016; published online 23 November 2016)

In simulations of electrical-acoustic stimulation (EAS), vocoded speech intelligibility is aided by preservation of low-frequency acoustic cues. However, the speech signal is often interrupted in everyday listening conditions, and effects of interruption on hybrid speech intelligibility are poorly understood. Additionally, listeners rely on information-bearing acoustic changes to understand full-spectrum speech (as measured by cochlea-scaled entropy [CSE]) and vocoded speech ($CSE_{CI}$), but how listeners utilize these informational changes to understand EAS speech is unclear. Here, normal-hearing participants heard noise-vocoded sentences with three to six spectral channels in two conditions: vocoder-only (80–8000 Hz) and simulated hybrid EAS (vocoded above 500 Hz; original acoustic signal below 500 Hz). In each sentence, four 80-ms intervals containing high-$CSE_{CI}$ or low-$CSE_{CI}$ acoustic changes were replaced with speech-shaped noise. As expected, performance improved with the preservation of low-frequency fine-structure cues (EAS). This improvement decreased for continuous EAS sentences as more spectral channels were added, but increased as more channels were added to noise-interrupted EAS sentences. Performance was impaired more when high-$CSE_{CI}$ intervals were replaced by noise than when low-$CSE_{CI}$ intervals were replaced, but this pattern did not differ across listening modes. Utilizing information-bearing acoustic changes to understand speech is predicted to generalize to cochlear implant users who receive EAS inputs. © *2016 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4967445]

[JFL] Pages: 3971–3979

## I. INTRODUCTION

Many studies have shown that cochlear implant (CI) users' speech perception may be enhanced by low-frequency acoustic cues available from residual hearing in the implanted or non-implanted ear (e.g., Tyler *et al.*, 2002; Turner *et al.*, 2004; Kong *et al.*, 2005; Gifford *et al.*, 2007; Dorman *et al.*, 2008; Brown and Bacon, 2009a; Dorman and Gifford, 2010; Zhang *et al.*, 2010). This phenomenon has been termed *electric-acoustic stimulation (EAS) benefit*. When presented alone, the low-frequency acoustic signal does not support proficient speech perception. However, when combined with the output from a CI, listeners are able to utilize one or more low-frequency cues, including fundamental frequency, voicing, the first formant ($F_1$), and amplitude envelope information, to enhance the perception of degraded cues from the vocoded ear. This enhancement leads to improved speech perception for the combined signal compared to the CI alone (Qin and Oxenham, 2006; Kong and Carlyon, 2007; Brown and Bacon, 2009a,b, 2010; Zhang *et al.*, 2010; Kong and Braida, 2011; Kong *et al.*, 2015; Oh *et al.*, 2016a,b), although the magnitude of this benefit is variable (Dorman *et al.*, 2015).

Residual low-frequency hearing preserves fine-structure information, which allows better encoding of *dynamic* spectral cues. It is known that normal-hearing listeners utilize rapid spectral changes in the speech signal to understand speech. The use of interrupted speech is a way to understand the relative importance of different acoustic information for speech intelligibility. Interruptions that follow a constant duty cycle date back to Miller and Licklider (1950) and are widely used even today (e.g., Powers and Wilcox, 1977; Shafiro *et al.*, 2011; Benard and Başkent, 2015; etc.). Others have explored effects of interruptions that are time-locked to specific events in the speech signal, such as consonant or vowel sounds (e.g., Cole *et al.*, 1996; Kewley-Port *et al.*, 2007; Fogerty *et al.*, 2015). Stilp *et al.* (2013) introduced a measure of information-bearing acoustic change in vocoded speech (cochlea-scaled

[a)]Results were presented at the 169th Meeting of the Acoustical Society of America, Pittsburgh, PA, USA.
[b)]Electronic mail: christian.stilp@louisville.edu
[c)]Current address: Audiology Institute, Hallym University of Graduate Studies, 405, Yeoksam-no, Gangnam-gu, Seoul, 135-841 South Korea

entropy in cochlear implants; $CSE_{CI}$) to evaluate the effect of reduced spectral resolution on the perception of dynamic changes in the speech signal. Unlike the traditional method of speech interruption, where focuses have been on the nature of the interrupter (noise versus silence, duty cycle frequency and period), this metric was based on the information-theoretic principle that uncertainty or unpredictability in the stimulus conveys potential information for perception (Kluender and Alexander, 2008; Kluender *et al.*, 2013). Speech intervals that rated as high $CSE_{CI}$ (high unpredictability, i.e., higher information) were predicted to be more important for perception, while low-$CSE_{CI}$ intervals (lower unpredictability; i.e., lower information) were predicted to be far less important. Sentence recognition was significantly poorer when high-$CSE_{CI}$ intervals were replaced by noise compared to when an equal number and duration of low-$CSE_{CI}$ intervals were replaced. This result has been replicated and extended across a wide range of spectral resolutions in noise-vocoded speech (Stilp and Goupell, 2015), suggesting broad importance for information-bearing acoustic changes when understanding vocoded speech. Thus, while CIs transmit degraded speech signals, these signals are still replete with rapid acoustic changes that are important for understanding speech. Notably, listeners in the Stilp *et al.* (2013) study performed more poorly than listeners in other studies of interrupted vocoded sentence intelligibility where constant duty cycle interruption was used (Başkent and Chatterjee, 2010; Chatterjee *et al.*, 2010; Başkent, 2012) despite having a smaller proportion of the speech signal replaced by noise. This outcome highlighted the importance of *what* got replaced by noise (i.e., regions of high information), not just *how much* of the signal got replaced, a distinction that was later supported experimentally by Stilp (2014). Given the importance of information-bearing acoustic changes for understanding noise-vocoded speech and full-spectrum speech when presented separately (Stilp and Kluender, 2010; Stilp *et al.*, 2013; Stilp, 2014; Stilp and Goupell, 2015), informational changes might be expected to be important for understanding hybrid speech as well.

While information-bearing acoustic changes are predicted to be important for understanding hybrid speech, the *degree* to which listeners utilize information-bearing changes in the hybrid listening condition is unclear. Information-bearing acoustic changes might prove *more* important for understanding hybrid speech than they are for understanding CI-alone (vocoded) speech. Hybrid speech conveys spectral cues across a broader range of cochlear locations, and may more effectively transmit dynamic low-frequency cues to place-appropriate neurons in the cochlear apex (Smith *et al.*, 2002; Oxenham *et al.*, 2004; Middlebrooks and Snyder, 2010). Spectral changes in lower-frequency regions of the speech signal play an important role in accurate speech recognition (Stevens, 2002). For example, many English vowels feature perceptually significant changes in $F_1$ across time (Hillenbrand and Nearey, 1999), and disrupting these changes can perturb vowel recognition (Nearey and Assmann, 1986). $F_1$ transitions are also important for consonant recognition such as final fricative voicing (/s/ vs /z/, e.g., "loss" versus "laws"). Given that hybrid speech better captures perceptually significant low-frequency changes in the speech signal, replacing these acoustic changes with noise might produce greater perceptual impairment than replacing corresponding segments in vocoder-only speech. On the other hand, information-bearing acoustic changes might prove *less* important for understanding hybrid speech than for understanding CI-alone speech. This is based on the notion that higher-quality speech signals are more perceptually resilient to degradation (i.e., noise replacement) than lower-quality signals. With higher signal quality (e.g., hybrid speech), listeners have better access to speech cues that are not available or accessible in impoverished speech (e.g., vocoder-alone speech). This explains the finding that information-bearing acoustic changes are more important for understanding vocoded sentences with few spectral channels versus many channels (at least 6 channels versus up to 24 channels; Stilp and Goupell, 2015).

The goals of the present study were (1) to investigate the effect of sound quality (i.e., vocoder speech versus hybrid speech) on the relative importance of information-bearing acoustic changes for speech recognition, and (2) to examine the effects of noise interruption on the perception of vocoder speech versus hybrid speech. To address these goals, we tested normal-hearing listeners in simulations of CI-alone (noise vocoder) and hybrid (noise vocoder plus low-pass-filtered speech) stimulation. Two key hypotheses were examined. First, we hypothesized that listeners would utilize information-bearing acoustic changes to understand both vocoded and hybrid sentences, but to different degrees. Specifically, we predicted that listeners would assign comparatively more importance to information-bearing acoustic changes for understanding vocoded speech as compared to hybrid speech. This outcome was expected because listeners relied more heavily on information-bearing acoustic changes to understand poorer-quality speech (noise-vocoded) than higher-quality speech (spectrally intact; Stilp, 2014). While the same general pattern of results was anticipated across listening conditions (poorer sentence intelligibility when high-$CSE_{CI}$ intervals were replaced by noise compared to when low-$CSE_{CI}$ intervals were replaced), we predicted a bigger difference between the high- and low-$CSE_{CI}$ conditions for vocoded sentences than for bimodal sentences. Second, we hypothesized that noise interruption would have a bigger effect on perception of hybrid speech than vocoded speech. The intact low-frequency portion of the speech signal contains voice pitch and voicing cues that facilitate speech segmentation, resulting in an EAS benefit (e.g., Spitzer *et al.*, 2009; Zhang *et al.*, 2010; Kong *et al.*, 2015). Interrupting these cues with noise (whether due to replacing low or high $CSE_{CI}$ intervals) was expected to impair the intelligibility of hybrid sentences relatively more than interrupting vocoded sentences, which lack these cues.

## II. METHODS

### A. Participants

Thirty-six normally hearing young adults between 18 and 35 years of age (30 females, average age 22.9 years) participated in this experiment. All were native speakers of American English and passed a hearing screening ($<$25 dB hearing level at octave frequencies between and including

250 to 8000 Hz) in both ears. Listeners provided informed consent and were compensated on an hourly basis for participation. Study procedures were approved by the University of South Florida Institutional Review Board.

## B. Stimuli

### 1. Materials

The experiment utilized IEEE sentences recorded by a female talker, as described in Kong *et al.* (2015). The speaker held the speaking rate and articulation effort constant across sentence lists and sentence sets. List 1 was used as familiarization, lists 2–13 were used for training, lists 14–15 were used for baseline testing, and lists 16–63 were used for testing. Mean sentence duration was 2340 ms (range = 1597–3289 ms).

### 2. Vocoder and electric-acoustic simulations

Vocoder speech was generated by first dividing the sentence bandwidth of 80–8000 Hz into three, four, five, or six spectral channels according to the Greenwood (1990) formula (see Table I). These spectral resolutions were selected to eliminate floor and ceiling effects and to model the range of speech perception in quiet observed among CI users. Each channel was bandpass filtered using a third-order elliptical filter with 2 dB passband ripple and −50 dB gain in the stopband. The amplitude envelope of each channel was extracted via Hilbert transform followed by a sixth-order Butterworth low-pass filter with a 400-Hz cutoff. Amplitude envelopes were used to modulate samples of white noise bandpass-filtered with the same corner frequencies, then recombined to form the vocoded sentence.

Hybrid speech was created by vocoding the signal as described above, then high-pass-filtering using a tenth-order Butterworth filter with 500-Hz cutoff. The original speech signal was low-pass-filtered using the same tenth-order Butterworth filter with 500-Hz cutoff. The root-mean-square (RMS) amplitude of the low-pass speech was adjusted to match that of the vocoded speech it was replacing, then the two signals were combined to create hybrid speech.

### 3. $CSE_{CI}$ calculation

The spectrum of hybrid speech contains both spectrally intact and degraded portions, raising the question as to whether information-bearing acoustic changes are better measured using a metric that assumes normal cochlear processing (CSE; changes across 33 ERB-spaced filters from 26 to 7743 Hz) or a metric based on channel vocoding ($CSE_{CI}$; changes across a small number of spectral channels from 80 to 8000 Hz). Both metrics capture spectral changes across the full speech spectrum in vocoder-alone and hybrid speech, but given that the spectrum was divided into vocoder channels and was mostly degraded due to noise-vocoding, $CSE_{CI}$ was determined to be the more appropriate metric of informational changes. $CSE_{CI}$ was calculated using the approaches described by Stilp *et al.* (2013) and Stilp and Goupell (2015). Sentences were divided into 16-ms slices to measure spectral changes on a fine timescale. For each 16-ms slice, RMS amplitude was measured in each vocoder channel, converting this speech interval into a vector of amplitude values.[1] $CSE_{CI}$ was parameterized as the Euclidean distance between vectors of RMS amplitudes for all pairs of neighboring spectral slices. These measures were referenced to the spectral channel corner frequencies used in vocoded speech signals, disregarding the 500-Hz cutoff used to delimit the intact portion of the signal in the hybrid condition.

Euclidean distances were summed in boxcars of five successive slices (80 ms) then sorted into ascending (low $CSE_{CI}$) or descending (high $CSE_{CI}$) order.[2] The boxcar that ranked first (lowest or highest $CSE_{CI}$) was replaced by speech-shaped noise (white noise filtered by a 100th-order finite impulse response filter, resulting in a flat spectrum to 500 Hz and −9 dB/octave decrease above that point) that had 5-ms linear onset/offset ramps and was matched to mean sentence level. Eighty-millisecond intervals immediately before and after the replaced segments, and at the beginning of the sentence, were always left intact. The procedure proceeded iteratively to the next-highest-ranked boxcar, which was replaced only if its contents had not already been replaced or preserved. A total of four 80-ms intervals were replaced by noise in each sentence, following Stilp and Goupell (2015), as this amount of noise replacement produces significant impairment in sentence intelligibility without reaching floor levels of performance. This corresponded to an average of 13.64% of the sentence being replaced by noise. Continuous sentences with no noise replacement were also generated at all levels of spectral resolution for both vocoder and hybrid configurations.

TABLE I. Channel frequencies in vocoder processing. Center frequencies ("Center") and corner frequencies ("Corner") for each channel are listed in Hz. The top row indicates the channel number out of three (first set of rows), four (second set), five (third set), or six (fourth set).

| Channel | | One | | Two | | Three | | Four | | Five | | Six | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Center | | 275 | | 1250 | | 4388 | | | | | | | |
| Corners | 80 | | 624 | | 2373 | | 8000 | | | | | | |
| Center | | 215 | | 748 | | 2028 | | 5103 | | | | | |
| Corners | 80 | | 424 | | 1250 | | 3234 | | 8000 | | | | |
| Center | | 183 | | 537 | | 1250 | | 2688 | | 5586 | | | |
| Corners | 80 | | 329 | | 832 | | 1844 | | 3886 | | 8000 | | |
| Center | | 163 | | 424 | | 892 | | 1730 | | 3234 | | 5932 | |
| Corners | 80 | | 275 | | 624 | | 1250 | | 2373 | | 4388 | | 8000 |

## C. Procedure

### 1. Training

Listeners first participated in a familiarization task where they listened to one list of ten unmodified sentences to acquaint them with the talker and sentence complexity. All listeners heard the same sentences at familiarization, and no responses were collected.

Next, listeners completed training with feedback on 12 experimental conditions: three levels of $\text{CSE}_{\text{CI}}$ (replacing high-$\text{CSE}_{\text{CI}}$ intervals with noise, replacing low-$\text{CSE}_{\text{CI}}$ intervals with noise, and continuous sentences with no noise-replacement) that were fully crossed with two levels of EAS simulation (vocoder-only, hybrid) and two levels of spectral resolution (half of the listeners heard sentences with three or five spectral channels, the other half heard sentences with four or six channels). Spectral resolutions were blocked so that listeners heard all higher-resolution conditions first, then progressed to all lower-resolution conditions. Training consisted of one list of ten sentences per condition. Listeners heard each sentence only once and reported as much of the sentence as they could, with guessing encouraged. Following the presentation of each stimulus, the subject gave his/her verbal response, then pressed a key to display the written sentence on the computer monitor.

Listeners then completed baseline testing on a single experimental condition without feedback to determine whether their performance was adequate to proceed to the next portion of the experiment. Listeners heard two lists of vocoded sentences (20 total) without any noise replacement with either three or four spectral channels. Performance criteria were set to 25% correct (three channels) and 30% correct (four channels) to avoid floor effects in the noise replacement conditions. Twelve listeners did not meet criterion and did not participate further in the experiment, resulting in a final sample of 24 listeners (19 females and 5 males, average age 22.5 years). Different lists of IEEE sentences were used for training and baseline testing; however, list assignments were the same for each listener.

### 2. Testing

Testing was divided into two sessions. Each session was evenly divided between the lower (three or four spectral channels) and higher level of spectral resolution (five or six channels). Spectral resolution was tested as a between-subjects variable (12 listeners heard sentences with three and five spectral channels; 12 listeners heard sentences with four and six channels). EAS and $\text{CSE}_{\text{CI}}$ were fully crossed and tested as within-subjects variables, being counterbalanced across listeners to appear equally often at each block of testing. Each experimental condition was tested using four IEEE sentence lists: one list for practice without feedback (10 sentences) followed by three lists for testing (30 sentences). All listeners heard the same 48 IEEE sentence lists at testing, but the experimental condition in which each sentence appeared was counterbalanced as described above. Listeners heard each sentence only once during testing.

### 3. Protocol

Stimuli were presented from a personal computer with a Lynx L22 sound card (Lynx Studio Technology, Inc. Costa Mesa, CA) and a PA-5 programmable attenuator (Tucker Davis Technology, Alachua, FL), then presented diotically at 70 dB SPL over Sennheiser HD 600 headphones (Sennheiser Electronic GmbH and Co. KG, Germany) to the listener, who was seated in a double-walled sound room. Listeners verbally repeated all words they understood from each sentence. Responses were recorded using Adobe Audition, then later scored offline for keywords correct (five per sentence). The recorded responses were scored separately by two native-English speakers, with a third scorer serving as a tiebreaker when scores differed between the first two scorers.

## III. RESULTS

For each listening condition, mean percent-correct scores were calculated across the three testing blocks, then arcsine-transformed for statistical analysis (Studebaker, 1985) (Fig. 1). Mean scores ranged from 19.6% [19.6 rationalized arcsine units (RAU)] for the three-channel high-$\text{CSE}_{\text{CI}}$ vocoder-only condition to 89.2% (91.9 RAU) for the six-channel continuous hybrid condition. Results were analyzed using a three-way mixed analysis of variance (ANOVA). Within-subject factors were noise replacement (three levels: high-$\text{CSE}_{\text{CI}}$ intervals replaced by noise, low-$\text{CSE}_{\text{CI}}$ intervals replaced by noise, continuous sentences with no noise replacement), listening mode (two levels: vocoder-only, hybrid), and spectral resolution [two levels broadly classified as "higher" (five or six spectral channels) versus "lower" resolution (three or four spectral channels), depending on which conditions each

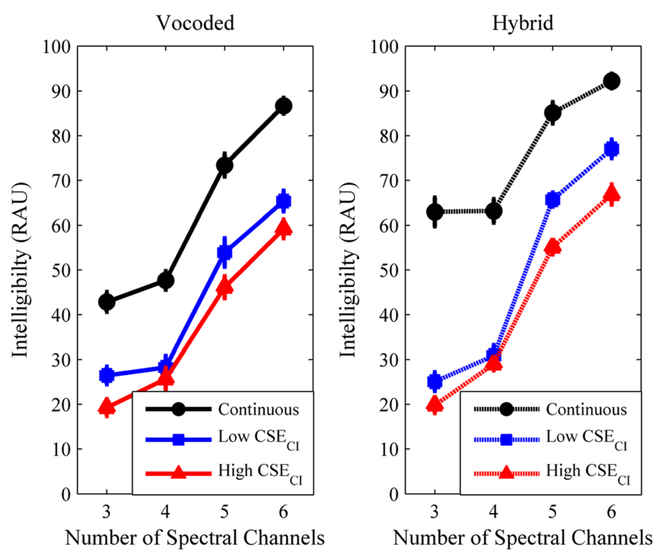


FIG. 1. (Color online) Mean sentence intelligibility (as measured in RAU) as a function of the number of spectral channels. Each line depicts one level of noise replacement: circles represent continuous (uninterrupted) sentences, squares represent sentences with low-$\text{CSE}_{\text{CI}}$ intervals replaced by noise, and triangles represent sentences with high-$\text{CSE}_{\text{CI}}$ intervals replaced by noise. Results for vocoder-alone sentences are portrayed at left in solid lines; results for hybrid sentences are portrayed at right in dashed lines. Error bars indicate standard error of the mean.

listener group heard]. The between-subjects variable was listener group, sorting listeners according to whether they heard sentences with three and five spectral channels, or heard sentences with four and six channels. Results showed significant main effects of noise replacement ($F_{2,44} = 427.20$, $p < 0.001$, $\eta_p^2 = 0.95$), listening mode ($F_{1,22} = 68.98$, $p < 0.001$, $\eta_p^2 = 0.76$), and spectral resolution ($F_{1,22} = 1165.76$, $p < 0.001$, $\eta_p^2 = 0.98$). Results also showed significant interactions between noise replacement and listening mode ($F_{2,44} = 11.85$, $p < 0.001$, $\eta_p^2 = 0.35$), noise replacement and spectral resolution ($F_{2,44} = 11.60$, $p < 0.001$, $\eta_p^2 = 0.35$), and the three-way interaction between noise replacement, listening mode, and spectral resolution ($F_{2,44} = 19.71$, $p < 0.001$, $\eta_p^2 = 0.47$). No other interactions were statistically significant. Given the large scale of the omnibus analysis, these results will be interpreted in turn below.

### A. Spectral resolution

Averaging across other factors (three levels of noise replacement, two levels of listening mode, and the two listener groups), performance varied as a function of spectral resolution. However, this main effect in the omnibus ANOVA collapses across listener groups, making this a fairly coarse comparison of performance with more spectral channels (five and six) versus fewer spectral channels (three and four). The influence of spectral resolution was better revealed by the significant interaction between the within-subjects factor of number of spectral channels and the between-subjects factor of listener group ($F_{1,22} = 11.14$, $p < 0.01$, $\eta_p^2 = 0.34$). This interaction was further investigated using independent-samples *t*-tests with Bonferroni correction for multiple comparisons ($\alpha = 0.05/3 = 0.0167$). Performance did not significantly differ across three-channel [mean = 32.73 RAU, standard error (SE) = 1.63] and four-channel sentences (mean = 37.44, SE = 1.46; $t_{22} = 2.15$, $p = 0.04$), but there was significant improvement in changing from four spectral channels to five (mean = 63.22, SE = 1.45; $t_{22} = 12.53$, $p < 0.001$), and changing from five channels to six (mean = 74.55, SE = 0.85; $t_{22} = 6.75$, $p < 0.001$). Given that listening mode and noise replacement were the primary effects of interest, interactions between spectral resolution and these factors are described below.

### B. EAS listening mode

Figure 1 displays performance as a function of number of spectral channels for vocoder-only (left) and hybrid sentences (right). Averaging across other factors (three levels of noise replacement, two levels of spectral resolution, and the two listener groups), sentence intelligibility was significantly higher for hybrid sentences (mean = 56.07 RAU, SE = 0.86) than vocoder-alone sentences (mean = 47.90, SE = 1.07), consistent with prior research. However, this EAS benefit varied across noise replacement conditions, as supported by the significant interaction between these factors. This relationship was explored by calculating EAS benefit as the RAU score difference between hybrid and vocoder-alone conditions for each of the three levels of noise replacement. Figure 2 illustrates the significant three-way interaction between listening mode, noise replacement, and spectral resolution. For continuous speech, EAS benefits are larger for sentences with fewer (three to four) spectral channels than sentences with more spectral channels (five to six). Conversely, for $CSE_{CI}$ conditions, EAS benefit showed the opposite pattern by being larger for sentences with more (five to six) spectral channels than sentences with fewer spectral channels (three to four).

### C. Effects of listening mode on noise replacement

As expected, speech intelligibility decreased significantly when moving from continuous speech to noise-interrupted speech. The degree of decrement, however, varied depending on the spectral channels and listening mode, promoting closer investigation of the effects of noise replacement. Figure 3 shows the decrement in performance (RAU decrease relative to continuous speech) for noise-interrupted sentences in each channel condition and listening mode. Following Stilp and Goupell (2015), these decrements serve as a proxy for the perceptual importance of speech intervals that were replaced by noise for sentence understanding. As seen in this figure, there was a differential effect of noise replacement on vocoder-alone versus hybrid speech. First, the effect of noise replacement was relatively constant across spectral channels for vocoder-alone speech. However, noise replacement impaired intelligibility of hybrid speech to a greater extent for three- and four-channel sentences, and by progressively smaller amounts as the number of spectral channels increased. Finally, decrements were significantly larger for hybrid sentences than vocoder-alone sentences at three spectral channels (paired-samples *t*-tests at $\alpha = 0.05/2 = 0.025$ for multiple comparisons at each level of spectral resolution; $t_{11} > 6.53$, $p < 0.001$) and four channels ($t_{11} > 2.85$, $p < 0.02$), but decrements were not different at

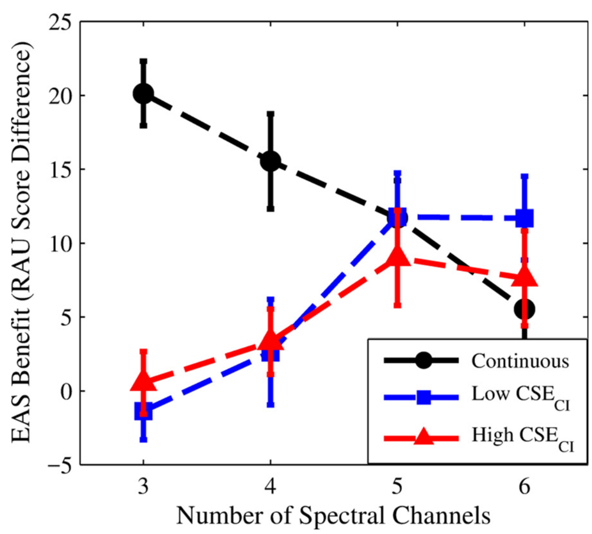


FIG. 2. (Color online) Mean EAS benefit (measured by RAU score difference across hybrid and vocoder-alone sentence conditions) as a function of number of spectral channels. Each line depicts one level of noise replacement: circles represent continuous (uninterrupted) sentences, squares represent sentences with low-$CSE_{CI}$ intervals replaced by noise, and triangles represent sentences with high-$CSE_{CI}$ intervals replaced by noise. Error bars indicate standard error of the mean.

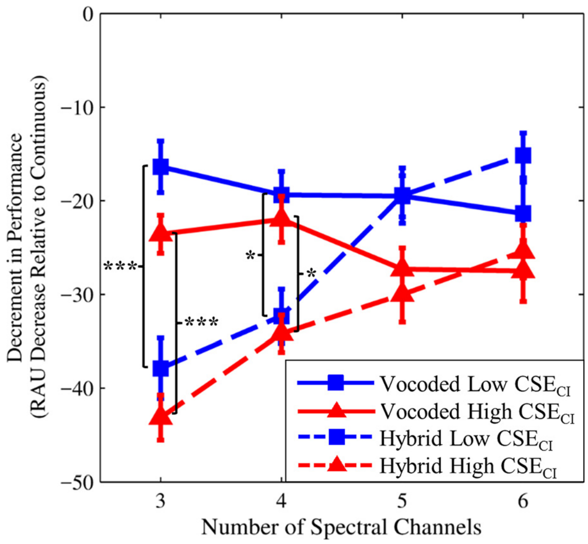


FIG. 3. (Color online) Mean decreases in performance owing to noise replacement in different listening modes and numbers of spectral channels. These measures serve as a proxy for the perceptual importance of speech intervals that were replaced by noise for sentence understanding (see Stilp and Goupell, 2015). Performance decrements are shown as RAU decreases relative to performance with continuous sentences (similar patterns are observed when calculated as normalized percentage point decrease). Squares represent sentences with low-CSE$_{CI}$ intervals replaced by noise, and triangles represent sentences with high-CSE$_{CI}$ intervals replaced by noise. Solid lines represent vocoded sentences while dashed lines represent hybrid sentences. Error bars indicate standard error of the mean. * $p < 0.05$, *** $p < 0.001$.

five channels ($t_{11} < 0.94$, $p > 0.37$) or six channels ($t_{11} < 1.31$, $p > 0.21$).

### D. Information-bearing acoustic changes

The data from Fig. 1 are replotted in Fig. 4 to illustrate effects of information-bearing acoustic changes on performance. The original omnibus ANOVA analyzed three levels of noise replacement: two levels of actual noise replacement (low CSE$_{CI}$, high CSE$_{CI}$) and one level without any noise replacement (continuous sentences). This obscures investigations of differences among conditions where sentence intervals were replaced by noise (i.e., the influence of information-bearing acoustic changes on sentence intelligibility). Therefore, a second omnibus ANOVA was conducted of the same structure as the first except noise replacement consisted of only two levels: low CSE$_{CI}$ and high CSE$_{CI}$. All main effects were still statistically significant (listening mode: $F_{1,22} = 20.65$, $p < 0.001$, $\eta_p^2 = 0.48$; spectral resolution: $F_{1,22} = 1745.48$, $p < 0.001$, $\eta_p^2 = 0.99$) including information-bearing acoustic changes ($F_{1,22} = 48.09$, $p < 0.001$, $\eta_p^2 = 0.69$). This result indicates that sentences with high-CSE$_{CI}$ intervals replaced by noise were understood more poorly than sentences with low-CSE$_{CI}$ intervals replaced, replicating past investigations measuring the intelligibility of noise-vocoded speech (Stilp *et al.*, 2013; Stilp, 2014; Stilp and Goupell, 2015).

There was a significant interaction between information-bearing acoustic changes and spectral resolution ($F_{1,22} = 8.79$, $p < 0.01$, $\eta_p^2 = 0.29$). Given that the omnibus ANOVA collapses across the two listener groups, this is again a coarse comparison of "higher" spectral resolutions (five and six channels) versus "lower" spectral resolutions (three and four channels). To examine these results at different levels of spectral resolution, a paired-sample *t*-test was conducted to compare the high- and low-CSE$_{CI}$ scores obtained at each channel condition, averaged across vocoder and hybrid materials. Information-bearing acoustic changes influenced the intelligibility of sentences with three spectral channels (intelligibility difference of 6.22 RAU worse when high-CSE$_{CI}$ intervals were replaced compared to when low-CSE$_{CI}$ intervals were replaced; $t_{11} = 6.12$, $p < 0.001$), five-channel sentences (difference of 9.19 RAU; $t_{11} = 5.08$, $p < 0.001$), and six-channel sentences (difference of 8.20 RAU; $t_{11} = 3.56$, $p < 0.01$), but not four-channel sentences (difference of 2.22 RAU; $t_{11} = 1.62$, $p = 0.13$).

### E. Relationship between information-bearing acoustic changes and listening mode

Finally, of key theoretical interest is whether listeners utilized information-bearing acoustic changes differently to understand vocoded versus hybrid sentences. In the omnibus ANOVA reported in Sec. III D, the interaction between

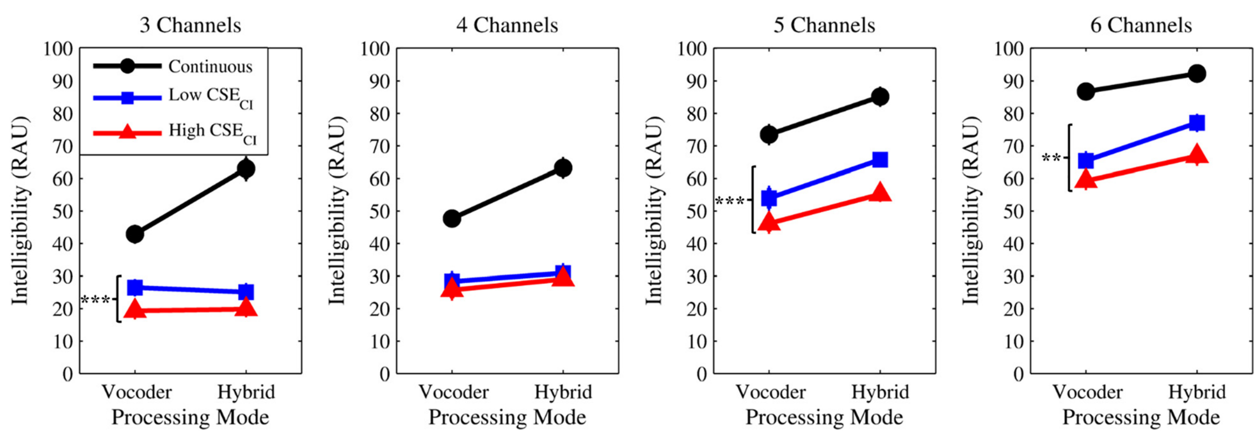


FIG. 4. (Color online) Mean sentence intelligibility (measured in RAU) depicted as a function of processing mode (vocoder, hybrid). Results are the same as those shown in Fig. 1 but rearranged to highlight differences in performance at each number of spectral channels. Each panel illustrates mean performance for the three noise-replacement conditions: circles represent continuous (uninterrupted) sentences, squares represent sentences with low-CSE$_{CI}$ intervals replaced by noise, and triangles represent sentences with high-CSE$_{CI}$ intervals replaced by noise. Error bars depict one standard error of the mean. ** $p < 0.01$, *** $p < 0.001$.

information-bearing acoustic changes and listening mode was not statistically significant ($F_{1,22} = 0.41$, $p = 0.53$). Thus, there was no statistical evidence that listeners relied on information-bearing acoustic changes differently to understand vocoder-alone and hybrid sentences (i.e., parallel lines in Fig. 3 for vocoder and hybrid conditions).

## IV. DISCUSSION

Information-bearing acoustic changes, spectral resolution, and listening mode have each been shown to influence speech intelligibility individually, and in some cases jointly (spectral resolution and listening mode: Kong *et al.*, 2015; spectral resolution and information-bearing acoustic changes: Stilp and Goupell, 2015). The present experiment investigated relationships between factors to answer two questions: the degree to which listeners rely on information-bearing acoustic changes to understand interrupted hybrid versus vocoded sentences, and how noise interruption affects the intelligibility of hybrid sentences versus vocoded sentences more broadly.

While listeners did show a significant effect of information-bearing acoustic changes (poorer sentence intelligibility when high-$CSE_{CI}$ intervals were replaced by noise compared to when low-$CSE_{CI}$ intervals were replaced, replicating Stilp *et al.*, 2013; Stilp, 2014; Stilp and Goupell, 2015), this effect did not differ across hybrid and vocoded listening modes. This outcome may be explained by a similarity in the low-frequency information-bearing acoustic changes that exist in speech materials for both modes. Stilp *et al.* (2013) found that measures of information-bearing acoustic change in full-spectrum (CSE) and noise-vocoded speech ($CSE_{CI}$) were fairly consistent in identifying the same perceptually significant sentence intervals. From this result, Stilp *et al.* conjectured that these measures of information-bearing acoustic change were more sensitive to the amplitude envelope than temporal fine structure, given the preservation of envelope information in channel vocoding. As such, changes in the envelope (i.e., the basis for CSE and $CSE_{CI}$) were suggested to be very similar as well. The present experiment provided a direct test of this suggestion, as speech information below 500 Hz in hybrid and vocoder-only materials had similar envelopes but very different temporal fine structures. While similar effects of $CSE_{CI}$ across vocoded and hybrid conditions might be due to similarity in low-frequency envelope changes, it bears mention that spectra above 500 Hz were identical in hybrid and vocoded speech, resulting in highly correlated measures of information-bearing acoustic changes across materials. Additionally, $CSE_{CI}$ was calculated using the same formula across both sets of materials for the sake of consistency. As a result, speech information in the low-frequency channel was not weighted differently than any other channel, contributing equally to the broadband measures of spectral change. While listeners are clearly aided by the low-frequency channel containing fine structure information (i.e., EAS benefit), it is unclear whether this alters the nature of information-bearing acoustic changes in this frequency region. Listeners might increase their reliance on low-frequency information-bearing acoustic changes to understand hybrid speech if these changes were weighted more heavily in $CSE_{CI}$ calculations, but further research is needed to test this possibility.

The combination of vocoded and low-pass-filtered speech is generally more intelligible than vocoded-only speech. However, an understanding of how this EAS benefit helps listeners combat interruptions in the speech signal is incomplete. The present results reveal a complex interaction of the EAS benefit with noise interruption and spectral resolution. For continuous sentences, the EAS benefit was largest for three-channel sentences and steadily decreased as more spectral channels were added. Fine-structure cues in the low-frequency channel were highly beneficial at lower spectral resolutions (given the relative paucity of other available cues for understanding such degraded sentences), resulting in large EAS benefits. The decrease in EAS benefits at higher spectral resolution is likely due to the elevation of baseline performance with vocoded speech. For noise-interrupted sentences, however, the opposite trend was observed. EAS benefits were eliminated for sentences with few spectral channels, and grew modestly as more spectral channels were added. The same pattern was observed when low- and high-$CSE_{CI}$ intervals were replaced by noise, so this pattern is specific to noise interruption and not information-bearing acoustic changes. This outcome suggests that noise disrupts the mechanisms that support EAS benefit, such as periodicity and harmonicity cues that help mark syllable and word boundaries (e.g., Spitzer *et al.*, 2009; Kong *et al.*, 2015) and is consistent with the notion that speech segmentation cues may take on greater importance for more severely degraded speech. That is, interrupting the continuity of harmonicity cues with noise disrupts segmentation and impairs sentence intelligibility, bringing hybrid sentence intelligibly down to vocoder-alone levels in lower spectral resolution conditions (Figs. 1 and 4) (see also Oh *et al.*, 2016a,b).

The interaction observed here between EAS benefit and spectral resolution contrasts with previous published reports on the intelligibility of interrupted hybrid sentences. Başkent (2012; Başkent and Chatterjee, 2010) presented listeners with hybrid sentences interrupted by 50% duty-cycle noise at an interruption rate of 1.5 Hz (333 ms period). EAS benefits in these studies were largest for sentences with few spectral channels and smaller as spectral resolution improved (similar to continuous sentences both here and in her reports, but opposite to noise-interrupted sentences in the present study). However, several significant methodological points differentiate these studies. First, the earlier studies employed 50% duty-cycle noise to replace half of the sentence with noise in a regular manner, while the present experiment employed four 80-ms interruptions that were not bound to occur with any such temporal regularity. Second, the two earlier studies examined a wide range of spectral resolutions (4, 8, 16, 32 channels) while the present study more densely sampled the lower end of spectral resolution (3, 4, 5, 6 channels). Third, the earlier studies used relatively high-context sentences whereas the present experiment used more challenging IEEE sentences. Fourth, studies by Başkent (2012; Başkent and Chatterjee, 2010) trained listeners on only some conditions and provided no feedback, while participants in the present experiment received training and feedback on every experimental condition. While these methodological differences make direct comparisons of the sizes of EAS benefits difficult, it is noteworthy that the results differ widely in their direction (magnitude of EAS benefit across different spectral resolutions). While interactions between the nature of noise interruption and the magnitudes of EAS benefits have yet to be explored systematically, the present results encourage such investigations as they appear to be more complex than previously considered.

The present results additionally shed light on the intelligibility of highly degraded speech. At moderate to poor levels of signal quality, speech intelligibility can be aided by the addition of a variety of factors, but these same factors might not provide any benefit if speech is further degraded. For example, Başkent (2012) reported phonemic restoration effects at a wide range of spectral resolutions in vocoded and EAS speech, but four-channel sentences showed no restoration effects. Similarly, vocoded sentences are more intelligible when high-$CSE_{CI}$ changes are available (low-$CSE_{CI}$ changes replaced by noise) than when low-$CSE_{CI}$ changes are available (high-$CSE_{CI}$ changes replaced by noise); however Stilp and Goupell (2015) reported that differences observed across a wide range of spectral resolutions (6 to 24 channels) were absent for four-channel vocoding. The present results follow a similar pattern: the addition of low-frequency speech information improves intelligibility of continuous sentences with as few as three spectral channels, but provides no benefit when sentences are further degraded through noise interruption (Fig. 2). Taken together with previous results, this finding suggests a lower limit for when factors that are known to improve speech intelligibility actually do so: beyond this limit, task difficulty is relatively extreme and performance is no longer aided.

The present report adds to demonstrations of information-bearing acoustic changes contributing to normal-hearing listeners' speech perception (Stilp and Kluender, 2010; Stilp *et al.*, 2013; Stilp, 2014; Stilp and Goupell, 2015), but it is important to note that such informational changes are utilized by listeners with impaired hearing as well. While sensorineural hearing loss impairs the ability to detect rapid acoustic changes in narrow frequency regions (e.g., formant transitions), listeners still utilize rapid changes in broad spectral shape to identify speech sounds (Alexander and Kluender, 2009). Similar to normal hearing listeners, CI users exhibit enhancement effects where acoustic changes are perceptually emphasized (Goupell and Mostardi, 2012; Wang *et al.*, 2012). What these diverse listening populations have in common is an auditory system that is primarily sensitive to changes in the input (Kluender *et al.*, 2003). While encoding at the sensory periphery may differ, the rest of the auditory system is similarly designed to be optimally sensitive to changes in the input. Thus, normal-hearing listeners' utilization of information-bearing acoustic changes to understand hybrid speech is predicted to generalize to CI users who receive EAS speech inputs. An age difference exists between listeners in the present report and typically older CI users, which might result in decreased reliance on information-bearing acoustic changes due to poorer temporal resolution (e.g., Frisina *et al.*, 2001) but might not affect integrating bimodal inputs (Dorman *et al.*, 2012). An important outcome of the present research is the observation that complex interactions exist between listening mode, signal interruptions, and spectral resolution that can differentially alter speech intelligibility. Future research should consider the implications of these complex relationships on speech perception by listeners with impaired hearing.


## ACKNOWLEDGMENTS

This work was supported by National Institutes of Health–National Institute on Deafness and Other Communication Disorders DC012300 (Y.Y.K., principal investigator). The authors thank Craig Spandau for assistance with data collection.


[1]Euclidean distances between vectors of RMS-amplitude values form a unitless metric of signal change. Previous research indicates that this metric is an excellent correlate of psychological distance, as various pairs of novel complex sounds with equal spectral differences (separated by equal CSE) were discriminated equally well (Stilp *et al.*, 2010). Additionally, highly comparable patterns of noise-interrupted sentence intelligibility were observed when sentence intervals were replaced on the basis of spectral changes using linear (RMS Volts), logarithmic (dB), or loudness-based (sones) scales of amplitude (Stilp, 2011).

[2]Note that information-bearing acoustic changes (as measured by CSE) cannot be equated with changes in signal amplitude. Chen and Loizou (2012) directly compared the intelligibility of noise-corrupted sentences with measures of CSE and RMS amplitude. Correlations between sentence recognition and speech intelligibility indices were higher for high CSE compared to low CSE, consistent with depictions of CSE as reflecting potential information for speech perception (Stilp and Kluender, 2010; Kluender *et al.*, 2013). As for measures of RMS amplitude, speech intelligibility indices exhibited the highest correlations with behavior for medium-amplitude measures of speech. Many of these medium-amplitude regions mark the transitions between low-amplitude and high-amplitude regions (i.e., C/V transitions), and these rate as high-change intervals. For example, a CVC syllable such as "deed" exhibits an inverted-U shape for its amplitude profile, with lower intensity levels at consonantal onset and offset but an extended peak for the vowel nucleus. A CSE profile of this syllable exhibits an M shape. There are large spectral changes at the transitions from C to V (first peak) and V to C (second peak). During the vowel nucleus, however, far less spectral change is occurring, leading to a sharp decrease in CSE that lasts until the vowel transitions into the final consonant.


Alexander, J. M., and Kluender, K. R. (**2009**). "Relativity of spectral tilt change in stop consonant perception by hearing-impaired listeners," J. Speech Hear. Res. **52**, 653–670.

Başkent, D. (**2012**). "Effect of speech degradation on top-down repair: Phonemic restoration with simulations of cochlear implants and combined electric–acoustic stimulation," JARO **13**(5), 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**(1-2), 127–133.

Benard, M. R., and Başkent, D. (**2015**). "The effect of visual cues on top-down restoration of temporally interrupted speech, with and without further degradations," Hear. Res. **328**, 24–33.

Bhargava, P., Gaudrain, E., and Başkent, D. (**2014**). "Top-down restoration of speech in cochlear-implant users," Hear. Res. **309**, 113–123.

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

Brown, C. A., and Bacon, S. P. (**2009b**). "Achieving electric-acoustic benefit with a modulated tone," Ear Hear. **30**(5), 489–493.

Brown, C. A., and Bacon, S. P. (**2010**). "Fundamental frequency and speech intelligibility in background noise," Hear. Res. **266**(1), 52–59.

Chatterjee, M., Peredo, F., Nelson, D., and Başkent, D. (**2010**). "Recognition of interrupted sentences under conditions of spectral degradation," J. Acoust. Soc. Am. **127**(2), EL37–EL41.

Chen, F., and Loizou, P. C. (**2012**). "Contributions of cochlea-scaled entropy and consonant-vowel boundaries to prediction of speech intelligibility in noise," J. Acoust. Soc. Am. **131**(5), 4104–4113.

Cole, R. A., Mak, B., Fanty, M., and Bailey, T. (**1996**). "The contribution of consonants versus vowels to word recognition in fluent speech," Int. Conf. Ac. Speech Signal Proc. **2**, 853–856.

Dorman, M. F., Cook, S., Spahr, A., Zhang, T., Loiselle, L., Schramm, D., and Gifford, R. (**2015**). "Factors constraining the benefit to speech understanding of combining information from low-frequency hearing and a cochlear implant," Hear. Res. **322**, 107–111.

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

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

Dorman, M. F., Spahr, A., Gifford, R. H., Cook, S., Zhanga, T., Loiselle, L., Whittingham, J., and Schramm, D. (**2012**). "Bilateral and bimodal benefits as a function of age for adults fitted with a cochlear implant," J. Hear. Sci. **2**(4), EA37–39.

Fogerty, D., Ahlstrom, J. B., Bologna, W. J., and Dubno, J. R. (**2015**). "Sentence intelligibility during segmental interruption and masking by speech-modulated noise: Effects of age and hearing loss," J. Acoust. Soc. Am. **137**(6), 3487–3501.

Frisina, D. R., Frisina, R. D., Snell, K. B., Burkard, R., Walton, J. P., and Ison, J. R. (**2001**). "Auditory temporal processing during aging," in *Functional Neurobiology of Aging*, edited by P. R. Hof and C. V. Mobbs (Academic Press, San Diego, CA), pp. 565–579.

Gifford, R. H., Dorman, M. F., McKarns, S. A., and Spahr, A. J. (**2007**). "Combined electric and contralateral acoustic hearing: Word and sentence recognition with bimodal hearing," J. Speech Lang. Hear. Res. **50**(4), 835–843.

Goupell, M. J., and Mostardi, M. J. (**2012**). "Evidence of the enhancement effect in electrical stimulation via electrode matching," J. Acoust. Soc. Am. **131**(2), 1007–1010.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species- 29 years later," J. Acoust. Soc. Am. **87**(6), 2592–2606.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized/hVd/utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**(6), 3509–3523.

Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (**2007**). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **122**(4), 2365–2375.

Kluender, K. R., and Alexander, J. M. (**2008**). "Perception of speech sounds," in *The Senses: A Comprehensive Reference, Vol.3, Audition*, edited by P. Dallos and D. Oertel (Academic Press, San Diego, CA), pp. 829–860.

Kluender, K. R., Coady, J. A., and Kiefte, M. (**2003**). "Sensitivity to change in perception of speech," Speech Comm. **41**(1), 59–69.

Kluender, K. R., Stilp, C. E., and Kiefte, M. (**2013**). "Perception of vowel sounds within a biologically realistic model of efficient coding," in *Vowel Inherent Spectral Change*, edited by G. Morrison and P. Assmann (Springer, Berlin, Germany), pp. 117–151.

Kong, Y. Y., and Braida, L. D. (**2011**). "Cross-frequency integration for consonant and vowel identification in bimodal hearing," J. Speech Lang. Hear. Res. **54**(3), 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**(6), 3717–3727.

Kong, Y. Y., Donaldson, G., and Somarowthu, A. (**2015**). "Effects of contextual cues on speech recognition in simulated electric-acoustic stimulation," J. Acoust. Soc. Am. **137**(5), 2846–2857.

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

Middlebrooks, J. C., and Snyder, R. L. (**2010**). "Selective electrical stimulation of the auditory nerve activates a pathway specialized for high temporal acuity," J. Neurosci. **30**(5), 1937–1946.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**(2), 167–173.

Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**(5), 1297–1308.

Oh, S., Donaldson, G. S., and Kong, Y. Y. (**2016a**). "Top-down processes in simulated electric-acoustic hearing: The effect of linguistic context on bimodal benefit for temporally interrupted speech," Ear Hear. **37**, 582–592.

Oh, S., Donaldson, G. S., and Kong, Y. Y. (**2016b**). "The role of continuous low-frequency harmonicity cues for interrupted speech perception in bimodal hearing," J. Acoust. Soc. Am. **139**(4), 1747–1755.

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**(5), 1421–1425.

Powers, G. L., and Wilcox, J. C. (**1977**). "Intelligibility of temporally interrupted speech with and without intervening noise," J. Acoust. Soc. Am. **61**(1), 195–199.

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

Shafiro, V., Sheft, S., and Risley, R. (**2011**). "Perception of interrupted speech: Effects of dual-rate gating on the intelligibility of words and sentences," J. Acoust. Soc. Am. **130**(4), 2076–2087.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature **416**(6876), 87–90.

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**(6), EL236–EL241.

Stilp, C. E. (**2011**). "Efficient coding in perception of complex sounds." Doctoral dissertation, University of Wisconsin - Madison, Madison, WI.

Stilp, C. E. (**2014**). "Information-bearing acoustic change outperforms duration in predicting intelligibility of full-spectrum and noise-vocoded sentences," J. Acoust. Soc. Am. **135**(3), 1518–1529.

Stilp, C. E., and Goupell, M. J. (**2015**). "Spectral and temporal resolutions of information-bearing acoustic changes for understanding vocoded sentences," J. Acoust. Soc. Am. **137**(2), 844–855.

Stilp, C. E., Goupell, M. J., and Kluender, K. R. (**2013**). "Speech perception in simulated electric hearing exploits information-bearing acoustic change," J. Acoust. Soc. Am. **133**(2), EL136–EL141.

Stilp, C. E., and Kluender, K. R. (**2010**). "Cochlea-scaled spectral entropy, not consonants, vowels, or time, best predicts speech intelligibility," Proc. Natl. Acad. Sci. U.S.A. **107**(27), 12387–12392.

Stilp, C. E., Rogers, T. T., and Kluender, K. R., (**2010**). "Rapid efficient coding of correlated complex acoustic properties," Proc. Natl. Acad. Sci. U.S.A. **107**(50), 21914–21919

Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**(4), 1872–1891.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**(3), 455–462.

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

Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (**2002**). "Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization," Ear Hear. **23**(2), 98–105.

Wang, N., Kreft, H., and Oxenham, A. J. (**2012**). "Vowel enhancement effects in cochlear-implant users," J. Acoust. Soc. Am. **131**(6), EL421–EL426.

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**(1), 63–69.