Relations between speech rhythm and segmental deletion

Sam Tilsen
University of California, Berkeley
tilsen@berkeley.edu

Questions

- ***Does the rhythm (periodicity) of speech influence the likelihood of segmental deletion?***

- **Rhythm is not speech rate—*how do we measure it?***

Background

- **Cross-linguistic approaches to speech rhythm:**
  - **syllable vs. stress-timing: relative isochrony of interval durations (Pike 1945; Abercrombie 1965)**
  - **failure to find evidence for syllable or stress-timing, as well as cross-linguistic differences in syllable/foot isochrony (Bolinger 1965; Lehiste 1977; Dauer 1983)**
  - **Ramus et. al. (1999): ratios and variabilities of consonantal and vocalic interval durations; capture the distinction between syllable-timed and stress-timed languages**
  - **in syllable-timed languages, less deletion, less vowel-reduction, and simpler syllable structure, than in stress-timed languages (cf. Dauer 1983).**

- **Commonalities of interval-based approaches**
  - **measure rhythmic properties with durations of intervals (between syllables, stressed syllables, moras, sequences of C and V, etc.)**
  - **interval endpoints are defined by pre-conceived units: C, V, μ, σ, Ft**
  - **units are conceptualized as "containers"—size (duration) of container represents the container and the contents are either ignored or highly abstracted.**
  - **in the speech signal, intervals not so well-defined**

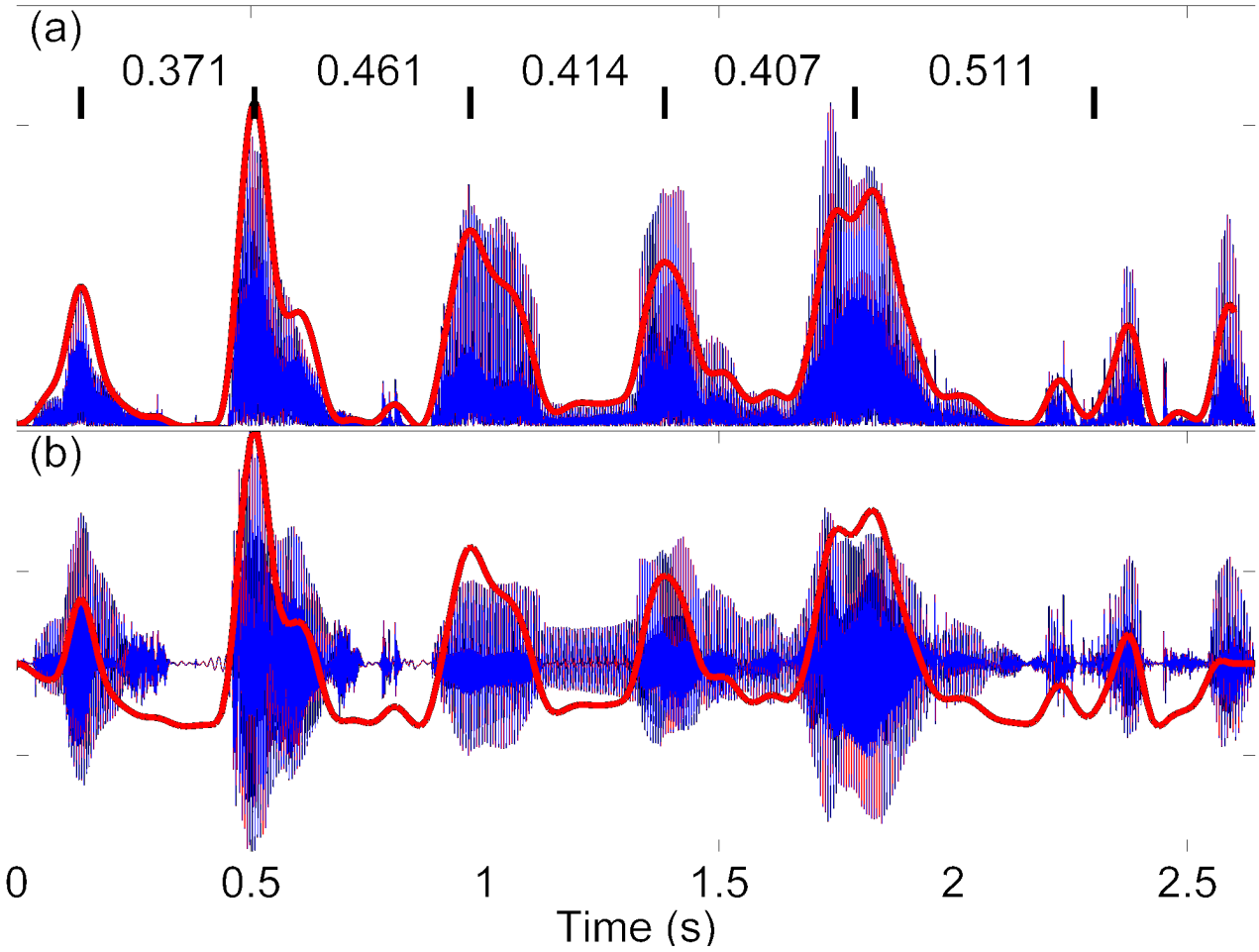
Method

- **Low-frequency Fourier Analysis of Speech Amplitude Envelope**
  ***developed in collaboration with* Keith Johnson**

  Tilsen, S. & Johnson, K. (to appear). Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America.*

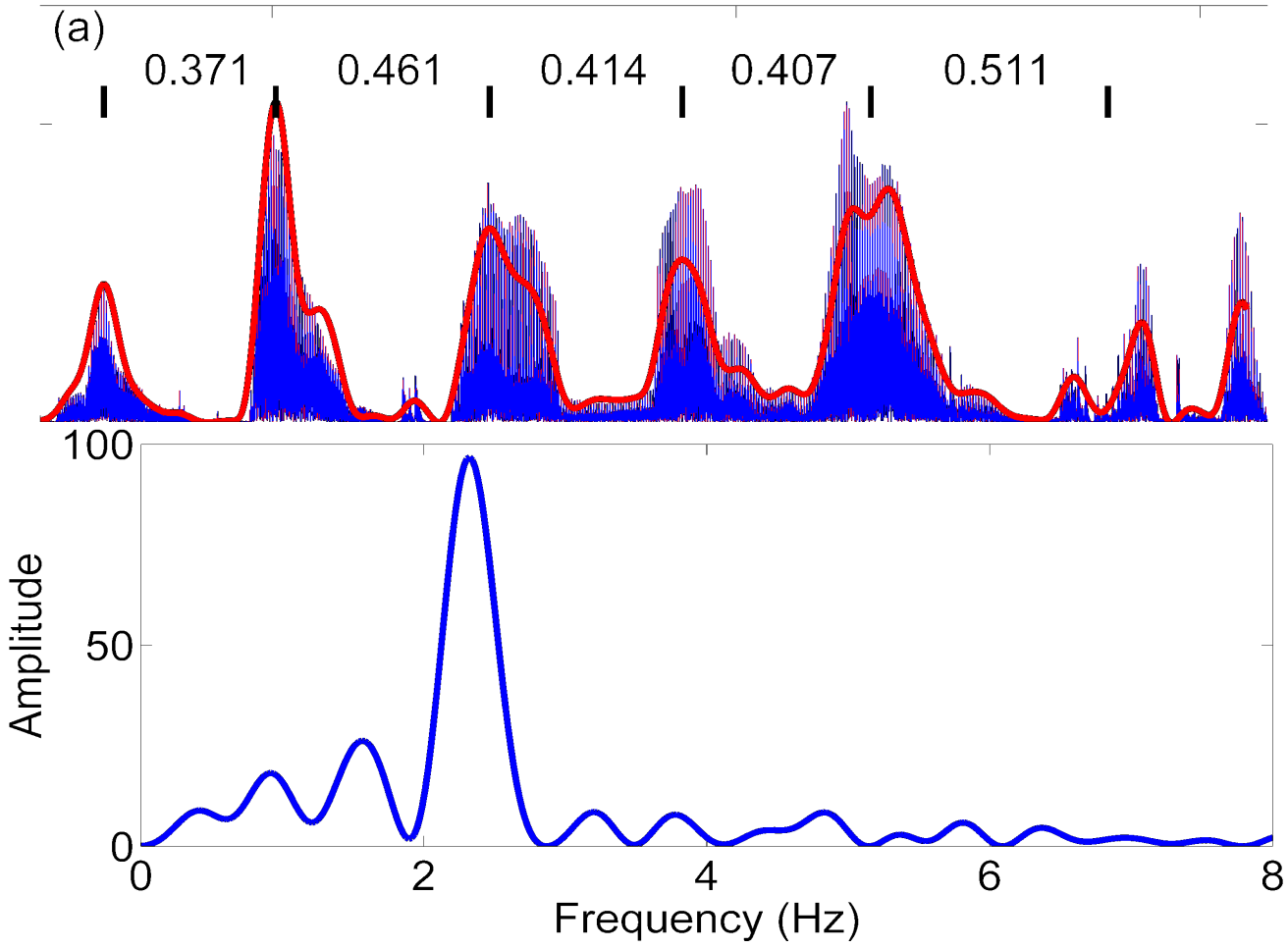
- **Use a representation of speech that does not rely upon intervals**

- **Vocalic energy amplitude envelope:**
  - **vocalic energy louder and more perceptually salient**
  - **children pay more attention to vocalic cues (Mehler et. al. 1996)**
  - **locations of syllable beats (p-centers) near onset of vowel (Allen 1972, 1975; Morton, Marcus, & Frankish 1976)**
  - **p-centers approximated by acoustic energy associated with vowel formants (Scott 1993; Pompino-Marschall 1989)**
  - **use Butterworth filter (gradual roll-offs in frequency domain) with 700-1300 Hz passband (Cummins & Port 1998)**
  - **amplitude envelope: lowpass filter the magnitude (absolute value) of the bandpass-filtered signal**



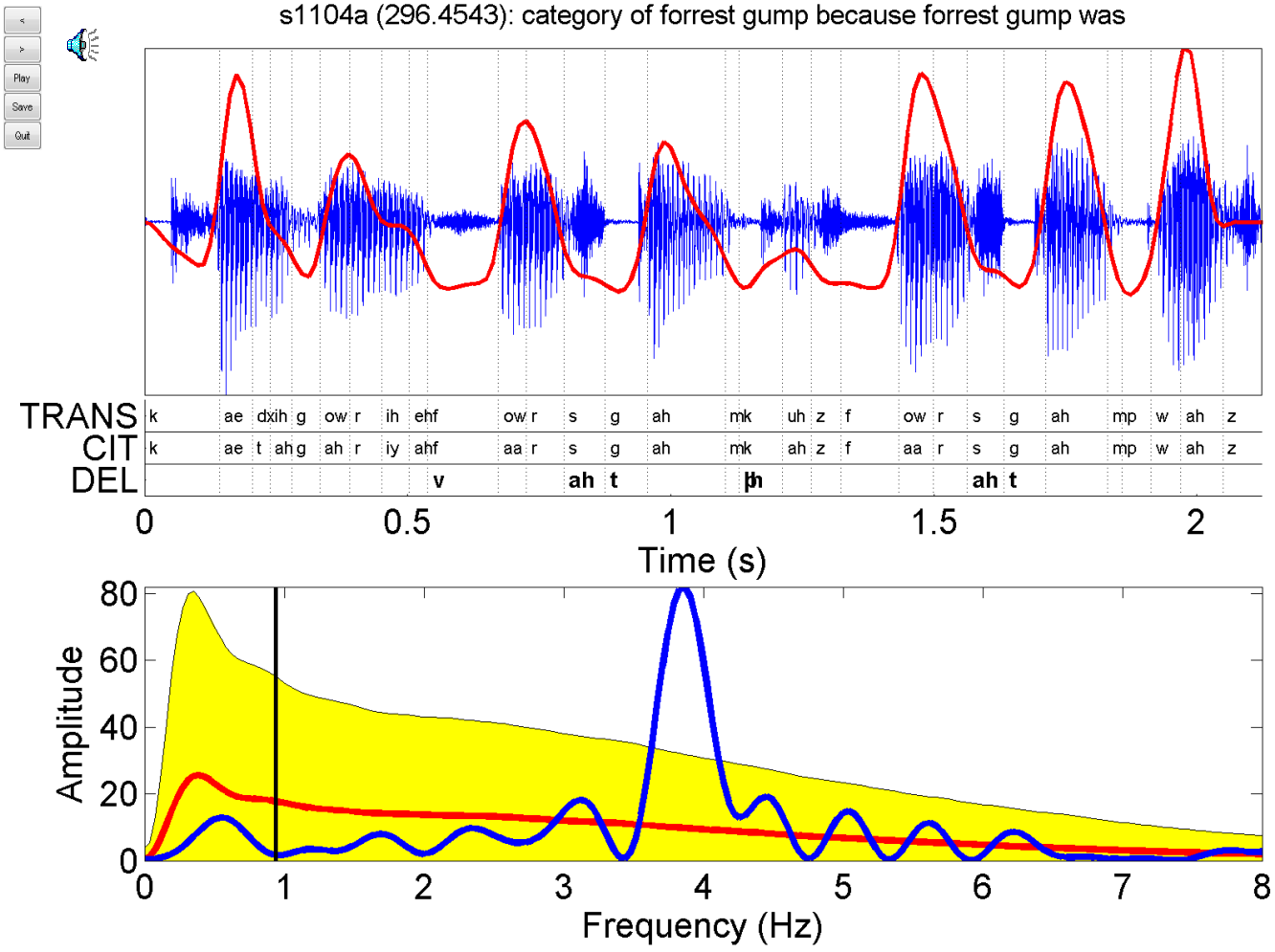
Method

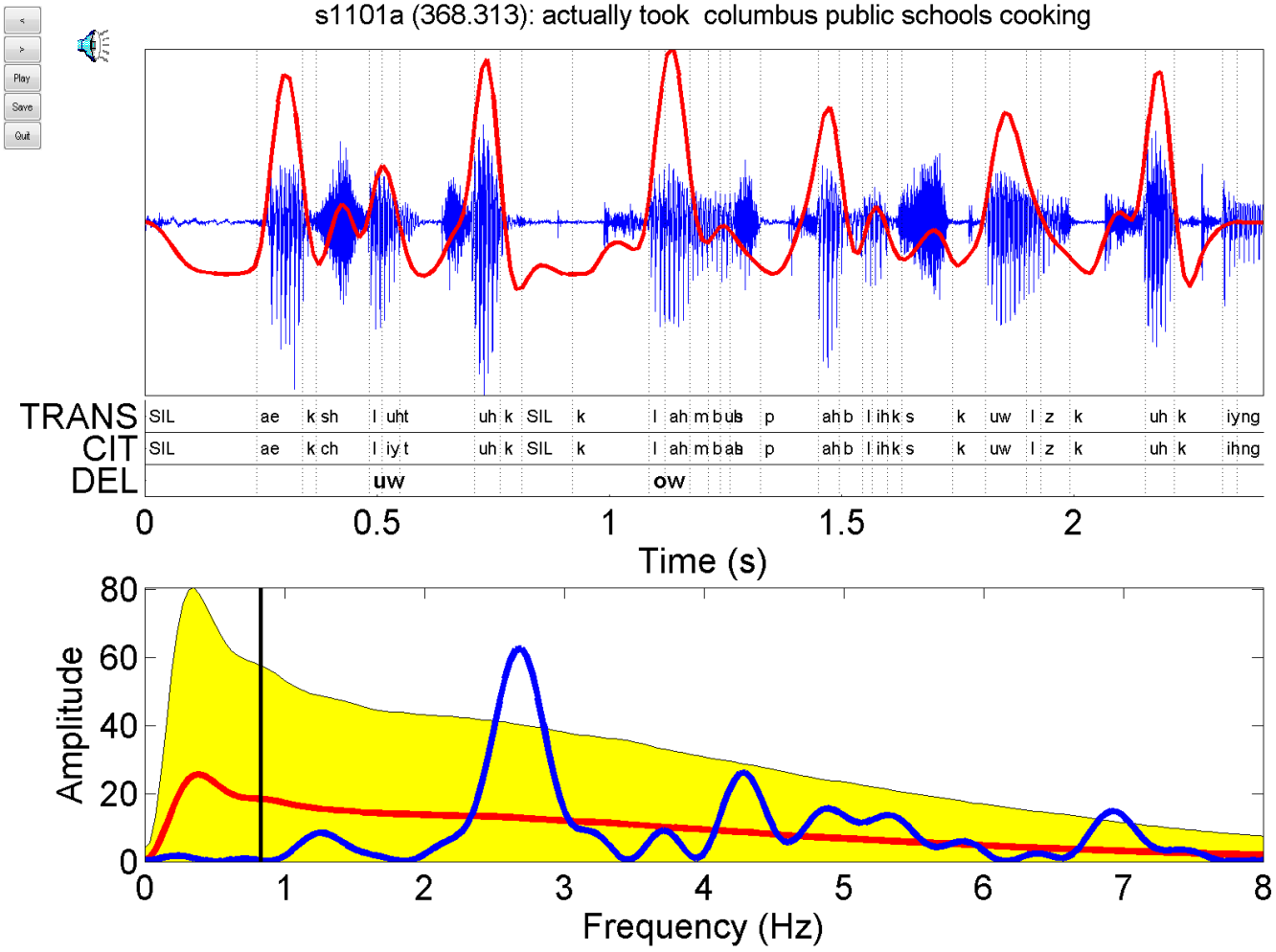
- **Low-frequency Fourier analysis of amplitude envelope:**
  - **normalize, window (Tukey r=0.1), and zero-pad amplitude envelope**
  - **take squared magnitude of the FFT to obtain the power spectrum**
  - **analyze significant peaks (amplitude & frequency) in spectrum**



Method

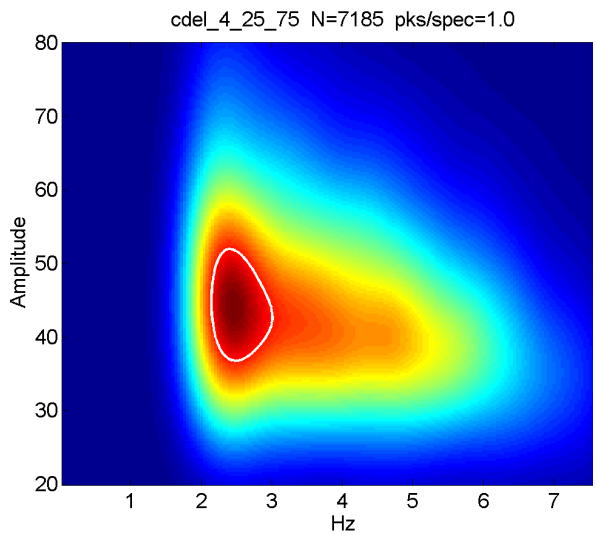
- **Buckeye Corpus:**
  - **hand-labeled phonetic transcription, 40 native central Ohio English speakers from balanced set of ages and genders. Pitt, Johnson, Hume, Kiesling, Raymond (2005).**
  - **approx. 300,000 words**
  - **deletions are determined by comparison citation form to phonetic transcription**

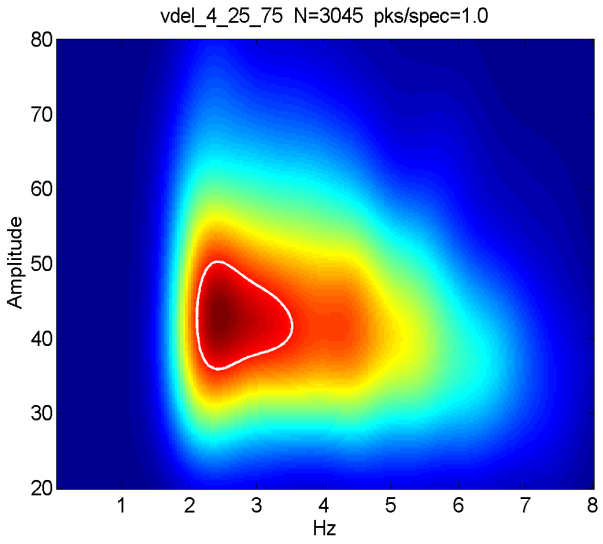

s1104a (296.4543): category of forrest gump because forrest gump was
Play
Save
Quit
TRANS k ae dx ih g ow r ih ehf ow r s g ah mk uh z f ow r s g ah mp w ah z
CIT k ae t ah g ah r iy ahf aa r s g ah mk ah z f aa r s g ah mp w ah z
DEL v ah t th ah t
0 0.5 1 1.5 2
Time (s)
Amplitude
0 20 40 60 80
0 1 2 3 4 5 6 7 8
Frequency (Hz)


s1101a (368.313): actually took columbus public schools cooking
Play
Save
Quit
TRANS SIL ae k sh l uht uh k SIL k l ah m b uh s p ah b l ih k s k uw l z k uh k iyng
CIT SIL ae k ch l iy t uh k SIL k l ah m b ah s p ah b l ih k s k uw l z k uh k ihng
DEL uw ow
0 0.5 1 1.5 2
Time (s)
Amplitude
0 20 40 60 80
0 1 2 3 4 5 6 7 8
Frequency (Hz)

Analysis

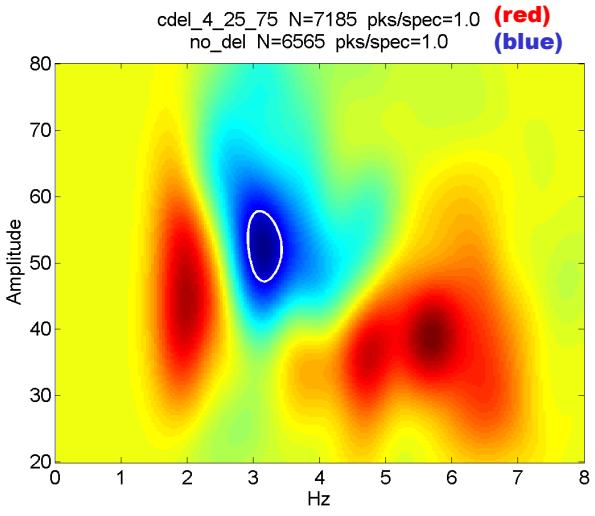
- **Take 2-3 s stretches (“chunks”) of speech centered upon a deletion**
- **Take the highest peak from each spectrum**
- **2-dimensional distribution of frequency-amplitude values**
- **consider “active” deletions (occur 25% - 75% of the time)**

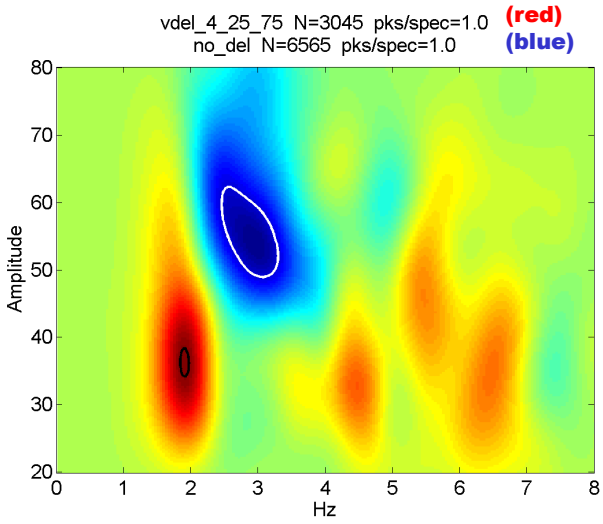




Analysis

- **comparison of chunks with and without deletions**

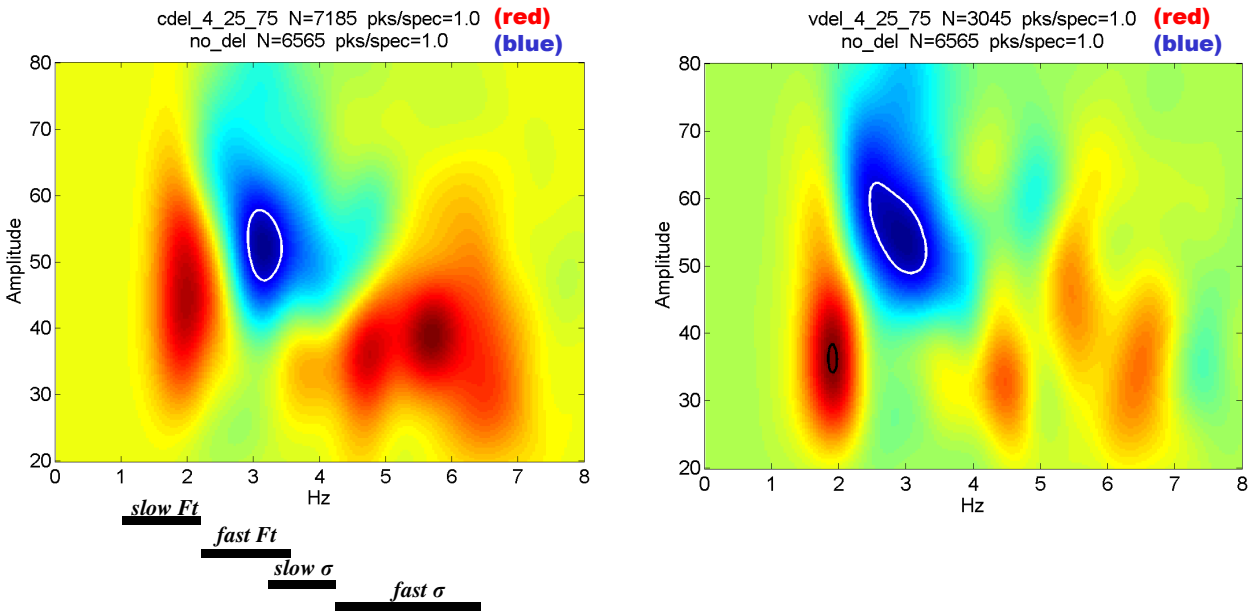




- **both C- and V-deletion are associated with rhythms in the 1.5-2.5 Hz range**
- **C-deletion is more strongly associated with rhythms in the 4.5-6 Hz range**
- **segmental preservation (absence of deletion) is associated with high-amplitude rhythms in the 2.5-3.5 Hz range**

Analysis

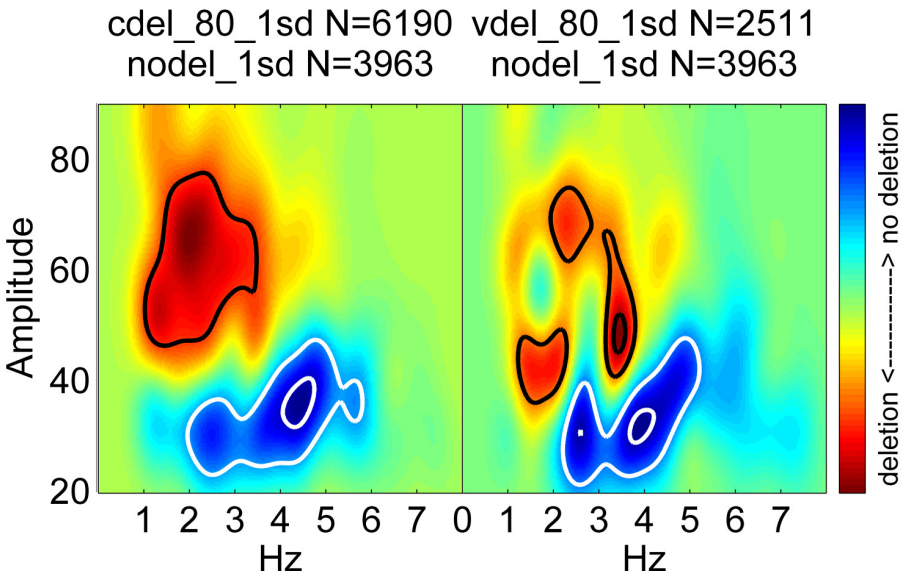
- **comparison of chunks with and without deletions**



- **deletion: slow-foot (C,V) and fast-syllable (C) timescales**
- **preservation: fast-foot timescale**
- **preservation may suggest a "rhythmic sweet-spot" for stability in gestural phasing (with certain assumptions)**

Analysis

- **comparison of chunks with and without deletions, controlled for speech-rate**



- **weak tendency for C-deletion to occur with rhythms in slow and fast-Ft range**
- **V-deletion associated with fast-Ft and slow-σ periodicity.**
- **V and C preservation associated with less rhythmic speech, especially in σ range**
- **more rhythmic speech contains more deletion**
- **C and V deletion associated with lower and higher frequency rhythms, respectively.**

Conclusion / Speculations

- **There exists a frequency range in the rhythm spectrum—corresponding to fast Ft-timing—where the likelihood of deletion is lower.**
- **More rhythmic speech contains more deletion.**
- **Lower frequency (Ft-timescale) rhythmic speech contains more C-deletion, higher frequency (σ-timescale) rhythmic speech contains more V-deletion.**

Why is rhythmicity associated with deletion?

- ***Gestural phasing:*** **gestural systems are coupled to rhythmic systems; when metrical/rhythmic coupling becomes stronger, intergestural coupling is less autonomous and/or more variable, resulting in a greater likelihood of overlap/deletion. (rhythmic systems drive gestural systems)**
- ***Transcriber perceptual bias:*** **Perception of rhythm caused corpus transcribers not to perceive segments.**
- ***Emergent rhythm:*** **segmental deletion occurred randomly, the effect of this happens to be more rhythmic speech. Or perhaps, speakers employ deletion in order to make speech more rhythmic (the teleological version).**

Why do consonants and vowels pattern differently?

- **Some vowel deletions result in loss of syllable, others do not. Consonant deletions less often occur with loss of syllable.**

References

- **Abercrombie, D. (1967).** ***Elements of general phonetics*****. Chicago: Aldine.**
- **Allen, G. D. (1972).** ***The location of rhythmic stress beats in English: An experimental study, parts I and II*****. Language and Speech, 15:72--100,179--195.**
- **Allen, G. D. (1975).** ***Speech rhythm: its relation to performance and articulatory timing*****. Journal of phonetics. 3: 75-86.**

  Chatfield, C. 1975. ***The analysis of time series*****. London: Chapman and Hall.**
- **Cummins, F. and Port, R. (1998). Rhythmic constraints on stress timing in English.** ***Journal of Phonetics*****, 26, 145-171.**
- **Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed.** ***Journal of Phonetics*****, 11, 51-62.**
- **Howell P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope.** ***Perceptual Psychophysics*****. Jan; 43(1):90-3.**
- **Jenkins, G.M. & Watts, D.G. 1968.** ***Spectral analysis and its applications*****. San Francisco: Holden-Day.**
- **Ladd, D. R. (1996).** ***Intonational phonology*****. (Cambridge Studies in Linguistics 79.) Cambridge: Cambridge University Press.**
- **Lehiste, I. (1977) Isochrony reconsidered.** ***Journal of Phonetics*****, 5 (3), 253-263.**
- **Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: the infant's viewpoint. In J. L. Morgan, & K. Demuth, Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition. p.101-116. Mahwah, NJ: Lawrence Erlbaum Associates.**
- **Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual Centers (P-centers).** ***Psychological Review*****. Vol. 83, No. 5, 405-408.**
- **Nakatani, L. H., O'Connor K. D., Aston, C. H. (1981) Prosodic aspects of American English speech rhythm.** ***Phonetica*** **38 (1-3), 84-106.**
- **Pike, K. L. (1945).** ***The intonation of American English*****, Ann Arbor, MI: University of Michigan Press.**
- **Pitt, M.A., Johnson, K., Hume, E., Kiesling, S., Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability.** ***Speech Communication*** **45 (1), 89-95.**
- **Pompino-Marschall, B. (1989). On the psychoacoustic nature of the P-center phenomenon.** ***Journal of Phonetics*****, 17, 175–192.**
- **Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora-timing in Japanese.** ***Journal of the Acoustical Society of America*****, 81 (5), 1574±1585.**
- **Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal.** ***Cognition*** **73: 265-292.**
- **Roach, P. (1982).** ***On the distinction between `stress-timed' and `syllable-timed' languages.*** **In D. Crystal, Linguistic controversies, London: Edward Arnold.**