# Spectral and Temporal Resolution of Information-Bearing Acoustic Changes in Vocoded Sentences

PS-667


UNIVERSITY OF LOUISVILLE

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


## INTRODUCTION

Information-bearing acoustic changes in the speech signal are highly important for speech perception. This has been demonstrated when acoustic changes were measured in full-spectrum sentences using cochlea-scaled entropy (CSE; Stilp & Kluender, 2010) and later extended to noise-vocoded sentences using an adapted version of the CSE metric ($CSE_{CI}$; Stilp, Goupell, & Kluender, 2013). Across studies, sentence intelligibility was poorer when speech intervals containing high-CSE / high-$CSE_{CI}$ acoustic changes were replaced by noise compared to replacing an equal number of low-CSE / low-$CSE_{CI}$ intervals. Information-bearing acoustic changes have been suggested to be fundamental to speech perception most broadly (Stilp *et al.*, 2013).

Recent results reveal that these acoustic changes take on increased importance for understanding speech in poorer listening conditions. Sentence intelligibility decreased more quickly when information-bearing acoustic changes were replaced by noise in vocoded sentences than in full-spectrum sentences (Stilp, in press). This result suggests that CI users may rely more heavily on information-bearing acoustic changes to understand speech than normal-hearing listeners. However, acoustic simulations of CI processing have thus far used only a single set of vocoder parameters (eight spectral channels, 150-Hz cutoff for amplitude envelope extraction). This limited set of conditions obscures the breadth and depth of perceptual reliance upon these acoustic changes to understand vocoded speech, complicating generalizations to CI users' speech perception.

Spectral and temporal resolutions of noise-vocoded sentences were manipulated to reveal contributions of information-bearing acoustic changes to speech understanding across wide ranges of signal quality. The importance of these acoustic changes was measured using the decrement in performance (relative to control conditions with no noise replacement) when high-information (high-$CSE_{CI}$) intervals were replaced by noise compared to when low-information (low-$CSE_{CI}$) intervals were replaced.

## METHODS

**Sentences**

- 192 sentences from the TIMIT database (mean duration = 2064 ms)
- Noise vocoding: filtering into $x$ spectral channels (4$^{th}$-order Butterworth) equally spaced from 300-5000 Hz, with amplitude envelopes extracted by half-wave rectification and low-pass filtering at $y$ Hz (2$^{nd}$-order Butterworth)
  - Zero-phase filtering doubled filter order while preserving temporal characteristics
  - Experiment 1: $x$ = 4, 6, 8, 10, 12, 16, 20, 24 channels, $y$ = 150 Hz
  - Experiment 2: $x$ = 8 channels, $y$ = 4, 8, 16, 32, 64 Hz

**$CSE_{CI}$**

- $CSE_{CI}$ = 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 (Figs. 1, 4); 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

**Analysis**

- Scores arcsine-transformed (Studebaker, 1985), then analyzed using repeated-measures ANOVA and Bonferroni-corrected $t$-tests

## EXPERIMENT 1

Participants: 24 normal-hearing, native English speakers

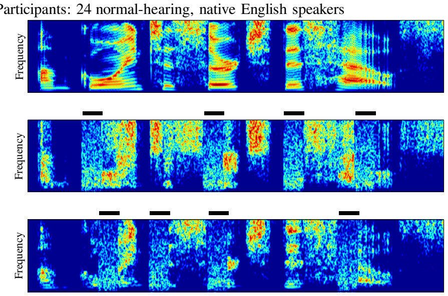


FIGURE 1. The sentence "I'm going to search this house" at full spectral resolution (top), with 4 (middle) and 24 spectral channels (bottom). Replacement of high-$CSE_{CI}$ intervals with noise is indicated by black lines above each spectrogram.

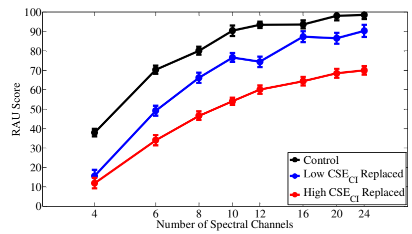


FIGURE 2. Mean sentence intelligibility in Expt. 1, measured in RAU scores. Error bars depict SEM.

- 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

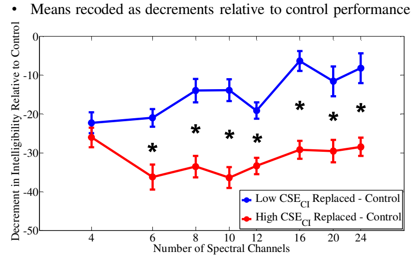


FIGURE 3. Decrements in performance for Expt. 1. Error bars depict SEM. * = significantly larger decrements when high-$CSE_{CI}$ intervals were replaced.

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

## EXPERIMENT 2

Participants: 30 normal-hearing, native English speakers (none from Expt. 1)

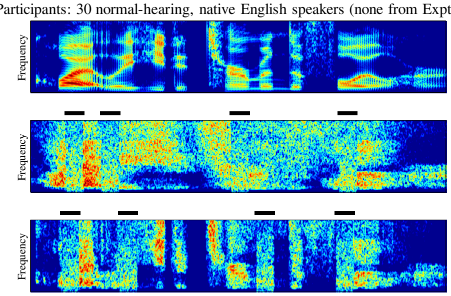


FIGURE 4. The sentence "Quietly, he determined to foil her" at full temporal resolution (top), with 4 Hz (middle) and 64 Hz envelope cutoff frequencies (bottom). Replacement of high-$CSE_{CI}$ intervals with noise is indicated by black lines above each spectrogram.

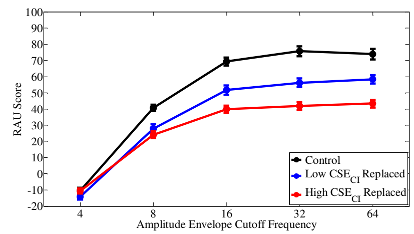


FIGURE 5. Mean sentence intelligibility in Expt. 2, measured in RAU scores. Error bars depict SEM.

- Control > low-$CSE_{CI}$-replaced > high-$CSE_{CI}$-replaced ($p < .001$)
- Performance improved with better temporal resolution ($p < .0001$)
- Significant interaction ($p < .001$)
  - Means again recoded as decrements relative to control performance

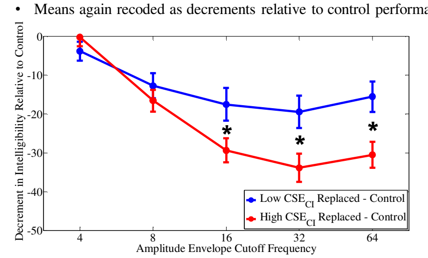


FIGURE 6. Decrements in performance for Expt. 2. Error bars depict SEM. * = significantly larger decrements when high-$CSE_{CI}$ intervals were replaced.

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

## DISCUSSION

Intelligibility of sentences with 16-24 channels nearly overcame replacement of low-$CSE_{CI}$ intervals, but performance was broadly and substantially impaired when high-$CSE_{CI}$ intervals were replaced (Fig. 3). High-$CSE_{CI}$ changes were most important (*i.e.*, produced the largest decrement when replaced with noise) for understanding sentences with 6-10 spectral channels. Peak importance of information-bearing acoustic changes overlaps the number of effective channels of information that CI users receive (*e.g.*, Friesen *et al.*, 2001; Dorman & Spahr, 2006).

Temporal resolution exhibited a different relationship with information-bearing acoustic changes in Experiment 2. Sentence intelligibility plateaued (and performance decrements asymptoted) at 16 Hz (Figs. 5-6). Not only was this well below the 50-Hz resolution Fu and Shannon (2000) deemed sufficient to understand vocoded sentences, but performance did not improve at higher resolutions as was observed in Experiment 1. Stilp's (in press) prediction that information-bearing acoustic changes become more important for speech perception in poorer listening conditions is partially supported, as it captures some aspects of signal quality (proportion of sentence duration replaced by noise, full-spectrum vs. noise-vocoded speech, spectral resolution) but not all (temporal resolution).

At very low signal qualities (4 channels in Experiment 1, 8 Hz in Experiment 2), replacing low- or high-$CSE_{CI}$ intervals resulted in similar decrements in performance (Figs. 3, 6), suggesting lower limits of spectral and temporal resolution that are necessary for defining information-bearing acoustic changes. This result also reiterates the interdependence of these acoustic properties. Spectral and temporal resolution trade off against one another for understanding vocoded speech (*e.g.*, Xu *et al.*, 2002; 2005; Nie *et al.*, 2006). With one acoustic property at relatively low resolution, increasing resolution of the complementary property may restore the perceptual importance of $CSE_{CI}$. Future research will explore how information-bearing acoustic changes interact with spectral-temporal tradeoffs.

Information-bearing acoustic changes prove highly important for understanding vocoded speech across wide ranges of signal quality barring fairly extreme acoustic distortion. This bears considerable importance for CI processing strategies that only transmit high-amplitude portions of the speech signal. Results suggest potential for additional benefit by also incorporating perceptually significant changes in amplitude.

## REFERENCES


Dorman, M.F., & Spahr, A.J. (2006) in Waltzman, S.B. & Roland, J.T. Jr. (eds.) *Cochlear Implants, Second Edition* (Thieme Medical Publishers: New York), pp. 193-204.

Friesen, L.M., Shannon, R.V., Başkent, D., & Wang, X. (2001). *JASA*, 110(2), 1150-1163.

Fu, Q.-J., & Shannon, R.V. (2000). *JASA*, 107(1), 589-597.

Nie, K., Barco, A., & Zeng, F. G. (2006). *Ear Hear.*, 27(2), 208-217.

Stilp, C.E. (in press, 2014). *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.

Studebaker, G.A. (1985). *JSLHR*, 28, 455-462.

Xu, L., Tsai, Y.J., Pfingst, B.E. (2002). *JASA*, 112(1), 247-258.

Xu, L., Thompson, C.S., Pfingst, B.E. (2005). *JASA*, 117(5), 3255-3267.