Information-Bearing Acoustic Changes Are Important for Understanding Vocoded Speech in a Simulation of Cochlear Implant Processing Strategies

5aSC9

Christian Stilp, Department of Psychological and Brain Sciences, University of Louisville

INTRODUCTION

- Sensory systems are optimally sensitive to changes in the input. This sensitivity plays a foundational role in perception of stimuli in the environment including speech (Kluender *et al.*, 2003).
- Information-bearing acoustic changes (IBACs) in the speech signal are important for understanding noise-vocoded speech:
  - IBACs were measured using cochlea-scaled entropy for cochlear implants (CSE_{CI}; Stilp *et al.*, 2013)
  - IBACs were of comparable importance for speech perception whether measured in noise-vocoded or full-spectrum speech (Stilp *et al.*, 2013)
  - IBACs were a better predictor of interrupted sentence intelligibility than proportion of sentence duration replaced by noise (Stilp, 2014)
  - Perceptual importance of IBACs maintained across wide ranges of spectral and temporal resolutions (Stilp & Goupell, under review)
- Noise vocoding is analogous to cochlear implant processing strategies that present acoustic information in all channels at all times (*e.g.*, Continuous Interleaved Sampling [CIS]). However, vocoding significantly departs from other processing strategies that present only the *n*-highest-amplitude channels out of *m* at any given time (*n*-of-*m* processing; *e.g.*, Advanced Combination Encoder [ACE]).
- Does the importance of IBACs for understanding speech maintain for sparser spectra such as those delivered by *n*-of-*m* processing? Also, does the metric capture perceptually significant changes in the speech spectrum on more rapid timescales such as those used by *n*-of-*m* processing (changes between 1-ms spectra, as opposed to 16 ms used in previous studies)?

METHODS

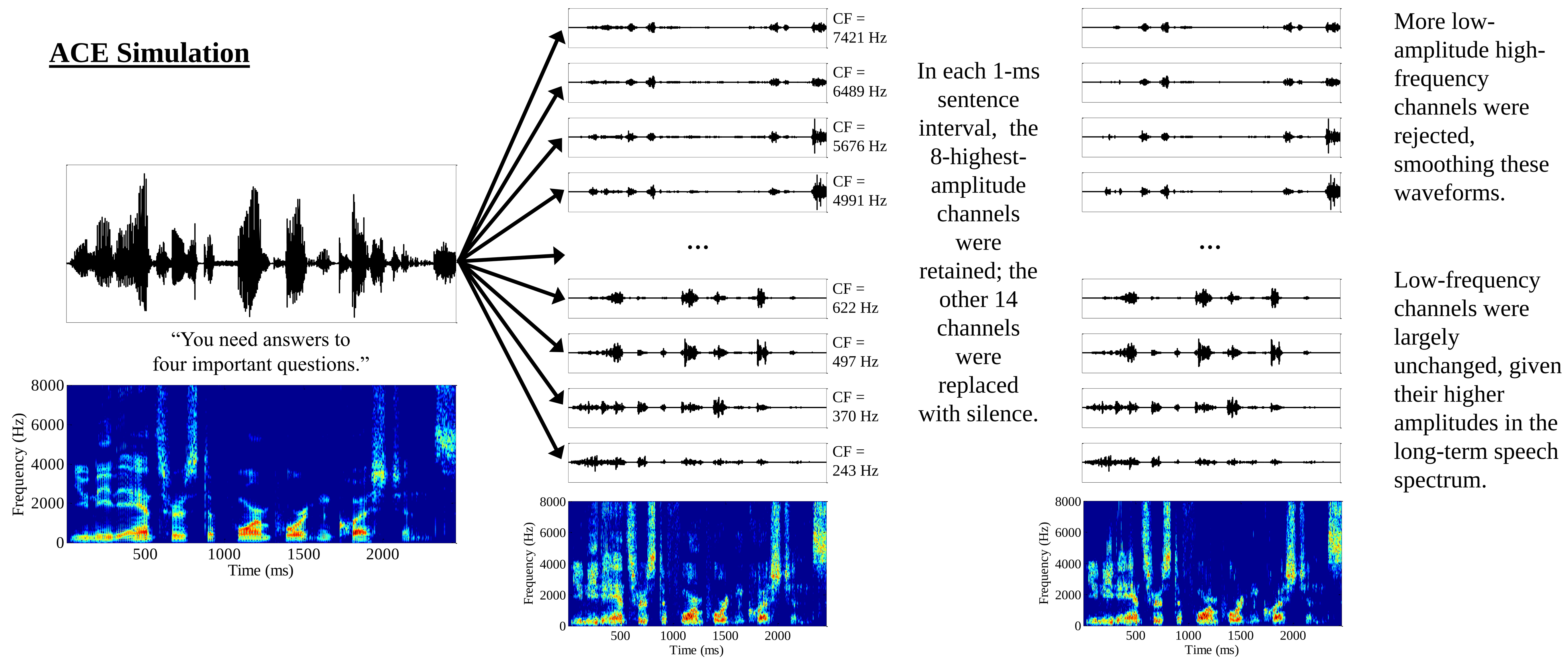
Participants
- 20 native English speakers with normal hearing

Stimuli
- 50 sentences from the TIMIT database (same stimuli as Stilp *et al.*, 2013)
- Sentences were noise-vocoded with 22 channels spanning 188-7938 Hz according to Greenwood's formula (see Goupell & Litovsky, 2014 for channel center and cutoff frequencies)
  - Channels were extracted using 4th order Butterworth filters, then half-wave rectified and low-pass filtered by 2nd-order Butterworth filters at 150 Hz to obtain amplitude envelopes
- Vocoded sentences were divided into 1-ms segments. In each segment, only the 8 highest-amplitude channels were retained, simulating *n*-of-*m* (8 channels out of 22) ACE processing at 1000 pulses/second stimulation rate

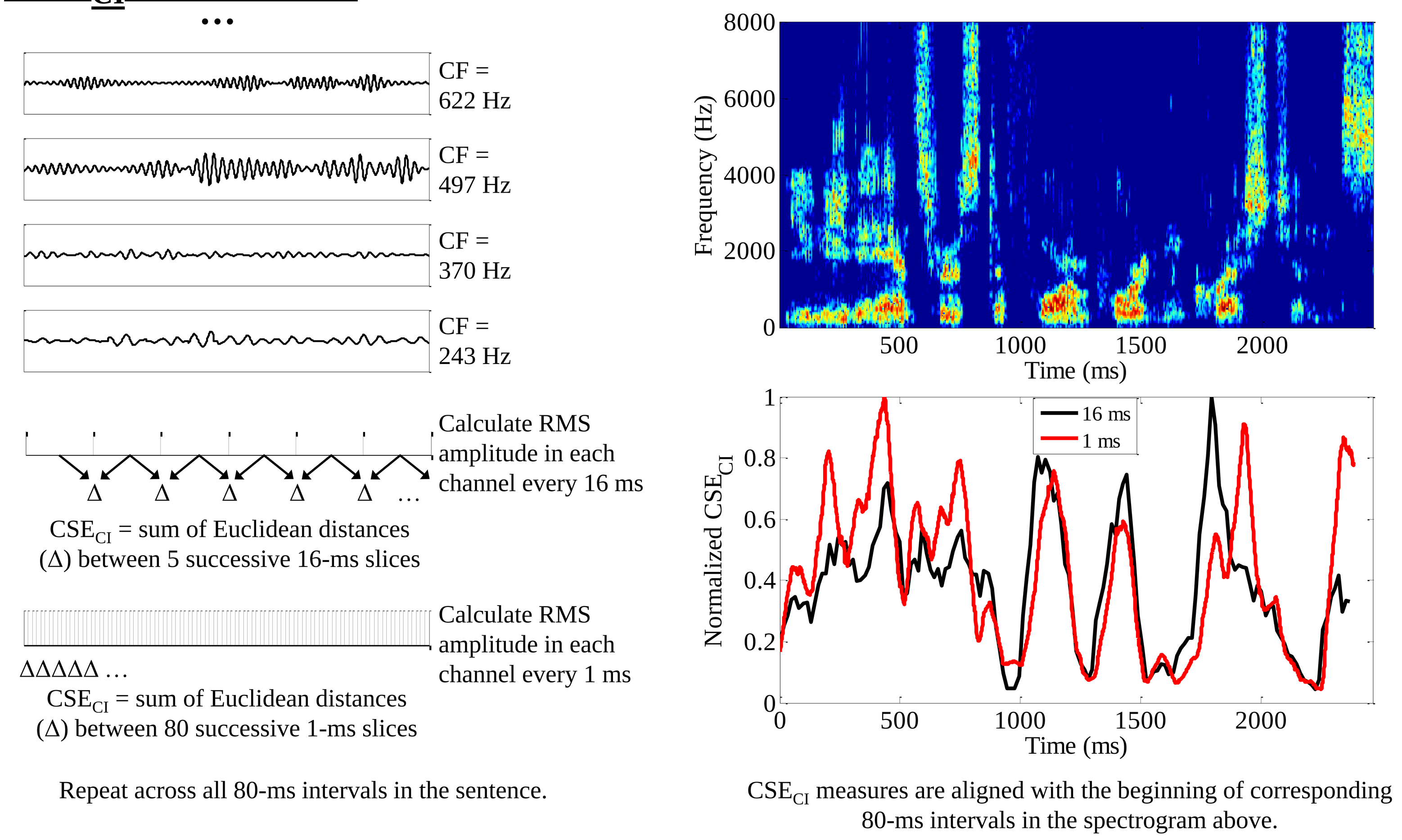
CSE_{CI}
- Spectral slices were either 16 ms (following Stilp *et al.* and others) or 1 ms
- Euclidean distances were calculated between all successive spectral slices (RMS-amplitude-profiles across all 22 spectral channels)
- CSE_{CI} was the summed distances between 5 16-ms spectral slices or 80 1-ms spectral slices
- Four 80-ms intervals with either the highest or lowest CSE_{CI} were replaced with speech-shaped noise; control sentences had no noise replacement

Procedure
- On each trial, one sentence was presented diotically at 70 dB SPL over circumaural headphones; no listener heard any sentence twice

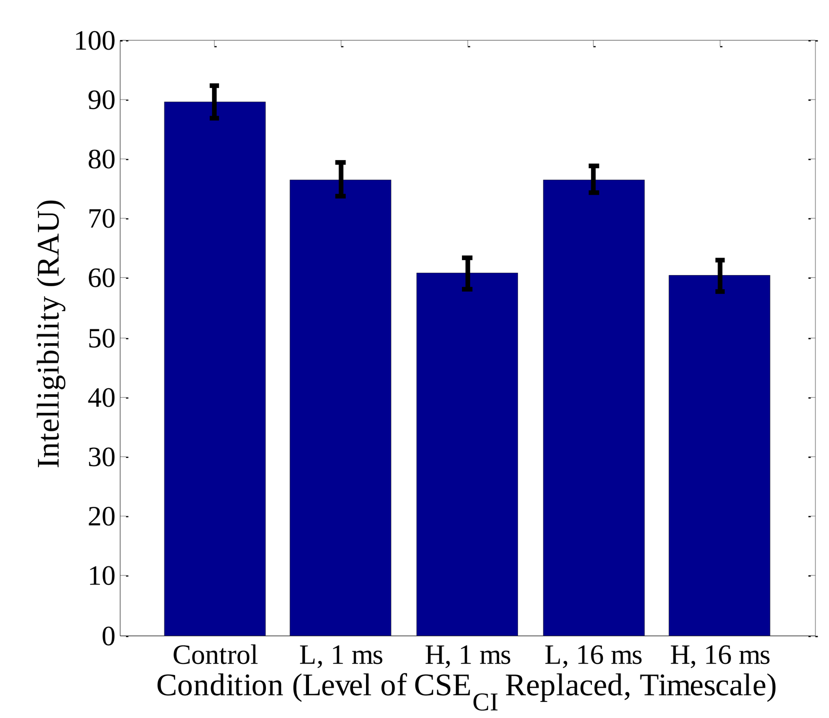
ACE Simulation



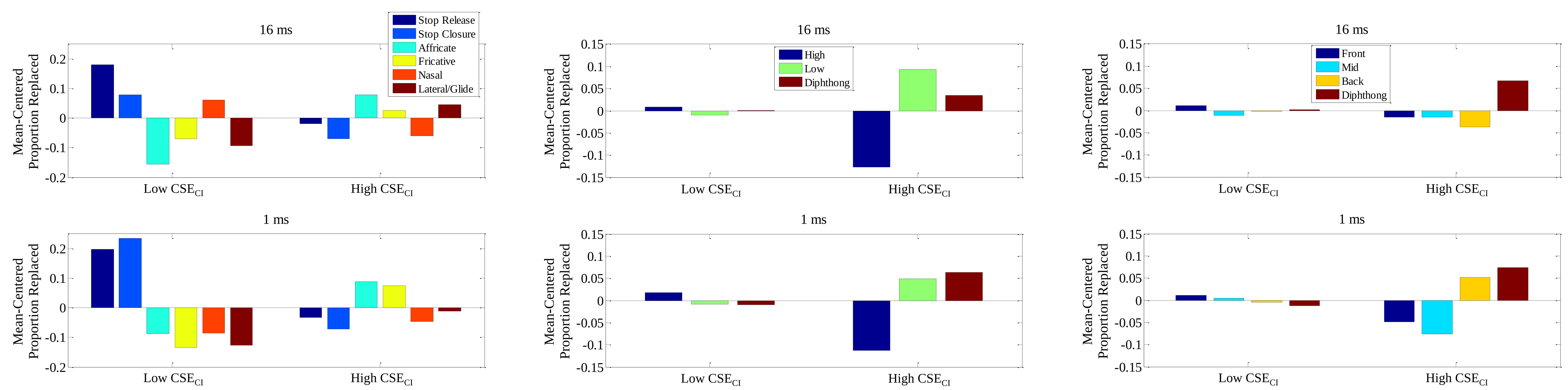
CSE_{CI} Calculation



Results



Phonemic Analysis: Which speech sounds were being replaced by noise?



RESULTS AND DISCUSSION

- A repeated-measures ANOVA revealed a significant main effect of level of CSE_{CI} replaced by noise ($F_{1,19} = 53.51$, $p < .001$, $\eta_p^2 = 0.74$). Relative to the control condition, performance decreased by 13 RAU when low-CSE_{CI} intervals were replaced by noise, and decreased by an additional 16 RAU when high-CSE_{CI} intervals were replaced. This pattern occurred for CSE_{CI} measured on both 1-ms (paired-samples *t*-test on high vs. low results: $t_{19} = 5.63$, $p < .001$) and 16-ms timescales ($t_{19} = 6.15$, $p < .001$). Window duration and the interaction were not statistically significant (both $F < 0.1$).
- Perceptually significant changes in the speech signal are maintained in a simulation of *n*-of-*m* processing. Similar to previous studies, replacing high-CSE_{CI} intervals impaired sentence intelligibility more than replacing low-CSE_{CI} intervals. While promising, this is a relatively primitive emulation of ACE-style processing. Relationships between IBACs and CI parameters / sequelae such as compression, stimulation rate, and spread of excitation require further investigation.
- Results were consistent across 1-ms and 16-ms slice durations, generalizing the timescale of IBACs to very rapid spectral changes. In both cases, however, 80-ms sentence intervals were replaced by noise. Different combinations of slice duration and interval duration may modulate the importance of IBACs for speech understanding.
- Patterns of phoneme replacement are largely consistent with those in full-spectrum speech on the basis of CSE (Stilp & Kluender, 2010):
  - Stops rated as lower CSE_{CI}. The broadband nature of CSE_{CI} identified lesser absolute changes given stops' lower amplitudes, despite local spectral changes in formant transitions.
  - Affricates and fricatives rated as higher-CSE_{CI}, reflecting spectral variability in frication noise over short timescales, especially 1 ms.
  - Low and diphthongal vowels rated as higher-CSE_{CI}, reflecting considerable formant kinematics. High vowels are much less kinematic, and rated as lower-CSE_{CI}.
  - Further analyses are needed to ascertain why front and mid vowels rated as lower CSE_{CI} and back vowels as higher CSE_{CI}, especially in 1-ms conditions. These vowels showed no clear pattern in Stilp and Kluender (2010).
- Results and patterns of phoneme replacement reflect IBACs measured on a spectrally broad scale (*i.e.*, across all vocoder channels). Revising CSE_{CI} to assess perceptually significant intervals in speech on a narrowband (within-channel) basis will be illuminating.
- IBACs are important for understanding full-spectrum speech (Stilp & Kluender, 2010; Stilp, 2014), CIS-style vocoded speech (Stilp *et al.*, 2013; Stilp, 2014; Stilp & Goupell, under review), and ACE-style vocoded speech. This further suggests that IBACs are likely available and important for speech perception by CI users. Results may lend new insights to CI processing strategies and improved speech perception.

REFERENCES

Goupell, M.G., & Litovsky, R.Y. (2014). *JARO.* 15(1), 115-129.
Kluender, K.R., Coady, J.A., & Kiefte, M. (2003). *Sp. Comm.* 41(1), 59-69.
Stilp, C.E. (2014). *JASA.* 135(3), 1518-1529.
Stilp, C.E., & Goupell, M.J. (revision under review). *JASA.*
Stilp, C.E., Goupell, M.J., & Kluender, K.R. (2013). *JASA,* 133(2), EL136-141.
Stilp, C.E., & Kluender, K.R. (2010). *PNAS,* 107(27), 12387-12392.