Low-frequency spectral analysis as a metric of speech rhythm

Sam Tilsen (tilsen@berkeley.edu)
Keith Johnson (keithjohnson@berkeley.edu)

Dept. of Linguistics
Univ. of California, Berkeley

1. Rhythm as frequency.

Our method is to use Fourier analysis to discover frequency components in the amplitude envelope of speech.

To produce a “rhythm spectrum”:

Step 1. Calculate the amplitude envelope

(of energy between 700-1300 Hz - Cummins & Port, 1998)

Step 2. Calculate a Fourier spectrum of the amp. envelope

(Tukey window, r=0.1, zero-padded, normalized to unit variance).

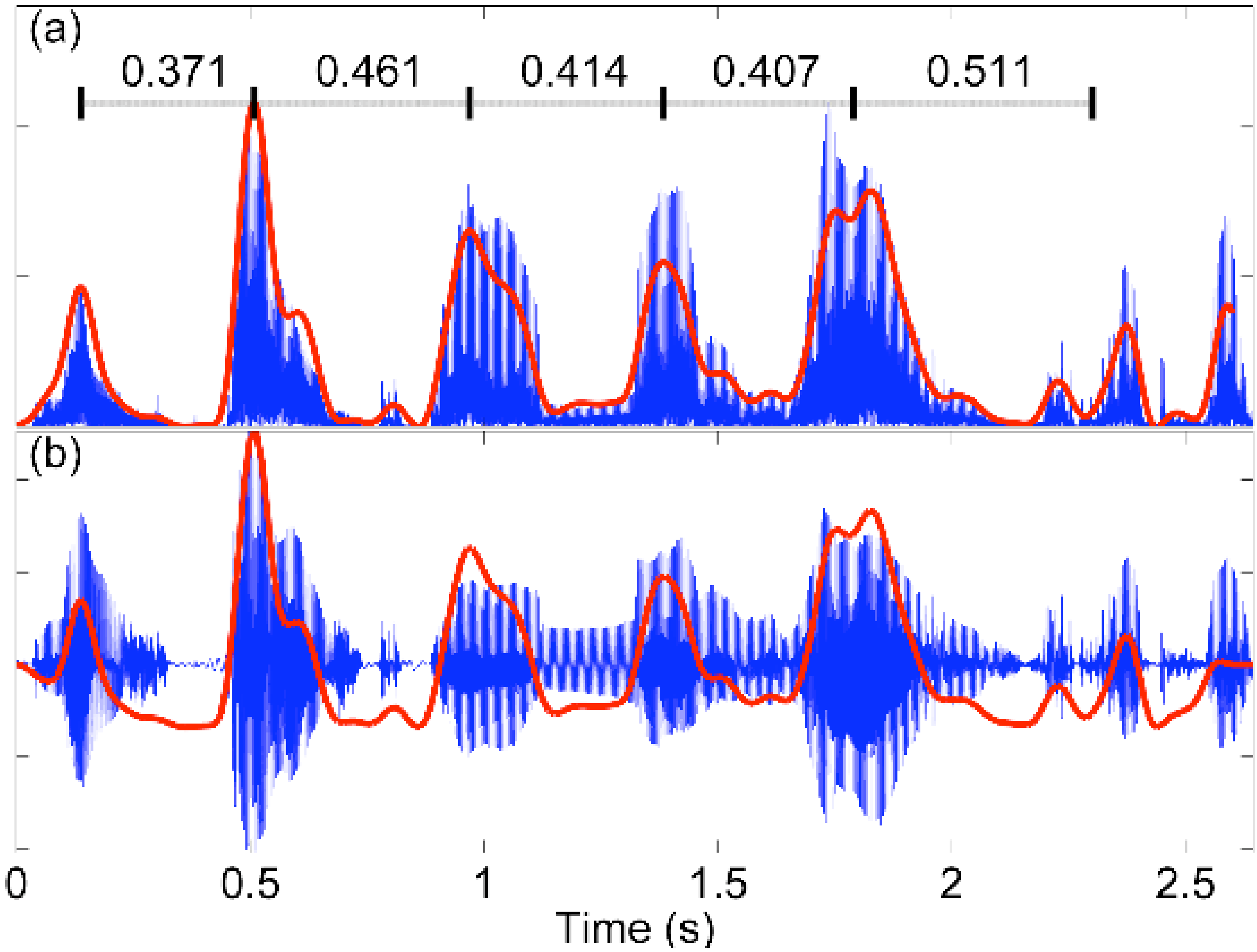

Figure 1. Taking the amplitude envelope of a 2.5 sec. stretch of speech - “at least based on money raised it looks like..”. (a) rectified signal with envelope. (b) original signal with windowed envelope. [Listen to this utterance.]

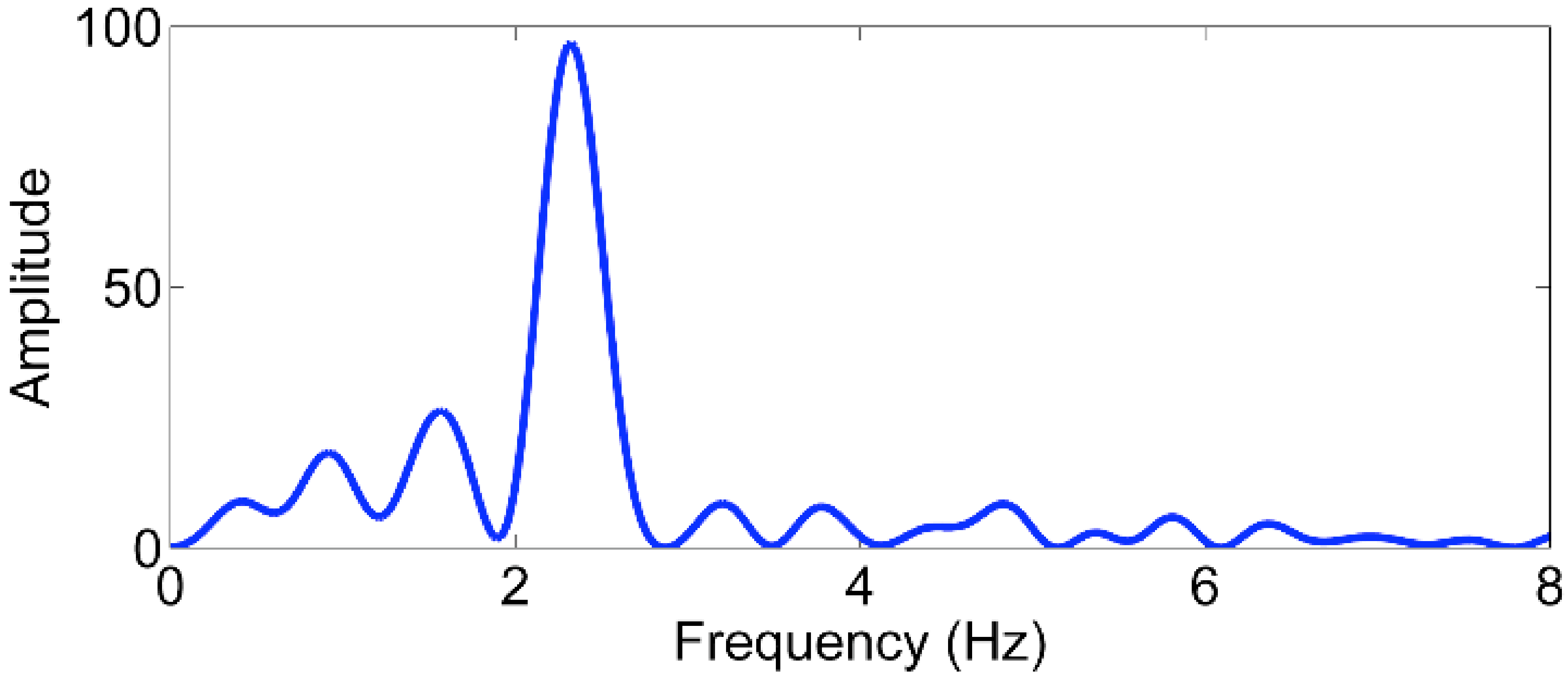

Figure 2. Fourier transform of the amplitude envelope shown in figure 1 - the “rhythm spectrum”. The peak at 2.3 Hz indicates a component with a period of about 435 ms (compare this to the interval durations in figure 1).

The “rhythm spectrum” captures degree of rhythmicity (vertical axis) and the rhythmic frequency components (horizontal axis).

Advantages of this method.

1. No intervals to measure.
2. No phonology-dependent decisions to make.
3. Reference to rhythm in terms of frequency.
4. Rhythms at multiple time-scales are represented.

Testing the method - the remaining sections of this poster.

2. Cross-linguistic comparisons
3. Rhythm in conversational speech
4. Rhythm in political speech

2. Rhythmic Typology

Many thanks to Franck Ramus for sharing with us the sound files used in Ramus, Nespor & Mehler (1999).

Method.

- For each sentence, calculate a single rhythm spectrum over the entire sentence (2-3 second chunk of speech).

- Pick the two highest amplitude peaks in the spectrum. The lowest frequency of these is Peak 1 and the next is Peak 2.

- Plot the median Peak 1 and Peak 2 frequencies.

- Plot median amplitudes of Peak 1 and Peak 2.

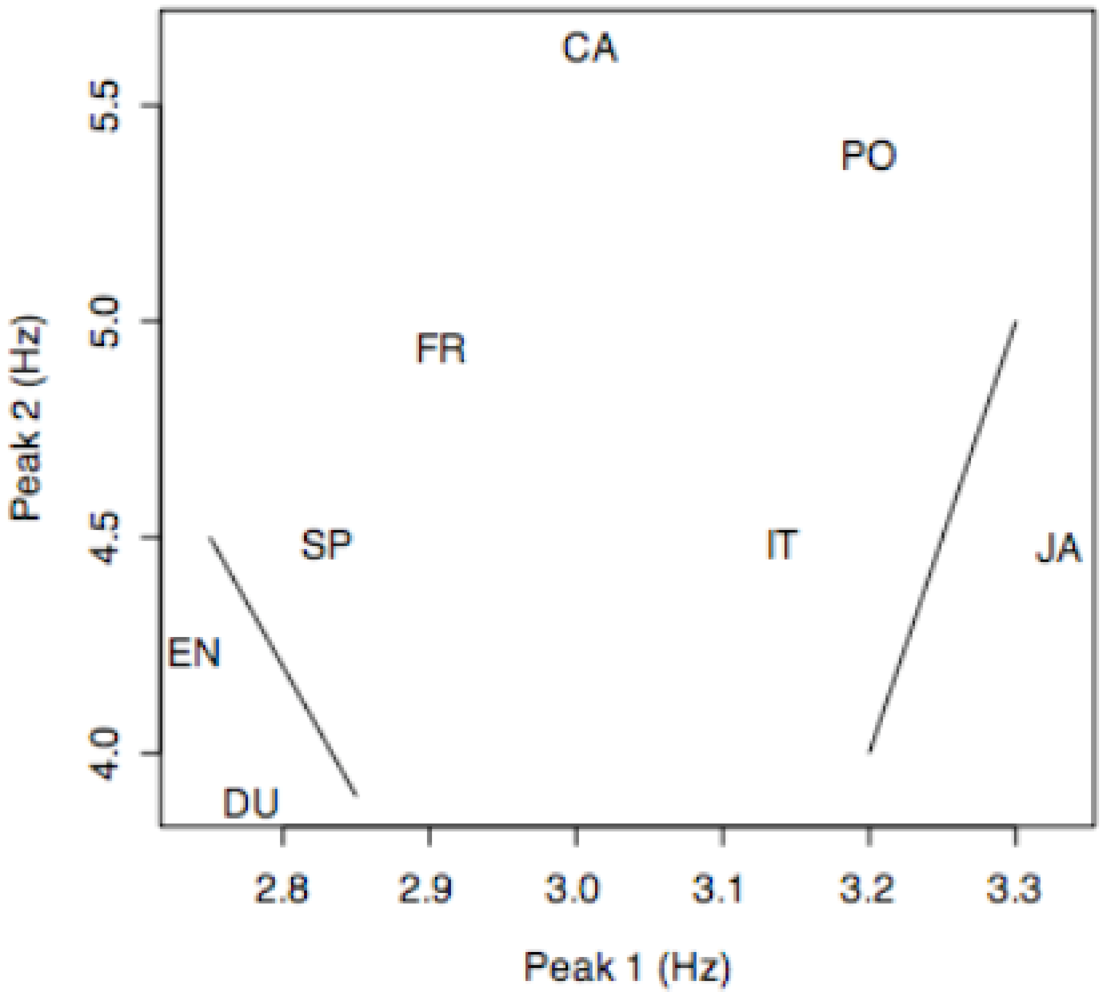

Figure 3. Stress-timed, syllable-timed and mora-timed languages occupy separate regions in the Peak 1 and Peak 2 space.

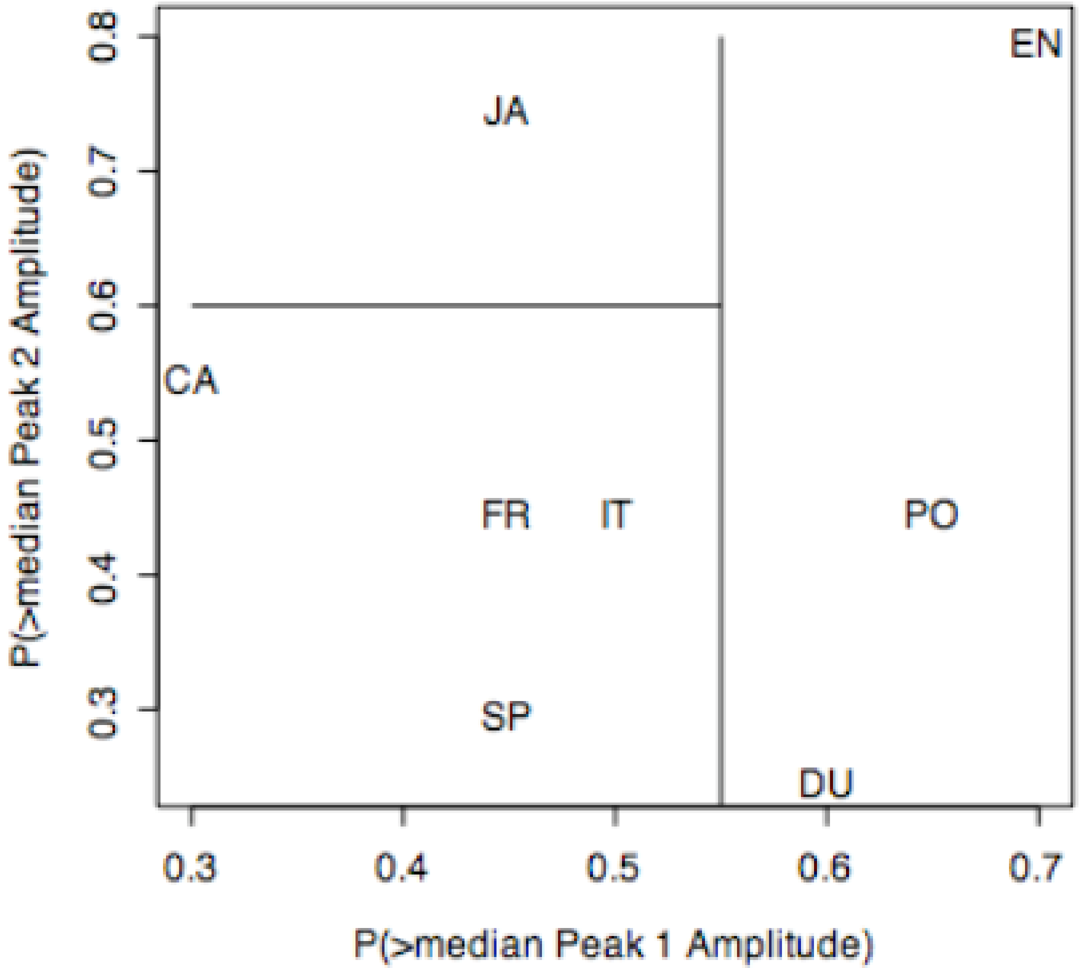

Figure 4. The amplitudes of Peak 1 and Peak 2 also separate stress-timed, syllable-timed, and mora-timed languages.

- There is some degree of separation of languages along traditional timing categories, with English and Dutch having the lowest frequency (median) Peak 1 and Japanese having the highest.

- Stressed-timed languages tend to have stronger Peak 1 and Mora-timed languages tend to have stronger Peak 2.

- But, as Ramus et al. found, there is a substantial amount of within-language variability in these data.

We turn now to a study of within-language rhythmic variation.

3. Rhythm in conversational speech.

Here we use the “Buckeye” corpus of conversational American English (Pitt et al., 2005). 40 hours of sociolinguistic interviews, phonetically transcribed by hand.

Rationale:

- Explore within-language variation.
- Look for rhythms as multiple time-scales.
- Determine how “rhythmic” conversational speech is.

Aside: using vowel and consonant interval parameters over the corpus (figure 5), we find that the distribution centers on 50% V and is very broad.

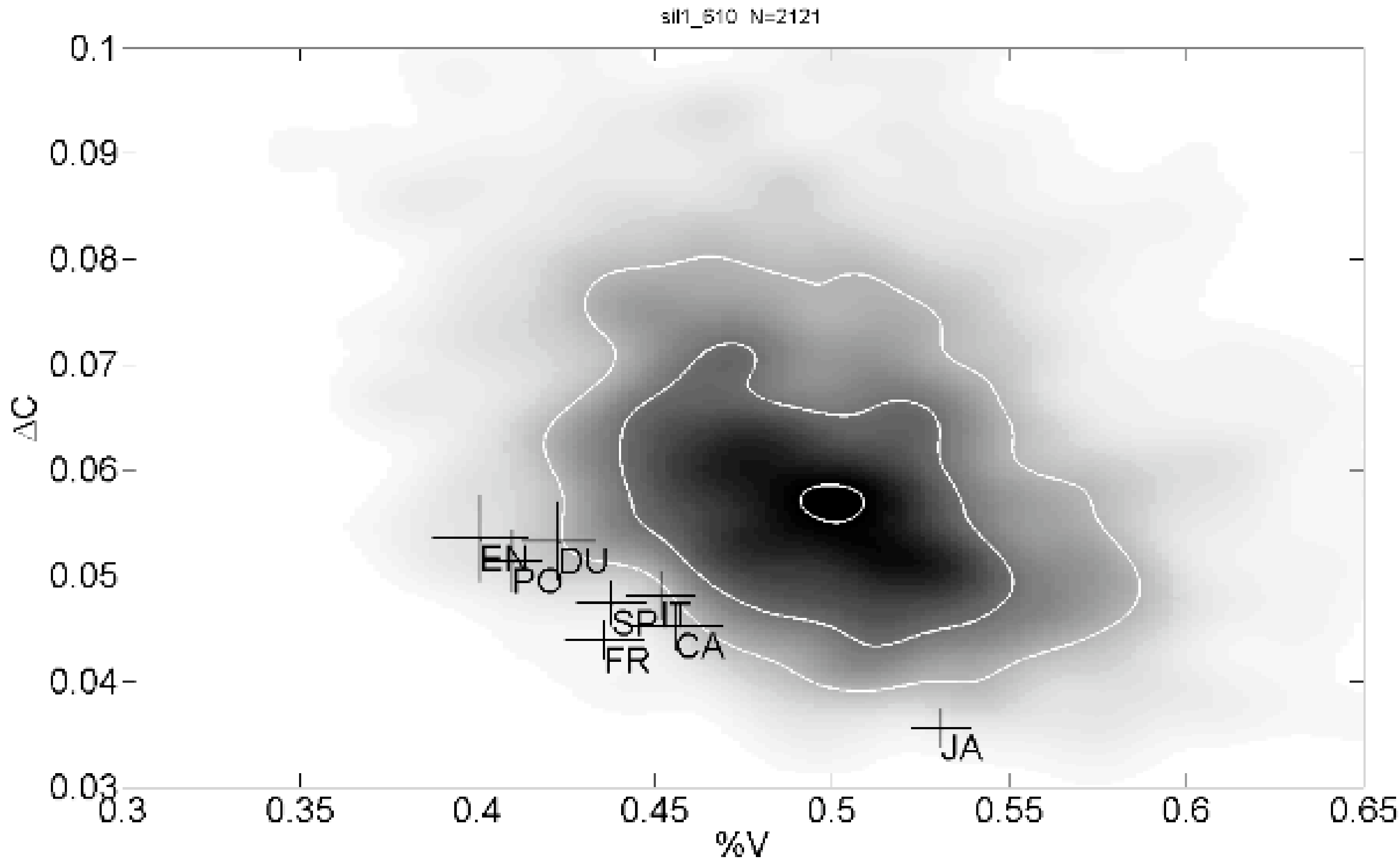

Figure 5. %V and ΔC values from Ramus et. al. against a gaussian kernel density plot of %V and ΔC values from 2,121 6-10 s chunks of speech from the Buckeye corpus.

Rhythm spectrum analysis of conversational speech.

- One additional example (figure 6) of the rhythm spectrum - this time of a faster rhythm (compare to figure 1).

- Speech in this corpus is often not particularly rhythmic - table 1

- Distribution of rhythm peaks (figure 7, left) has a low frequency center of gravity at 1.6 Hz. This is lower than found in isolated read sentences.

- Rhythms differ for utterances with consonant or vowel deletion (figure 7, middle and right).

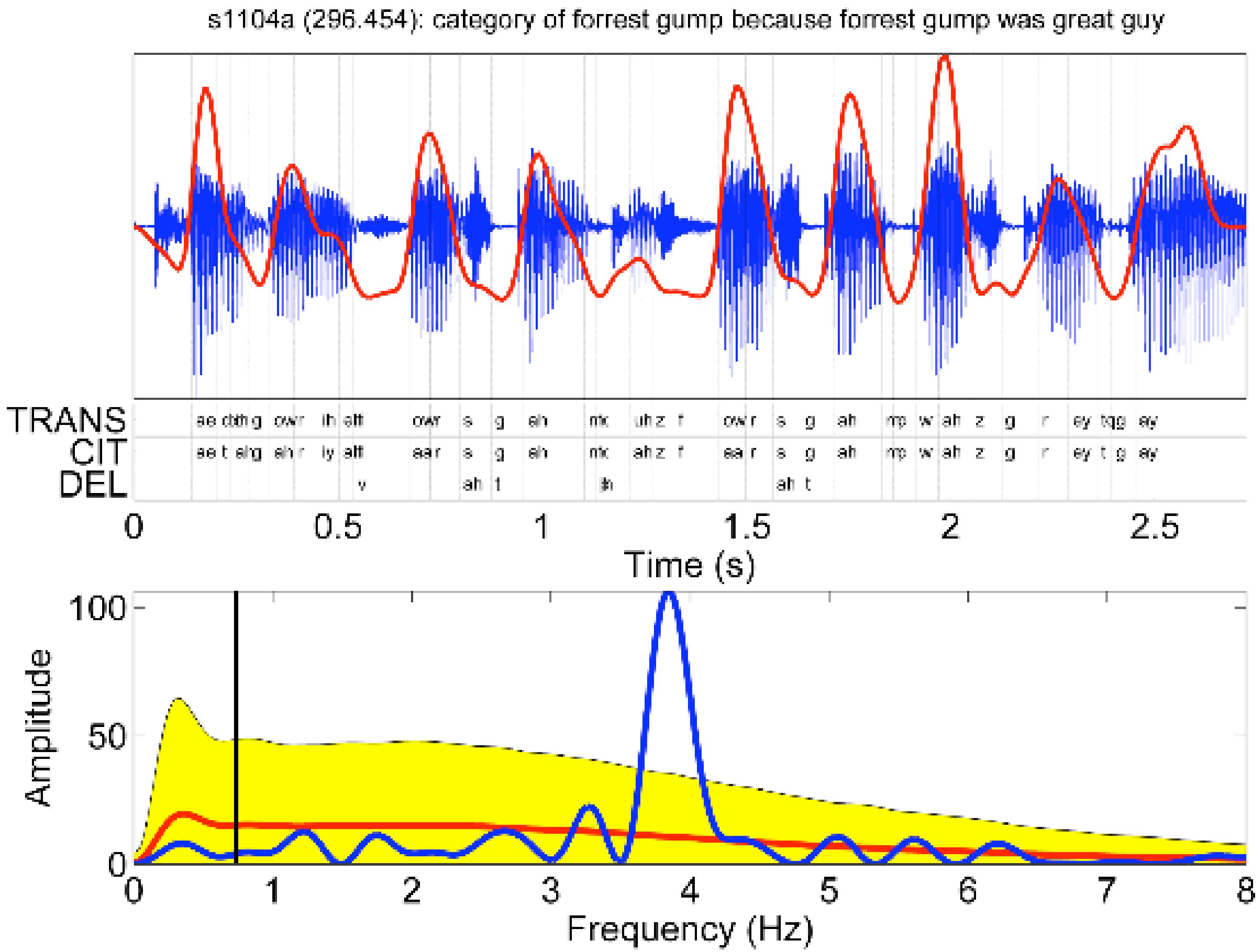

Figure 6. Rhythm spectrum of “category of forrest gump because forrest gump was great guy”. Example of a strong 4 Hz rhythm. [Listen to this utterance].

How often is conversational speech rhythmic?

Table 1: Counts of rhythmic chunks in several frequency ranges

Rhythmic chunks	0-1 Hz	1-2 Hz	2-3 Hz	3-4 Hz	4-5 Hz	5-6 Hz	Total (> 1 Hz)
count	1354	897	871	396	109	27	2303
percent	13.7%	9.1%	8.8%	4.0%	1.1%	< 0.1%	23.2%

- Chunks of 2-3 seconds in duration.

- Count a chunk as rhythmic if the rhythm spectrum has a peak with (normalized) amplitude over 50.

- If we include chunks with rhythms of 0-1 Hz the percentage of rhythmic chunks is still only 36.9%.

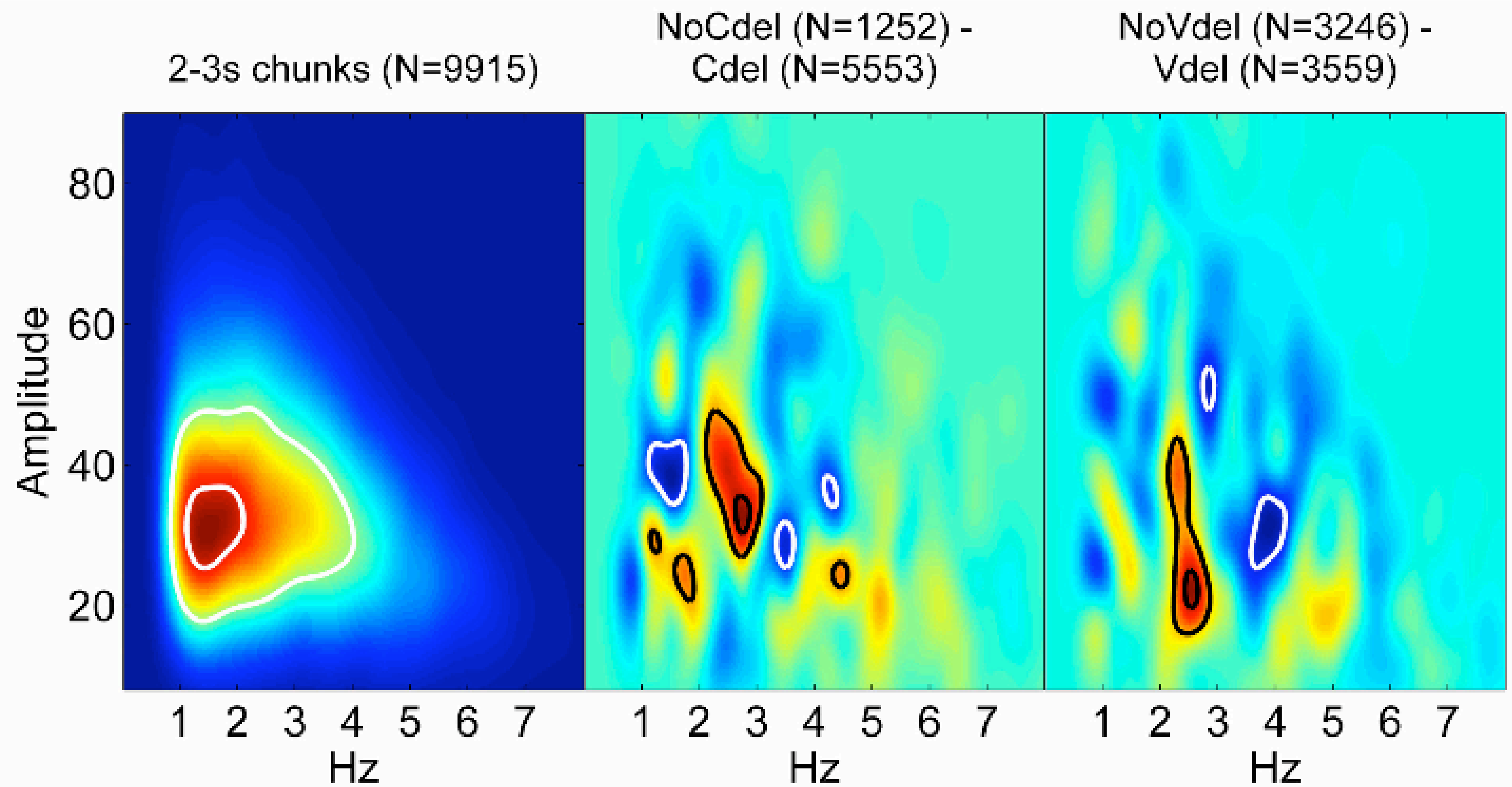

Figure 7. Gaussian density plots of the lowest two peaks in rhythm spectra (left), and density differences for utterances with or without consonant deletions (middle), or vowel deletions (right). No deletion = red, deletion = blue.

- Low frequency rhythms predominate (1.6 Hz)

- Rhythms around 1-2 Hz are associated with C deletion.

- Rhythms around 2-3 Hz are associated with a relative absence of C and V deletions.

- Rhythms around 3-4 Hz are associated with both C and V deletions.

4. Rhythm in political speech.

Look at within-speaker variation in rhythm.

Four speeches by US presidential candidate Barak Obama

Two policy speeches.

17 Sept. 2007 - New York (Speech at NASDAQ on economic policy)
7 Oct. 2007 - Portsmouth, New Hampshire (Speech on energy policy)

Two “inspirational” speeches.

29 Nov. 2007 - New York (Fund-raiser speech in Apollo Theater, Harlem)
3 Jan. 2008 - Des Moines, Iowa (Victory speech after the Iowa primary election)

Method.

- Mark stretches of speech (vs. clapping, cheering, etc.).
- Select chunks (speech and pause) of 6-10 seconds.
- Compute rhythm spectrum.
- Find the frequency and amplitude of the strongest peak.

Results.

- Distributions of peaks differ from speech to speech.

- “Inspirational” speeches have higher amplitude below 1 Hz.
- Policy speeches have more peaks between 2-3 Hz.

Use of rhythm in effective political speech in American English involves very slow rhythm - at 0.6 Hz (1.66 sec period).

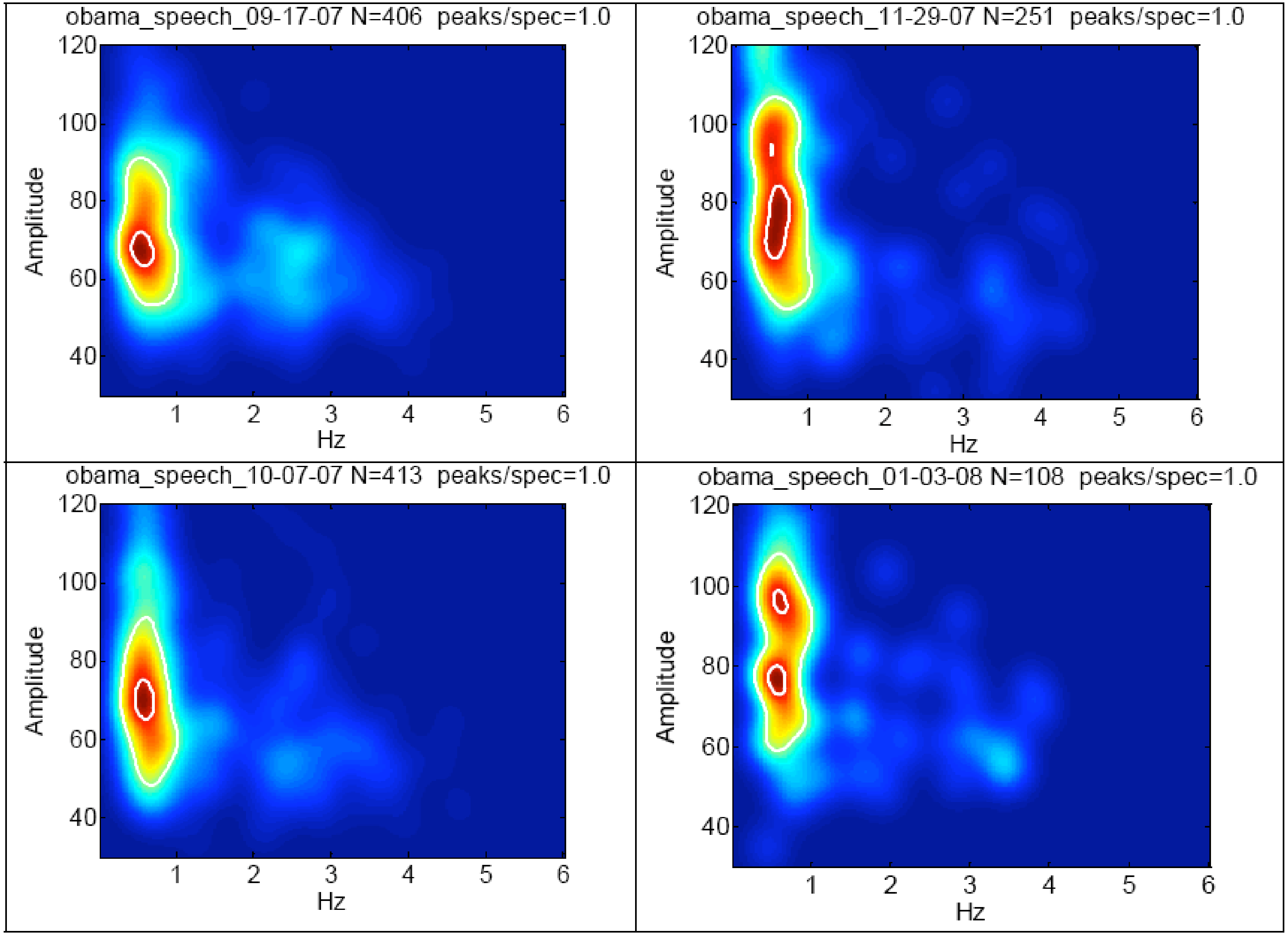

Figure 8. Rhythm profiles (Gaussian density plots) from policy speeches on the left and “inspirational speeches” on the right [listen to examples].

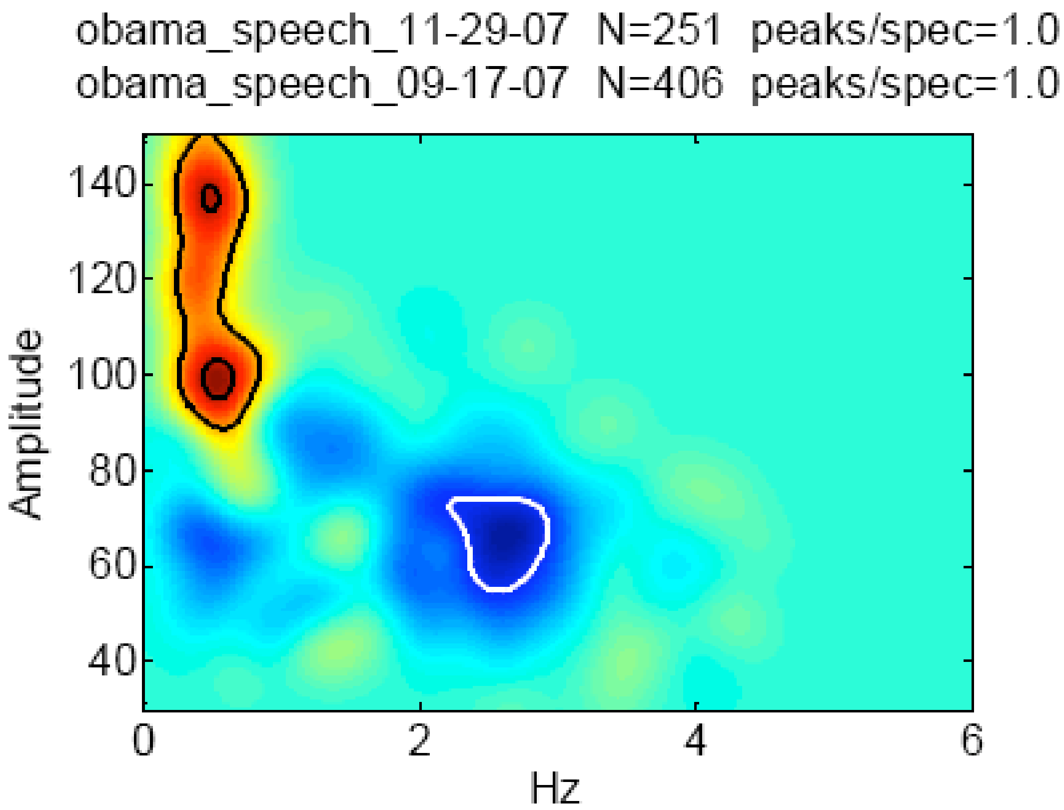

Figure 9. Difference between the distributions of the Harlem fund-raiser speech and the economic policy speech. Red = more in fund-raiser, Blue = more in policy speech.

Conclusions

- Rhythm can be studied in the frequency domain
  - in terms of repeating patterns in the amplitude envelope.

- Cross-linguistic differences emerge from such an approach.

- Some, but not all, conversational speech utterances are strongly rhythmic.

- Strongly rhythmic conversational or political speech in American English is slower than that observed in isolated read sentences.

References.

Allen, G. D. (1975). Speech rhythm: its relation to performance and articulatory timing. *Journal of Phonetics* **3**, 75-86.

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.

Lee, C.S. & Todd, N.P.M. (2004) Towards an auditory account of speech rhythm: Application of a model of the auditory 'primal sketch' to two multi-language corpora. *Cognition* **93(3)**, 225-254.

Morton, J., Marcus, S., and Frankish, C. (1976). Perceptual Centers (P-centers). *Psychological Review* **83(5)**, 405-408.

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.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition* **73**, 265-292.