QUANTIFYING MACRO-RHYTHM IN ENGLISH AND SPANISH

Christine Prechtel

University of California, Los Angeles
cprechtel@ucla.edu

ABSTRACT

This study measured and quantified macro-rhythm in English and Spanish. According to Jun's prosodic typology [1], macro-rhythm is phrase-medial tonal rhythm whose domain is equal to or slightly greater than a word. Macro-rhythm strength is determined by the regularity of f0 slope shape, the regularity of peak/valley distance intervals, and the number of peaks per word per sentence. Jun's model predicts that Spanish has stronger macro-rhythm than English because the most common pitch accent is L+<H* in Spanish and H* in English, and Spanish tends to accent content words with greater regularity than English. Seven speakers of each language read twenty sentences, which were measured for slope shape and peak/valley distance intervals (variability measures), and peak frequency. The preliminary results showed that the variability measures did not differ between languages, but that the peak frequency did. Therefore, the findings quantitatively support the prediction that Spanish has stronger macro-rhythm than English.

Keywords: prosodic typology, intonation, macro-rhythm, AM model, intonational phonology

1. INTRODUCTION

Over the years, researchers have sought to understand how and why prosodic features vary greatly across languages. Jun [1, 2] proposed a model of prosodic typology based on the Autosegmental-Metrical (AM) framework of intonational phonology. According to the AM model, intonation marks two major properties: prominence and phrasing [e.g. 3, 4, 5]. Intonational tunes are composed of pitch accents, which are prominent pitch targets or movements that mark the head of a word (e.g. a stressed syllable), and boundary tones, which are pitch targets or movements that mark the edge of a prosodic unit. Therefore, [1, 2] includes prominence and phrasing as parameters in the prosodic typology model. The prosodic properties of an utterance are a combination of word-level and phrase-level prosody in both of these parameters. The prosodic typology model [1] categorizes languages based on prominence, phrasing, and a third parameter, macro-rhythm (tonal rhythm).

The first parameter, prominence, is marked at the lexical level through one or a combination of the following: pitch accent, stress, and tone. Some languages may not mark lexical prominence at all, e.g. Mongolian and Seoul Korean. At the post-lexical or phrasal level, prominence is categorized based on whether it is marked by the head of the phrase (Head), such as a nuclear pitch accent, by a boundary tone at the phrase edge (Edge), or by both (Head/Edge). Languages can be Head-prominent like English and Spanish, Edge-prominent like Seoul Korean, or both like Bengali and Japanese [2].

The second parameter, phrasing, is categorized by the lexical and post-lexical prosodic units of a language [2]. At the lexical level, these units include morae, syllables, and feet, which contribute to traditional notions of speech rhythm (syllable-timed vs stressed-timed). Post-lexical units include the Accentual Phrase (AP), Intermediate Phrase (ip), and Intonational Phrase (IP).

The third parameter, macro-rhythm, is phrase-medial tonal rhythm, i.e. the regularity of high/low f0 alternations, whose unit is equal to or slightly greater than a Prosodic Word [1]. It is defined by the degree of rhythmic strength in f0; languages with frequent high/low f0 alternations, similar f0 rising and falling slopes, and similar distance intervals between peaks and valleys are said to have stronger macro-rhythm than other languages with less frequent alternations, less similar slopes, and less similar peak and valley intervals. The inclusion of macro-rhythm as a parameter for cross-linguistic comparison is based on the following phonological criteria: the number of phrase-level tones in a language's tonal inventory, the most common type of phrase-medial tone, and the frequency of f0 rise per word in a phrase. The model can therefore predict the strength of macro-rhythm in any language based on the prosodic structure as described in the AM framework.

The purpose of this study is to quantify and compare the macro-rhythm strength of two languages, American English and Mexican Spanish (henceforth English and Spanish). Although the strength of macro-rhythm for Spanish proposed in [1] was based on Castilian Spanish, results from Mexican Spanish (or any other variety) would also make language-general predictions about the macro-rhythm strength of Spanish, given the similarities in the intonational models across dialects [6]. English and Spanish were chosen for comparison because they are both Head-prominent languages with lexical stress.

While both languages have multiple types of pitch accents in their respective tonal inventories, the most common pitch accent in English is H* [7, 8], while the most common prenuclear pitch accent in Spanish is a rising pitch accent, L+<H* [9, 10, 11]. Therefore, Spanish is expected to have more f0 alternations than English.

In addition, the two languages differ with regards to the regularity at which content words (CWords) are pitch accented. Both Spanish and English tend to deaccent some types of CWords such as verbs [5, 12, 13, 14], and this varies by speech style; spontaneous speech is more likely to have deaccenting than lab speech [15]. However, according to [16], Spanish intonation, with some exceptions, is generally analysed with the expectation that every CWord bears a pitch accent. Furthermore, [17] found that Spanish places pitch accents on both new and old information, in contrast to languages such as English, where old information is deaccented [18]. Therefore, Spanish is expected to accent CWords with greater regularity than English.

To summarize, Spanish is predicted to have stronger macro-rhythm than English because the most common pitch accent in Spanish is L+<H* while it is H* in English, and CWords are pitch accented with greater regularity in Spanish than in English. The current study tests this prediction quantitatively.

2. METHODS

Participants were recruited from an undergraduate population, and they received course credit for their participation. They were either monolingual native speakers of American English or bilingual speakers of Mexican Spanish. Eligibility was determined through a language questionnaire before the start of the experiment. For the monolingual English group, participants who indicated that they had learned a language other than American English in their childhood were excluded from analysis. Speakers in either group who were disfluent readers were also excluded. A total of fourteen speakers were analyzed, seven speakers for each language.

Twenty sentences, each containing five CWords with a varying number of function words in between, were created for each language. The number of unstressed syllables between the stressed syllables, i.e., interstress interval (ISI), varied so that sentences would vary in the location of pitch accents within a sentence. Sentences were designed so that the total number of different ISI was similar between languages. This was to prevent any difference in pitch accent realizations between the two languages to be the result of the difference in the sentence material, especially the distance (in the number of syllables) between any two adjacent pitch accents. Since each sentence had five CWords, it was predicted to have a maximum of five pitch accents, thus five f0 peaks. Therefore, each speaker would produce a maximum of 100 CWords (5 CWords x 20 sentences).

To reduce the likelihood of disfluencies, participants were first given the list of sentences to read silently to themselves. They were then presented with each sentence one at a time on a computer screen and asked to read the sentence aloud fluently and without any pauses. Each sentence appeared twice, and two filler sentences were shown at the beginning of the experiment to familiarize the participants with the reading task. Each group was only presented with sentences in their target language, i.e. monolingual English speakers only read English sentences, and bilingual Spanish speakers only read Spanish sentences. All recordings were made in a sound-attenuated room at a sampling rate of 44.1 kHz (32 bit) using SM10A ShureTM microphone and headset.

Jun [1] proposed two ways to quantify macro-rhythm. The first way is to calculate the Macro-rhythm Variation Index (MacR_Var), which is the sum of the standard deviations of the rising slope (rSD) and falling slope (fSD), peak-to-peak distance (pSD), and valley-to-valley distance (vSD), summarized in (1).

$$\text{MacR\_Var} = \text{rSD} + \text{fSD} + \text{pSD} + \text{vSD} \quad (1)$$

A high number of MacR_Var is considered weakly macro-rhythmic because the large variability suggests irregularly shaped peaks and/or variable distance intervals between peaks. Therefore, English is predicted to have a higher MacR_Var value than Spanish.

The second way to calculate macro-rhythm is to count the frequency of low/high alternations in a phrase, known as the Frequency Index (MacR_Freq). These alternations should roughly correspond to the size of a Prosodic Word (PWord), i.e., a Cword plus surrounding unaccented function words and/or clitics. The MacR_Freq is calculated by dividing the number of f0 peaks per sentence by the number of PWords in the sentence, as summarized in (2). A language with stronger macro-rhythm will have a MacR_Freq value close to 1, meaning each PWord will have one f0 peak. Therefore, Spanish is predicted to have a MacR_Freq value closer to 1 than English.

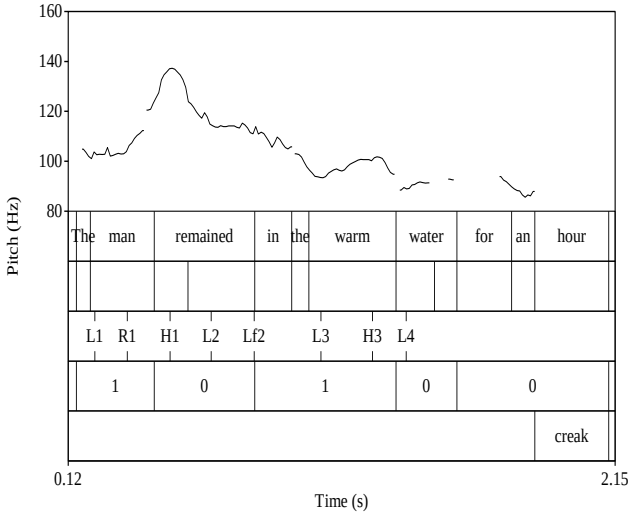
$$\text{MacR\_Freq} = \frac{\text{Number of f0 peaks per sentence}}{\text{Number of PWords per sentence}} \quad (2)$$

The first repetition of each sentence was chosen for analysis unless it was too disfluent, in which case the second repetition was analysed instead. Sentences

were excluded from analysis if they did not contain a minimum of three consecutive non-disfluent CWords. The recordings were annotated by hand in Praat [19] and labelled for words, syllables, turning points in the pitch track (f0 labels), and the number of peaks in the sentence. A script was used to extract the time and height values of the f0 labels, which were used to calculate peak-to-peak distance (ms), valley-to-valley distance (ms), rising slope, and falling slope.

Figures 1 and 2 show examples of the labelling for each language. Tier 3 marks f0 turning points with the following labels: L for low (valley), R for rise, H for high (peak), and Hf and Lf for a fall after a high or low f0 plateau, respectively. L was determined by the lowest point before the next f0 rise; R was labelled at the end of a low plateau just before the start of the rise for the following high target; H was determined by the highest point in the peak; Hf marked the end of a high plateau before falling to a low f0 point; and Lf marked the end of a low plateau before falling to even lower f0 point. The number after the tone label indicates the order in which it occurred in the sentence. For example, in Figures 1 and 2, L1 is followed by H1, which is followed by L2, etc. Because macro-rhythm is defined as phrase-medial tonal rhythm, sentences were only labelled up to the final H to avoid influence from the boundary tone. If there was no final H, which was common in the English data, the last L point was labelled before the f0 dropped again.

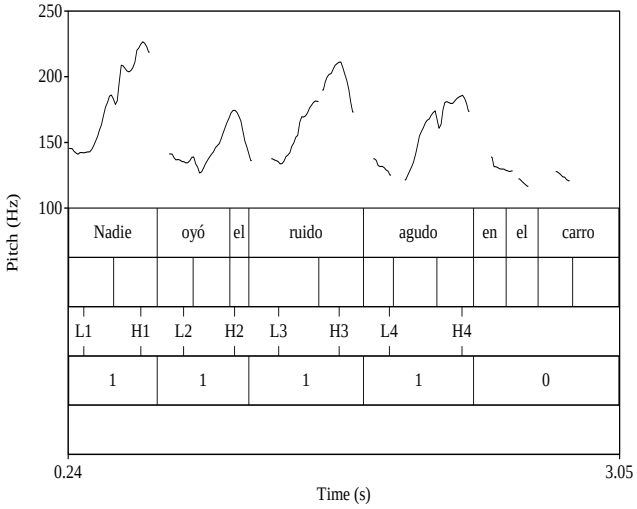
Figure 1: Example of an annotated sentence read by a male English speaker. H=high peak, L= low valley, R=rise, Lf=fall at the end of a plateau to an even lower f0. The labels are numbered in the order in which they occur in the utterance. There is no H2 label because of the plateau fall from L2 to L3.



Rising slope was calculated by taking the difference between the H label and the preceding L label, or the R label if the L target was followed by a low plateau. Similarly, falling slope was calculated by taking the difference between the L label and the preceding H (or Hf) label.

The tier below the f0 labels captures the number of peaks per word per sentence. The presence of a peak within the PWord interval was marked with a '1' and the absence of a peak with a '0.' A sentence with a greater number of '1' labels is predicted to have stronger macro-rhythm than a sentence with a fewer number of '1' labels. The bottom tier was for comments, noting creakiness or truncated syllables, which could affect f0 perturbation and alignment.

Figure 2: Example of an annotated sentence read by a male Spanish speaker. The labels are numbered in the order in which they occur in the utterance.



3. RESULTS

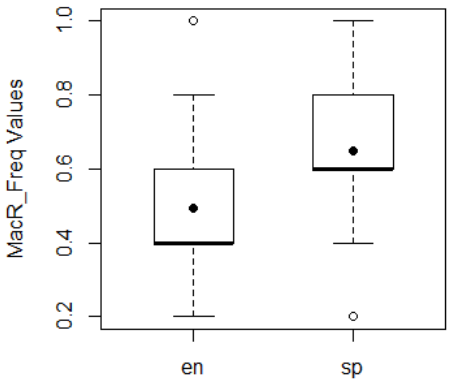
Some of the speakers were more disfluent readers than others, so not every speaker contributed the maximum 100 CWords. A total of 695 pitch accent-bearing words were analysed in English and a total of 659 words were analysed in Spanish.

To calculate the MacR_Var values, the standard deviations were taken for rising slope, falling slope, peak-to-peak distance, and valley-to-valley distance. The raw data were then transformed into z-scores and added together. A two-sample one-tailed t-test showed that the Spanish speakers did not have less overall variation than English speakers ($t(12) = 0.45$, $p = 0.33$). Because the measures were combined into a single score, it was unclear if certain measures differed by language. To address this, linear mixed effects models were run for each of the four measures individually, with group as the predictor and speaker

as the random intercept. None of the models showed a significant effect for group.

MacR_Freq values were also calculated, and Figure 3 summarizes the distribution of the data for each language group. On average, the values for Spanish were higher than English, indicating that Spanish has more peaks per word per sentence than English. However, there was some within-group variation. Although most English speakers had MacR_Freq values between 0.4 and 0.45, two speakers had higher values, indicating a more rhythmic speech pattern in their reading style. Similarly, two Spanish speakers had values below 0.6, indicating a less rhythmic speech pattern in their reading style.

Figure 3: Distribution of the MacR_Freq values by language group. The means are represented by black dots.



The results of a two-sample one-tailed t-test showed that English had fewer peaks per sentence than Spanish ($t(12) = -2.42$, $p = 0.02$). To further test these differences, a linear mixed effects model was run in Stata [20]. The dependent variable was the number of peaks per sentence, with group as the categorical predictor and speaker as the random intercept. Sentence was not included as a random effect because there was not enough variability for the model to converge. Results showed that group was a significant predictor ($\beta = 0.78$, SE = 0.32, z = 2.42, $p = 0.02$), meaning that Spanish speakers had more peaks per sentence than English speakers. These results support the hypothesis that Spanish has stronger macro-rhythm than English.

4. DISCUSSION

The results indicate that Spanish has stronger macro-rhythm than English in peak frequency per word per sentence. Specifically, the MacR_Freq index is a useful measure for quantifying macro-rhythm. In the mixed effects model, there was a significant difference between Spanish and English based on the number of peaks. As expected, Spanish had higher values overall, reflecting the greater number of phonological low/high alternations than English and contributing to the perception that Spanish has more regular tonal alternations and therefore stronger macro-rhythm.

In contrast, the MacR_Var index did not capture the strength of macro-rhythm. The rationale behind MacR_Var was to reflect the overall variation of the distance interval and shape of f0 of each tonal unit within a speaker. However, the differences in variability were not enough to be significant across language group. Similarly, comparing individual measures such as rising slope and peak-peak distance across language also did not yield significant results. This suggests that quantifying overall variability may not capture meaningful differences between slope shape and timing information. This is perhaps surprising for peak distance and valley distance, which one might expect to differ if Spanish has more low/high alternations and more regularly pitch-accented CWords than English. The lack of an effect could be partially attributed to the small number of subjects in the study. One might predict that there is a correlation between the number of peaks and the peak-to-peak distance in a sentence, where more peaks would decrease the distance intervals between them, and thus reduce the variability in distance. Since the MacR_Var index did not consider the number of peaks, this remains an open question for future investigation.

There are a few potential confounds in this study. First, it should be noted that the lack of an annotated peak on many utterance-final CWords did not necessarily mean that the CWord was unaccented. In some cases, creakiness at the end of the utterance made it impossible to extract the f0 information, even if it was perceptually a H target or rise. Second, this study compared monolingual English speakers to bilingual Spanish speakers. Since most participants are not perfectly balanced bilinguals, there is likely an interaction of language. However, despite the potential influence of English, the bilingual group's Spanish is still more macro-rhythmic than the English group. Future work should compare monolingual English and Spanish speaking groups.

Finally, future work should examine macro-rhythm in other speech styles. The stimuli for the current experiment were produced as read speech, which may be more or less macro-rhythmic than other styles such as spontaneous speech. One possibility is to compare similar types of natural speech corpora of English and Spanish to see if macro-rhythm changes when the data are not controlled for syllable stress placement and number of CWords per sentence.

5. REFERENCES

[1] Jun, S-A. 2014. Prosodic typology: by prominence type, word prosody, and macro-rhythm. In: Jun, S-A. (ed), *Prosodic Typology II*. Oxford University Press, 520-539.

[2] Jun, S-A. 2005. Prosodic Typology. In: Jun, S-A. (ed), *Prosodic Typology*. Oxford: Oxford University Press, 430-453.

[3] Beckman, M. E. 1996. The Parsing of Prosody. *Language and Cognitive Processes* 11, 17-67.

[4] Shattuck-Hufnagel, S., Turk, A. 1996. A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research*, 25, 193-247.

[5] Ladd, D. R. 1996/2008. *Intonational Phonology*. Cambridge: Cambridge University Press.

[6] Prieto, P., Roseano, P. (eds). 2010. *Transcription of Intonation of the Spanish Language*. Munich: Lincom Europa.

[7] Dainora, A. 2001. An Empirically Based Probabilistic Model of Intonation in American English. Ph.D. dissertation, University of Chicago.

[8] Dainora, A. 2006. Modelling Intonation in English. In: Goldstein, L., Whalen, D. H., Best, C. T. (eds), *Laboratory Phonology* 8. Berlin: Mouton de Gruyter, 107-132.

[9] de-la-Mota, C., Butragueno, P. M., & Prieto, P. 2010. Mexican Spanish intonation. In: P. Prieto & P. Roseano (eds), *Transcription of Intonation of the Spanish Language*. Munich: Lincom Europa, 319-350.

[10] Aguilar, L., de la Mota, C., Prieto, P. 2009. SP_ToBI training materials
http://prosodia.upf.edu/sp_tobi/en/index.php

[11] Estebas-Vilaplana, Prieto, P. 2010. Castilian Spanish intonation. In: P. Prieto, P. Rosano (eds), *Transcription of Intonation of the Spanish Language.* Munich: Lincom Europa, 17-48.

[12] Schmerling, S.F. 1976. *Aspects of English Sentence Stress.* Austin: University of Texas Press.

[13] Ortega-Llebaria, M., Prieto, P. 2009. Perception of word stress in Castilian Spanish: the effects of sentence intonation and vowel type. In: M. Vigário, S. Frota, M.J. Freitas (eds), *Phonetics and Phonology: Interactions and Interrelations.* Amsterdam: Benjamins, 35-50.

[14] Face, T.L. 2003. Intonation in Spanish declaratives: differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics* 2: 115-131.

[15] Rao, R. 2009. De-accenting in spontaneous speech in Barcelona Spanish. *Studies in Hispanic and Lusophone Linguistics* 2(1): 31-75.

[16] Hualde, J.I., Prieto, P. 2015. Intonational variation in Spanish: European and American varieties. In S. Frota, P. Prieto (eds), *Intonation in Romance*. Oxford: Oxford University Press, 350-391.

[17] Cruttenden, A. 1993. The de-accenting and re-accenting of repeated lexical items. In: *Proceedings of the ESCA Workshop on Prosody, Lund*, 16-19.

[18] Katz, J., Selkirk, E. 2011. Contrastive focus vs. discourse-new: evidence from phonetic prominence in English. *Language* 87(4), 771-816.

[19] Boersma, P., Weenink, D. 2018. Praat [Computer program]. Version 6.0.31.

[20] StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.