EFFECTS OF LANGUAGE, SPEAKING STYLE AND AGE ON PROSODIC RHYTHM

Gayatri Rao[a] *& Rajka Smiljanic*[b]

[a]Department of Psychology; [b]Department of Linguistics, University Of Texas at Austin, USA
raog@mail.utexas.edu; rajka@mail.utexas.edu

ABSTRACT

The present study uses the envelope modulation spectrum (EMS) measure to assess the effect of language, speaking style, and age on rhythm. Three experiments using a weighted mean of the EMS called the spectral centroid were conducted. The first examined rhythm in English and Croatian across conversational and clear speaking styles. The second experiment compared rhythmic characteristics of monolingual English and Croatian speakers to Croatian-English bilingual speakers. Finally, we compared the rhythmic characteristics of clear and conversational speech utterances produced by young and elderly adult native speakers of English. Nested factorial analyses revealed that spectro-temporal characteristics of rhythm were affected by language, speaking style, and age. Talker emerged as a significant factor as well. These results suggest that low-rate amplitude modulations can be used to characterize rhythmic distinctions due to the factors examined here. Moreover, these spectral measures are capable of bypassing phonologically dependent duration measurements that are laborious and time consuming to extract.

Keywords: rhythm, envelope modulation spectrum, spectral centroid

1. INTRODUCTION

Previous attempts to classify language rhythm in terms of stress and syllable-timed patterns have yielded mixed results [2, 4, 6]. Arvaniti [1] argues that the measures used in such studies confound rhythm and timing by only considering consonantal and vocalic duration variability. As a result, these metrics are able to successfully classify prototypical languages but fall short with non-prototypical and mixed languages.

Recent spectral approaches [5, 12] have explored the use of low frequency modulation information of speech to describe rhythmicity. Tilsen and Johnson [12] found that the amplitudes of the two most prominent peaks of the power spectrum differentiated stress-, syllable- and mora-timed languages. Liss, LeGendre and Lotto [5] found that the EMS distinguished between the rhythm of dysarthric and control speech, of various types of dysarthric speech and across gender and talker.

In the current study, we use a ‘center of mass’ measure called the spectral centroid in combination with EMS (called the EMS + centroid approach) to explore rhythmic variation in English and Croatian. This approach builds on the spectral measure previously used in [5, 12] by looking at the spectral centroid instead of the most prominent peak, thus preventing loss of pertinent information in the surrounding regions. Additionally, spectral measures include more than syllabic information by including disfluencies and pauses. As a result, they provide an account of syllable prominence, stressed and unstressed syllable variation and their distribution. Moreover, they do away with a need for phonologically dependent criteria and tedious manual measurements.

English and Croatian are well suited for this analysis since they differ in their sound and prosodic properties. While English is a stress-timed language, Croatian is considered a rhythmically mixed language with complex consonant clusters and no vowel reduction. In addition to language, speaking style was examined. Clear speech is a distinct, listener-oriented speaking style that differs in acoustic-articulatory features from conversational speech [10]. The two speaking styles are expected to differ rhythmically in all three experiments.

2. EXPERIMENT 1: ENGLISH AND CROATIAN RHYTHM

Experiment 1 examines the effects of language (English vs. Croatian), and speaking style (conversational vs. clear) on rhythm.

2.1. Stimuli

The stimuli were 20 short semantically anomalous sentences in English and Croatian (e.g., *Your* *tedious beacon lifted our cab*) previously used in [9].

2.2. Subjects

The recordings of two native speakers of American English (AE) and two native speakers of Croatian were analyzed. Their ages ranged from 24-32 years of age (mean = 28). All speakers were female and had no known speech or hearing impairment.

2.3. Procedure

The native monolingual AE speakers recorded 20 sentences in English (ENG_{ENG} sentences). Two native speakers of Croatian recorded 20 corresponding sentences in Croatian (CRO_{CRO} sentences). All speakers were asked to read the sentences in clear (CL) and conversational (CO) speaking styles. All acoustical analyses were conducted using Praat 5.2.05 software and statistical analysis was conducted using R and Matlab. The peak amplitude of all waveforms was first normalized to 1. Matlab was used to extract the EMS for each waveform.

2.4. Measurements

For the purposes of this study, EMS was used to extract sentence-level rhythm information. Since modulation frequencies in the range 0-16 Hz contribute to sentence intelligibility [3], the cutoff frequency used in this study was 16 Hz. In order to include temporal information provided by less prominent peaks, this study used the spectral centroid in combination with EMS to calculate rhythm. The spectral centroid is a weighted mean of the spectrum and is calculated using the following formula:

(1) $$Centroid = \frac{\sum_{n=0}^{N-1} Frequency(n) Amplitude(n)}{\sum_{n=0}^{N-1} Amplitude(n)}$$

2.5. Results

A 2X2X2, nested factorial analysis of variance (ANOVA) was conducted. The language (ENG_{ENG}, CRO_{CRO}), speaker id (SP1 or SP2) and style of speech (CL or CO) served as independent variables and the spectral centroid of the EMS of each sentence served as the dependent variable. The ANOVA revealed significant main effects of language ($F(1, 19) = 6.0154$, $p < 0.05$), style ($F(1, 76) = 78.2431$, $p < 0.001$) and speaker ($F(1, 38) = 51.615$, $p < 0.001$). Two interactions between language and speaker ($F(1, 38) = 37.291$, $p < 0.05$) and language and style ($F(1, 76) = 5.3808$, $p < 0.001$) were significant. Post-hoc analyses revealed significant differences between English and Croatian rhythm ($p < 0.05$) as well as conversational and clear speaking styles ($p < 0.05$). Furthermore, speaker 1 and speaker 2 were also significantly different ($p < 0.05$).

Table 1: Table of centroid means for language and style (in Hz).

Factor	Factor level	Mean (SD)
Language:	English	8.8474 (1.1475)
	Croatian	8.2909 (2.0736)
Style:	Conversational	9.30460 (1.1792)
	Clear	7.8337 (1.8131)

2.6. Discussion

The results showed that rhythm significantly varied for the two languages. English speech had a higher spectral centroid than Croatian speech. In addition, rhythmic properties varied across the two speaking styles. Conversational speech demonstrated a larger centroid value than clear speech. These higher centroid values may reflect differences in syllable structure and distribution of prominent units across languages and speaking styles. Finally, speaker emerged as a significant factor in our analyses. The results suggest that the rhythm of a language is implemented differently by individual speakers and across speaking styles.

3. EXPERIMENT 2: BILINGUAL RHYTHM

In experiment 2, the effects of native language background (English vs. Croatian vs. Bilingual), and speaking style on rhythm were analyzed. It was hypothesized that there would be a significant effect of native language background on rhythm. Croatian-English bilingual speech rhythm is expected to vary significantly from that of the monolingual speakers of the target language [13].

3.1. Stimuli

Stimuli were the same as in experiment 1.

3.2. Subjects

Two highly proficient Croatian-English bilinguals recorded the same stimuli as in 1. They learned English in Zagreb, Croatia and were undergraduate students in the US at the time of the recording [8]. Their ages ranged from 24-32 years of age (mean = 28) and had no known speech or hearing impairment.

3.3. Procedure and measurements

Procedure and measurements were the same as in experiment 1.

3.4. Results

A 3X2X2, nested factorial analysis of variance (ANOVA) was conducted. The background language (ENG_{ENG}, CRO_{CRO} or ENG_{CRO}), speaker id (SP1 or SP2) and style of speech (CL or CO) served as independent variables and the spectral centroid of the EMS of each sentence served as the dependent variable. The ANOVA revealed significant main effects of language ($F(2, 38) = 10.1140$, $p < 0.001$), speaker ($F(1, 57) = 45.7920$, $p < 0.001$) and style ($F(1,114) = 90.9652$, $p < 0.001$). There was also significant interaction between language and speaker ($F(2, 57) = 19.7520$, $p < 0.001$). Additional post-hoc Bonferroni tests ($p = 0.05$) showed that ENG_{CRO} rhythm differed significantly from CRO_{CRO} but not from ENG_{ENG}. ENG_{ENG} did not differ significantly from CRO_{CRO}. Bonferroni tests ($p = 0.05$) also showed that speakers 1 and 2 differed significantly.

Table 2: Table of centroid means for language and style.

Factor	Factor level	Mean (SD)
Language:	Bilingual	9.4041 (1.6498)
	English	8.8474 (1.1475)
	Croatian	8.2909 (2.0736)
Style:	Conversational	9.5685 (1.2356)
	Clear	8.1264 (1.8376)

3.5. Discussion

The results of experiment 2 revealed that all three factors, language, speaking style and speaker significantly affected spoken language rhythmic properties. Bilingual speech showed the largest spectral centroid values, followed by English and then Croatian signaling language timing and syllable structure differences across monolingual and bilingual speech. Post-hoc tests showed that although ENG_{CRO} rhythm differed from CRO_{CRO}, it did not differ from ENG_{ENG} suggesting that these non-native speakers approximated the rhythm of the target L2 language. In this analysis, ENG_{ENG} did not differ from CRO_{CRO}. It is possible that the introduction of large variability in the rhythmic characteristics of the bilingual speech masked the cross-language differences. It is likely that the two Croatian speakers came from two distinct dialectal areas while the AE speakers belonged to the same dialectal area. Previous research has shown that rhythmic variation within a language can be as large as across languages [4]. CO speech had a significantly higher spectral centroid than CL speech. The results also revealed that the precise rhythmic characteristics within a language vary across speakers.

4. EXPERIMENT 3: ELDERLY RHYTHM

In experiment 3, the effects of age, and style on rhythm were examined. Elderly speech is slower and contains more disfluencies [6] which will affect EMS and result in rhythmic differences in the two age groups.

4.1. Stimuli

Stimuli used were the same as in experiments 1 and 2.

4.2. Subjects

Two elderly adult native speakers of English who participated in [11] were used. Their age range was between 66-78 years old (mean = 72.6). They had no known speech or hearing impairment. Two young adults were the same as in 1.

4.3. Procedure and measurements

Procedure and measurements were the same as in experiment 1 and 2.

4.4. Results

A nested factorial, 2X2X2 ANOVA was used with age group (young or elder), style (CL or CO) and speaker (SP1 or SP2) as independent variables and the spectral centroid as the dependent variable. The results showed significant main effects of age ($F(1, 19) = 18.040$, $p < 0.001$) and style ($F(1, 38) = 59.5311$, $p < 0.001$) on rhythm and a significant interaction between age and speaker ($F(1, 76) = 5.5690$, $p < 0.001$). Thus, the differences in rhythm across age groups were affected by individual differences of talkers within each group. Post-hoc pairwise comparisons (Bonferroni tests, $p = 0.05$) showed a significant difference between the young and elder age groups but no significant differences between the speakers.

Table 3: Table of centroid means for age and style.

Factor	Factor level	Mean (SD)
Age:	Old	9.6064 (1.4492)
	Young	8.8474 (1.1475)
Style:	Conversational	9.7474 (1.0350)
	Clear	8.7065 (1.4453)

4.5. Discussion

The results revealed a significant effect of age on utterance-level rhythm. Elderly adults had a higher centroid value compared to young adults suggesting greater variability in properties that affected rhythm such as the rate of vowel and consonant reduction and deletion in their speech. Rhythm also differed across the two speaking styles for all talkers. Clear speech demonstrated a lower spectral centroid than conversational speech. In addition to speaking style and age, the rhythm across age groups was also affected by individual differences. Even though there weren't notable individual differences in rhythm in the entire sample, within each age group, speaker was a significant factor.

5. GENERAL DISCUSSION AND CONCLUSIONS

Combined, the above results demonstrate that the centroid of the EMS is capable of distinguishing the rhythmic characteristics of native speakers of Croatian and English and native and non-native speakers of English. In addition, it is also capable of detecting differences in speaking styles, age groups and individual talkers.

The following trends were observed for the magnitude of the centroid from the analyses above:

$\text{Centroid}_{\text{ENG}} > \text{Centroid}_{\text{CRO}}$
$\text{Centroid}_{\text{bilingual CRO}} > \text{Centroid}_{\text{monolingual CRO}}$
$\text{Centroid}_{\text{conversational}} > \text{Centroid}_{\text{clear}}$
$\text{Centroid}_{\text{elder}} > \text{Centroid}_{\text{young}}$

A larger centroid can arise from increased energy in the higher frequencies of the spectrum. This increased information in the latter portion of the spectrum signals more prominent faster components. These high frequency components may be related to stress patterns as well as segmental information such as consonant and vowel reductions and deletions [5, 12].

The above trends demonstrate that Croatian and English are rhythmically distinct. Furthermore, high proficiency bilingual speakers in this experiment produced rhythmic patterns that closely resembled the target L2 rhythm. The results also showed that within-language utterances produced by younger and older adults and in two different speaking styles differ in their rhythmic properties.

Unlike Smiljanić and Bradlow [9] who found temporal stability across conversational and clear speech using durational measurements, the spectral measure used here detected rhythmic differences across the two speaking styles. This discrepancy points to the differences in the two measurements. That is, the EMS measure captures rhythmic patterns due to voicing, stop bursts, fricatives, etc. allowing insight into the rhythmic features that extend beyond durational properties. Finally, individual speaker differences were found in the above analysis. However, since speaker 1 and 2 weren't held constant across conditions, further analysis is necessary to interpret the specific effects of speaker on the rhythm centroid and will be explored in future work.

Given the small sample size of subjects and languages, these results offer a preliminary but promising description of the use of the EMS + centroid approach in quantifying rhythm. Future studies will examine the use of this approach in quantifying languages, speakers etc. in greater depth.

6. ACKNOWLEDGEMENTS

We are grateful to Andrew Lotto for guidance with EMS.

7. REFERENCES

[1] Arvaniti, A. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66, 46-63.
[2] Dauer, R.M. 1983. Stress-timing and syllable-timing reanalyzed. *JP* 11, 51-62.
[3] Drullman, R., Festen, J.M., Plomp, R. 1994. Effect of temporal envelope smearing on speech reception. *JASA* 95, 1053-1064.
[4] Grabe. E., Low, E.L. 2003. Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology* 7, 515-546.
[5] Liss, J.M., LeGendre, S.J., Lotto, A.J. 2010. Discriminating dysarthria type from envelope modulation Spectra. *J. Speech Lang. and Hearing Sci.* 53, 1246-1255.
[6] Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265-292.
[7] Smith, B.L., Wasowicz, J., Preston J. 1987. Temporal characteristics of the speech of normal elderly adults. *J. Speech and Hearing Research* 30, 522-529.
[8] Smiljanić, R., Bradlow, A.R. 2007. Clear speech intelligibility: Listener and talker effects. *Proc. of the 16th ICPhS* Saarbrucken, Germany.
[9] Smiljanić, R., Bradlow, A.R. 2008. Stability of temporal contrasts across speaking styles in English and Croatian *JP* 36(1) 91-113.
[10] Smiljanić, R., Bradlow, A.R. 2009. Speaking and hearing clearly: talker and listener factors in speaking style changes. *Linguistics & Language Compass* 3(1), 236-264.
[11] Smiljanić, R. In preparation. *Effect of Age on Clear Speech Intelligibility*.
[12] Tilsen, S., Johnson, K. 2008. Low-frequency fourier analysis of speech rhythm. *JASA* 124(2), EL34-EL39.
[13] White, L., Mattys, S.L. 2007. Calibrating rhythm: First language and second language studies. *JP* 35, 501-522.