Getting Rhythm:
A Methodological Comparison of Quantitative Speech Rhythm Analysis

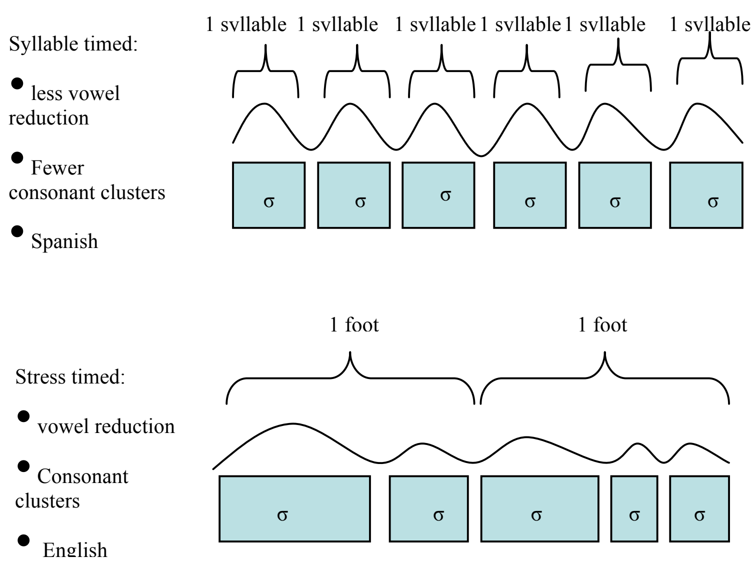
Mary Kohn and Melissa Farr

Presented at SECOL LXXVI

I. Introduction

Impressionistic descriptions of speech rhythm have long been used to distinguish syllable-timed languages from stress-timed languages (see **Figure 1**). Pike (1945) was the first to define these two categories by stating that languages in which the length of each syllable has approximately the same duration are syllable-timed, and stress-timed languages are languages which hold consistent the duration of each syllabic foot, thereby ensuring that the duration between one foot and the next is also approximately equal. Dauer (1983) refined these observations by concluding that the distinctive perceptual differences between syllable-timed languages like Spanish and stress-timed languages like English are actually a result of phonological patterns, with stress-timed languages accepting more complex syllable structures and consonant clusters and also employing vowel reduction. Syllable-timed languages are considered to lie on the opposite end of the spectrum with few consonant clusters and little vowel reduction.

Figure 1: Graphic depiction of differences between stress and syllable timing

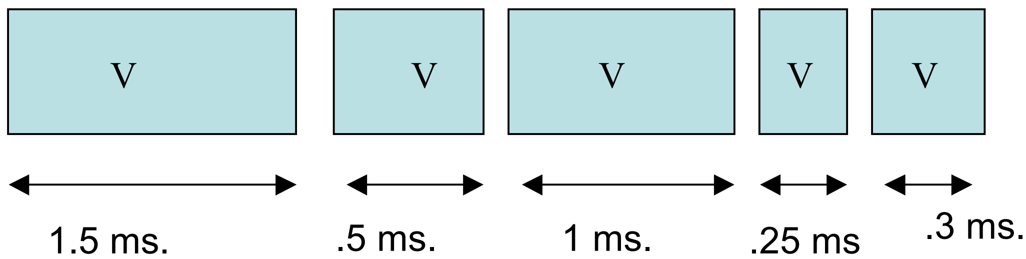


Based on these distinctions, several methods of quantifying rhythm have been developed, including the Pairwise Variability Index (PVI) established by Lowe, Grabe and Nolan (2000). This index measures the absolute value of the difference of vowel length in consecutive syllables, normalized for speech rate[1], averaged over speech (See **Figure 2**). When applying the metric to languages perceived as stress-timed or syllable-timed, Lowe, et al (2000) were able to demonstrate that PVI appropriately separated out these perceived distinctions with syllable-timed languages having lower PVI values than stress-timed language samples. This would be expected as lower PVI values would indicate more similarity in vowel duration over speech, resulting in more similar syllable durations when compared with stress-timed languages, as shown in Figure 1. Such conclusions lend credence to the idea that impressionistic distinctions such as "syllable-timing" and "stress-timing" are founded in phonetic fact and have led to a host of sociolinguistic studies employing PVI to analyze the social distribution of rhythmic variation (Carter 2005, Thomas and Carter 2006, Coggshall 2008).

---

[1] The metric normalizes for speech rate by dividing the absolute value of the difference of the consecutive syllables by the mean duration of the two syllables.

Figure 2: PVI is calculated by taking the absolute value in the difference of vowel pairs in consecutive syllables. One can then compare speakers by taking the mean or median of these values.



PVI 1: |(1.5 - .5)|/ ((1.5 +.5)/2) = 1
PVI 2: |(.5 – 1)|/((.5+1)/2) = .67

PVI 3: |(1-.25)|/((1+.25)/2) = 1.2
PVI 4: |(.25-.3)|/((.25+.3)/2) = .18

. However, certain areas of concern arise with the PVI, or any interval-based approach. Steiner (2003), for example, demonstrates that nasals and approximants can bias results using such measures. Further, many of these phonemic units are difficult to parse from acoustic output, especially when the consonants that intervene between vowels are liquids, rhotics, or approximates, thus making analyses difficult to replicate. There are also theoretical concerns regarding the use of pre-conceived units such as syllables or consonant and vowels as the basis for the measurement for rhythm (Johnson and Tilsen 2008). Concerns arise regarding what is being measured in pair wise assessments. The method measures the content of the higher structures (namely, vowel duration) rather than directly measuring the overarching structure that organizes the content (namely, rhythm). These authors propose that a better measure would not only include temporal segmenting based on purely acoustic features, but also amplitude and pitch. Such a method would focus on perceptual centers (see Cummins and Port 1998), also called “beats,” which may be loosely described as the perceptual boundary of a syllable. These usually correspond with the onset of a vowel but shift in syllables with consonant clusters. Such information is derived using a low pass band filter that limits the wave-form display to showing mainly sonorant energy. Measuring from sonorant peak to sonorant peak thus eliminates the bias of determining phonemic boundaries in phonetic output. These sonorant envelopes are referred to as amplitude envelopes by Johnson and Tilsen (2008).

In order to refine this method, Johnson and Tilsen (2008) developed a method called a "Low Frequency Fourier Analysis of Speech Rhythm" (here after referred to as LFFA) that extracts information regarding the periodicity of the amplitude wave extracted from a low band pass filtered spectrogram thus providing information on the overall rhythm of a phrase. This method avoids issues regarding the imposition of phonemic boundaries through extracting an acoustic measure of rhythm and considers the entire body of the phrase at one time. This method remedies two important critiques of PVI; namely, that measurements with PVI are often subjective and artificially impose phonemic contrasts onto acoustic output; and that comparing lower order units to each other is not an appropriate way to analyze the higher order structure, which is, after all, the rhythm that we want to study. However, a confounding factor for the LFFA is that it does not control for speech rate.

This study aims to contribute to the body of speech rhythm research by directly comparing PVI with LFFA to see if differences in the description of speech rhythm occur between the two methods; or, if no difference emerges. If LFFA successfully reveals differences in iconically distinct varieties, then it would be worth pursuing research that identifies if the perceptual predictions that LFFA make regarding rhythmicity are more true to perception data than PVI predictions. If information regarding periodicity of the rhythm obtained by the LFFA differs from PVI distinctions, linguists may prefer to explore which method most accurately describes how humans perceive speech rhythm. Further, if found to accurately align with impressionistic characterizations of rhythm, the LFFA may be used to compare individual phrases in order to more closely explore shifts in rhythm across a running speech. Such comparisons are less than ideal for interval measures which rely on means of large samples for their overall description of rhythm.

Speech rhythm is a perceptually salient feature that deserves attention in sociolinguistics, but our research is only as good as our methods. In order for meaningful work to continue on speech rhythm variation the methods available to sociolinguists must be more fully understood through direct comparison, theoretical critique, and finally, perceptual experimentation. Finding a method that can overcome differences in both speech rate and speech content is necessary in order to understand how rhythm patterns in the languages of the world, and whether the feature can index social identity within speech communities.

In order to compare the PVI and LFFA methods of analyzing speech rhythm, both methods will be used to analyze the speech of six English speakers. Two of these English speakers have Spanish as a heritage language and thus are expected have more syllable-timed speech. For each speaker at least 100 syllable pairs were demarcated on a text grid in PRAAT at the beginning and end of the central vowel. The median variation in absolute difference of consecutive vowel length normalized for speech rate was extracted following the methodology for calculating PVI established in the literature (Thomas and Carter 2006). The sound sample was then low band pass filtered to identify sonorant contributions to the amplitude of the waveform. Finally, phrases were extracted from each participant's sample. The value of the rise and fall of the amplitude tracker were analyzed using a Fourier's Analysis to find the periodicity of the changes in amplitude. The speakers will be compared using each metric in order to identify whether the distribution for each metric emerges as similar.

The results from this analysis are surprising in that the speakers do not separate into the iconic groups expected for either measure. However, this deviation is not necessarily due to a failure of the measures as each method ranked the participants in the same order. These results prove encouraging for continued analysis of speech rhythm.

Figure 3: The PVI values and Fourier Hz values should be in an inverse relationship if the measures separate the rhythm of the speakers' speech values in the same way. Thus, a low fourier Hz value and a high PVI value should correlate with more stress-timed speech, while a high fourier Hz value and a low PVI value should correlate with more syllable-timed speech

More stress timed		Fourier (Hz)		PVI
Low fourier Hz/ High PVI	**Latino Male**	1.938	**Older Anglo F**	0.57
	Older Anglo F	2.1	**Latino Male**	0.44
	Latino Female	2.346	Latino Female	0.48
	Older Anglo M	2.466	Older Anglo M	0.43
	Younger Anglo F	2.641	Younger Anglo F	0.38
High fourier Hz/ Low PVI	Younger Anglo M	3.7	Younger Anglo M	0.31
More syllable timed				

II. Methods

II. a. Materials and Participants[2]

In order to compare the two methodologies, this study will analyze the speech of six individuals (see **Figure 4**). All participants took part in sociolinguistic interviews that lasted for approximately 1 hour. Two participants are speakers of Latino English, a variety that has been both traditionally described as syllable-timed, as well as

[2] These sections are combined because the choice of materials is the choice of participants for this analysis. The decision to not manufacture stimuli follows (Swerts, Stangert, and Heldner 1996) who demonstrate that read speech is prosodically distinct from natural speech.

quantifiably so (Carter 2005). The female speaker grew up in Venezuela and moved to the US 2.5 years before the time of the interview, although she had brief periods of prior exposure to English through week-long visits to family members in the US starting at about sixteen years of age. The male speaker moved to the US from Chile 6.5 years before the interview and had lived in North Carolina for two periods of about 3 years each. Because neither participant had exposure to native American English before puberty, there is expected to be prosodic influence from their mother tongue. Both Latino participants had been living in Hickory, North Carolina for at least 2.5 years at the time of the interview.

Four other speakers are taken from a study analyzing the speech of Anglo participants in Raleigh, North Carolina. All Raleigh participants were born and grew up in Raleigh, NC, although the older male went to college in Virginia and DC. Two of these speakers were over the age of forty at the time of the interview, while two of the speakers are young college students. Because the rhythm of White Anglo Vernacular (WAV) speakers is traditionally described as stress-timed, both PVI and LFFA should separate the WAV speakers from the Latino English speakers. Presenting two age groups of WAV speakers will also allow for a generational comparison of rhythm used by these two groups.

Figure 4. Participants

Initials	Date of Recording	Recording number	Gender	Ethnicity	Birth year
MC	02/22/2008	Ral 121	male	Anglo	1988
LD	04/06/2008	Ral 127	female	Anglo	1985
JJ	03/01/2008	Ral 122	male	Anglo	1962
DS	02/22/2008	Ral 120	female	Anglo	1927
Y	10/12/2007	Hck 094	male	Latino	1976
R	10/14/2006	Hck 08	female	Latina	1974

II. b. Data Collection

All interviews were collected through the North Carolina Language and Life Project in North Carolina between the years 2006 and 2008[3]. Spontaneous speech derived from interviews were analyzed instead of read speech, since read speech is known to exhibit different prosodic patterns from spontaneous speech (Swerts, Stangert, and Heldner 1996). Questions were constructed to encourage informal conversation and focused on aspects of everyday life, such as daily routines and holiday traditions, and participants were encouraged to openly direct the course of the conversation. Interviews were collected within the communities in question to reduce the level of formality. In order to encourage informal speech, interviews were not conducted in a sound-proof booth, so special considerations were made to ensure optimal recording quality. Electronic devices, when possible, were turned off in the room and test recordings were performed in order to ensure acceptable sound quality. When such measures failed, the interview was disregarded and another comparable interview was selected from the database.

All interviews were digitally recorded on a Marantz Professional Solid State digital recorder model PMD660 with a Sony electrat condenser lavaliere microphone model ECM – 44B at a sampling rate of 44,100 Hz using 16 bit quantization. The recordings were then checked for quality in Audacity and then uploaded as WAV files into an electronic database that is maintained in conjunction with the North Carolina State University library system and the North Carolina Language and Life Project. The

[3] Information regarding data collection is known because Mary Kohn participated in both projects and thus was informed of the methodological protocol associated with the data collection.

files were analyzed using PRAAT version 5.0.42 for Windows on an HP Pavillion Entertainment Notebook PC and an IBM Thinkpad.

II c. Data Measurements

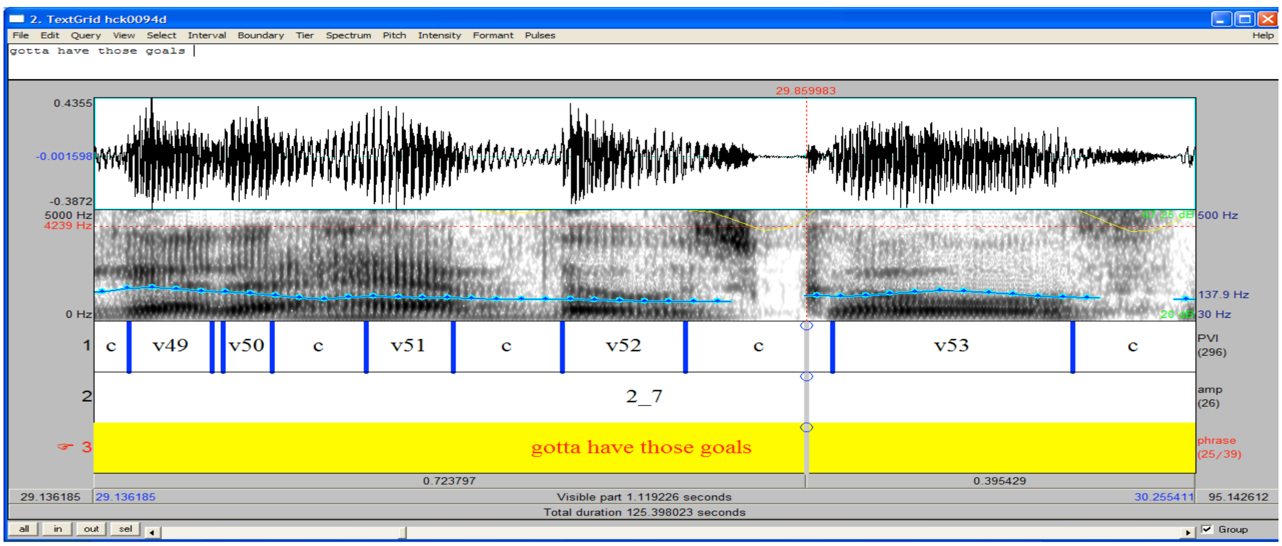
We began our PVI analysis at the first complete phrasal utterance after 600 seconds, or ten minutes, for each speaker in an attempt to avoid initial careful speech that is common as speakers adjust to the sociolinguist interview. Our methodology was taken from Carter (2005) and Thomas and Carter (2006) and is listed in detail in the chart below:

Figure 5: PVI measurement guidelines

PVI Segmenting guidelines
• All vowels, including diphthongs, are treated the same way.
• Formant transitions for obstruents were included as part of the vowel
• Vowels contained within feet before a pause of 70 ms or more, hesitation, repair or restart were marked as such and not included in the analysis
• Glides, including syllable coda /l/ and /r/, were included in the vowel measure
• vowel-medial /l/ and /r/ were assigned to the syllable to which they morphemically belonged
• when /r/ follows a stressed syllable, it is included in that vowel and syllable measure
• if the following vowel is stressed, /r/ is not included in measure and is counted as belonging to the following syllable
• if all else fails, assign /r/ to the preceding syllable
• syllabic consonants count as the “vowel” of that syllable
• For preceding /w/ and /r/, the signal was spliced so that the part of the signal after the splice sounded impressionistically like [b]. /j/ and /l/, preceding, were spliced so that the part after the splice no longer sounded like consonantal /j/ and /l/.

Generally vowels were segmented from the start of clear high amplitude formants to the end of such formants, as demonstrated in V49, V50, and V52 in figure 6. These represent the vowels in *gotta* and *those*. The stop consonants and fricatives in these words provide distinct boundaries that can be visually located by either drops in intensity in the wave form, or a lack of periodicity. The dipthongal /o/ is treated in the same way as the other two vowels. However, distinctions are not always so clear. The guidelines in Figure 2 help resolve some of these issues. For example, token V53 represents the vowel duration for *goals* in the statement *gotta have those goals though in life.* For this segment the /l/ is included in the length of the vowel as it would be difficult to determine a boundary. The boundary of a preceding /l/, though, had to be determined aurally and is not included in the vowel. The subjectivity of this method provides an example of why the PVI methodology may be considered flawed.

Figure 6: Example Spectogram



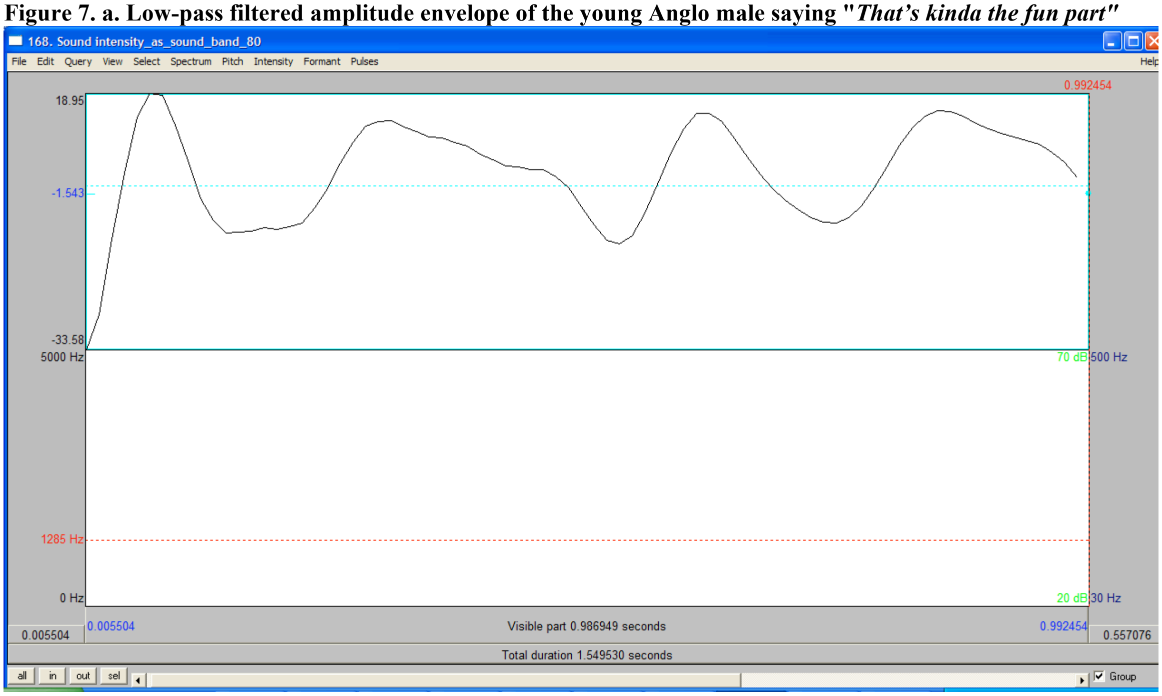
In order to prevent transcriber bias, each researcher segmented 50 tokens per speaker and tokens were double checked by both researchers. Vowel segments located in the foot before a pause were discarded according to the method proposed by Thomas and Carter (2006). A perl script was written to calculate PVI values for each syllable pair. This calculation takes the absolute value of the difference in the length of the syllable pairs divided by the total mean length of the syllables. Dividing the absolute difference of the mean length of the syllables normalizes the metric for speech rate. The script performed this calculation for each consecutive pair (i.e. V1&V2, V2&V3, V3&V4) and each PVI value was included in subsequent analyses. When referring to the average PVI per speaker, we will be referring to the median value of the total tokens taken.

To complete the LFFA, a second and third tier were added to the praat text grid. Phrases longer than 1 second that did not include pauses greater than 70 milliseconds were marked off starting at 600 seconds into the interview. The longest token taken was 4.3 seconds. Any phrase longer than that was subdivided into an acceptable length. At least ten phrases were delineated for each speaker. When possible the analysis was restricted to phrases that were contained within the PVI analysis section of the interview, although due to slow speech rate, the time of analysis had to be extended for the Latino male and Anglo female participant. Each phrase was labeled by speaker and token number in the second tier and then transcribed in the third tier of the text grid.

The manner for processing the speech signal in order to complete the fourier analysis on the amplitude envelope was adapted from Johnson and Tilsen's (2008) methodology. First, each sound sample was extracted from the file. Then the sample was band pass filtered using a pass hann band filter from 700 Hz to 1300 Hz in order to extract sonorant energy from energy resulting from frication and other extraneous sounds that are not related to perceptual "beats" as defined by Cummins and Port (1998). This differs somewhat from Johnson and Tilsen's (2008) method which employed a Butterworth filter, but allows for analysis to be completed on PRAAT software. The window for analysis was set to a Welch parabolic window, instead of the Tukey window used by Johnson and Tilsen (2008) in order to make the analysis compatible with PRAAT software. The window length in the spectrogram was set to 0.1 in order to better analyze slow variations in amplitude over the course of each utterance. The intensity contour was then extracted from the sound file. PRAAT was used to create a sound file with the amplitude wave representing the shape of the "sound wave" of the new "sound." Writing the filtered amplitude wave as a sound wave allows the software to perform a spectral analysis on the wave. Before performing the spectral analysis, the new "sound wave" that represents the amplitude envelope was once again low band pass filtered using a hann[4] filter with a cutoff of 10 Hz. This final filter prevented rapid fluctuations in amplitude such as those caused by glottal pulses from contributing irrelevant information to the spectrum. The hertz value of the highest peak for the filtered amplitude envelope should thus provide information about the periodicity of the amplitude wave.

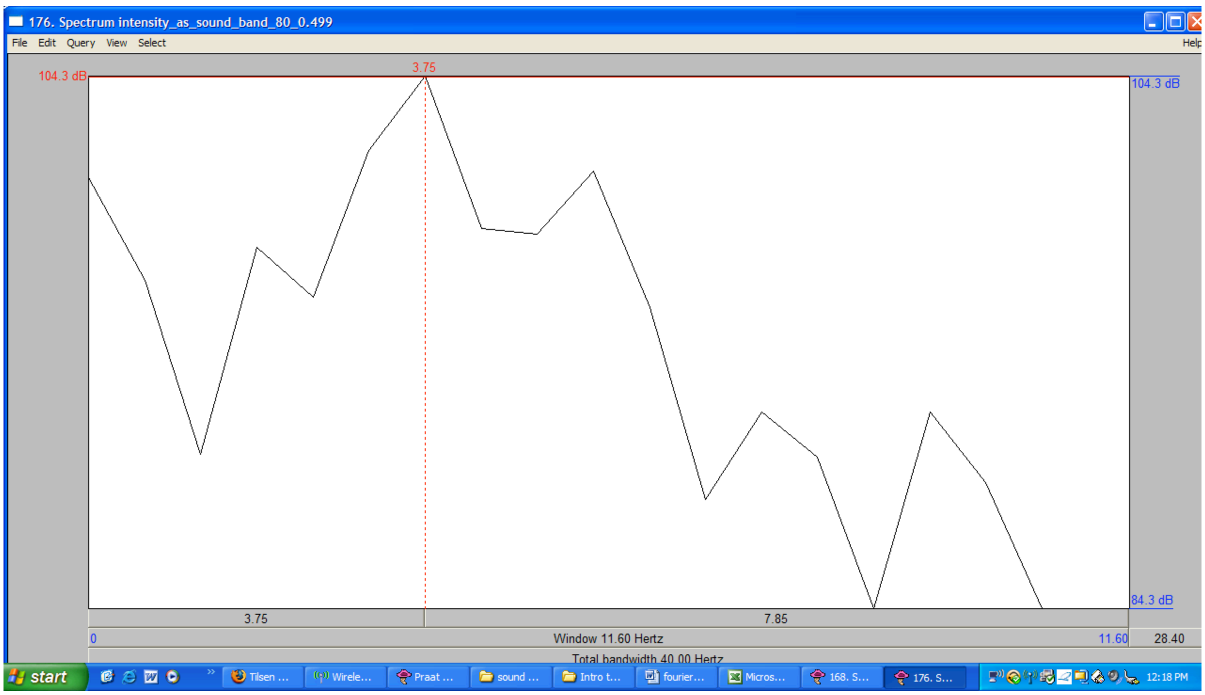
Figures 7.a. and 7.b. demonstrate how the filtered amplitude pocket corresponds to the spectral analysis of the periodicity of the signal.

Figure 7. a. Low-pass filtered amplitude envelope of the young Anglo male saying "*That's kinda the fun part*"



[4] Once again, this filter was used in place of the Butterworth filter used by Johnson and Tilsen (2008) in order to make the analysis compatible with PRAAT software.

Figure 7.b. Spectral Slice of the same signal



The amplitude pocket depicted in 4.a. comes from a phrase that is just under one second long of the young Male participant saying *that's kinda the fun part.* Within this phrase *that* receives main stress while *fun* receives secondary stress. When the non-sonorant energy is filtered out of the signal four roughly evenly distributed amplitude peaks emerge. The pattern is roughly repeating four times in one second. The fourier analysis mathematically represents a complex wave by demonstrating the hertz value and amplitude of the simple waves that can be combined to compose the complex wave. The highest peak in the spectrum should represent the largest contributing periodic sine wave that can be derived from the amplitude pocket. Because the largest repeated pattern in example 4.a. is a series of peaks that are approximately .25 seconds apart, the periodicity of the amplitude wave should be right around 4 Hz (indicating that the pattern repeats 4 times a second). Looking at the spectral analysis of this signal in figure 7.b., there is indeed a peak in amplitude at 3.75 Hz indicating that the periodicity of the amplitude pocket is captured by the method utilized above. To confirm that the method does replicate Johnson and Tilsen's (2008) methods, sound files from their paper were downloaded and the results were compared. For the sound files analyzed using this method, the results were within .10 Hz of Johnson and Tilsen's (2008) results. Having thus confirmed the methodology, the filtered amplitude pocket of the ten low band pass filtered phrases per speakers were viewed in a fourier analysis and the hertz value of the highest peak was extracted for each as a representation of the periodicity of the signal. Any peak that was lower than twice the frequency of the phrase length was disregarded, following Johnson and Tilsen's (2008) assertion that these peaks represent amplitude imbalances in the signal that are possibly caused by natural speech features like pitch accent and fillers. Finally, a rate of speech was calculated for each of the phrases analyzed by dividing the number of syllables per token by the duration of the phrase.

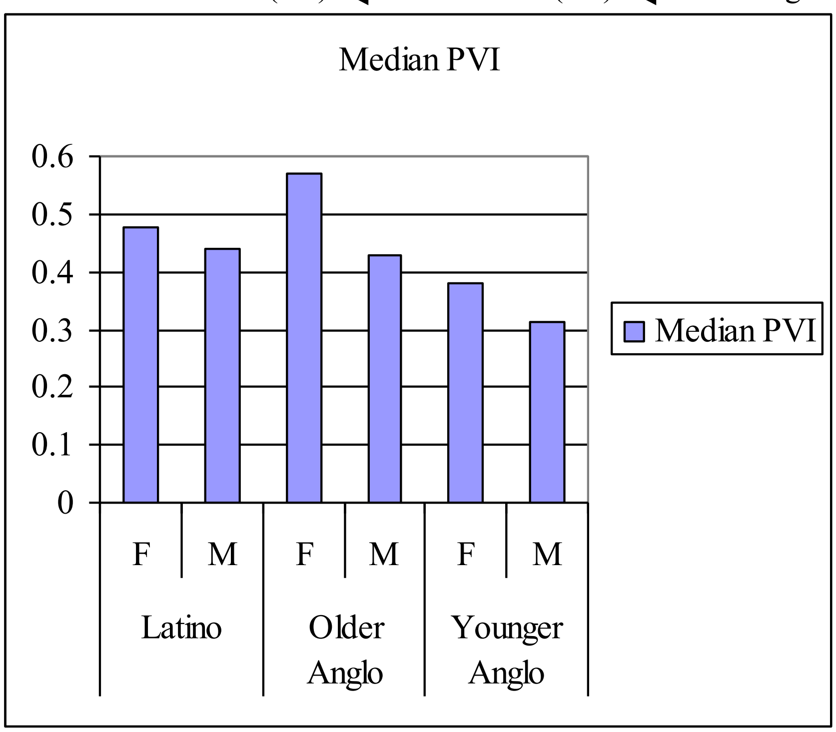
III. Results

As stated in section 1, we hypothesized that the Latino speakers would be separated out from their Anglo cohort as more syllable-timed due to the influence of their mother tongue. Spanish is considered iconic of syllable timing (Pike 1945) and Latino English dialects in the US still maintain greater syllable timing than other dialects of

English including AAVE and WAV according to Carter (2005). However, the PVI analysis did not support this generalization. The speaker with the lowest PVI value was the young white male, as demonstrated in figure 8. The Latino female had the second highest PVI value in the cohort. In fact, both the Latino speakers fell on the less syllable-timed end of the continuum when compared with other speakers in the study. That being said, the older Anglo female is the most stress-timed in the study. This finding was statistically significant when compared with all the other speakers in the study in a series of t-tests (results shown in figure 9). The only other statistically-significant difference was between the young Anglo female and male[5] indicating that the group is relatively homogenous in their production of English speech rhythm.

Figure 8: Median PVI values for each participant

LOW PVI Y. Anglo M. (.31) ← Y. Anglo F. (.38) ← O. Anglo M. (.43)
HIGH PVI Latino M. (.44) ← Latino F. (.48) ← O. Anglo F. (.57)



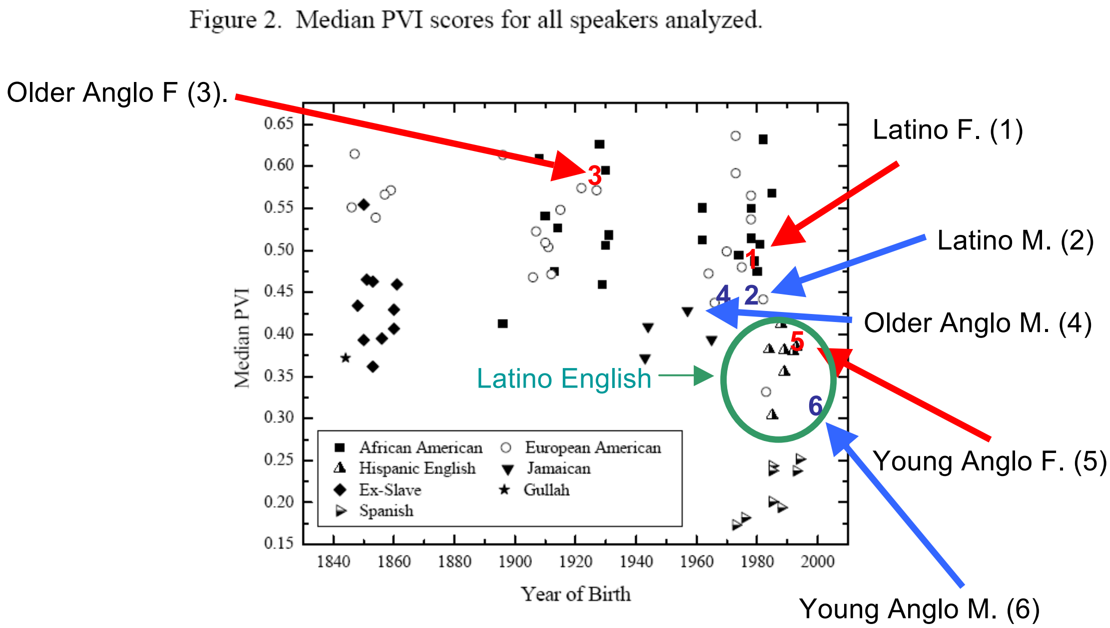
[5] P $\leq$.043, T = 2.04, sdev =.369

Figure 9: Statistical Analysis demonstrating difference between the Older Anglo Female and other participants using a 2-tailed t-test.

	P ≤	t =	stdev
Latino female	0.008	-2.67	0.364
Latino male	0.005	0.381	0.362
Anglo older male	0.008	2.69	0.373
Anglo younger female	0.03	2.16	0.406
Anglo younger male	0.001	2.04	0.369

In order to contextualize these findings, the results were overlaid on a graph representing a range of PVI values differentiated by ethnicity and age as presented in Thomas and Carter (2006). It must be noted that while the older Anglo female in our study has a PVI value that appears quite similar to several of her peers, the Latinos in this study fall almost a full tenth of a point above their cohorts. This finding may be due to differences in the community in which the data was collected. Carter (2005) notes that the Raleigh community from which the Latino English PVI values originate is an insular community. However, the Latino community in Hickory, North Carolina is not. In fact, a previous analysis has demonstrated surprisingly high levels of accommodation to local Anglo dialects, including pre-nasal ash raising (Kohn 2008), a feature that is thought to be resisted in Latino English (Thomas Carter and Coggshall 2006). The Latino female's speech was analyzed in Kohn (2008) and was found to exhibit signs of accommodation to WAV such as pre-nasal ash raising. As such, there exists the possibility that the Latinos from Hickory, North Carolina who are included in this study exhibit rhythmic accommodation to local norms as well.

Figure 10: Comparison with Thomas and Carter (2006) findings



Also, although the low PVI values for the young Anglo participants appears surprising in that WAV is expected to be rather stress-timed when compared to Latino English, there is an Anglo cohort in the chart provided above that exhibits similar low PVI values. This cohort was also born after 1980. This indicates that there is a possible rise in variation of speech rhythm among younger Anglo WAV speakers, with some showing more syllable-timing tendencies than older generations.

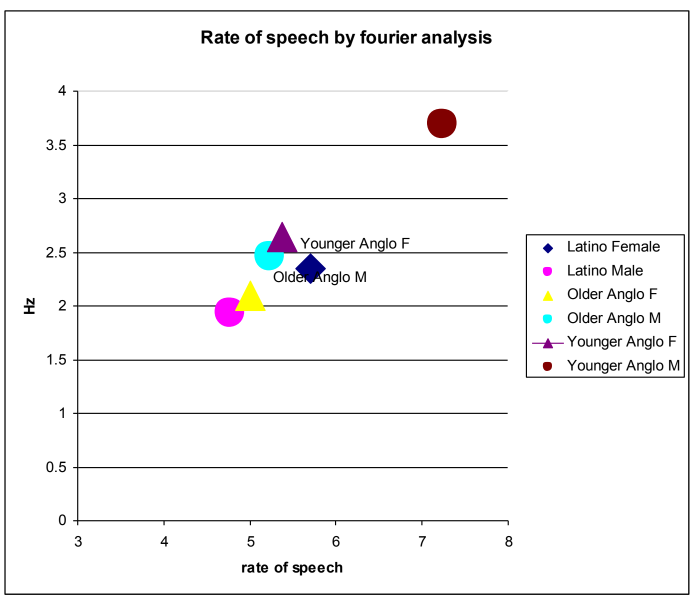
Supporting these findings, the LFFA analysis ordered the speakers in an almost identical hierarchy with the exception that the Latino male is now listed as the most stress-timed (see **Figure 11**). These results are surprising and encouraging in that they reinforce the PVI results following dramatically different methodologies. The difference between the young Anglo male and the older Anglo female and the Latino male was found to be statistically significant using a t-test[6]. All other distinctions were not significant using this statistical analysis.

Figure 11: LFFA results

High LFFA (Hz) Y. Anglo M. (3.7) ← Y. Anglo F. (2.6) ← O. Anglo M. (2.4)
High LFFA (Hz) Latino F. (2.3) ← O. Anglo F. (2.1) ← Latino M. (.1.9)

However, there is concern that such results may be biased by speech rate. As such, the speakers were plotted by the mean duration of the number of syllables divided by length of utterance against the LFFA metric in figure 12. It is not surprising that a clear correlation between speech rate and LFFA values emerge with the fastest talking speakers having higher LFFAs and the slower speakers having lower LFFAs.

Figure 12: Graphic Representation of Speech Rate



[6] Results of t-test between young Anglo male and older Anglo female: t= -2.12, sdev= 1.69, degrees of freedom = 18 The probability of this result, assuming the null hypothesis, is 0.048. Results of t-test between young Anglo male and Latino male: t= -2.21, sdev= 1.78, degrees of freedom = 18 The probability of this result, assuming the null hypothesis, is 0.040

When reflecting on the nature of the measure, such a correlation is expected as the repeating fluctuations in the amplitude pocket have a greater opportunity to repeat within the speech sample if the syllables containing the sonorant peaks are of a shorter duration. This may even explain the switch in ranking of the Latino male as he is the slowest speaker in the group with an average rate of speech of 4.77 syllables per second. This is compared with the total average speaking rate of the group of 5.55 syllables per second.

IV. Discussion and Further Directions

The results of this analysis have encouraging implications for the study of speech rhythm in that two vastly different metrics predicted similar hierarchies of rhythm for the speakers in this study. The validity of this correlation is supported by the fact that the hierarchy that emerged ran contrary to expectations, indicating that the similarities between the results are unlikely to be due to bias on the part of the researchers. As discussed in the introduction, a positive correlation between the two methods may prove encouraging for the adoption of the LFFA analysis for comparing speech rhythm in that the method avoids the tedious and difficult segmenting process involved in calculating PVI. However, speech rate appears to have a great affect on LFFA. This observation proves problematic in that speech rate is difficult to control in naturalistic speech settings. Results can potentially be biased by unusually rapid or slow speakers. As such, before adopting this methodology for wide-spread use in a sociolinguistic analysis, a method for normalizing for speech rate should be established. This is our next step towards establishing a rhythm analysis method that reduces a reliance on forcing phonemic boundaries on phonetic output.

We would also like to create stimuli based on each methodology in order to test perceptions against said stimuli. It would be especially interesting to vary this stimuli by speech rate as well as by the method used to produce the rhythm, as speech rate may be important to the perception of rhythm which would indicate that such normalization is not appropriate for speech rhythm measures.

References

Carter, P. (2005). Prosodic variation in SLA: Rhythm in an urban NC Hispanic Community. Penn Working Papers in Linguistics, 11.2.

Coggshall, E. L. (2008). The prosodic rhythm of two varieties of native American English. Penn Working Papers in Linguistics 14.2.

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.

Johnson, K. and Tilsen, F. (2008). Low frequency Fourier analysis of speech rhythm. Journal of the Acoustic Society of America. 124 (2): EL34-EL39.

Kohn, M. (2008) "Latino English in North Carolina: A Comparison of Emerging Communities" Unpublished M.A. thesis..

Low E.L, Grabe E and Nolan F. (2000) Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. Language and Speech 43 (3):229 – 259.

Pike, K. (1945). The Intonation of American English. Ann Arbor: University of Michigan Press.

Steiner, I. (2003) A refined acoustic analysis of speech rhythm. LingColl 2003, Budapest.

Swerts, Marc, Eva Strangert, and Mattias Heldner. 1996. “F0 declination in read-aloud and spontaneous speech.” In Proceedings of the Fourth International Conference on spoken Language Processing, Philadelphia, October 1996. Vol. 3. Piscataway, NJ: IEEE, 1501-04.

Thomas, Erik and Phillip Carter (2006). Prosodic Rhythm and African American English. English World Wide. 27: 331-355.

Thomas, Erik; Phillip Carter, Elizabeth Coggshall. 2006. Acoustic evidence for new dialect formation. Paper presented at the American Dialect Society Annual Meeting:Albuquerque.