Spectral and temporal resolutions of information-bearing acoustic changes for understanding vocoded sentences[a)]

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

Matthew J. Goupell
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

(Received 23 May 2014; revised 12 December 2014; accepted 27 December 2014)

Short-time spectral changes in the speech signal are important for understanding noise-vocoded sentences. These information-bearing acoustic changes, measured using cochlea-scaled entropy in cochlear implant simulations [CSE_{CI}; Stilp *et al.* (**2013**). J. Acoust. Soc. Am. **133**(2), EL136–EL141; Stilp (**2014**). J. Acoust. Soc. Am. **135**(3), 1518–1529], may offer better understanding of speech perception by cochlear implant (CI) users. However, perceptual importance of CSE_{CI} for normal-hearing listeners was tested at only one spectral resolution and one temporal resolution, limiting generalizability of results to CI users. Here, experiments investigated the importance of these informational changes for understanding noise-vocoded sentences at different spectral resolutions (4–24 spectral channels; Experiment 1), temporal resolutions (4–64 Hz cutoff for low-pass filters that extracted amplitude envelopes; Experiment 2), or when both parameters varied (6–12 channels, 8–32 Hz; Experiment 3). Sentence intelligibility was reduced more by replacing high-CSE_{CI} intervals with noise than replacing low-CSE_{CI} intervals, but only when sentences had sufficient spectral and/or temporal resolution. High-CSE_{CI} intervals were more important for speech understanding as spectral resolution worsened and temporal resolution improved. Trade-offs between CSE_{CI} and intermediate spectral and temporal resolutions were minimal. These results suggest that signal processing strategies that emphasize information-bearing acoustic changes in speech may improve speech perception for CI users.
[http://dx.doi.org/10.1121/1.4906179]

[DB] Pages: 844–855

I. INTRODUCTION

Rapid changes in the acoustic spectrum play an important role in speech perception. Spectral changes between the offset of a precursor sound and onset of a subsequent target sound are perceptually emphasized. This process produces spectral contrast effects that influence speech identification (e.g., Lotto and Kluender, 1998; Holt, 2005; Alexander and Kluender, 2008; Kingston *et al.*, 2014). Contrast effects are also observed when spectral properties in precursor sounds are reliable or predictable across time, then change upon introduction of the subsequent speech target (e.g., Ladefoged and Broadbent, 1957; Watkins, 1991; Sjerps *et al.*, 2011; Laing *et al.*, 2012; see Stilp *et al.*, 2010a for similar results with non-speech sounds). When spectral properties do not change and are instead predictable across both precursor and target sounds, the auditory system deemphasizes them and increases its reliance on changing (thus, more informative) properties (Darwin *et al.*, 1989; Kiefte and Kluender, 2008; Alexander and Kluender, 2010; Stilp and Anderson, 2014). These studies demonstrate that speech sound identification is clearly influenced by changes (or lack thereof) in the acoustic input.

Information-theoretic approaches to speech perception (Kluender and Alexander, 2007; Kluender *et al.*, 2013) argue for the perceptual importance of acoustic changes in the speech signal. These approaches stem from Shannon information theory (Shannon, 1948), in which unpredictability or change conveys potential information to the receiver (here, the perceiver), while predictability or stasis conveys no new information. Stilp and Kluender (2010) tested the degree to which sentence intelligibility directly relied upon temporal intervals containing highly or minimally informational acoustic changes. They developed a metric of information-bearing acoustic change named cochlea-scaled spectral entropy (CSE). When successive spectra were dissimilar to each other (i.e., high CSE), they were predicted to be highly important for sentence understanding; successive spectra that were similar (i.e., low CSE) were predicted to be less important for perception. Replacing high-CSE intervals with noise resulted in poorer sentence intelligibility than replacing an equal number of low-CSE intervals, establishing the prominent role of information-bearing acoustic changes for sentence intelligibility.

Cochlear implant (CI) processing maintains speech envelope information (albeit using fewer spectral channels than are available in someone with healthy and normal hearing). CSE measures changes in the spectral envelope of

[a)]Portions of this research were presented at the 37th MidWinter Meeting of the Association for Research in Otolaryngology and the 167th Meeting of the Acoustical Society of America.
[b)]Author to whom correspondence should be addressed. Electronic mail: christian.stilp@louisville.edu

speech, suggesting information-bearing acoustic changes will also be available for speech perception in CI processing. Stilp and colleagues (Stilp *et al.*, 2013; Stilp, 2014) adapted CSE to measure information-bearing acoustic changes in noise-vocoded speech (CSE_{CI}). CSE and CSE_{CI} both make the same prediction for perceptual performance: replacing highly informational sentence intervals (based on CSE or CSE_{CI}) with noise would produce poorer sentence intelligibility than replacing an equal number of minimally informational intervals. Performance confirmed this prediction, revealing the importance of information-bearing acoustic changes for perception of full-spectrum and noise-vocoded sentences.

With CSE and CSE_{CI} capturing aspects of the speech signal that are important for understanding full-spectrum and noise-vocoded speech, Stilp and colleagues (2013) suggested that information-bearing acoustic changes are fundamental to speech perception most broadly. Generalizing results from Stilp and Kluender (2010) to other normal-hearing listeners may be reasonable, but generalizing results for materials that approximate spectral degradation of CI processing may be problematic. No two CI users are alike, given differences in neural survival, insertion depth, duration of deafness, age at implantation, and a host of other factors. CI users' perceptual performance exhibits high intersubject variability, and reasons for this variability are not fully understood (e.g., Blamey *et al.*, 2013). Further, Stilp *et al.* (2013) and Stilp (2014) used a single set of vocoder parameters to define spectral (eight spectral channels) and temporal resolutions (channels low-pass filtered at 150 Hz for extracting amplitude envelopes) of speech. Separate and joint effects of spectral and temporal resolutions on intelligibility of vocoded speech are well known (Xu *et al.*, 2002, 2005; Xu and Zheng, 2007), and these results have been very useful for better understanding CI users' speech perception.

Intelligibility of vocoded speech generally increases with the number of available spectral channels. While estimates vary slightly across studies,[1] intelligibility of sentences presented in quiet asymptotes at 6–8 channels (Dorman and Loizou, 1997; Loizou *et al.*, 1999; Friesen *et al.*, 2001; Goupell *et al.*, 2008). Presenting sentences in background noise increases the point at which the asymptote occurs, with estimates depending on the signal-to-noise ratio and characteristics of the noise, such as envelope modulations (Friesen *et al.*, 2001; Qin and Oxenham, 2003; Shannon *et al.*, 2004). The noise need not be continuous to impair understanding of vocoded speech. Intelligibility of sentences with segments replaced by noise at periodic intervals improves with increasing spectral resolution up to (and possibly beyond) 32 channels (Nelson and Jin, 2004; Başkent, 2012).

Stilp (2014) proposed that information-bearing acoustic changes become more important for speech intelligibility in worse listening conditions. Replacing low- or high-CSE_{CI} changes with noise produced larger decrements in intelligibility of eight-channel noise-vocoded sentences than full-spectrum sentences (Stilp, 2014). However, these results might be influenced by ceiling effects. Smaller decrements in intelligibility might have been due to the high intelligibility of full-spectrum sentences [mean = 108 rationalized arcsine units (RAU; Studebaker, 1985) or 96% for sentences without noise interruption, mean = 80 RAU or 80% for uninterrupted noise-vocoded sentences] rather than information-bearing acoustic changes being relatively less important for understanding these materials. Comparisons of speech intelligibility across spectrally rich and degraded materials are highly sensitive to such factors. Comparing intelligibility of similar speech materials, such as different instances of noise-vocoded speech, would provide a more sensitive measure of the importance of information-bearing acoustic changes. Nelson and Jin (2004) proposed greater perceptual importance is attributed to envelope cues as spectral resolution decreases, and CSE_{CI} more closely reflects envelope cues than temporal fine structure (Stilp *et al.*, 2013). Thus, these informational changes may become more important for sentence intelligibility as the number of spectral channels in vocoded speech decreases. Such comparisons would be freer from potential ceiling effects than those in Stilp (2014).

Intelligibility of vocoded sentences also increases with improved representation of amplitude envelope modulations (e.g., higher cutoff frequencies for low-pass filters that extract amplitude envelopes).[2] Fu and Shannon (2000) proposed that when adequate spectral cues are available, envelope modulations up to 16–20 Hz are sufficient for speech intelligibility (e.g., Drullman *et al.*, 1994; Fu *et al.*, 2004), but when spectral cues are reduced as in vocoder processing, listeners utilize envelope modulations up to 50 Hz (e.g., Shannon *et al.*, 1995).[3] These limits depend on the spectral resolution of speech: higher low-pass envelope cutoff frequencies produce better speech intelligibility at poorer spectral resolutions, but performance asymptotes at lower cutoff frequencies as spectral resolution improves (Xu *et al.*, 2005; Xu and Zheng, 2007; Stone *et al.*, 2008).

The perceptual importance of information-bearing acoustic changes is expected to maintain at temporal resolutions below the 50-Hz limit proposed by Fu and Shannon (2000). Speech contains significant amplitude modulations at 3–8 Hz (Houtgast and Steeneken, 1985), and spectra of speech spoken at a medium rate are maximally different from each other when separated by 128 ms, or a modulation rate of ≈8 Hz (Stilp *et al.*, 2010b). Sentence intelligibility at these lower temporal resolutions is predicted to suffer by larger amounts when high-CSE_{CI} intervals are replaced by noise compared to replacing low-CSE_{CI} intervals.

The present experiments explore the spectral and temporal resolutions of information-bearing acoustic changes that are important for understanding noise-vocoded sentences. Spectral and temporal resolutions of speech were first manipulated independently to define perceptual reliance upon information-bearing acoustic changes across wide ranges of signal quality (Experiment 1: number of spectral channels, Experiment 2: low-pass filter cutoff frequency for envelope extraction). Spectral and temporal resolutions were then manipulated simultaneously to explore spectrotemporal trade-offs and their relationship with these informational changes (Experiment 3). In each experiment, replacing high-CSE_{CI} sentence intervals with noise is predicted to impair sentence intelligibility more than replacing low-CSE_{CI} intervals. Additionally, information-bearing acoustic changes are

predicted to become more important for sentence intelligibility (i.e., produce larger decrements in performance when they are replaced by noise) as signal quality worsens, consistent with Stilp (2014).

II. EXPERIMENT 1: SPECTRAL RESOLUTION

A. Methods

1. Participants

Twenty-four undergraduates were recruited from the Department of Psychological and Brain Sciences at the University of Louisville. All participants reported being native English speakers with normal hearing and received course credit for their participation.

2. Stimuli

Experimental materials were 192 sentences from the TIMIT database [91 talkers from the North Midland dialect region (23 women, 68 men), each 5–9 words, mean duration = 2064 ms]; sentences included all items tested in Stilp *et al.* (2013) and Stilp (2014). Sentences had a native sampling rate of 16 kHz. Sentences were root-mean-square (RMS) amplitude-normalized and then processed by a channel vocoder. Sentences were vocoded using 4, 6, 8, 10, 12, 16, 20, and 24 channels with center frequencies equally spaced between 300 and 5000 Hz according to Greenwood's (1990) formula (Table I). Fourth-order bandpass Butterworth filters were used for channel analysis and synthesis. Halfwave rectification and second-order low-pass Butterworth filters with 150-Hz cutoff extracted amplitude envelopes, which modulated white Gaussian noise carriers. Zero-phase filtering doubled the filter order while maintaining temporal characteristics.

TABLE I. Center frequencies (in Hz) for spectral channels in vocoder processing. Each column depicts the number of spectral channels tested in different experimental conditions, and channel number is listed at the beginning of each row. Total sentence bandwidth spanned 300–5000 Hz.

		Number of channels tested							
		4	6	8	10	12	16	20	24
Channel number	1	463	403	376	360	349	336	329	324
	2	982	684	565	502	463	418	392	376
	3	1929	1103	822	684	603	513	463	433
	4	3658	1729	1169	915	774	623	544	496
	5		2665	1637	1209	982	750	635	565
	6		4061	2269	1583	1237	899	737	643
	7			3124	2059	1549	1072	852	728
	8			4279	2665	1929	1273	982	822
	9				3435	2395	1506	1129	926
	10				4414	2963	1778	1295	1041
	11					3658	2093	1481	1169
	12					4507	2460	1692	1309
	13						2886	1929	1465
	14						3381	2197	1637
	15						3957	2499	1827
	16						4626	2840	2037
	17							3224	2269
	18							3658	2526
	19							4147	2810
	20							4698	3124
	21								3471
	22								3855
	23								4279
	24								4747

Information-bearing acoustic changes were investigated using the noise replacement paradigm, which assesses the perceptual importance of certain acoustic intervals for speech intelligibility by replacing them with noise (e.g., Cole *et al.*, 1996; Kewley-Port *et al.*, 2007). This allows comparison of results to previous investigations that have also used this paradigm (Stilp *et al.*, 2013; Stilp, 2014). Sentences were divided into 16-ms slices to measure spectral changes on a fine timescale. Following Stilp *et al.* (2013) and Stilp (2014), CSE_{CI} was parameterized as the Euclidean distance between RMS amplitude profiles across vocoder channels for all pairs of neighboring spectral slices. Distances were summed in boxcars of five successive slices (80 ms, following Stilp and Kluender, 2010; Stilp *et al.*, 2013; Stilp, 2014) then sorted into ascending (low CSE_{CI}) or descending (high CSE_{CI}) order. The boxcar that ranked first (lowest or highest CSE_{CI}) was replaced by speech-shaped noise (white noise processed by a 100th-order finite impulse response filter to achieve a flat spectrum up to 500 Hz and −9 dB/octave decrease above that point) matched to mean sentence level (5-ms linear onset/offset ramps). Eighty-millisecond intervals immediately before and after replaced segments, as well as 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. Four 80-ms intervals were replaced by noise in each sentence, as replacing four low-CSE_{CI} intervals with noise in eight-channel sentences produced performance significantly worse than control levels (69 RAU versus 80 RAU) and replacing four high-CSE_{CI} intervals produced performance well above chance levels (44 RAU versus <0 RAU) (Stilp, 2014). With mean sentence duration of 2064 ms, an average of 16% of total sentence duration was replaced with noise.[4]

3. Procedure

All stimuli were resampled at 44 100 Hz sampling rate in MATLAB before digital-to-analog conversion by an RME HDSPe AIO sound card (RME Audio, Haimhausen, Germany). Stimuli were passed through a programmable attenuator (TDT PA4, Tucker-Davis Technologies, Alachua, FL) and headphone buffer (TDT HB6), and then were presented diotically at 70 dB sound pressure level via circumaural headphones (Beyer-Dynamic DT-150, Beyerdynamic Inc. USA, Farmingdale, NY). Listeners participated in single-wall sound-isolating booths (Acoustic Systems, Inc., Austin, TX). Following acquisition of informed consent, listeners were given instructions and told to expect that some sentences would be difficult to understand, so guessing was encouraged. The experiment was run using custom MATLAB scripts, and listeners typed their responses. Listeners first completed 12 practice sentences arranged in increasing difficulty without feedback (sentences with no noise, low-CSE_{CI}

changes replaced by noise, and then high-CSE_{CI} changes replaced at 20 spectral channels, then 12, 8, and finally 4 channels). Listeners then completed 192 experimental sentences [8 spectral resolutions (4, 6, 8, 10, 12, 16, 20, 24 channels) × 3 levels of CSE_{CI} replaced by noise (high, low, none) × 8 repetitions]. One sentence was presented per trial and no listener heard any sentence more than once. While listeners heard sentences in the same order, the order of experimental conditions was pseudo-randomized: within each of eight 24-trial blocks, listeners heard one sentence in each of the 24 conditions in random order. Each of the 24 versions (experimental conditions) of a given sentence was presented only once across all listeners.

Responses were scored offline by two raters blind to experimental conditions using guidelines listed in Stilp *et al.* (2010b). Inter-rater reliability, measured by intraclass correlation, was 0.99. Percent of words correctly identified in each sentence was averaged across raters and then arcsine-transformed (Studebaker, 1985) for analysis.

B. Results

Sentence intelligibility improved with more spectral channels, improving more quickly for control and low-CSE_{CI}-replaced sentences, but improving more slowly for high-CSE_{CI}-replaced sentences [Fig. 1(a)]. While intelligibility of four-channel, noise-interrupted sentences was rather low (12–16 RAU), it was well above chance performance in an open-set sentence recognition task (which approaches −23 RAU or 0%) and, thus, not a floor effect. Mean intelligibility of 24-channel control sentences was 98 RAU, which was significantly lower than intelligibility of full-spectrum TIMIT sentences in Stilp (2014) [open circle in Fig. 1(a); mean = 108 RAU, independent-samples t-test: $t_{44} = 3.11$, $p < 0.01$]. This suggests performance would continue to improve at spectral resolutions >24 channels, so this result is not a ceiling effect. Therefore, all results were analyzed in a three (level of CSE_{CI} replaced by noise) by eight (number of spectral channels) repeated-measures analysis of variance (ANOVA). Performance differed as a function of the level of CSE_{CI} replaced by noise ($F_{2,46} = 463.45$, $p < 0.0001$, $\eta^2_p = 0.95$). Intelligibility of control sentences was higher than that for low-CSE_{CI}-replaced sentences, which was higher than that for high-CSE_{CI}-replaced sentences (Bonferroni-corrected paired-samples t-tests, all $p < 0.0001$). Performance also improved as spectral resolution increased ($F_{4.65,106.91} = 285.93$, $p < 0.0001$, $\eta^2_p = 0.93$; degrees of freedom and mean square error adjusted using Greenhouse-Geisser correction for violation of sphericity).

Of primary interest, the interaction between information-bearing acoustic change and spectral resolution was statistically significant ($F_{14,322} = 3.66$, $p < 0.0001$, $\eta^2_p = 0.14$). However, interpretation is complicated by the fact that some results are for noise-interrupted sentences (low CSE_{CI}, high CSE_{CI}), while other results are for sentences without any noise replacement (control). Therefore, differences from control performance were calculated (low CSE_{CI} replaced minus control, high CSE_{CI} replaced minus control) to highlight the relative importance of low versus high information-bearing acoustic changes when they were replaced by noise [Fig. 1(b)]. A two (low-CSE_{CI}-replaced minus control, high-CSE_{CI}-replaced minus control) by eight (number of spectral channels) repeated-measures ANOVA was conducted. Significant main effects of level of CSE_{CI} ($F_{1,23} = 203.62$, $p < 0.0001$, $\eta^2_p = 0.90$) and number of spectral channels ($F_{7,161} = 2.51$, $p < 0.025$, $\eta^2_p = 0.10$) largely recapitulate those reported in the omnibus ANOVA above. The interaction was statistically significant ($F_{7,161} = 4.79$, $p < 0.0001$, $\eta^2_p = 0.17$). Paired-samples t-tests examined performance decrements at each level of spectral resolution (Bonferroni-corrected for multiple comparisons, $\alpha = 0.05/8 = 0.0063$). Decrements did not significantly differ at four spectral channels, but in all other cases, performance decreased by larger amounts when high-CSE_{CI} intervals were replaced by noise (all $p < 0.0007$).

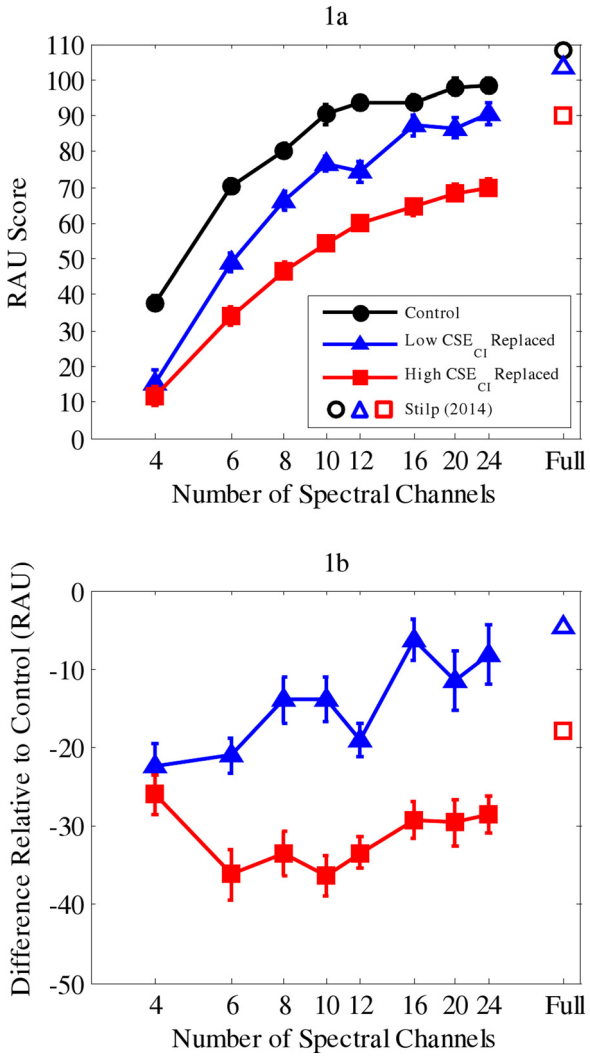

FIG. 1. (Color online) Results from Experiment 1. (a) Mean sentence intelligibility is plotted as a function of the number of spectral channels. Filled circles depict control conditions, filled triangles depict conditions where low-CSE_{CI} changes were replaced by noise, and filled squares depict conditions where high-CSE_{CI} changes were replaced. Unfilled shapes depict corresponding conditions in full-spectrum sentences tested in Stilp (2014). (b) Results are recalculated as mean decrements in performance relative to the control condition [e.g., low CSE_{CI} replaced points in (b) = low CSE_{CI} replaced performance in (a) minus control performance in (a)]. Unfilled shapes depict corresponding decrements in intelligibility of full-spectrum sentences from Stilp (2014). Error bars depict standard errors.

C. Discussion

Information-bearing acoustic changes were highly important for understanding sentences across a very broad range of spectral resolutions. Replacing low-CSE_{CI} intervals produced large decrements in performance at low spectral resolutions and smaller decrements as spectral resolution improved, almost fully overcoming noise-replacement of these speech intervals at 24 spectral channels. Replacing high-CSE_{CI} intervals had far greater consequences on performance, producing significantly larger decrements in performance at all spectral resolutions above four channels. Even with 24 channels of spectral resolution, performance only reached 70 RAU (decrement of 29 RAU compared to the control condition).

Results mostly support the prediction that information-bearing acoustic changes become more important for speech perception as listening conditions worsen (Stilp, 2014). In Stilp (2014), performance decrements progressively increased as more high-CSE_{CI} intervals were replaced with noise; decrements were also larger when replacing information-bearing acoustic changes in noise-vocoded versus full-spectrum sentences. In Experiment 1, performance decrements largely increased as spectral resolution decreased [Fig. 1(b)]. Performance with four spectral channels did not follow this pattern, producing similar decrements whether low-CSE_{CI} or high-CSE_{CI} intervals were replaced by noise. This is consistent with research demonstrating extreme difficulty understanding four-channel vocoded sentences when noise-interruption is introduced (Nelson and Jin, 2004; Başkent, 2012). For example, using 50% duty cycle noise and an interruption rate of 1.5 Hz, Başkent (2012) reported mean scores of ≈5% correct[5] (−2 RAU). Results indicate that some degree of spectral resolution is necessary in order for perception to rely upon information-bearing acoustic changes to understand sentences.

For sentences with 6–24 spectral channels, replacing high-CSE_{CI} intervals with noise impaired performance more than replacing an equal number of low-CSE_{CI} intervals. These results were observed with temporal resolution held constant using a 150-Hz low-pass cutoff for extracting amplitude envelopes (as in Stilp *et al.*, 2013; Stilp, 2014). Experiment 2 investigated the importance of these acoustic changes across a broad and lower range of temporal resolutions while spectral resolution was held constant.

III. EXPERIMENT 2: TEMPORAL RESOLUTION

A. Methods

1. Participants

Thirty undergraduates were recruited from the Department of Psychological and Brain Sciences at the University of Louisville. All participants reported being native English speakers with normal hearing and received course credit for their participation. None participated in Experiment 1.

2. Stimuli

Experiment 2 used 120 sentences selected from those used in experiment 1. Stimulus processing proceeded as in Experiment 1 with two changes. First, spectral resolution was held constant at eight spectral channels, using the same center frequencies as those tested in Experiment 1. Second, temporal resolution was manipulated by varying the cutoff frequency of low-pass filters that extracted amplitude envelopes in vocoding. Cutoff frequencies spanned a broad range (4, 8, 16, 32, 64 Hz), both above and below the 50 Hz deemed sufficient for high levels of vocoded sentence intelligibility by Fu and Shannon (2000). Four 80-ms, low-CSE_{CI} intervals were replaced by noise at each level of temporal resolution as were four high-CSE_{CI} intervals.[6] These ten experimental conditions were accompanied by five control conditions, where no noise-replacements were conducted at each of these five temporal resolutions.

3. Procedure

The procedure for Experiment 2 was similar to that in Experiment 1. Listeners first completed 12 practice sentences arranged in increasing difficulty without feedback (no noise, low-CSE_{CI} changes replaced by noise, and then high-CSE_{CI} changes replaced in sentences with envelope modulations low-pass filtered at 32 Hz, then 16, 8, and, finally, 4 Hz). In the main experiment, within each of 8 15-trial blocks, listeners heard one sentence in each of the 15 conditions in random order. Each of the 15 versions of a given sentence was presented twice across all listeners. Responses were scored offline by the same two raters. Inter-rater reliability, measured by intraclass correlation, was 0.99.

B. Results

Sentence intelligibility improved with higher temporal resolution before plateauing at 75 RAU for control sentences, 58 RAU for low-CSE_{CI}-replaced sentences, and 43 RAU for high-CSE_{CI}-replaced sentences [Fig. 2(a)]. Intelligibility of 4-Hz sentences was at floor levels irrespective of noise replacement (−14 to −10 RAU, or 2%–3% correct across conditions). Because they were at floor performance, these results were removed from statistical analyses. Remaining results were analyzed in a three (level of CSE_{CI} replaced by noise) by four (low-pass filter cutoff frequency) repeated-measures ANOVA. Performance varied as a function of the level of information replaced ($F_{2,58} = 88.03$, $p < 0.001$, $\eta_p^2 = 0.75$). As in Experiment 1, performance in the control conditions was higher than the low-CSE_{CI}-replaced conditions, which was higher than the high-CSE_{CI}-replaced conditions (Bonferroni-corrected paired-samples *t*-tests, all $p < 0.0001$). Performance also increased as envelope cutoff frequency increased ($F_{3,87} = 135.16$, $p < 0.001$, $\eta_p^2 = 0.82$).

The interaction between CSE_{CI} and temporal resolution was significant ($F_{6,174} = 2.83$, $p < 0.05$, $\eta_p^2 = 0.09$), but interpretation is again complicated due to the mixture of control and experimental conditions. Results were recalculated as decrements relative to control performance [Fig. 2(b)]. A two (low-CSE_{CI}-replaced minus control, high-CSE_{CI}-replaced minus control) by four (low-pass filter cutoff frequency) repeated-measures ANOVA revealed main effects of CSE_{CI} ($F_{1,29} = 32.85$, $p < 0.0001$, $\eta^2_p = 0.53$) and

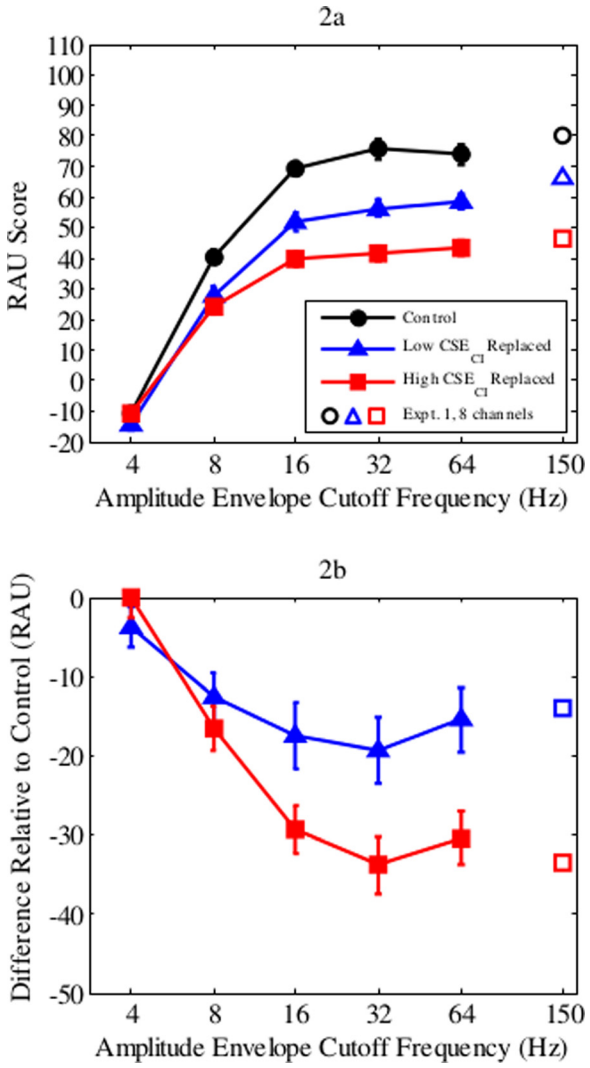

FIG. 2. (Color online) Results from Experiment 2. (a) Mean sentence intelligibility is plotted as a function of low-pass filter cutoff frequency. Filled circles depict control conditions, filled triangles depict conditions where low-CSE_{CI} changes were replaced by noise, and filled squares depict conditions where high-CSE_{CI} changes were replaced. Unfilled shapes depict intelligibility of eight-channel, 150-Hz sentences in Experiment 1. (b) As in Fig. 1(b), results are recalculated as decrements relative to control performance. Unfilled shapes depict corresponding decrements for eight-channel sentences from Experiment 1. Error bars depict standard errors.

temporal resolution ($F_{3,87} = 3.16$, $p < 0.05$, $\eta^2_p = 0.10$), recapitulating those reported in the omnibus ANOVA above. The interaction approached statistical significance ($F_{3,87} = 2.48$, $p = 0.07$, $\eta^2_p = 0.08$; the interaction remained a nonsignificant trend when cutoff frequencies were restricted to 8–32 or 8–16 Hz to account for potential ceiling effects). At 16 Hz and higher envelope cutoff frequencies, replacing high-CSE_{CI} intervals resulted in larger decrements than replacing low-CSE_{CI} intervals, but decrements were comparable at 8 Hz (Bonferroni-corrected paired-samples t-tests using $\alpha = 0.05/4 = 0.0125$; $p < 0.004$ at/above 16 Hz and $p > 0.25$ at 8 Hz).

C. Discussion

Sentences with only 4-Hz temporal resolution were largely unintelligible, resulting in floor effects. At 8-Hz resolution, sentence intelligibility was well above floor performance, but replacing low-CSE_{CI} or high-CSE_{CI} intervals with noise produced comparable performance. When vocoded sentence envelopes were low-pass filtered with a 16-Hz cutoff frequency, replacing high-CSE_{CI} intervals produced larger decrements in performance than replacing low-CSE_{CI} intervals. Information-bearing acoustic changes contributed to sentence intelligibility once a sufficient level of temporal resolution was available, similar to Experiment 1 for sentences with a sufficient level of spectral resolution (i.e., six or more spectral channels).

Results from Experiment 2 violated the prediction of Stilp (2014) that information-bearing acoustic changes become more important for speech intelligibility as listening conditions worsen. Performance decrements in Fig. 2(b) were larger at higher temporal resolutions, plateauing at 16 Hz. This cannot be attributed to testing too narrow a range of temporal resolutions, as performance plateaued well below the 50-Hz resolution that Fu and Shannon (2000) deemed sufficient for noise vocoded sentence intelligibility. Performance remained at asymptotic levels for 150-Hz cutoffs tested in Experiment 1 [unfilled shapes in Fig. 2(a)]. While information-bearing acoustic changes became more important for sentence intelligibility when spectral resolution was decreased (Experiment 1, but only down to six spectral channels) and as total proportion of sentence duration replaced by noise was increased (Stilp, 2014), this prediction does not hold when only temporal resolution is manipulated. As currently parameterized, information-bearing acoustic changes do not universally become more important for understanding speech as listening conditions worsen.

Listeners exploited information-bearing acoustic changes to understand vocoded sentences above certain lower limits of spectral resolution (four spectral channels, tested with a 150-Hz envelope cutoff frequency in Experiment 1) and temporal resolution (8-Hz envelope cutoff, tested with eight spectral channels in Experiment 2). These lower limits highlight the interdependency of spectral and temporal resolutions for sentence intelligibility. Several studies reveal perceptual trade-offs where speech intelligibility is aided by improvement in one acoustic property (spectral or temporal resolution) when the other property is of poor quality (Xu *et al.*, 2002, 2005; Kong and Zeng, 2006; Nie *et al.*, 2006; Xu and Zheng, 2007; Xu and Pfingst, 2008). These demonstrations have largely been for recognition of consonants or vowels only. Relationships between information-bearing acoustic changes, temporal resolution, and spectral resolution may be greatly informed by examining trade-offs in perception of noise-vocoded sentences. Experiment 3 tests this possibility by varying all three properties simultaneously.

IV. EXPERIMENT 3: SPECTROTEMPORAL TRADE-OFFS

A. Methods

1. Participants

Twenty-four undergraduates were recruited from the Department of Psychological and Brain Sciences at the

University of Louisville. All reported being native English speakers with normal hearing, and received course credit for their participation. None participated in any other experiments.

2. Stimuli

Experiment 3 used the same 192 sentences from Experiment 1. Stimulus processing followed that in Experiments 1 and 2, but for a subset of conditions: 6, 8, 10, and 12 spectral channels (see Table I for channel center frequencies), and amplitude envelope cutoff frequencies of 8, 16, and 32 Hz. These conditions captured the steepest portions of psychometric functions plotted in Figs. 1(a) and 2(a) while simultaneously avoiding ceiling and floor levels of performance. Conditions were fully crossed, creating 12 combinations of spectral and temporal resolution. For each combination, four 80-ms intervals were replaced with speech-shaped noise on the basis of having low-CSE_{CI} or high-CSE_{CI} changes, creating 24 conditions in all. No control conditions were included in Experiment 3.

3. Procedure

The procedure for Experiment 3 was the same as in Experiment 1. Listeners completed 12 practice sentences that progressively decreased in both spectral and temporal resolutions. In the main experiment, conditions were tested in a different random order from that tested in Experiment 1. Each of the 24 versions of a given sentence was presented only once across all listeners. Responses were scored offline by two raters, one of whom analyzed results from Experiments 1 and 2. Inter-rater reliability, measured by intraclass correlation, was 0.99.

B. Results

Consistent with Experiments 1 and 2, sentence intelligibility increased as spectral and temporal resolutions increased, and replacing high-CSE_{CI} intervals produced poorer performance than replacing low-CSE_{CI} intervals. However, interactions among these factors were modest. Results were analyzed in a three-way repeated-measures ANOVA. Sentence intelligibility was again worse when high-CSE_{CI} intervals were replaced by noise ($F_{1,23} = 103.15$, $p < 0.001$, $\eta_p^2 = 0.82$). Performance significantly improved at each increase of spectral resolution ($F_{3,69} = 173.15$, $p < 0.001$, $\eta_p^2 = 0.88$; three *post hoc* Bonferroni-corrected *t*-tests, all $p < 0.006$) and each increase of temporal resolution ($F_{2,46} = 206.14$, $p < 0.001$, $\eta_p^2 = 0.90$; two *post hoc* Bonferroni-corrected *t*-tests, both $p < 0.016$). As in Experiment 2, the interaction between CSE_{CI} and temporal resolution was statistically significant ($F_{2,46} = 7.46$, $p < 0.01$, $\eta_p^2 = 0.25$). Interactions between CSE_{CI} and spectral resolution ($F_{3,69} = 2.10$, $p = 0.11$) and spectral and temporal resolutions were not statistically significant ($F_{6,138} = 0.79$, $p > 0.05$).

Effects of spectral resolution, temporal resolution, and level of CSE_{CI} in replaced speech intervals are illustrated in contour plots, with low-pass envelope cutoff frequency on the ordinate and number of spectral channels on the abscissa (Fig. 3). While the three-way interaction between factors was not statistically significant ($F_{6,138} = 0.79$, $p > 0.05$), local trends in performance are evident. Vertical or horizontal boundaries between regions would suggest utilizing only one acoustic cue for sentence recognition, but these were not observed. Oblique contour boundaries indicate integration of spectral and temporal properties for sentence intelligibility. Contour boundaries for low-CSE_{CI}-replaced sentences were largely parallel to the minor diagonal of the contour plot (connecting 6 channels/32 Hz with 12 channels/8 Hz), suggesting some degree of perceptual facility in trading lower spectral resolution for higher temporal resolution and vice versa. Contour regions were much wider for high-CSE_{CI}-replaced sentences, indicating larger trade-offs were necessary in order to improve sentence intelligibility.

C. Discussion

Results illuminate sufficient amounts of spectral and temporal resolutions required in order for perception to exploit information-bearing acoustic changes. In Experiment 1, performance with six spectral channels significantly differed when low- versus high-CSE_{CI} changes were replaced, but these sentences had 150-Hz temporal resolution. Here, the difference in performance was smaller in six-channel sentences with only 32-Hz temporal resolution. Similarly, in Experiment 2, intelligibility of eight-channel, 8-Hz temporal resolution sentences was comparable whether low- or high-CSE_{CI} changes were replaced. Experiment 3 revealed at least 12 channels were required in order for performance to diverge at 8-Hz temporal resolution. Suggested lower limits of spectral resolution necessary for exploiting information-bearing acoustic changes must consider the temporal resolution as well, and vice versa.

Surprisingly, the interaction between spectral and temporal resolutions (i.e., spectrotemporal trade-offs) and the three-way interaction with information-bearing acoustic changes were not statistically significant. This is contrary to spectrotemporal trade-offs reported in Mandarin tone recognition (Xu *et al.*, 2002; Kong and Zeng, 2006), talker gender identification (Fu *et al.*, 2004; Schvartz and Chatterjee, 2012), phoneme identification (Fu *et al.*, 2004; Xu *et al.*, 2005; Xu and Zheng, 2007; Xu and Pfingst, 2008), and sentence recognition (Nie *et al.*, 2006). However, comparisons to the present results are difficult for two reasons. Nie and colleagues (2006) reported equivalent sentence recognition across an increasing number of active electrodes and a decreasing stimulation rate for CI subjects, but manipulating stimulation rate in CIs and varying envelope cutoff frequency in vocoded sentences for NH listeners have very different effects on the speech signal. Further, open set sentence recognition is a far more difficult task than closed set identification of consonants, vowels, talker gender, or Mandarin tones (2–20 response options, as opposed to the entirety of the English language). Task difficulty was also higher due to interrupting the speech with noise (mean performance for 12-channel, 32-Hz sentences: high-CSE_{CI}

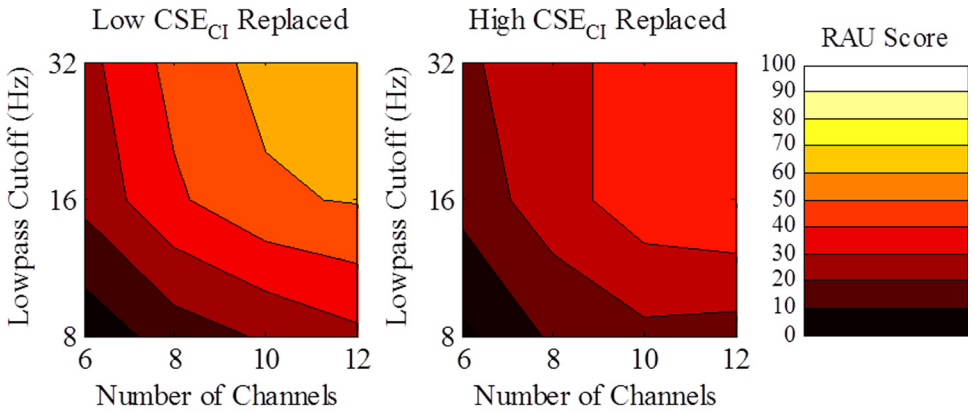

FIG. 3. (Color online) Contour plots depicting results from Experiment 3. Amplitude envelope cutoff frequency is plotted as a function of the number of spectral channels for low-CSE$_{CI}$-replaced (left) and high-CSE$_{CI}$-replaced sentences (right; contour legend presented at far right). Moving toward the upper-right corner of the plots indicates better performance (spanning the range of 8–67 RAU), but in different fashions depending on whether low- or high-CSE$_{CI}$ intervals were replaced by noise.

replaced = 49 RAU; low-CSE$_{CI}$ replaced = 67 RAU), a manipulation that other investigations did not employ.

While imperfect, the best comparison of Experiment 3 results may be to Xu and colleagues (Xu *et al.*, 2005; Xu and Zheng, 2007), who examined phoneme identification in quiet across broad ranges of spectral and temporal resolutions (1–16 spectral channels, 1 - to 512-Hz low-pass filter cutoff frequencies), including many of the same conditions as Experiment 3. In these matched conditions, Xu and colleagues reported extremely subtle spectrotemporal trade-offs for consonant identification and minimal trading for vowel identification. Trade-offs were more pronounced at very low spectral (1–4 channels) and/or temporal (1–4 Hz) resolutions, which are lower than those tested here. While spectral and temporal resolutions required for phoneme identification are fairly dissociable (temporal resolution being more important for consonant identification, spectral resolution for vowel identification), but not for sentence recognition (where they are integrated), trade-offs appear to be very modest for adequate levels of spectral and temporal resolutions in both tasks.

The results of this experiment further challenge Stilp's (2014) proposal that information-bearing acoustic changes become more important as listening conditions worsen. Experiment 1 produced a significant interaction between level of CSE$_{CI}$ replaced and number of spectral channels when they ranged from 4 to 24 (envelope cutoff frequency = 150 Hz); here, the interaction was only a trend when the number of channels were restricted to 6–12 (envelope cutoff frequencies = 8–32 Hz). Experiment 2 showed only a trend for the interaction between CSE$_{CI}$ and envelope cutoff frequencies when varied from 8 to 64 Hz (number of spectral channels = 8), but here the interaction was significant when cutoff frequencies spanned 8–32 Hz (number of spectral channels = 6–12). Thus, the importance of information-bearing acoustic changes does not uniformly increase as signal resolution decreases, as these results are sensitive to the ranges of signal parameters presented. This is particularly true when performance approaches floor levels at low signal resolutions (e.g., 4 Hz in Experiment 2) or ceiling levels at higher signal resolutions (e.g., plateaus in performance at 16 Hz and above in Experiment 2). However, challenging Stilp's (2014) proposal does not undermine the overall importance of information-bearing acoustic changes for speech perception; all three experiments showed worse performance when high-CSE$_{CI}$ intervals were replaced with noise than when low-CSE$_{CI}$ intervals were replaced.

V. STIMULUS ANALYSIS

The present experiments investigated the importance of information-bearing acoustic changes in speech at different spectral and temporal resolutions, providing modest evidence for spectrotemporal trade-offs in sentence intelligibility (differences across contour plots in Fig. 3). This encourages closer analysis of how this metric varies across wide ranges of spectral and temporal resolution at all possible combinations. Sentences were noise-vocoded at each level of spectral resolution at every temporal resolution tested. For each sentence, CSE$_{CI}$ was calculated and summed into 80-ms boxcars. Rather than replacing boxcars with noise, mean CSE$_{CI}$ was calculated for the four highest-ranked (high CSE$_{CI}$) and four lowest-ranked (low CSE$_{CI}$) boxcars following the methodology described in Experiment 1. Grand means calculated across all sentences are presented in Fig. 4.

Means in high-CSE$_{CI}$ [Fig. 4(a)] and low-CSE$_{CI}$ [Fig. 4(b)] analyses increased with higher temporal resolution. This is intuitive, as increasing low-pass envelope cutoffs more faithfully transmitted rapid spectral changes that contributed to larger measures of information-bearing acoustic changes, especially for high CSE$_{CI}$. The relationship between CSE$_{CI}$ and spectral resolution, however, showed the opposite pattern: mean CSE$_{CI}$ *decreased* when more spectral channels were used in noise vocoding. While perhaps unexpected, this relationship is well explained by how vocoding affected calculation of CSE$_{CI}$. In vocoding, channel bandwidths decreased as the number of spectral channels increased. For a vocoded sentence with many (e.g., 24) spectral channels, RMS amplitude in any given channel pooled acoustic energy across a narrow frequency extent. As a result, RMS amplitudes were often smaller values with few

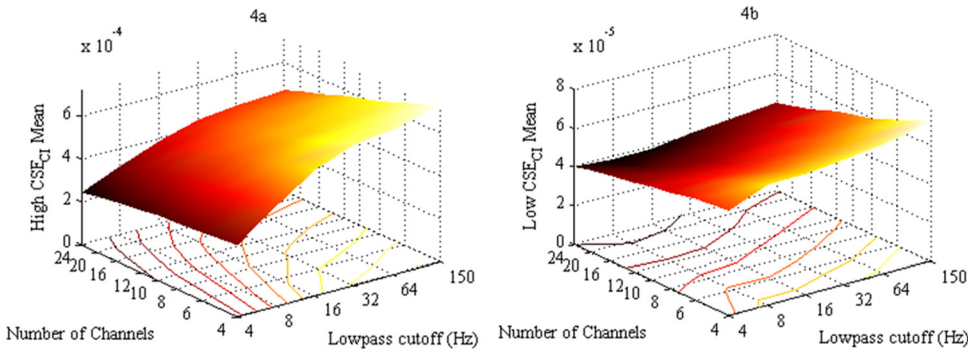

FIG. 4. (Color online) Acoustic analyses of experimental sentences. Plots portray mean CSE_{CI} for the four highest-ranked (a) and four lowest-ranked (b) boxcars, averaged across all sentences. Mean CSE_{CI} is on each vertical axis, number of spectral channels on each ordinate, and low-pass filter cutoff frequency for amplitude envelope extraction along each abscissa. Corresponding contour plots are cast onto the ground plane. Note the scale change in vertical axes across (a) and (b).

large excursions from zero. Smaller means and variances among RMS amplitudes (confirmed in separate analyses) resulted in relatively lower values of CSE_{CI}. Conversely, for a sentence with few (e.g., four) spectral channels, vocoding divided the frequency spectrum into few channels with much larger bandwidths. RMS amplitude of a given channel was pooled across a much broader frequency extent, resulting in larger means and variances for acoustic energy than those observed for sentences with many channels (confirmed in separate analyses). Larger means and variances among RMS amplitudes resulted in larger values of CSE_{CI}, especially for measures of high CSE_{CI} [Fig. 4(a)]. Thus, for a given segment of rapidly changing speech, measures of CSE_{CI} were higher for fewer vocoder channels than for greater numbers of vocoder channels.

Performance in Experiments 1 and 2 closely tracked measures of CSE_{CI} in Fig. 4. Information-bearing acoustic changes numerically increased in sentences with poorer spectral resolution, and decreases in performance relative to the control condition increased in magnitude as fewer channels were included in vocoding [Fig. 1(b)]. Similarly, information-bearing acoustic changes numerically increased in sentences with higher low-pass filter cutoffs, and performance decrements increased as the low-pass cutoff frequency increased [Fig. 2(b)]. Mean CSE_{CI} values from the stimulus analysis are strongly correlated with mean decrements in performance across Experiments 1 and 2 ($r = -0.78$, $r^2 = 0.61$, $p < 0.001$), providing further support that perception exploits information-bearing acoustic changes in the speech signal.

VI. GENERAL DISCUSSION

Information-bearing acoustic changes in the speech signal are highly important for accurate speech perception. When sentence intervals containing high-informational acoustic changes are replaced with noise, sentence intelligibility decreases by greater amounts than when an equal number of low-informational-change intervals are replaced. This has been demonstrated for perception of both full-spectrum and noise-vocoded sentences (Stilp and Kluender, 2010; Stilp *et al.*, 2013; Jiang *et al.*, 2013; Stilp, 2014). From these results, Stilp (2014) proposed that information-bearing acoustic changes become more important for speech perception in more challenging listening conditions. However, this proposal might be overly influenced by ceiling effects, making information-bearing acoustic changes appear less important in better listening conditions (i.e., for full-spectrum sentences, where ceiling effects are present) and more important for more challenging materials (noise-vocoded sentences). The present experiments tested the validity and generalizability of this prediction by examining the role of information-bearing acoustic changes for understanding noise-vocoded sentences with widely varying spectral and temporal resolutions.

Results from Experiment 1 mostly supported this prediction. Figure 1(b) shows larger decrements in sentence intelligibility as spectral resolution decreased, especially when high-CSE_{CI} intervals were replaced by noise. The lone exception was intelligibility of four-channel sentences, which produced comparable decrements in performance whether low- or high-CSE_{CI} intervals were replaced. Four-channel, noise-vocoded sentences interrupted by noise are extremely difficult to understand (Başkent, 2012), perhaps regardless of what acoustic information was originally in noise-replaced intervals. Nevertheless, listeners increasingly relied on information-bearing acoustic changes to understand sentences with poorer spectral resolution.

Results from Experiment 2 did not support Stilp's (2014) prediction, as perceptual reliance on information-bearing acoustic changes did not increase with decreasing temporal resolution. Lower levels of temporal resolution produced floor effects (4 Hz) or comparable decrements in performance whether low-CSE_{CI} or high-CSE_{CI} intervals were replaced (8 Hz). Above 8 Hz, replacing high-CSE_{CI} intervals with noise produced larger decrements in performance than replacing low-CSE_{CI} intervals, and this difference was relatively constant across all higher temporal resolutions tested, including 150 Hz in Experiment 1. The importance of information-bearing acoustic changes for speech understanding does not uniformly increase as signal resolution decreases: decrements in performance increase as larger proportions of total sentence duration are replaced with noise (Stilp, 2014) and largely as the number of spectral channels decreases (Experiment 1), but not as temporal resolution

decreases (Experiment 2). These results are also sensitive to the ranges of parameters tested, as presenting narrower ranges of spectral and temporal resolutions in Experiment 3 attenuated interactions between signal parameters and information-bearing acoustic changes.

Across experiments, information-bearing acoustic changes proved most important (i.e., produced the largest decreases in performance relative to the control condition) at lower spectral resolutions and higher temporal resolutions, seemingly opposite patterns of results. Analyses of information-bearing acoustic changes at different spectral and temporal resolutions (Fig. 4) illuminated these results. CSE_{CI} values (specifically, the difference between high and low CSE_{CI} values) increase at higher temporal resolutions and at lower spectral resolutions of noise-vocoded sentences, suggesting greater importance for speech perception in these situations. Listener performance supported this interpretation, as decreases in performance (relative to control conditions) following noise-replacement were highly correlated with CSE_{CI} values at different spectral and temporal resolutions. This corroborates Stilp and Kluender (2010) and Stilp (2014), who reported significant correlations between CSE (full-spectrum speech) or CSE_{CI} (noise-vocoded speech with eight channels and 150-Hz envelope cutoffs) and sentence intelligibility when spectral and temporal resolutions were held constant.

Stilp and colleagues (2013) proposed that information-bearing acoustic changes are fundamental to speech perception. This proposal was based on similar patterns of results when information-bearing acoustic changes were replaced by noise in noise-vocoded or full-spectrum sentences, and comparable intelligibility of eight-channel noise-vocoded sentences when intervals were replaced with noise on the basis of CSE (changes measured in full-spectrum speech) or CSE_{CI} (changes measured in vocoded speech) (Stilp and Kluender, 2010; Stilp *et al.*, 2013). Two aspects of the present investigation suggest revisiting this proposal. First, measures of CSE_{CI} increase at higher temporal resolutions and at lower spectral resolutions (Fig. 4). This appears to be due to the vocoding process and not the acoustic changes themselves (see Sec. V), as largely the same sentence intervals were identified as low- or high-CSE_{CI} across conditions despite overall CSE_{CI} values changing.[4,6] Raw values of CSE_{CI} are important within an experimental condition (i.e., for determining high versus low changes), but need not necessarily be comparable across conditions. While sentence intelligibility did not improve at both higher temporal resolutions and lower spectral resolutions, the importance of information-bearing acoustic changes (as quantified by decrements in performance) did follow this pattern. Second, the importance of information-bearing acoustic changes for speech perception has only been investigated using the noise-replacement paradigm to date. For an acoustic property of speech (information-bearing acoustic changes or otherwise) to be truly fundamental to speech perception, its importance must maintain across a variety of experimental methods and approaches. One might expect information-bearing acoustic changes to maintain their perceptual importance in different approaches (e.g., see discussion of contrast effects in the Introduction), but direct investigations are needed to confirm this prediction. Whether information-bearing acoustic changes are truly fundamental to speech perception or not, the present results reiterate their importance for speech intelligibility.

Research on the intelligibility of interrupted sentences dates back at least to Miller and Licklider (1950). This method has been tested extensively over the last several decades, including measuring intelligibility of interrupted, noise-vocoded sentences (Nelson *et al.*, 2003; Nelson and Jin, 2004; Başkent and Chatterjee, 2010; Chatterjee *et al.*, 2010; Başkent, 2012; Bhargava *et al.*, 2014). Nelson and Jin (2004) interrupted vocoded sentences with a 25% duty cycle (8 ms of noise followed by 24 ms of intact speech in each cycle), which most closely approximates the total amount of noise replacement in the present experiments (16%). Participants correctly identified ≈38% of keywords in four-channel sentences at +8 and +16 dB signal-to-noise ratios. In Experiment 1, less of total sentence duration was replaced by noise in four-channel sentences, yet worse performance was observed (12–16 RAU, or 15%–18% correct). Some of this difference is likely due to different levels of baseline intelligibility (≈72% correct in Nelson and Jin, 2004; 38 RAU or 38% correct here) and number of talkers presented (ten in Nelson and Jin, 2004; 91 here). Nevertheless, results stress the importance of *what* got replaced by noise (i.e., information-bearing acoustic changes) more than merely *how much* got replaced (proportion of total sentence duration).

Alexander and colleagues (Alexander and Hariram, 2013; Hariram and Alexander, 2014) extended the present model to calculate neural-scaled entropy (NSE), which quantifies information in the firing patterns of simulated auditory nerve responses. While CSE and CSE_{CI} measure changes in the speech spectrum over 16-ms intervals, NSE measures the degree to which neural firing rates change at 1-ms resolution. Measures of NSE are excellent predictors of the intelligibility of nonlinearly frequency-compressed speech, especially to the reduction in information owing to lower start frequencies and higher compression ratios (Alexander and Hariram, 2013; Hariram and Alexander, 2014). The predictive power of CSE, CSE_{CI}, and NSE demonstrates the broad utility of information-theoretic approaches for understanding low-level signal and neural contributions to speech perception.

Perceptually significant acoustic changes have been identified in a wide range of speech materials, from full-spectrum sentences to noise-vocoded sentences across wide ranges of spectral and temporal resolutions (Stilp and Kluender, 2010; Jiang *et al.*, 2013; Stilp *et al.*, 2013; Stilp, 2014; Experiments 1 and 2). This promotes extending the present approach to listeners with hearing impairment, especially CI users. While speech that CI users perceive is impoverished compared to what normal-hearing listeners perceive, both signals are still replete with informational acoustic changes. Subcortical and cortical processing in normal and electrical hearing is similarly predicated on optimizing sensitivity to changes in the input. This fact predicts that information-bearing acoustic changes will be perceptually important for speech perception by CI users as well. Future research should directly establish the importance of information-bearing acoustic changes in speech for CI users.

Once this has been established, signal processing strategies that emphasize or acoustically enhance these perceptually significant changes can be developed. Some signal processing strategies in CIs use amplitude as the primary criterion for determining channel selection and stimulation. This approach is agnostic to perceptually significant acoustic changes in the signal. Incorporating and/or emphasizing information-bearing acoustic changes in the speech signal might provide additional perceptual benefits for CI users, but further research is needed to test this possibility.

ACKNOWLEDGMENTS

We thank Ijeoma Okorie, Andrew McPherson, and Elizabeth Niehaus for assistance in data collection and analysis. This work was partially supported by Grant No. R00-010206 from National Institute on Deafness and Other Communication Disorders (M.J.G.).

[1]Experimental factors that vary widely across these investigations include: test materials [relatively easy Hearing in Noise Test sentences, easy-to-moderate City University of New York sentences, moderate-to-difficult Texas Instruments/Massachusetts Institute of Technology (TIMIT) sentences]; amount of practice (from only nine trials, e.g., Stilp *et al.*, 2013, to 8–10 h of practice, e.g., Shannon *et al.*, 1995); and, design (randomized presentation order, e.g., Stilp *et al.*, 2013, to fully blocked in decreasing order of difficulty, e.g., Dorman *et al.*, 1997; Loizou *et al.*, 1999). These factors might inflate or deflate the number of spectral channels required to understand speech accordingly, warranting caution when comparing results across studies.

[2]It bears mentioning that experiments using CI simulations frequently test the lower limits of temporal resolution required for understanding speech. Current CI processing strategies typically deliver temporal modulation information up to 400 Hz or higher.

[3]Intelligibility of vocoded speech has been shown to asymptote with $<$50 Hz cutoff frequencies, but experimental design may have played a role in these findings. Participants benefitted from up to 10 h of practice and blocked experimental designs that progressively increased in difficulty (e.g., Shannon *et al.*, 1995; Xu *et al.*, 2005; Xu and Zheng, 2007).

[4]Analyses of which consonants and vowels were replaced largely follow those of Stilp and Kluender (2010): as greater amounts of CSE_{CI} were replaced by noise, fewer stop consonants, nasals, front vowels, back vowels, and high vowels were replaced, and more affricates, laterals/glides, low vowels, and diphthongs were replaced. Some differences are to be expected given similar, but not identical, parameterizations of CSE and CSE_{CI} and elimination of acoustic information $<$300 Hz in vocoding. This pattern was largely unaffected by the number of spectral channels in vocoding. Selection of 16-ms slice duration and 80-ms boxcar duration might influence these patterns of replacement, but values were selected to permit comparison of behavioral results to previous findings through common measures of CSE_{CI} (Stilp *et al.*, 2013; Stilp, 2014).

[5]Performance in Experiment 1 exceeded that of other studies exploring intelligibility of noise-interrupted, vocoded sentences (Nelson and Jin, 2004; Başkent, 2012). In the present study, noise replacement was limited to four 80-ms intervals, replacing an average of 16% of total sentence duration with noise. Earlier studies replaced 25%–75% of total sentence duration with noise (most commonly 50%) and at variable intervals, depending on the duration of the duty cycle (8–333 ms).

[6]Analyses of which consonants and vowels were replaced replicated results from Experiment 1 (see footnote 4). Increasing the envelope cutoff frequency in vocoding made trends more pronounced for consonants and slightly less pronounced for vowels (i.e., same patterns of phoneme replacement, but to greater or lesser degrees).

Alexander, J. M., and Hariram, V. (**2013**). "Neural-scaled entropy as a model of information for speech perception," Proc. Meet. Acoust. **19**, 050179.

Alexander, J. M., and Kluender, K. R. (**2008**). "Spectral tilt change in stop consonant perception," J. Acoust. Soc. Am. **123**(1), 386–396.

Alexander, J. M., and Kluender, K. R. (**2010**). "Temporal properties of perceptual calibration to local and broad spectral characteristics of a listening context," J. Acoust. Soc. Am. **128**(6), 3597–3613.

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. Otolaryn. **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.

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

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., O'Leary, S., Rousset, A., Schauwers, K., Skarzynski, H., Skarzynski, P. H., Sterkers, O., Terranti, A., Truy, E., Van de Heyning, P., Venail, F., Vincent, C., and Lazard, D. S. (**2013**). "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients," Audiol. Neurootol. **18**(1), 36–47.

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.

Cole, R., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (**1996**). "The contribution of consonants versus vowels to word recognition in fluent speech," in *Proc. ICASSP'96*, Atlanta, GA, pp. 853–856.

Darwin, C. J., McKeown, J. D., and Kirby, D. (**1989**). "Perceptual compensation for transmission channel and speaker effects on vowel quality," Speech Commun. **8**(3), 221–234.

Dorman, M. F., and Loizou, P. C. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants," Am. J. Otol. **18**(6 Suppl), S113–S114.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**(4), 2403–2411.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**(5), 2670–2680.

Friesen, L. M., Shannon, R. V., Başkent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**(2), 1150–1163.

Fu, Q. J., Chinchilla, S., and Galvin, J. J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryn. **5**(3), 253–260.

Fu, Q. J., and Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**(1), 589–597.

Goupell, M. J., Laback, B., Majdak, P., and Baumgartner, W.-D. (**2008**). "Current-level discrimination and spectral profile analysis in multichannel electrical stimulation," J. Acoust. Soc. Am. **124**(5), 3142–3157.

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

Hariram, V., and Alexander, J. M. (**2014**). "Neural-scaled entropy predicts the effects of nonlinear frequency compression on speech perception," J. Acoust. Soc. Am. **136**(4), 2311.

Holt, L. L. (**2005**). "Temporally nonadjacent nonlinguistic sounds affect speech categorization," Psychol. Sci. **16**(4), 305–312.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**(3), 1069–1077.

Jiang, Y., Stilp, C. E., and Kluender, K. R. (**2013**). "Cochlea-scaled entropy predicts intelligibility of Mandarin Chinese sentences," Proc. Meet. Acoust. **18**, 060006.

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.

Kiefte, M., and Kluender, K. R. (**2008**). "Absorption of reliable spectral characteristics in auditory perception," J. Acoust. Soc. Am. **123**(1), 366–376.

Kingston, J., Kawahara, S., Chambless, D., Key, M., Mash, D., and Watsky, S. (**2014**). "Context effects as auditory contrast," Atten. Percept. Psychophys. **76**, 1437–1464.
Kluender, K. R., and Alexander, J. M. (**2007**). "Perception of speech sounds," in *The Senses: A Comprehensive Reference, Vol. 3, Audition*, edited by P. Dallos and D. Oertel (Academic, San Diego), pp. 829–860.
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), pp. 117–151.
Kong, Y.-Y., and Zeng, F.-G. (**2006**). "Temporal and spectral cues in Mandarin tone recognition," J. Acoust. Soc. Am. **120**(5), 2830–2840.
Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**(1), 98–104.
Laing, E. J., Liu, R., Lotto, A. J., and Holt, L. L. (**2012**). "Tuned with a tune: Talker normalization via general auditory processes," Front. Psychol. **3**, 203.
Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**(4), 2097–2103.
Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**(4), 602–619.
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**(2), 167–173.
Nelson, P. B., and Jin, S. H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**(5), 2286–2294.
Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**(2), 961–968.
Nie, K., Barco, A., and Zeng, F. G. (**2006**). "Spectral and temporal cues in cochlear implant speech perception," Ear Hear. **27**(2), 208–217.
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**(1), 446–454.
Schvartz, K. C., and Chatterjee, M. (**2012**). "Gender identification in younger and older adults: Use of spectral and temporal cues in noise-vocoded speech," Ear Hear. **33**(3), 411–420.
Shannon, C. E. (**1948**). "A mathematical theory of communication," Bell Sys. Tech. J. **27**, 379–423 623–656.
Shannon, R. V., Fu, Q. J., and Galvin, J. (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," Acta Otolaryngol. **124**, 50–54.
Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**(5234), 303–304.
Sjerps, M. J., Mitterer, H., and McQueen, J. M. (**2011**). "Constraints on the processes responsible for the extrinsic normalization of vowels," Atten. Percept. Psychophys. **73**(4), 1195–1215.
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., Alexander, J. M., Kiefte, M., and Kluender, K. R. (**2010a**). "Auditory color constancy: Calibration to reliable spectral properties across speech and nonspeech contexts and targets," Atten. Percept. Psychophys. **72**(2), 470–480.
Stilp, C. E., and Anderson, P. W. (**2014**). "Modest, reliable spectral peaks in preceding sounds influence vowel perception," J. Acoust. Soc. Am. **136**(5), EL383–EL389.
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., Kiefte, M., Alexander, J. M., and Kluender, K. R. (**2010b**). "Cochlea-scaled spectral entropy predicts rate-invariant intelligibility of temporally distorted sentences," J. Acoust. Soc. Am. **128**(4), 2112–2126.
Stilp, C. E., and Kluender, K. R. (**2010**). "Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility," Proc. Natl. Acad. Sci. **107**(27), 12387–12392.
Stone, M. A., Fullgrabe, C., and Moore, B. C. J. (**2008**). "Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region," J. Acoust. Soc. Am. **124**(4), 2272–2282.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**(3), 455–462.
Watkins, A. J. (**1991**). "Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion," J. Acoust. Soc. Am. **90**(6), 2942–2955.
Xu, L., and Pfingst, B. E. (**2008**). "Spectral and temporal cues for speech recognition: Implications for auditory prostheses," Hear. Res. **242**(1), 132–140.
Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**(5), 3255–3267.
Xu, L., Tsai, Y. J., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**(1), 247–258.
Xu, L., and Zheng, Y. F. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**(3), 1758–1764.