Defining Spectral and Temporal Resolutions of Information-Bearing Acoustic Changes for Understanding Noise-Vocoded Sentences

Christian E. Stilp, Department of Psychological and Brain Sciences, University of Louisville
Matthew J. Goupell, Department of Hearing and Speech Sciences, University of Maryland

2aSC18

INTRODUCTION

Information-bearing acoustic changes in the speech signal are highly important for speech perception. This has been demonstrated for:

- Informational changes measured in full-spectrum sentences using cochlea-scaled entropy (CSE; Stilp & Kluender, 2010; Stilp, 2014)
  - CSE measures differences between short-time spectra weighted and scaled to broadly mimic cochlear processing
- Informational changes measured in noise-vocoded sentences using a modified version of CSE (CSE_{CI}; Stilp, Goupell, & Kluender, 2013; Stilp, 2014)
  - CSE_{CI} measures differences between short-time spectra calculated across all channels of noise-vocoded speech
- Poorer sentence intelligibility when high-CSE / high-CSE_{CI} intervals were replaced by noise compared to replacing an equal number of low-CSE / low-CSE_{CI} intervals (Stilp & Kluender, 2010; Stilp *et al.*, 2013; Stilp, 2014)

Information-bearing acoustic changes have been suggested to be:

- More important for understanding speech in poorer listening conditions (Stilp, 2014)
- Utilized more heavily by CI users for understand speech (Stilp, 2014)

However, acoustic simulations of CI processing have used only a single set of vocoder parameters (8 spectral channels, 150-Hz cutoff for amplitude envelopes), limiting generalizability of results.

Here, spectral and temporal resolutions of noise-vocoded sentences were manipulated independently and jointly to reveal contributions of information-bearing acoustic changes to speech understanding across broad ranges of signal quality.

METHODS

Participants

- All native English speakers with no known hearing impairments
- n = 24 (Expt. 1), 30 (Expt. 2), or 24 (Expt. 3)

Stimuli

- TIMIT sentences filtered into spectral channels (4th-order Butterworth filters) equally spaced from 300-5000 Hz (Greenwood, 1990)
- Amplitude envelopes extracted by half-wave rectification and low-pass filtering (2nd-order Butterworth filter) and assigned to white noise carriers
  - Number of channels and low-pass filter cutoff varied by experiment
  - Zero-phase filtering doubled filter order while preserving temporal characteristics

CSE_{CI}

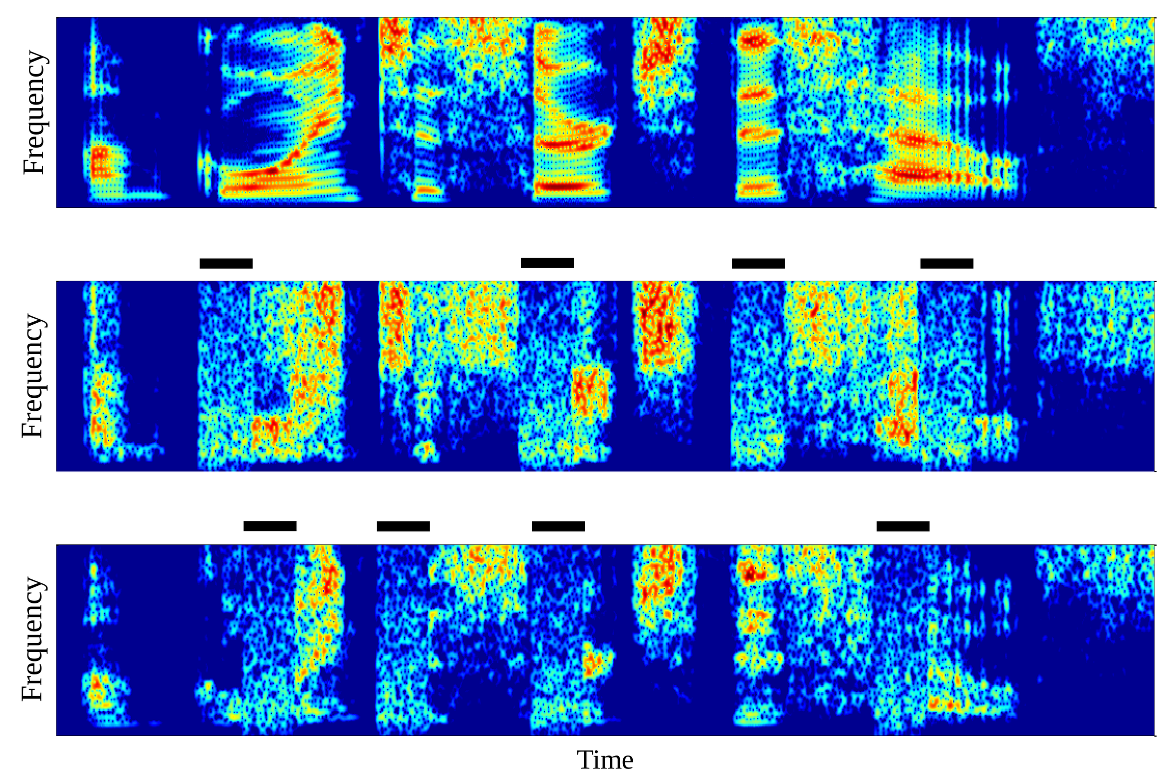
- CSE_{CI} = Euclidean distances between RMS-amplitude profiles of successive 16-ms sentences slices, summed into 80-ms intervals
- Four 80-ms intervals with high- or low-CSE_{CI} replaced with speech-shaped noise; control sentences had no noise replacement

Procedure

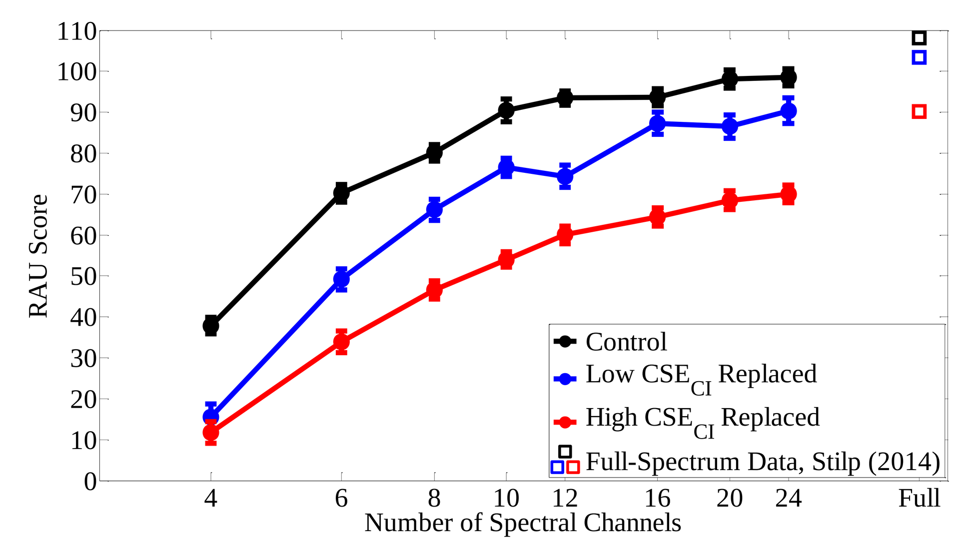
- Stimuli presented diotically at 70 dB SPL via circumaural headphones
- One sentence presented per trial; no listener heard any sentence twice
- Scores analyzed using repeated-measures ANOVA and Bonferroni-corrected t-tests

EXPT. 1: SPECTRAL RESOLUTION

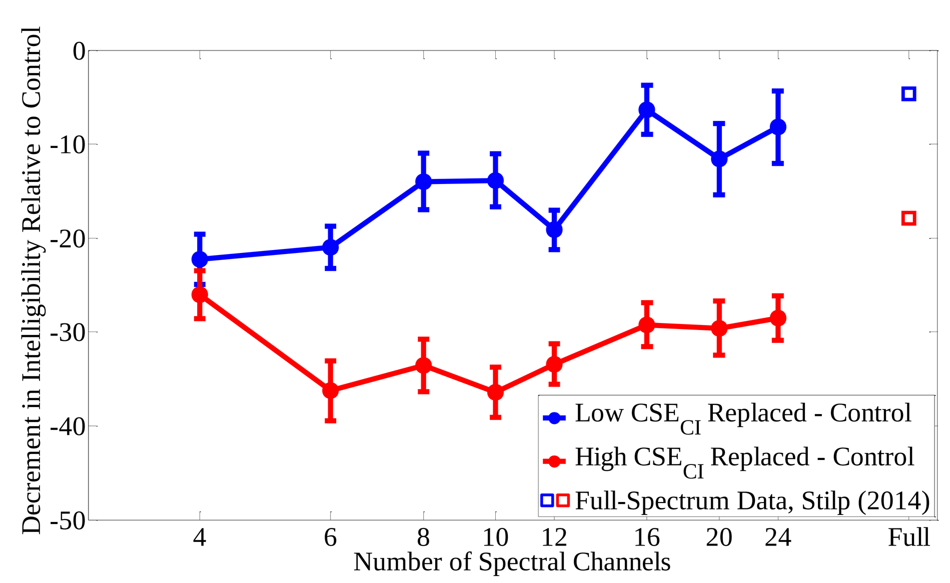
Spectral resolutions tested: 4, 6, 8, 10, 12, 16, 20, 24 channels
Temporal resolution tested: 150 Hz



The sentence "I'm going to search this house" at full spectral resolution (top), with 4 (middle) and 24 spectral channels (bottom). Replacement of high-CSE_{CI} intervals with noise is indicated by black lines above each spectrogram.



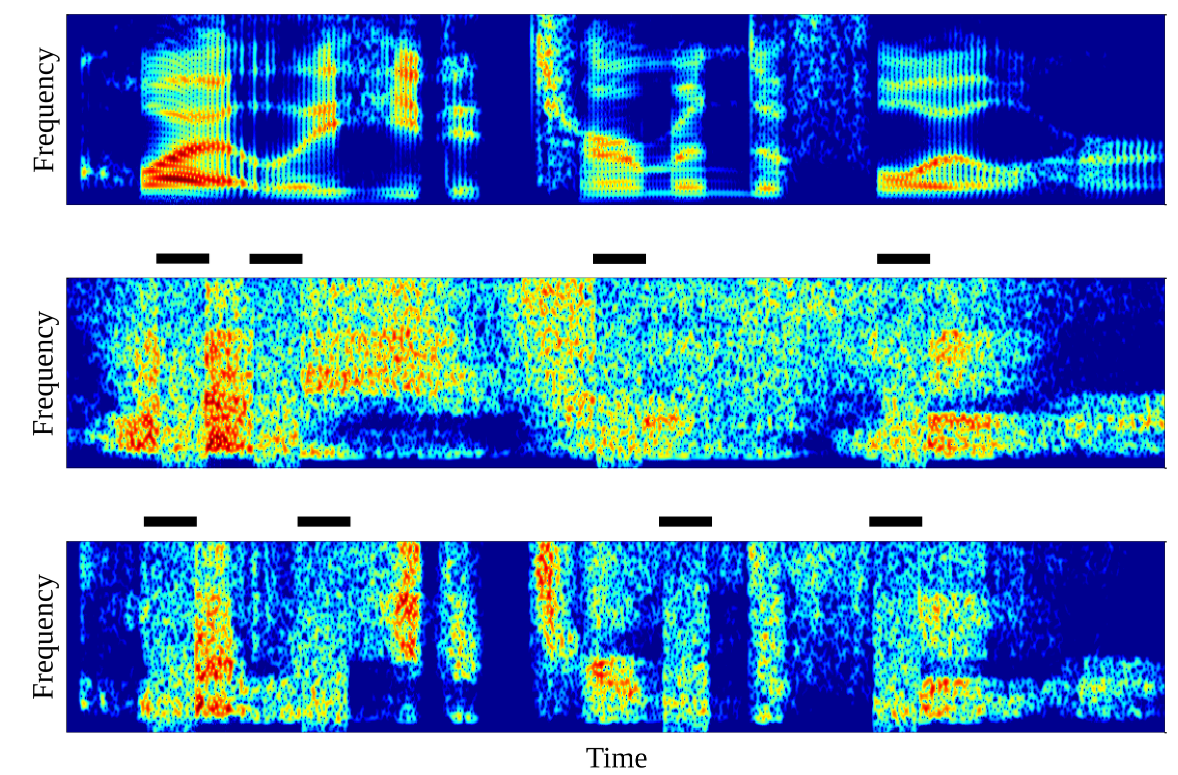
- Control > low-CSE_{CI}-replaced > high-CSE_{CI}-replaced ($p < .0001$)
- Performance improved with more spectral channels ($p < .0001$)
- Significant interaction ($p < .0001$)
  - Means recoded as decrements relative to control performance (larger decrement = greater perceptual importance of replaced segments)



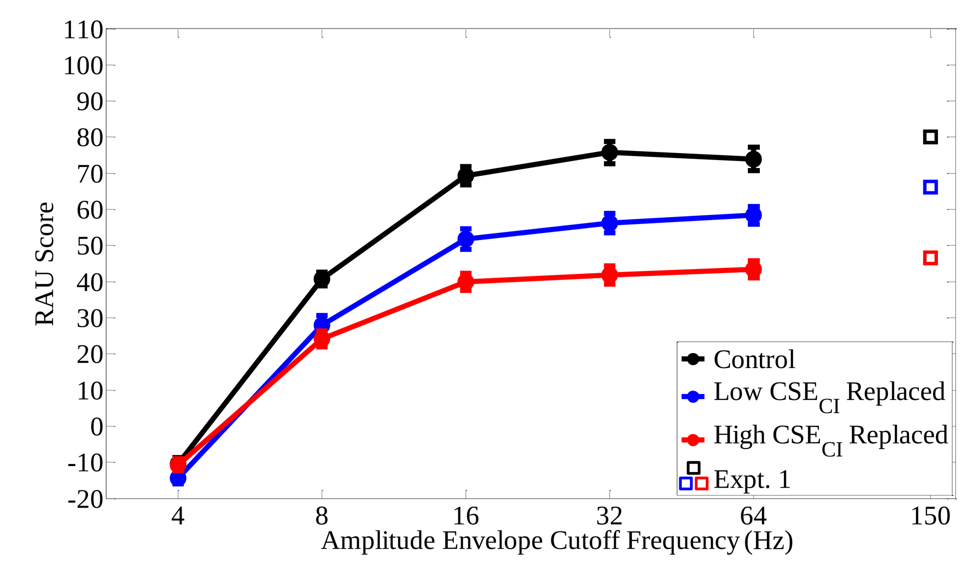
- Larger decrements when replacing high-CSE_{CI} changes than replacing low-CSE_{CI} changes at 6+ channels (all $p < .0007$)
- High-CSE_{CI} changes most important (produce largest decrements when replaced by noise) at 6-10 channels of spectral resolution

EXPT. 2: TEMPORAL RESOLUTION

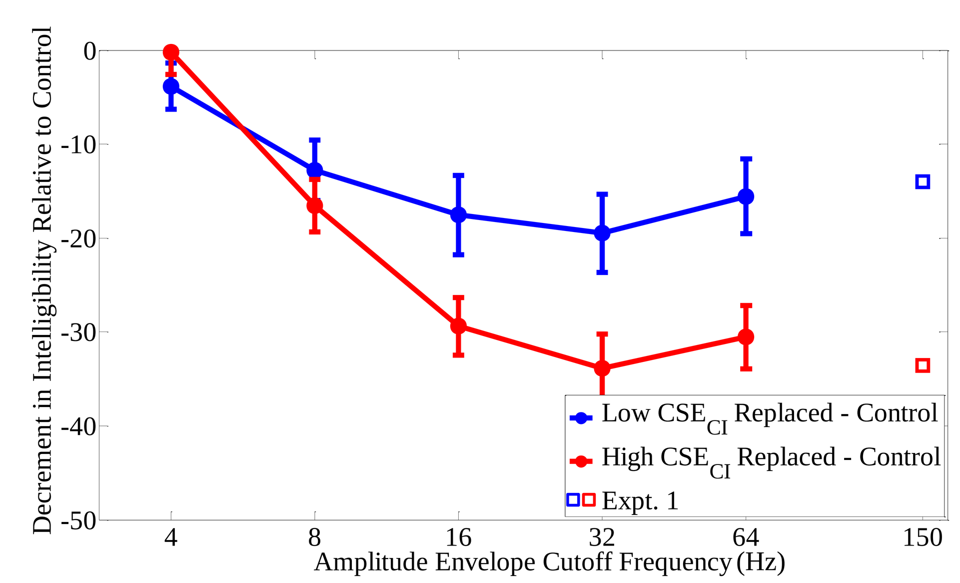
Spectral resolution tested: 8 channels
Temporal resolutions tested: 4, 8, 16, 32, 64 Hz



The sentence "Quietly, he determined to foil her" at full temporal resolution (top), with 4 Hz (middle) and 64 Hz envelope cutoff frequencies (bottom). Replacement of high-CSE_{CI} intervals with noise is indicated by black lines above each spectrogram.



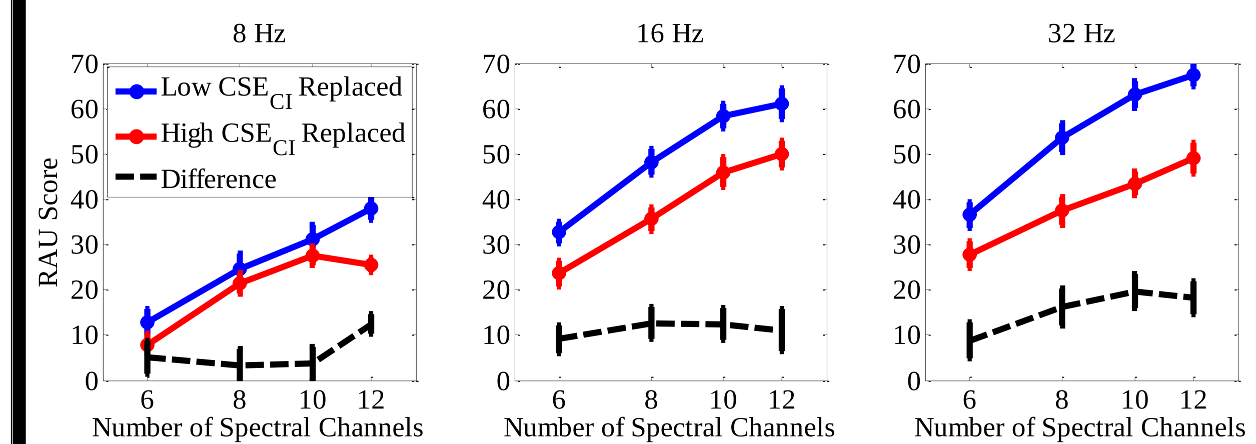
- Control > low-CSE_{CI}-replaced > high-CSE_{CI}-replaced ($p < .001$)
- Performance improved with better temporal resolution ($p < .0001$)
- Significant interaction ($p < .001$)
  - Means recoded as decrements relative to control performance (larger decrement = greater perceptual importance of replaced segments)



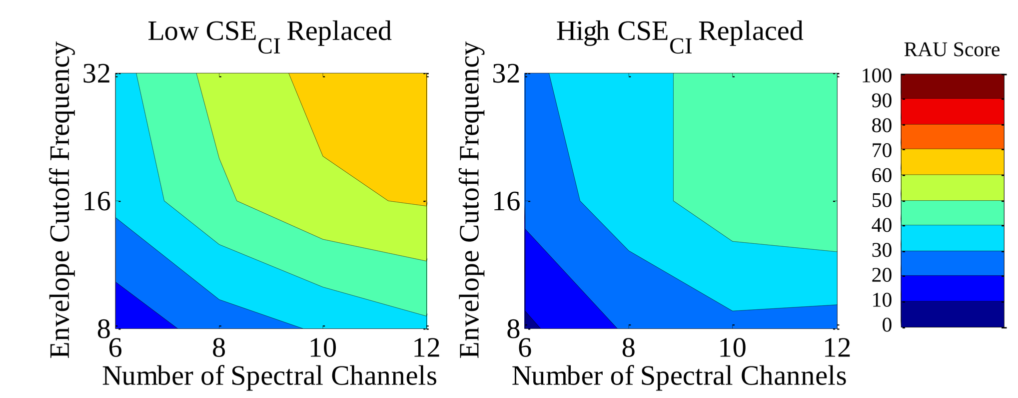
- Larger decrements when replacing high-CSE_{CI} changes than replacing low-CSE_{CI} changes at 16+ Hz (all $p < .004$)
- Apparent improvement at 64 Hz likely an artifact due to slight decrease for performance in 64-Hz control sentences; decrements not significantly different at 64 vs. 32 Hz

EXPT. 3: SPECTROTEMPORAL TRADEOFFS

Spectral resolutions tested: 6, 8, 10, 12 channels
Temporal resolutions tested: 8, 16, 32 Hz
No control sentences



- Across low- and high-CSE_{CI} conditions, greatest divergences in performance for 32-Hz temporal resolution or 8-10 spectral channels



- Greater spectrotemporal tradeoffs for low-CSE_{CI}-replaced sentences
- Larger tradeoffs observed at even lower spectral / temporal resolutions, but for phoneme recognition without noise replacements (Xu *et al.*, 2002; 2005), complicating comparisons to the present results.

CONCLUSIONS

Information-bearing acoustic changes were more important for speech intelligibility (*i.e.*, produced larger decrements) when:

- Spectral resolution decreased, consistent with Stilp (2014)
  - Greatest importance for understanding sentences with 6-10 channels
- Temporal resolution increased, inconsistent with Stilp (2014)
  - Floor effects below 16 Hz; flat performance at/above 16 Hz

Results indicate some low level of signal quality is necessary in order to distinguish low-CSE_{CI} intervals from high-CSE_{CI} intervals:

- Expt. 1: > 4 channels / 150 Hz
- Expt. 2: > 8 channels / 4-8 Hz
- Expt. 3: > 6-10 channels / 8 Hz

Modest evidence for greater spectrotemporal tradeoffs in low-CSE_{CI}-replaced sentences, but three-way interaction not statistically significant.

Results promote extending this approach to CI users. In both healthy and electrical hearing, the central auditory system similarly strives to be maximally sensitive to changes in the acoustic input. Results may lend new insights to CI processing strategies and improved speech perception.

REFERENCES

Greenwood (1990) *JASA*
Stilp (2014) *JASA*
Stilp, Goupell, & Kluender (2013) *JASA*
Stilp & Kluender (2010) *PNAS*
Xu, Tsai, & Pfingst (2002) *JASA*
Xu, Thompson, & Pfingst (2005) *JASA*