Information-bearing acoustic change outperforms duration in predicting intelligibility of full-spectrum and noise-vocoded sentences[a)]

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

(Received 3 May 2013; revised 15 December 2013; accepted 9 January 2014)

Recent research has demonstrated a strong relationship between information-bearing acoustic changes in the speech signal and speech intelligibility. The availability of information-bearing acoustic changes reliably predicts intelligibility of full-spectrum [Stilp and Kluender (2010). Proc. Natl. Acad. Sci. U.S.A. **107**(27), 12387–12392] and noise-vocoded sentences amid noise interruption [Stilp *et al.* (2013). J. Acoust. Soc. Am. **133**(2), EL136–EL141]. However, other research reports that proportion of signal duration preserved also predicts intelligibility of noise-interrupted speech. These factors have only ever been investigated independently, obscuring whether one better explains speech perception. The present experiments manipulated both factors to answer this question. A broad range of sentence durations (160–480 ms) containing high or low information-bearing acoustic changes were replaced by speech-shaped noise in noise-vocoded (Experiment 1) and full-spectrum sentences (Experiment 2). Sentence intelligibility worsened with increasing noise replacement, but in both experiments, information-bearing acoustic change was a statistically superior predictor of performance. Perception relied more heavily on information-bearing acoustic changes in poorer listening conditions (in spectrally degraded sentences and amid increasing noise replacement). Highly linear relationships between measures of information and performance suggest that exploiting information-bearing acoustic change is a shared principle underlying perception of acoustically rich and degraded speech. Results demonstrate the explanatory power of information-theoretic approaches for speech perception. © *2014 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4863267]

PACS number(s): 43.71.Es, 43.71.An, 43.66.Ba [BRM] Pages: 1518–1529

I. INTRODUCTION

Much in the world is predictable from time to time and place to place, but there is little information in such predictability or certainty. Instead, events that deviate from this predictability can be highly informative. These ideas were mathematically formalized in Claude Shannon's information theory (Shannon, 1948). In information theory, information is always defined by unpredictability, uncertainty, or change. The greater the uncertainty or unpredictability, the greater the amount of information that can potentially be transmitted. Whenever there is no uncertainty or complete predictability, there is no new information.

While these tenets were originally developed for practical applications, they are highly relevant for understanding sensation and perception (Attneave, 1954; Barlow, 1961). It is both true and fortunate that sensory systems respond primarily to change (Kluender *et al.*, 2003). When sensory inputs are changing or unpredictable, neural firing increases dramatically; when inputs are fixed or unchanging, neural firing decreases precipitously or ceases altogether (i.e., adaptation). Thus, sensory inputs that are changing or unpredictable should prove most informative for perception.

Information-theoretic approaches to vision have a long tradition (Attneave, 1954; Barlow, 1961) and continue to be productive in investigations motivated by the "efficient coding hypothesis." In brief, the efficient coding hypothesis proposes that sensory systems have evolved and adapted to encode predictable aspects of the environment in order to optimize sensitivity to unpredictable (more informative) aspects. The efficient coding hypothesis has been highly productive for understanding vision (e.g., Olshausen and Field, 1996; Barlow, 2001; Simoncelli and Olshausen, 2001; Simoncelli, 2003), but despite these tenets applying to sensation and perception most broadly, applications to audition (and other modalities) have only recently begun to be explored.

The auditory system strives to be optimally sensitive to change (i.e., information), and demonstrations have been clearest in perception of speech. Recent research reveals several instances where perceptual performance cannot be explained by frequency, modulation frequency, time, or whether speech intervals contain consonants or vowels, but performance closely corresponds to measures of information-bearing acoustic change (Alexander and Kluender, 2008; Stilp and Kluender, 2010; Stilp *et al.*, 2010). This relationship has remained strong across a wide range of speaking rates, amid various forms of temporal distortion, in spectrally rich and degraded materials, and across tonal and non-tonal languages (Stilp *et al.*, 2010; 2013; Jiang *et al.*, 2013). Information-bearing acoustic changes are important for speech perception by listeners with sensorineural hearing loss (Alexander and Kluender, 2009), simulated

---

[a)] Portions of this work were presented at the 166th Meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **134**, 4232 (2013)].

[b)] Author to whom correspondence should be addressed. Electronic mail: christian.stilp@louisville.edu

0001-4966/2014/135(3)/1518/12/$30.00

hearing loss (Alexander *et al.*, 2011), and for acoustic simulations of cochlear implant (CI) processing (Stilp *et al.*, 2013). Across all of these examples, information-bearing acoustic changes in the speech signal prove highly important for speech perception.

Stilp and Kluender (2010) investigated the extent to which sentence intelligibility directly relied upon information-bearing acoustic changes in the speech signal. They operationalized information-bearing acoustic change by developing the metric cochlea-scaled spectral entropy (CSE). In brief, CSE measures differences between successive short-duration spectra scaled to reflect the frequency spacing and weighting of the cochlea. Speech intervals that rated as high-CSE were predicted to be highly important for perception, while speech intervals that rated as low-CSE were predicted to be far less important. Sentence intervals were replaced with speech-shaped noise on the basis of low, medium, or high measures of CSE. Listeners understood fewer words in sentences when high-CSE intervals were replaced by noise (low-CSE intervals retained) compared to when low-CSE intervals were replaced by noise (high-CSE intervals retained). Most compelling, measures of CSE predicted a remarkable amount of sentence intelligibility ($r^2 = 0.80$). Results demonstrated the prominent role of information-bearing acoustic changes for sentence intelligibility.

Results of Stilp and Kluender (2010) have important implications for speech perception by listeners who use CIs. Initial encoding by the peripheral auditory system widely varies in acoustic versus electrical hearing, but subsequent subcortical and cortical processing is similarly predicated on optimizing sensitivity to changes in the input. While specific measures of information-bearing acoustic change are bound to differ in acoustic versus electrical hearing, its broad perceptual importance is expected to maintain. Stilp *et al.* (2013) tested this prediction by conducting an experiment similar to that of Stilp and Kluender (2010) but using noise-vocoded sentences to mimic the spectral degradation of CI processing. Intervals in vocoded sentences were replaced with noise on the basis of measures of CSE (measured before vocoding) or a revised measure of information available in vocoded speech, $\mathrm{CSE}_{\mathrm{CI}}$ [cochlea-scaled entropy for CIs; parameterized as Euclidean distances between successive root-mean-square (rms) amplitude profiles across eight vocoder channels]. Sentence intelligibility was poorer when high-information intervals were replaced by noise (low-information intervals retained) compared to when low-information intervals were replaced, replicating Stilp and Kluender (2010). Critically, performance did not differ when segments in vocoded sentences were replaced on the basis of measures of CSE or $\mathrm{CSE}_{\mathrm{CI}}$. Results suggested fundamental principles of perception that are shared across healthy hearing and acoustic simulations of electrical hearing.

While results supported a strong relationship between information-bearing acoustic changes in the speech signal and sentence intelligibility, questions remain regarding the exact nature of this relationship. Results of Stilp and Kluender (2010) suggest that this relationship is graded (i.e., linear), such that replacing incrementally more information-bearing acoustic changes with noise systematically worsens speech intelligibility. This suggestion is supported by the highly linear relationship and robust regression fit between these measures of information and performance. However, two points obscure whether the same graded relationship exists between $\mathrm{CSE}_{\mathrm{CI}}$ and sentence intelligibility. First, Stilp *et al.* (2013) tested only two conditions of $\mathrm{CSE}_{\mathrm{CI}}$ (replacing low- or high-$\mathrm{CSE}_{\mathrm{CI}}$ changes in 80-ms intervals) compared to six experimental conditions tested by Stilp and Kluender (2010) (replacing low-, medium-, or high-CSE changes in 80- or 112-ms intervals). This not only makes regression analysis of the results of Stilp *et al.* impractical, but minimal overlap with conditions tested by Stilp and Kluender (2010) is insufficient to infer the nature of the relationship between $\mathrm{CSE}_{\mathrm{CI}}$ and speech intelligibility (Fig. 1). Second, performance in Stilp *et al.* (2013) decreased dramatically when intervals in vocoded sentences were replaced with noise (from a mean RAU (rationalized arcsine units; Studebaker, 1985) score of 70.07 [control] to 18.74 [mean of all experimental conditions]), approaching floor levels when high-$\mathrm{CSE}_{\mathrm{CI}}$ intervals were replaced (mean = 9.92). This decline was markedly larger than that observed in Stilp and Kluender (2010) (from a mean RAU score of 96.38 [control] to 70.47 [mean of all experimental conditions]), who replaced the same mean proportion of full-spectrum sentence duration (0.347) with noise. This precipitous decrease in intelligibility without data from any intermediate conditions obscures the nature of the relationship between $\mathrm{CSE}_{\mathrm{CI}}$ and speech intelligibility.

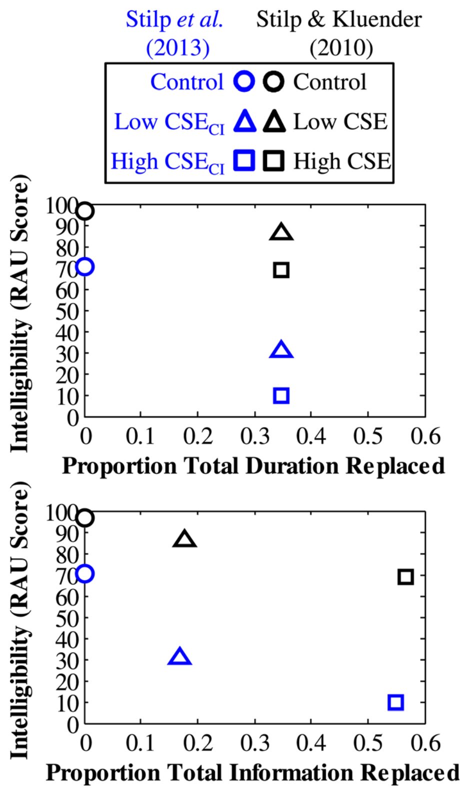

FIG. 1. (Color online) Mean intelligibility (in RAU scores) when information-bearing acoustic changes are replaced by noise in full-spectrum (Stilp and Kluender, 2010) and noise-vocoded sentences (Stilp *et al.*, 2013). Results are poorly differentiated when plotted by the total sentence duration replaced by noise (which was matched across experimental conditions; top), but far better differentiated when plotted by the proportion of total information-bearing acoustic changes replaced (bottom).

Moreover, another metric has already been shown to exhibit a graded relationship with speech intelligibility: Proportion of preserved signal duration. In a seminal paper, Miller and Licklider (1950) reported a highly nonmonotonic relationship between speech intelligibility and interruption rate, but across all interruption rates, intelligibility progressively improved with increasing duration of preserved signal. Results were highly consistent when speech interruptions were periodic or aperiodic. Later investigations explored speech intelligibility when interruptions corresponded to specific events in the speech signal (e.g., consonant or vowel sounds; Cole *et al.*, 1996). Recent studies propose that proportion of total duration (PTD) of preserved signal is the primary predictor of speech intelligibility when consonant or vowel sounds are replaced by noise (e.g., Fogerty and Kewley-Port, 2009; Lee and Kewley-Port, 2009; Wang and Humes, 2010; Fogerty *et al.*, 2012; Kidd and Humes, 2012; Fogerty, 2013). When PTD failed to predict sentence intelligibility when vowels were replaced by noise, Fogerty and colleagues (2012) labeled this result the "exception to this PTD rule" (p. 1675), as PTD predicted sentence intelligibility when consonants were replaced, and predicted word intelligibility when vowels or consonants were replaced. Robust linear regression fits indicate that the relationship between PTD and speech intelligibility is graded (Fogerty and Kewley-Port, 2009; Fogerty and Humes, 2010; Fogerty *et al.*, 2012). While other factors may interact with PTD (e.g., interruption rate: Miller and Licklider, 1950; frequency region: Li and Loizou, 2007; lexical difficulty: Wang and Humes, 2010; phonetic content: Fogerty and Kewley-Port, 2009; Fogerty *et al.*, 2012), its graded relationship with speech intelligibility has been demonstrated across a wide range of studies.

Proportional measures of information (CSE, CSE_{CI}) and time (PTD) each predict intelligibility of noise-interrupted speech, yet these metrics are ideologically at odds with each other. Investigations of CSE suggest it is the information in noise-replaced intervals that drives perceptual performance, not their mere duration. This argument is exemplified by significant changes in speech intelligibility depending on the proportion of CSE replaced when PTD was held constant (Stilp and Kluender, 2010; Stilp *et al.*, 2013). Conversely, investigations of PTD suggest that perceptual performance closely follows the total proportion of speech duration replaced, which effectively collapses across acoustically distinct events in the speech signal (e.g., regions of low and high information-bearing acoustic change). Investigations of how closely proportional measures of information or time correspond to perceptual performance are agnostic to the role the other factor plays: Investigations of CSE held PTD constant, while investigations of PTD collapsed across CSE. Direct comparison of the predictive power of these metrics for speech intelligibility would be highly revealing, but these factors have only ever been investigated independently. If measures of information describe perceptual performance better than measures of time (Alexander and Kluender, 2010; Stilp and Kluender, 2010; Stilp *et al.*, 2010, 2013), proportion of information-bearing acoustic changes replaced by noise should correspond more closely with speech intelligibility than proportion of signal duration replaced, even if both metrics significantly predict performance in the same experiment.

The goals of the present experiments are threefold. First, results will directly compare the predictive power of information-bearing acoustic change versus proportion of signal replaced by noise for sentence intelligibility. Following previous reports, measures of information are predicted to better explain perceptual performance than measures of time (Alexander and Kluender, 2010; Stilp and Kluender, 2010; Stilp *et al.*, 2010, 2013). Second, experiments will determine whether the relationship between information-bearing acoustic changes and intelligibility of noise-vocoded sentences is graded, as has been reported for perception of full-spectrum sentences (Stilp and Kluender, 2010). This goal directly tests the proposal of Stilp *et al.* (2013) that emphasizing sensitivity to information-bearing acoustic change is a common principle underlying speech perception in both acoustic and simulated electrical hearing. Third, matched experimental designs will reveal whether the perceptual importance of information-bearing acoustic changes diminishes, maintains, or increases in poorer listening conditions (i.e., amid noise replacements in noise-vocoded sentences [Experiment 1; CSE_{CI}] compared to full-spectrum sentences [Experiment 2; CSE]). While absolute levels of performance are expected to differ (i.e., higher overall intelligibility for full-spectrum sentences), relative levels of performance will reveal how perception weights information-bearing acoustic changes as greater proportions of sentences are replaced by noise.

II. EXPERIMENT 1

A. Methods

1. Participants

Twenty-two 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 110 sentences from the TIMIT database (DR3 region, each 5–8 words, mean duration = 2125 ms; Garofolo *et al.*, 1990). Signal processing followed the methods described in Stilp *et al.* (2013) (see Fig. 2). Sentences were rms-amplitude normalized, then processed by an eight-channel vocoder with a Gaussian noise carrier (second-order low-pass Butterworth filters with 150-Hz cutoff to extract amplitude envelopes; fourth-order bandpass Butterworth filters for channel analysis and synthesis). Channel center frequencies were equally spaced between 300 and 5000 Hz according to Greenwood's (1990) formula (center frequencies: 376, 565, 822, 1169, 1637, 2269, 3124, 4279 Hz).

Sentences were then divided into 16-ms slices. 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

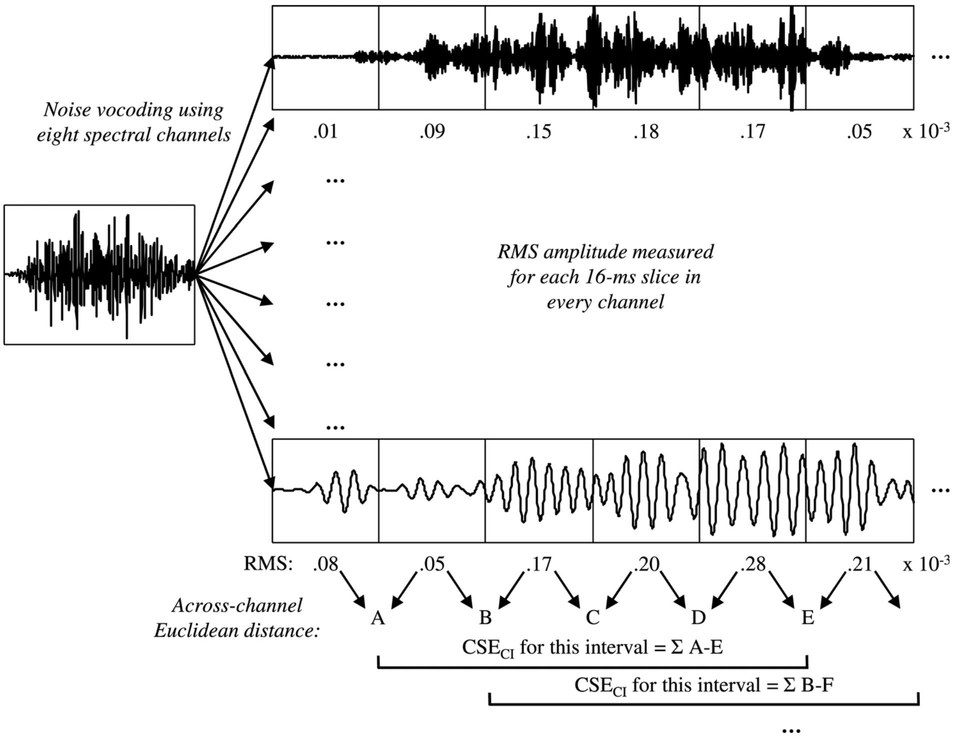

FIG. 2. Process for calculating CSE_{CI}. Sentences are noise-vocoded in eight channels with center frequencies equally spaced from 300 to 5000 Hz according to Greenwood's (1990) formula (lowest channel [CF = 376 Hz] and highest channel [CF = 4279 Hz] are shown at the bottom and top, respectively). Each channel is then divided into 16-ms slices, and the rms amplitude for each slice is recorded. Euclidean distances are calculated for rms outputs across all channels between all neighboring 16-ms intervals. Five successive distances are then summed in a single boxcar that reflects information-bearing acoustic changes in this speech interval. Boxcars are convolved across the entire sentence. See Sec. II A 2 for more details.

boxcars of five successive slices (80 ms) then sorted into ascending (low CSE_{CI} condition) or descending order (high CSE_{CI}). The boxcar that ranked first (lowest or highest CSE_{CI}) was replaced by speech-shaped noise 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-ranked boxcar, which was replaced only if its contents had not already been replaced or preserved. Ceiling performance was expected if only one 80-ms segment was replaced, as that condition resembles investigations of phonemic restoration (e.g., Warren, 1970), so this condition was omitted. A total of two, three, four, five, or six 80-ms intervals were replaced by noise in each sentence. Given mean sentence duration of 2125 ms, conditions replaced mean proportions of 0.075, 0.113, 0.151, 0.188, and 0.226 of total sentence duration with noise. Replacing six 80-ms segments with noise (0.226 of total sentence duration) is expected to produce performance well above floor levels, which was approached in Stilp *et al.* (2013) when 710 ms of sentences containing high-CSE_{CI} segments was replaced (0.347 of total sentence duration). For low-CSE_{CI} conditions, mean proportions of information-bearing acoustic changes replaced in sentences were 0.011, 0.021, 0.034, 0.051, and 0.073. For high-CSE_{CI} conditions, mean proportions of acoustic changes replaced were 0.220, 0.296, 0.358, 0.410, and 0.454.

3. *Procedure*

Sentences were up-sampled to 44 100 Hz and presented diotically at 70 dB sound pressure level via circumaural headphones (Beyer-Dynamic DT-150; Heilbronn, Germany). Listeners participated individually in single-wall sound-isolating booths (Acoustic Systems). 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. Listeners completed 12 practice sentences followed by 110 experimental sentences (10 trials in each of 10 experimental conditions plus 10 control trials presenting vocoded sentences without any noise replacement). 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 ten 11-trial blocks, listeners heard one sentence in each of the 11 conditions in random order.

Responses were scored offline by two raters blind to experimental conditions using guidelines listed in Stilp *et al.* (2010). Inter-rater reliability, measured by intraclass correlation, was 0.99. Intelligibility was measured as the average percent of words correctly identified in each sentence. Scores were arcsine-transformed for data analysis (Studebaker, 1985).

B. Results

1. *Analysis of variance*

Results of Experiment 1 are presented in Fig. 3. The first set of analyses compared intelligibility of sentences with intervals replaced by noise, so performance in the control condition [i.e., no noise replacement, mean = 80.24 RAU of words correctly identified,[1] standard error (s.e.) = 2.46] was omitted. Results were analyzed in a 2 (level of CSE_{CI} replaced by noise: low, high) × 5 (proportion of sentence duration replaced by noise: 0.075, 0.113, 0.151, 0.188, and 0.226 [160, 240, 320, 400, 480 ms divided by mean sentence duration]) repeated-measures analysis of variance (ANOVA). Replicating Stilp *et al.* (2013), intelligibility was higher when low-CSE_{CI} segments were replaced by noise (mean = 66.27 RAU,

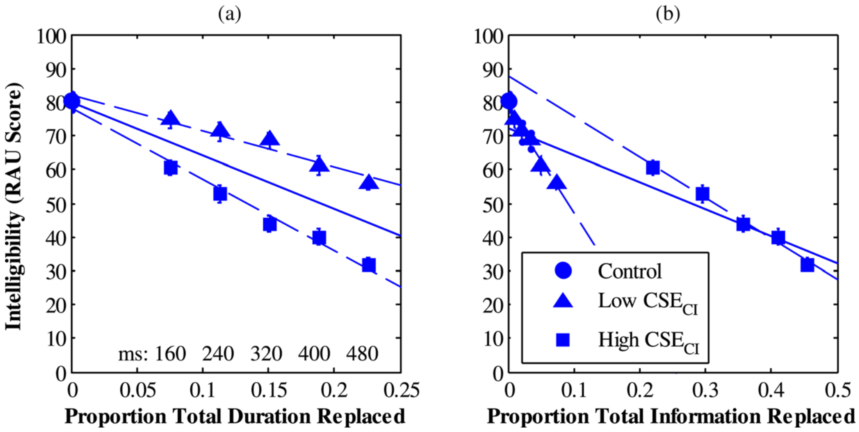

FIG. 3. (Color online) Results of Experiment 1. Error bars indicate standard errors; solid lines indicate linear regression fit to all data; dashed lines indicate regression fit to data from only that condition (low or high CSE$_{CI}$). Circles indicate the control condition; triangles indicate conditions where low-CSE$_{CI}$ intervals were replaced by noise; squares indicate conditions where high-CSE$_{CI}$ intervals were replaced. (a) Mean intelligibility is presented as a function of the proportion of total sentence duration replaced by noise. Raw durations of sentence replacement are indicated along the abscissa. (b) The same mean intelligibility scores from (a) are presented as a function of the proportion of information-bearing acoustic changes in sentences replaced by noise.

s.e. = 1.58) than when high-CSE$_{CI}$ segments were replaced (mean = 45.83, s.e. = 1.34) ($F_{1,21} = 336.57$, $p < 0.001$, $\eta_p^2 = 0.94$). Intelligibility also varied as a function of the proportion of sentence duration replaced by noise ($F_{4,84} = 33.03$, $p < 0.001$, $\eta_p^2 = 0.61$). *Post hoc* analyses using Tukey's Honestly Significant Differences (HSDs) revealed that conditions separated by 0.038 of total sentence proportion (80 ms) did not differ from one another; all other pairwise comparisons were statistically significant ($\alpha = 0.01$) (Table I).

The interaction between factors was statistically significant ($F_{4,84} = 2.66$, $p < 0.05$, $\eta_p^2 = 0.11$). Tukey HSD tests revealed that at each duration of noise replacement, intelligibility of high-CSE$_{CI}$-replaced sentences was poorer than that of low-CSE$_{CI}$-replaced sentences ($\alpha = 0.01$). Interestingly, intelligibility did not significantly differ when replacing 0.226 of total sentence duration (480 ms) containing low-CSE$_{CI}$ intervals (mean = 55.85 RAU, s.e. = 1.74) compared to replacing 0.075 of total sentence duration (160 ms) containing high-CSE$_{CI}$ intervals (mean = 60.63, s.e. = 2.16); this result was verified by paired-samples *t*-test ($t_{21} = 1.90$, $p = 0.07$). For this pair of conditions, performance differed only modestly when far larger sentence proportions containing less-perceptually-important signal and far smaller sentence proportions containing more-perceptually-important signal were replaced by noise.

2. Linear regressions

Initial regression analyses examined each predictor separately, following previous research where only one of these variables (PTD, CSE) was under investigation. The proportion of signal duration replaced was a significant predictor of sentence intelligibility [$r = -0.73$, $r^2 = 0.53$, $p < 0.05$; solid line in Fig. 3(a)]. While a single regression significantly predicted performance across all experimental conditions, visual inspection of Fig. 3(a) indicates that fitting separate regressions to the data would be more appropriate, especially given the systematic increase in prediction error at larger values of PTD. With control performance included in each regression, PTD predicts sentence intelligibility when low CSE$_{CI}$ and high CSE$_{CI}$ are analyzed separately [low CSE$_{CI}$: $r = -0.98$, $r^2 = 0.96$, $p < 0.001$; high CSE$_{CI}$: $r = -0.99$, $r^2 = 0.99$, $p < 0.001$; dashed lines in Fig. 3(a)]. Performance declined twice as quickly when high-CSE$_{CI}$ intervals were replaced by noise (low-CSE$_{CI}$ regression slope = −107.32, high-CSE$_{CI}$ slope = −211.03; $t_8 = 6.28$, $p < 0.001$). Even in analyses where proportion of signal duration replaced is the only predictor variable, the perceptual importance of information-bearing acoustic change is evident.

The proportion of total information (CSE$_{CI}$) replaced was an excellent predictor of sentence intelligibility [$r = -0.93$, $r^2 = 0.87$, $p < 0.001$; solid line in Fig. 3(b)]. This result is consistent with Stilp and Kluender (2010), who reported an r^2 value of 0.80 for the relationship between CSE and sentence intelligibility. While a single regression explains the vast majority of variability in listener performance, separate regressions on low-CSE$_{CI}$ and high-CSE$_{CI}$ results also provided reliable fits to the data [low CSE$_{CI}$: $r = -0.99$, $r^2 = 0.98$, $p < 0.001$; high CSE$_{CI}$: $r = -0.99$, $r^2 = 0.99$, $p < 0.001$; dashed lines in Fig. 3(b)]. Performance

TABLE I. Results from Experiment 1. Values depict mean sentence intelligibility in RAU scores. Standard errors of the mean are listed in parentheses. Values atop each column indicate proportions of sentence duration replaced by noise (160–480 ms in 80-ms increments).

	Proportion of sentence duration replaced by noise					
	0	0.075	0.113	0.151	0.188	0.226
Low CSE$_{CI}$ replaced	80.24 (2.46)	74.67 (2.43)	71.13 (2.88)	68.65 (2.47)	61.07 (2.87)	55.85 (1.62)
High CSE$_{CI}$ replaced		60.63 (2.16)	52.86 (2.47)	43.97 (2.47)	39.96 (2.39)	31.75 (1.90)

declined three times faster when low-CSE_{CI} intervals were replaced with noise (low-CSE_{CI} slope $= -327.14$, high-CSE_{CI} slope $= -104.51$; $t_8 = 7.62$, $p < 0.001$). This is opposite to the pattern observed in Fig. 3(a), revealing a strong role of PTD in the low-CSE_{CI} condition. The distribution of CSE_{CI} values in a sentence generally displays a large positive skew, indicating that a large portion of a given sentence contains minimal information-bearing acoustic changes. As larger proportions of sentence duration are replaced by noise, the total proportion of information-bearing acoustic change replaced grows very slowly for the low-CSE_{CI} condition [cf., small range of values along the abscissa in Fig. 3(b); maximum $= 0.073$] but grows very quickly for the high-CSE_{CI}-replaced condition (cf., wide range of values along the abscissa).

A final analysis compared unique contributions of duration versus information to perceptual performance. Multiple regressions were calculated on each listener's results using both proportion of total sentence duration replaced and proportion of total information replaced as predictors.[2] Standardized regression coefficients for each predictor (β_{duration}, $\beta_{\text{information}}$) were recorded from each analysis. Across participants, means for both coefficients were significantly different from zero ($\beta_{\text{duration}} = -0.39$, one-sample t-test: $t_{21} = 10.00$, $p < 0.001$; $\beta_{\text{information}} = -0.73$: $t_{21} = 7.86$, $p < 0.001$), indicating each factor significantly contributed to performance. More importantly, listeners weighted information-bearing acoustic changes more than they weighted duration of sentence replacement (paired-sample $t_{21} = 5.31$, $p < 0.001$), indicating measures of information better explained sentence intelligibility than measures of time.

C. Discussion

Results from Experiment 1 clarify the relationship between information-bearing acoustic changes and speech intelligibility. Stilp and Kluender (2010) revealed a graded (i.e., highly linear) relationship between information-bearing acoustic changes in full-spectrum speech and sentence intelligibility. Results of Stilp *et al.* (2013) were inconclusive whether this extended to information-bearing acoustic changes in noise-vocoded sentences (CSE_{CI}). Here, information-bearing acoustic changes clearly share a graded relationship with intelligibility of noise-vocoded sentences [Fig. 3(b)]. The relationship between information-bearing acoustic changes and speech intelligibility is highly linear for both full-spectrum ($r^2 = 0.80$; Stilp and Kluender, 2010) and noise-vocoded sentences ($r^2 = 0.87$; Experiment 1). The consistency of this relationship across acoustically diverse materials confirms that emphasizing sensitivity to information-bearing acoustic changes is a common principle underlying speech perception in healthy hearing and acoustic simulations of CI processing (Stilp *et al.*, 2013).

Experiment 1 provided the first direct test of whether measures of information or time better predicted intelligibility of noise-interrupted sentences. While proportion of total sentence duration replaced was a significant predictor of sentence intelligibility, three points reveal its inferiority to measures of information-bearing acoustic change for explaining perceptual performance. First, despite the global regression fit between PTD and performance being statistically significant, results in Fig. 3(a) are clearly best summarized by fitting separate regressions to low CSE_{CI} and high CSE_{CI} results. This is a significant weakness for the explanatory power of proportional sentence duration, as it implies something other than duration is driving performance. Second, when results are fit with a single regression function, measures of information explain a substantially larger amount of overall variability than measures of duration ($r^2 = 0.87$ versus $r^2 = 0.53$). Third and most compelling, multiple regressions revealed that listeners weighted information-bearing acoustic changes more than they weighted proportion of sentence duration replaced. In all, measures of information better explained perceptual performance than measures of time, consistent with previous literature (Alexander and Kluender, 2010; Stilp and Kluender, 2010; Stilp *et al.*, 2010, 2013).

Results illustrate how listeners weight information-bearing acoustic changes over a range of proportions of sentence duration replaced by noise. This range is intermediate to the comparatively extreme duration replaced in Stilp *et al.* (2013). However, how listeners rely on acoustic changes in full-spectrum sentences across this range of duration replacement is unknown. Similar to Stilp *et al.* (2013), Stilp and Kluender (2010) replaced an average of 0.347 of total sentence duration with noise without exploring any intermediate proportions. Additionally, their report of a graded relationship between information-bearing acoustic changes and speech intelligibility was demonstrated while fixing PTD at one of two similar values (0.347 using 80-ms intervals, 0.387 using 112-ms intervals). While graded relationships between information and speech perception are reported in Stilp and Kluender (2010) and Experiment 1, their significance for speech intelligibility would be considerably strengthened by testing matched experimental designs across noise-vocoded and full-spectrum speech materials.

When larger proportions of sentence duration were replaced by noise, performance decreased twice as quickly when high-CSE_{CI} intervals were replaced compared to low-CSE_{CI}-intervals. If listeners weight information-bearing acoustic changes more heavily as listening conditions worsen, regression slopes would be expected to be steeper for more challenging listening conditions (e.g., noise-vocoded speech) and shallower for more favorable listening conditions (e.g., full-spectrum speech). This prediction cannot be tested using existing data, as Stilp and Kluender (2010) fixed proportions of sentence duration replaced at much larger values than those investigated here, making direct comparisons problematic.

Experiment 2 fills these gaps by using the design of Experiment 1 to assess intelligibility of full-spectrum sentences when speech intervals are replaced on the basis of measures of CSE. Results permit direct comparison of the perceptual importance of information-bearing acoustic changes for understanding spectrally intact and degraded sentences. If information-bearing acoustic changes are weighted more heavily for understanding vocoded sentences (Experiment 1) than full-spectrum sentences (Experiment 2), performance will decrease more slowly with increasing noise replacement in

Experiment 2. Two additional predictions mirror those made for Experiment 1. First, CSE is predicted to exhibit a graded relationship with speech intelligibility over the range of sentence proportions replaced. Second, proportional measures of information are again predicted to better explain sentence intelligibility than proportional measures of time.

III. EXPERIMENT 2

A. Methods

1. *Participants*

Twenty-two 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*

The same raw TIMIT sentences used in Experiment 1 were used in Experiment 2 but without any noise-vocoding. Information-bearing acoustic change was measured using CSE (Stilp and Kluender, 2010; Stilp *et al.*, 2010). To calculate CSE, sentences were rms-amplitude normalized and divided into 16-ms slices as before, then each slice was passed through 33 simulated auditory filters (Patterson *et al.*, 1982) spaced 1 ERB (equivalent rectangular bandwidth) apart (Glasberg and Moore, 1990) from 26 to 7743 Hz. Euclidean distances were calculated between spectral profiles (i.e., energy across all bands) for all pairs of adjacent 16-ms spectra. Distances were again summed in 80-ms boxcars (five successive slices) then sorted into ascending or descending order. Replacing boxcars with noise followed the same procedure described in Experiment 1. A total of two, three, four, five, or six 80-ms intervals were replaced with noise in each sentence. As the same sentence materials were used as in Experiment 1, proportions of sentence duration replaced remained the same (0.075, 0.113, 0.151, 0.188, and 0.226). For low-CSE conditions, mean proportions of information-bearing acoustic changes replaced in sentences were 0.006, 0.011, 0.020, 0.032, and 0.051. For high-CSE conditions, mean proportions of acoustic changes replaced were 0.236, 0.318, 0.387, 0.444, and 0.492.

3. *Procedure*

The procedure for Experiment 2 was the same as in Experiment 1. Experimental conditions were tested in a different pseudo-randomized order, but again within each of ten 11-trial blocks, listeners heard one sentence in each of the 11 conditions in random order. Responses were scored offline by the same two raters, again blind to experimental conditions. Inter-rater reliability, measured by intraclass correlation, was 0.97. Average percentages of words correctly identified were again arcsine-transformed before data analysis.

B. Results

1. *Analysis of variance*

Results were analyzed in a 2 (level of CSE replaced by noise) × 5 (proportion of sentence duration replaced by noise) repeated-measures ANOVA, omitting control performance where no noise-replacement occurred (mean = 108.08 RAU, s.e. = 2.27). Consistent with Stilp and Kluender (2010) and Experiment 1, intelligibility was higher when low-CSE segments were replaced by noise (mean = 101.83, s.e. = 1.87) than when high-CSE segments were replaced (mean = 89.58, s.e. = 2.28) ($F_{1,21} = 83.60$, $p < 0.001$, $\eta_p{}^2 = 0.80$). Performance again varied as a function of the proportion of sentence replaced by noise ($F_{4,84} = 24.63$, $p < 0.001$, $\eta_p{}^2 = 0.54$). *Post hoc* analyses using Tukey's HSD revealed that no conditions separated by only 0.038 of total sentence proportion (80 ms) differed from each other; two pairs of conditions separated by 0.076 of total sentence proportion (160 ms) differed from each other at $\alpha = 0.05$ (0.113 vs 0.188, 0.151 vs 0.226); all other contrasts significantly differed at $\alpha = 0.01$ (Table II).

The interaction between factors approached statistical significance ($F_{4,84} = 2.45$, $p < 0.06$, $\eta_p{}^2 = 0.10$). Six Bonferroni-corrected paired-sample *t*-tests examined key contrasts between experimental conditions ($\alpha = 0.05/6 = 0.0083$). For the first five *t*-tests, replacing low-CSE intervals resulted in higher intelligibility than replacing high-CSE intervals at each level of sentence proportion replacement (all $p < 0.0057$). A sixth *t*-test indicated that sentence intelligibility did not significantly differ when replacing 0.226 of total sentence duration (480 ms) containing low-CSE intervals (mean = 96.36, s.e. = 2.89) compared to replacing 0.075 of total sentence duration (160 ms) containing high-CSE intervals (mean = 98.42, s.e. = 2.60) ($t_{21} = 0.91$, $p = 0.37$). Replacing far larger proportions of less-perceptually-important signal again produced comparable performance to replacing far smaller proportions of more-perceptually-important signal.

2. *Linear regressions*

Proportion of sentence duration replaced was again a significant predictor of sentence intelligibility [$r = -0.69$, $r^2 = 0.48$, $p < 0.025$; solid line in Fig. 4(a)]. As in Experiment 1, prediction error again increased at larger values of PTD, so

TABLE II. Results from Experiment 2. Values depict mean sentence intelligibility in RAU scores. Standard errors of the mean are listed in parentheses. Values atop each column indicate proportions of sentence duration replaced by noise (160–480 ms in 80-ms increments).

	Proportion of sentence duration replaced by noise					
	0	0.075	0.113	0.151	0.188	0.226
Low CSE replaced	108.08 (2.27)	106.16 (1.93)	103.57 (2.03)	103.43 (2.30)	99.62 (2.82)	96.36 (2.89)
High CSE replaced		98.42 (2.60)	94.29 (2.44)	90.19 (2.43)	83.89 (2.65)	81.14 (2.89)

separate regressions for low-CSE and high-CSE results summarized the data more effectively [low CSE: $r = -0.96$, $r^2 = 0.92$, $p < 0.01$; high CSE: $r = -0.99$, $r^2 = 0.99$, $p < 0.001$; dashed lines in Fig. 4(a)]. Performance declined more than twice as quickly when high-CSE intervals were replaced by noise (low-CSE slope $= -50.52$, high-CSE slope $= -121.36$; $t_8 = 8.36$, $p < 0.001$). The perceptual importance of information-bearing acoustic change is again evident even when proportion of sentence duration replaced serves as the predictor variable.

Consistent with Stilp and Kluender (2010), proportion of total information (CSE) replaced reliably predicted sentence intelligibility [$r = -0.93$, $r^2 = 0.87$, $p < 0.001$; solid line in Fig. 4(b)]. Individual regressions again provided excellent fits to the data [low CSE: $r = -0.98$, $r^2 = 0.97$, $p < 0.001$; high CSE: $r = -0.98$, $r^2 = 0.97$, $p < 0.001$; dashed lines in Fig. 4(b)]. Performance declined more quickly when low-CSE intervals were replaced compared to when high-CSE intervals were replaced (low-CSE slope $= -222.92$, high-CSE slope $= -54.60$; $t_8 = 4.48$, $p < 0.01$), again reflecting large proportions of total sentence duration replaced in low-CSE conditions without replacing a large proportion of total information (maximum for these conditions $= 0.051$).

Relative contributions of proportional time and proportional information to performance were again determined by multiple regressions calculated on each participant's data. Standardized regression coefficients both significantly differed from zero ($\beta_{\text{duration}} = -0.33$, one-sample t-test: $t_{21} = 7.41$, $p < 0.001$; $\beta_{\text{information}} = -0.56$: $t_{21} = 4.91$, $p < 0.01$). Critically, listeners again weighted information-bearing acoustic changes significantly more than they weighted duration of sentence replacement (paired-sample $t_{21} = 2.95$, $p < 0.01$).

Regression slopes were tested across experiments to compare changes in sentence intelligibility as greater sentence proportions (total sentence duration, total sentence information) were replaced by noise. Steeper regression slopes indicated that sentence intelligibility decreased more quickly for noise-vocoded materials (Experiment 1) than full-spectrum materials (Experiment 2). This pattern of results was observed when information-bearing acoustic changes served as the regression predictor (low CSE_{CI} vs low CSE: $t_8 = 3.31$, $p < 0.05$; high CSE_{CI} vs high CSE: $t_8 = 6.47$, $p < 0.001$) and when PTD served as the predictor (low CSE_{CI} vs low CSE: $t_8 = 4.18$, $p < 0.005$; high CSE_{CI} vs high CSE: $t_8 = 7.10$, $p < 0.001$). In all cases, performance declined faster when information-bearing acoustic changes were replaced in vocoded sentences, revealing their greater perceptual importance to understand spectrally degraded speech.

3. *Comparisons to previous data*

Figure 5 reveals results from Experiments 1 and 2 to be highly consistent with previous results from Stilp and Kluender (2010) and Stilp *et al.* (2013) where PTD was fixed at relatively high values. Regressions were recalculated for low-information (low CSE or low CSE_{CI}) conditions, high-information (high CSE or high CSE_{CI}) conditions, or low- and high-information conditions together to include related results from Stilp and Kluender (2010) (CSE analyses) or Stilp *et al.* (2013) (CSE_{CI} analyses). The most notable change in regression fits occurred when PTD (inappropriately) predicted performance using a single regression function. While these regression fits improved (from $r^2 = 0.53$ to 0.74 for CSE_{CI}, from $r^2 = 0.48$ to 0.65 for CSE), they were still inferior to separate regressions fit to low-information-replaced conditions or high-information-replaced conditions ($r^2 \geq 0.93$). Otherwise, regression results indicated high congruence across all experiments.

C. Discussion

Results from Experiment 2 are consistent with Experiment 1 and previous research in several important ways. First, replacing high-CSE intervals impaired sentence intelligibility more than replacing an equal number of low-CSE intervals (Stilp and Kluender, 2010; Jiang *et al.*, 2013). This is equally true when information-bearing acoustic changes are measured and replaced in noise-vocoded

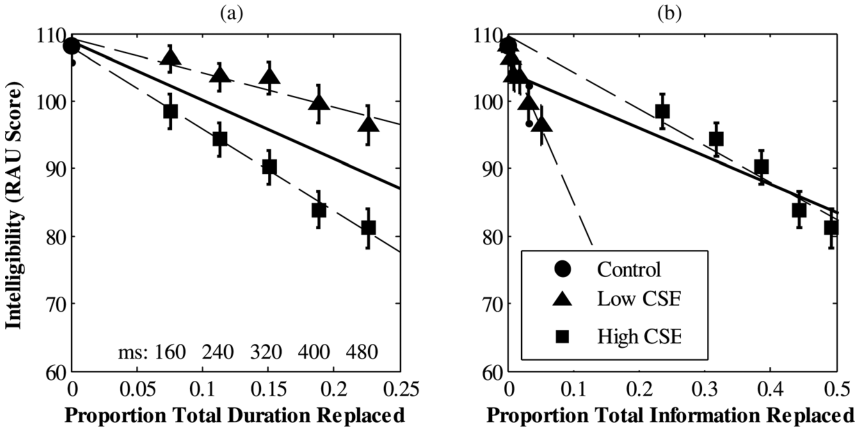

FIG. 4. Results of Experiment 2. Error bars indicate standard errors; solid lines indicate linear regression fit to all data; dashed lines indicate regression fit to data from only that condition (low or high CSE). Circles indicate the control condition; triangles indicate conditions where low-CSE intervals were replaced by noise; squares indicate conditions where high-CSE intervals were replaced. (a) Mean intelligibility is presented as a function of the proportion of total sentence duration replaced by noise. Raw durations of sentence replacement are indicated along the abscissa. (b) The same mean intelligibility scores from (a) are presented as a function of the proportion of information-bearing acoustic changes in sentences replaced by noise.

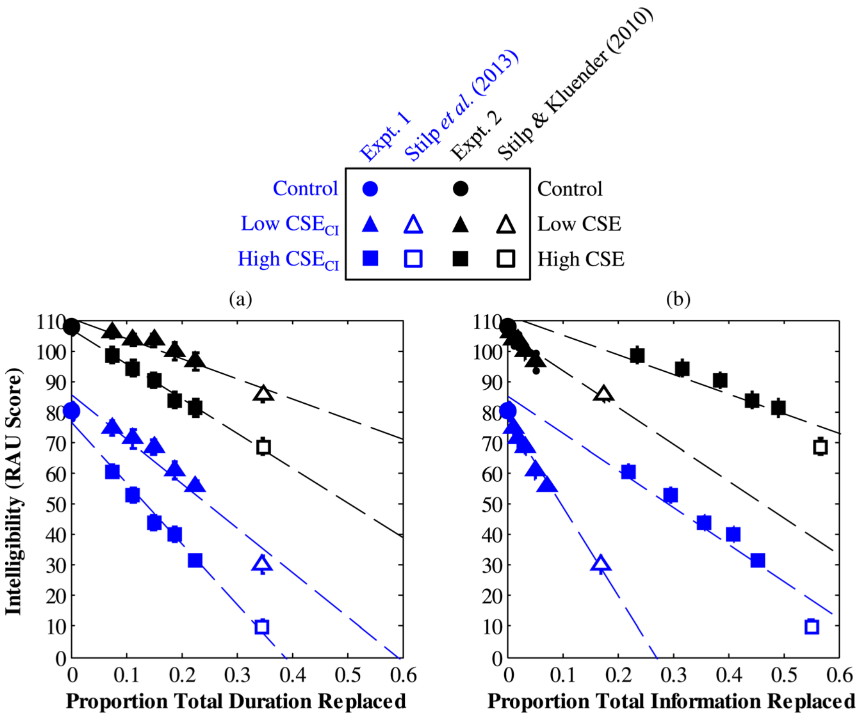

FIG. 5. (Color online) Present results are highly congruent with sentence intelligibility in previous investigations of information-bearing acoustic changes (Stilp and Kluender, 2010 [80-ms low-CSE and high-CSE results]; Stilp *et al.*, 2013). Error bars indicate standard errors; dashed lines indicate regression fits to data from a given condition (low or high CSE$_{CI}$, low or high CSE; filled shapes) including its paired condition from previous studies (hollow shapes, see legend). Circles indicate control conditions; triangles indicate conditions where low-CSE$_{CI}$/CSE intervals were replaced by noise; squares indicate conditions where high-CSE$_{CI}$/CSE intervals were replaced. Mean intelligibility scores are presented as a function of the proportion of total sentence duration replaced by noise (a) and as a function of the proportion of information-bearing acoustic changes replaced (b).

sentences (Stilp *et al.*, 2013; Experiment 1). Second, CSE shares a graded relationship with sentence intelligibility, such that replacing incrementally more information-bearing acoustic changes with noise progressively worsens performance. This was first reported by Stilp and Kluender (2010), who varied the amount of CSE replaced by noise but fixed the total proportion of sentence duration replaced. Here, this graded relationship maintains across a wide range of sentence proportions replaced, similar to the highly linear relationship observed in Experiment 1. Third, the proportion of sentence duration replaced was a significant predictor of intelligibility, consistent with investigations of intelligibility of noise-interrupted, full-spectrum speech (e.g., Miller and Licklider, 1950; Fogerty and Kewley-Port, 2009; Lee and Kewley-Port, 2009; Kidd and Humes, 2012; Wang and Humes, 2010; Fogerty *et al.*, 2012; Fogerty, 2013), but proportion of information-bearing acoustic changes replaced proved to be a superior predictor of performance.

Across experiments, information-bearing acoustic changes became more important for speech perception as listening conditions worsened. Regression slopes indicate that sentence intelligibility decreased faster with increasing noise replacements (whether on the basis of duration or information) in vocoded materials compared to full-spectrum materials. This observation is separable from differences in regression intercepts, which reflect global differences in performance for full-spectrum materials versus vocoded materials with minimal practice and no feedback. This difference in perceptual weighting may be highly revealing for listeners with impaired hearing such as CI users, as they may rely more heavily on information-bearing acoustic changes than normal-hearing listeners to understand speech.

Finally, results are highly consistent with previously published findings by Stilp and Kluender (2010) and Stilp *et al.* (2013) (Fig. 5). Perception behaved very systematically across wide ranges of proportions of total sentence duration replaced [Fig. 5(a)] and total information replaced [Fig. 5(b)]. Observing this systematicity in perception of full-spectrum and noise-vocoded sentences suggests that exploiting information (i.e., change) in the speech signal is a fundamental operating characteristic of auditory perception.

IV. GENERAL DISCUSSION

Sensory systems respond primarily to change, and the auditory system is no exception. Unpredictable or changing stimuli are expected to be most informative for perception, and the auditory system enhances sensitivity to change to maximize the amount of information that can be transmitted. Previous research revealed the high perceptual importance of information-bearing acoustic changes for perception of sentences that were spectrally intact (Stilp and Kluender, 2010; Jiang *et al.*, 2013) or degraded by noise vocoding (Stilp *et al.*, 2013). However, important questions remained regarding the relationship between these different measures of information-bearing acoustic change, the nature of their relationships with speech intelligibility, and how they compared to other well-established predictors of perceptual performance. The present experiments set three goals to address these questions.

The first goal was to directly compare the predictive power of information-bearing acoustic changes versus proportion of sentence duration replaced for sentence intelligibility. Information-bearing acoustic changes closely corresponded to perceptual performance amid different forms of temporal distortion, across variable speaking rates, in full-spectrum as well as noise-vocoded sentences, and in different languages (Stilp and Kluender, 2010; Stilp *et al.*, 2010, 2013; Jiang *et al.*, 2013; Experiments 1 and 2). The proportion of total sentence duration preserved (PTD) also reliably predicted intelligibility of sentences (e.g., Fogerty and Kewley-Port, 2009; Lee and Kewley-Port, 2009; Fogerty *et al.*, 2012) and words amid noise interruptions (Wang and Humes, 2010; Fogerty *et al.*, 2012; Kidd and Humes, 2012; Fogerty, 2013). However, investigations of how closely measures of information or time correspond to perceptual performance were relatively agnostic to the role the other factor played, making it impossible to determine whether one factor better explained listener performance.

Consistent with previous research, proportions of information-bearing acoustic changes (CSE, $\mathrm{CSE_{CI}}$) replaced and total sentence duration replaced (PTD) were each significant predictors of sentence intelligibility, but several indices revealed that measures of information corresponded more closely to listener performance. First, multiple regression analyses revealed larger coefficients for proportion of total information replaced than PTD replaced. Second, when a single regression was fit to all results from a given experiment (solid lines in Figs. 3 and 4), information-bearing acoustic change accounted for far more variability in listeners' responses than PTD ($r^2 = 0.87$ versus $r^2 = 0.52$ in Experiment 1; $r^2 = 0.85$ versus $r^2 = 0.46$ in Experiment 2). Third, plotting intelligibility as a function of PTD revealed the unsuitability of fitting a single regression to all results [solid lines in Figs. 3(a) and 4(a)]. Performance in low-$\mathrm{CSE_{CI}}$/CSE-replaced conditions consistently outperformed regression predictions while performance in high-$\mathrm{CSE_{CI}}$/CSE-replaced conditions consistently underperformed predictions. Results were more conducive to fitting low-$\mathrm{CSE_{CI}}$/CSE and high-$\mathrm{CSE_{CI}}$/CSE results separately, indicating systematic differences in performance that measures of PTD alone could not capture.

The second goal was to elucidate the relationship between information-bearing acoustic changes measured in noise-vocoded sentences ($\mathrm{CSE_{CI}}$) and intelligibility of these sentences amid noise interruption. Information-bearing acoustic changes measured in full-spectrum sentences (CSE) share a graded, highly linear relationship with sentence intelligibility, such that replacing progressively greater amounts of CSE with noise systematically worsened performance (Stilp and Kluender, 2010). Results of Stilp *et al.* (2013) intimated that this close correspondence may extend to $\mathrm{CSE_{CI}}$ and perception of noise-vocoded sentences as well. Experiment 1 confirmed that information-bearing acoustic changes share a graded relationship with intelligibility of vocoded sentences; this relationship maintained across broad ranges of sentence duration proportions replaced [Fig. 3(a)] as well as sentence information proportions replaced [Fig. 3(b)]. This linear relationship was also observed in Experiment 2, which tested the same experimental conditions but using full-spectrum sentences. Results confirm fundamental principles of perception that are shared across healthy hearing and acoustic simulations of electrical hearing.

The third goal was to determine whether the perceptual importance of information-bearing acoustic changes diminishes, maintains, or increases in poorer listening conditions. As larger proportions of sentence duration were replaced by noise, performance decreased more quickly when high-information intervals were replaced compared to when low-information intervals were replaced. This rate of decrease (as quantified by regression slopes) was more pronounced for vocoded sentences in Experiment 1, indicating perception relied more heavily on information-bearing acoustic changes for understanding these more challenging materials than the full-spectrum sentences of Experiment 2. This is consistent with the fact that noise vocoding removes fundamental frequency, harmonicity, temporal fine structure, and many other cues that aid in speech intelligibility. Removing these cues increases the perceptual importance of remaining cues for speech recognition, and Experiment 1 and Stilp *et al.* (2013) clearly indicate that information-bearing acoustic change is among these remaining cues. This may be highly significant for speech perception by CI users, as they may rely on information-bearing acoustic changes more heavily than normal-hearing listeners.

A growing body of research suggests that measures of information better explain speech perception than measures of time. Experiments by Stilp and colleagues (Stilp and Kluender, 2010; Jiang *et al.*, 2013; Stilp *et al.*, 2013) demonstrated the perceptual significance of information-bearing acoustic changes when proportion of sentence duration replaced was held constant; the present results reveal information to outperform duration in predicting sentence intelligibility when both factors are varied. Measures of amplitude modulations across time (i.e., modulation transfer function; Houtgast and Steeneken, 1985) cannot predict intelligibility of sentences with desynchronized spectral bands, but CSE reliably predicted intelligibility of these materials (Stilp *et al.*, 2010). Time-reversing short-duration segments (Saberi and Perrott, 1999) have dramatically different effects on sentence intelligibility depending on the speaking rate, but CSE was a rate-invariant metric of perceptually significant information for understanding these materials (Stilp *et al.*, 2010). Finally, Alexander and Kluender (2010) demonstrated that perceptual calibration to reliable spectral properties in a listening context is determined by the accumulation of evidence for that regularity, not its mere duration. Together, results emphasize the perceptual significance of what information is transmitted over any span of time.

Information-bearing acoustic changes in the speech signal are important for speech perception by listeners with healthy hearing (Alexander and Kluender, 2008; Stilp and Kluender, 2010; Stilp *et al.*, 2010; Jiang *et al.*, 2013; Experiment 2), sensorineural hearing loss (Alexander and Kluender, 2009), in simulations of hearing loss (Alexander *et al.*, 2011), and acoustic simulations of electrical hearing (Stilp *et al.*, 2013; Experiment 1). Optimizing sensitivity to change is a fundamental principle of how sensory systems operate (Kluender *et al.*, 2003); this is equally true for normal and impaired hearing. With its strong empirical foundation in speech perception by normal-hearing listeners, this information-theoretic perspective holds great promise for better understanding and ameliorating challenges presented by hearing impairment.

Information-theoretic approaches to speech perception may also bear considerable importance for assistive devices such as CIs. In current CI processing strategies, channels are selected for stimulation according to their (high) amplitudes, treating amplitude as the most important acoustic property for speech perception. This approach neglects other important details in the speech signal, leading to suggestions that something other than amplitude may be most effective for encoding speech in CIs (Loizou, 1998; Wilson, 2006). CI processing strategies do not process speech intervals with greater versus lesser information-bearing acoustic changes any differently, despite the fact that they contribute very differently to speech perception. This indifference to perceptually significant acoustic changes in speech inherently limits the amount of information that CIs transmit for speech perception. Recent findings reveal that, similar to NH (normal-hearing) listeners (e.g., Viemeister, 1980; Viemeister and Bacon, 1982), CI users exhibit auditory enhancement effects where acoustic changes between sounds are perceptually emphasized (Goupell and Mostardi, 2012; Wang *et al.*, 2012). This perceptual enhancement of acoustic changes suggests that CI users may also exploit information-bearing acoustic changes to understand speech. By measuring and emphasizing information-bearing acoustic changes, more information that is important for perception may be transmitted by the CI, which is predicted to improve CI users' speech perception.

Along with Stilp and colleagues (2013), the present results confirm a strong relationship between speech perception and information in speech signals that is available in acoustic simulations of electrical hearing. This relationship raises questions regarding the spectral and temporal properties of CSE_{CI}. A sizable literature has explored changes in speech intelligibility when varying the number of spectral channels (e.g., Friesen *et al.*, 2001; Loizou *et al.*, 1999; Shannon *et al.*, 1995) and cutoff frequency of amplitude envelopes (e.g., Drullman *et al.*, 1994 in full-spectrum speech; Xu *et al.*, 2005 and Xu and Zheng, 2007 in vocoded phonemes). In the present effort, all vocoded sentences had eight spectral channels and 150-Hz cutoffs for amplitude envelopes in each channel. One major conclusion drawn here is that information-bearing acoustic changes become more perceptually important in poorer listening conditions. This was supported by steeper regression slopes for perception of vocoded sentences compared to full-spectrum sentences, and larger decreases in performance when more high-CSE_{CI}/CSE intervals were replaced by noise compared to when more low-CSE_{CI}/CSE intervals were replaced. However, evaluating the role of CSE_{CI} in further degraded speech stimuli introduces competing predictions. If information-bearing acoustic changes in fact become more important for speech intelligibility as listening conditions worsen, as the present results suggest, differences in performance upon replacing low-CSE_{CI} versus high-CSE_{CI} intervals will increase (e.g., larger differences between regression slopes) as the number of spectral channels or amplitude envelope cutoff frequency decreases. Conversely, further reducing spectral or temporal resolution (i.e., fewer than eight spectral channels, envelope cutoff frequencies below 150 Hz) lowers overall signal quality and may make low-CSE_{CI} and high-CSE_{CI} changes less distinct from each other. In this situation, replacing low-CSE_{CI} versus high-CSE_{CI} intervals may have increasingly similar effects on performance. Further research will be necessary to tease apart these competing predictions for perceptual performance.

The present results make great strides in revealing fundamental principles of speech perception that are shared across spectrally rich and degraded materials. The close correspondence between information-bearing acoustic changes and sentence intelligibility further encourages information-theoretic approaches to speech perception (e.g., Kluender and Alexander, 2007; Stilp and Kluender, 2010; Stilp *et al.*, 2010, 2013; Jiang *et al.*, 2013; Kluender *et al.*, 2013) and perception most broadly. This approach holds great promise for better understanding and ameliorating challenges presented by hearing impairment.

ACKNOWLEDGMENTS

Research was supported by funds from the University of Louisville. We thank Ijeoma Okorie and Andrew McPheron for their efforts in data collection and analysis, and Pavel Zahorik and Terry Nearey for valuable feedback.

[1] Loizou *et al.* (1999) reported 90% (roughly 93 RAU) intelligibility for TIMIT sentences with only five spectral channels, but considerable differences in presentation order and practice complicates comparison of results.

[2] The interaction term between proportional time and proportional information was not included in multiple regression analyses, as it did not contribute any variability beyond the main effects. While these measures are near-perfectly correlated when analyzing low-CSE_{CI} or high-CSE_{CI} conditions alone ($r \geq 0.96$), this correlation is significantly weakened when analyzing both conditions together ($r = 0.44$, $p = 0.18$). This is due to the fact that values for proportional information increase monotonically across low- and high-CSE_{CI} conditions [see values along abscissa of Fig. 3(b)], but values for proportional sentence duration repeat [see values along abscissa of Fig. 3(a)]. Similar results are observed for Experiment 2 ($r = 0.40$, $p = 0.22$).

Alexander, J. M., Jenison, R. L., and Kluender, K. R. (**2011**). "Real-time contrast enhancement to improve speech recognition," PLoS One **6**(9), e24630.

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

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

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.

Attneave, F. (**1954**). "Some informational aspects of visual perception," Psychol. Rev. **61**, 183–193.

Barlow, H. B. (**1961**). "Possible principles underlying the transformations of sensory messages," in *Sensory Communication*, edited by W. A. Rosenblith (MIT Press, Cambridge, MA and John Wiley, New York), pp. 53–85.

Barlow, H. B. (**2001**). "Redundancy reduction revisited," Network: Comput. Neural Syst. **12**, 241–253.

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 *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'96)*, Atlanta, GA, pp. 853–856.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**(2), 1053–1064.

Fogerty, D. (**2013**). "Acoustic predictors of intelligibility for segmentally interrupted speech: Temporal envelope, voicing, and duration," J. Speech Lang. Hear. Res. **56**(5), 1402–1408.

Fogerty, D., and Humes, L. E. (**2010**). "Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors," J. Acoust. Soc. Am. **128**(5), 3114–3125.

Fogerty, D., and Kewley-Port, D. (**2009**). "Perceptual contributions of the consonant-vowel boundary to sentence intelligibility," J. Acoust. Soc. Am. **126**(2), 847–857.

Fogerty, D., Kewley-Port, D., and Humes, L. E. (**2012**). "The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss," J. Acoust. Soc. Am. **132**(3), 1667–1678.

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.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (**1990**). "DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM," *NIST Order No. PB91-505065, National Institute of Standards and Technology, Gaithersburg, MD*.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Deviation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

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), 2592–2605.

Houtgast, T. and Steeneken, H. J. M. (**1985**). "A review of the MTF-concept in room acoustics," 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.

Kidd, G. R., and Humes, L. E. (**2012**). "Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences," J. Acoust. Soc. Am. **131**(2), 1434–1448.

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 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 Commun. **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), pp. 117–151.

Lee, J. H., and Kewley-Port, D. (**2009**). "Intelligibility of interrupted sentences at sub-segmental levels in young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **125**(2), 1153–1163.

Li, N., and Loizou, P. C. (**2007**). "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am. **122**(2), 1165–1172.

Loizou, P. C. (**1998**). "Mimicking the human ear," IEEE Signal Process. Mag. **15**(5), 101–130.

Loizou, P. C., Dorman, M. F., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**(4), 2097–2103.

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

Olshausen, B. A., and Field, D. J. (**1996**). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature **381**, 607–609.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**(6), 1788–1803.

Saberi, K., and Perrott, D. R. (**1999**). "Cognitive restoration of reversed speech," Nature **398**, 760.

Shannon, C. E. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423, 623–656.

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

Simoncelli, E. P. (**2003**). "Vision and the statistics of the visual environment," Curr. Opin. Neurobiol. **13**, 144–149.

Simoncelli, E. P., and Olshausen, B. A. (**2001**). "Natural image statistics and neural representation," Ann. Rev. Neurosci. **24**, 1193–1216.

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. (**2010**). "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 spectral entropy, not consonants, vowels, or time, best predicts speech intelligibility," Proc. Natl. Acad. Sci. U.S.A. **107**(27), 12387–12392.

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

Viemeister, N. F. (**1980**). "Adaptation of masking," in *Psychophysical, Physiological, and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Noordwijkerhout, The Netherlands), pp. 190–198.

Viemeister, N. F., and Bacon, S. P. (**1982**). "Forward masking by enhanced components in harmonic complexes," J. Acoust. Soc. Am. **71**(6), 1502–1507.

Wang, X., and Humes, L. E. (**2010**). "Factors influencing recognition of interrupted speech," J. Acoust. Soc. Am. **128**(4), 2100–2111.

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

Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science **167**(3917), 392–393.

Wilson, B. S. (**2006**). "Speech processing strategies," in *Cochlear Implants: A Practical Guide*, 2nd ed., edited by H. R. Cooper and L. C. Craddock (Whurr Publishers Ltd., London), pp. 21–69.

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., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**(3), 1758–1764.