Journal of Language Research ZABANPAZHUHI

Journal of Language Research, *Alzahra University*
Vol. 17, No. 57, Winter 2026, pp.71-95
Research Article

Quantifying macro-rhythm in Persian: Evidence from Tehrani and Kermani varieties of Persian

Abedeh Mirzaee[1], Golnaz Modarresi Ghavami[2]

Received: 2024/10/04 Accepted: 2024/12/01

Abstract

Previous studies on Persian rhythm and prosodic typology have focused on micro-rhythm. However, micro-rhythm alone cannot explain the perceptible prosodic differences between language varieties such as Tehrani and Kermani varieties of Persian, which are classified under a single typological category, namely pitch-accent languages. In the current study, we examined the prosodic differences between these two Persian varieties using the parameter of macro-rhythm (Jun, 2014). We have quantified macro-rhythm using data from these varieties and compared their strength. This research aims to introduce a quantitative method for measuring macro-rhythm using Persian language data, and to classify Tehrani and Kermani varieties based on the strength of macro-rhythm. To achieve this, we analyzed audio samples of speakers from Tehrani and Kermani varieties using Praat software (version 6.2.12). We calculated three quantitative indices of macro-rhythm: nPVI (Normalized Pairwise Variability Index), MacR Variation Index (MacR_Var), and MacR Frequency Index (MacR_Freq). The values of these indices indicate that the macro-rhythm strength of Kermani variety is stronger than that of Tehrani variety, and quantifying macro-rhythm is a reliable method for studying the prosodic typology of languages.

Keywords: macro-rhythm, speech rhythm, prosodic typology, Persian language, Tehrani, Kermani.

How to Cite:
Mirzaee, A; Modarresi Ghavami, G (2026), Quantifying macro-rhythm in Persian: Evidence from Tehrani and Kermani varieties of Persian, *Journal of Language Research*, 17 (57), 71-95.
https://doi.org/10.22051/jlr.2024.48436.2496 homepage: https://zabanpazhuhi.alzahra.ac.ir

1. PhD student in Linguistics, Allameh Tabataba'i University, Tehran, Iran; bdh_mirzaee@yahoo.com
2. Associate Professor of Linguistics, Allameh Tabataba'i University, Tehran, Iran (corresponding author); modarresighavami@atu.ac.ir

1. Introduction

Rhythm is the temporal organization of speech, perceived as regular auditory or visual events. Speech rhythm has been extensively studied; however, its precise acoustic and perceptual correlates remain a topic of debate.

Early work on speech rhythm, commonly referred to as isochrony (Pike, 1945; Abercrombie, 1967; Ladefoged, 1975), posited that rhythm arises from similar or equal time intervals between linguistic units such as syllables, stresses, or morae. Based on this view, languages were categorized as stress-timed, syllable-timed, or mora-timed.

Experimental evidence, however, did not support isochrony (Arvaniti, 2009; 2012). For instance, Catalan shares a syllable structure similar to that of Spanish (syllable-timed) but also exhibits characteristics typical of stress-timed languages, namely shortened vowels. Such findings led researchers to focus on the timing of specific phonetic and phonological features, including consonant clusters, vowel duration, and vowel reduction (Dauer, 1983; 1987). Consequently, various metrics have been developed to measure rhythm through consonantal and vocalic time intervals (Ramus, Nespor & Mehler, 1999; Grabe & Low, 2002; White & Mattys, 2007). These studies, which measured the duration of segmental or syllabic elements, were able to clarify some aspects of speech rhythm; however, their results were not consistent. This inconsistency arises because rhythm perception does not solely rely on duration metrics. Instead, speech rhythm perception is likely the result of several acoustic correlates operating at multiple levels of prosodic structure (Prechtel, 2020).

Other approaches have classified speech rhythm based on the temporal and acoustic features of prominence (e.g., Lee & Todd, 2004; Jun, 2005; Tilsen & Johnson, 2008; Tilsen & Arvaniti, 2013). These studies have shown that in languages like Korean, where prominence (stress) is not indicated at the lexical level, there is still periodicity at the post-lexical level (Jun, 2005). The findings indicate that prominence at the phrasal level plays a significant role in the perception of rhythm. This helps explain the contradictory results of previous studies within the framework of isochrony, which suggest that the timing unit is limited and does not encompass the full complexity of speech rhythm (Tilsen & Johnson, 2008).

The acoustic correlates of prominence include duration, intensity, and fundamental frequency. Studies have shown that fundamental frequency is a stronger cue for prominence compared to syllable duration and energy; furthermore, the regularity of pitch movement plays a role in the perception of rhythm at the phrasal level (Barry, 1981; Thomassen, 1982; Lerdahl & Jackendoff, 1983; Handel, 1993; Dilley & Shattuck-Hufnagel, 1999; Dilley & McAuley, 2006, 2008; Andreeva, Barry & Steiner, 2007; Kohler, 2008; Barry, Andreeva & Koreman, 2009; Niebuhr, 2009; Cumming, 2011). The rhythm resulting from pitch movements is referred to as tonal rhythm. Since tonal rhythm plays a crucial role in word segmentation and in indicating prominence (Welby, 2007; Dilley & McAuley, 2008; Niebuhr, 2009; Kim & Cho, 2009; Warner, Otake & Arai, 2010; Jun, 2014), it is expected that this characteristic is a specific correlate of rhythmicity.

Rhythmicity encompasses both the frequency domain (the number of alterations in fundamental frequency (F0) within a phrase), and the temporal domain (i.e., the intervals between pitch targets). This approach integrates previous methodologies and applies them at the phrasal level (Prechtel, 2020).

The strength of tonal rhythm is language-specific and is determined by the prosodic structure of the language. In the prosodic hierarchy, F0 marks the boundaries of linguistic units at various lexical and post-lexical levels. The size and structure of these units vary across languages (Jun, 2014). These prosodic differences lead to some languages being perceived as more rhythmic than others. Jun (2005; 2014) proposes a model for prosodic typology to illustrate the cross-linguistic differences in prosodic structure as well as differences in tonal rhythm. This model compares languages analyzed within the Autosegmental-Metrical (AM) framework of intonational phonology.

In most studies of prosodic typology, languages have been classified into stress-accent and non-stress types based on the prosodic features of phonological words. However, the prosodic characteristics of units above the word level are often overlooked. Jun (2014) argues that classifying languages based solely on the type of lexical stress cannot fully account for their differences and similarities. For example, Masoumi's (2021) research observed that Standard Persian and Kermani dialect of Persian, despite their clear prosodic differences, are both categorized as stress-accent languages based on their phonetic features. This classification however, does not adequately explain the prosodic distinctions between these two varieties. Therefore, in Jun's (2014) approach, unlike previous classifications that focused only on word prosody, three parameters have been proposed for classifying the prosody of languages: word prosody, prominence type, and macro-rhythm. This prosodic typology framework is based on the interaction of two types of rhythm: micro-rhythm and macro-rhythm.

Micro-rhythm arises from the regular sequencing of syllables, morae, or strong and weak syllable patterns, evoking concepts such as syllable-timed, stress-timed, and mora-timed rhythms from previous studies. Macro-rhythm, on the other hand, refers to the organization of F0 alternations at the post-lexical level. Macro-rhythm, defined as the "phrase-medial tonal rhythm", is perceived through F0 movements associated with post-lexical prosody, with its unit size being equal to or slightly larger than a prosodic word, regardless of whether the tonal pattern consists of edge tones (such as Accentual Phrase (AP) or word tones), head tones (such as pitch accents or lexical tones), or both. This parameter can effectively illustrate the observed differences in languages that are similar in terms of stress and type of prominence (Jun, 2014).

The degree of macro-rhythm is language-specific and is determined based on three rules:

1. The macro-rhythm of a pitch contour composed of level tones (H* or L*) is weaker than a contour formed by sequences of falling (H*+L) or rising (L+H*) tones (Jun, 2014:525).
2. A pitch contour that consists of similar sub-tonal units has a stronger macro-rhythm compared to a pitch contour with less similar sub-tonal units (Jun, 2014:525).
3. A pitch contour that it’s intervals between sub-tonal units are more regular exhibits a stronger macro-rhythm (Jun, 2014:525).

Figure 1.

(a) where L and H are alternating, is more macro-rhythmic than (b) (Jun, 2014:525).

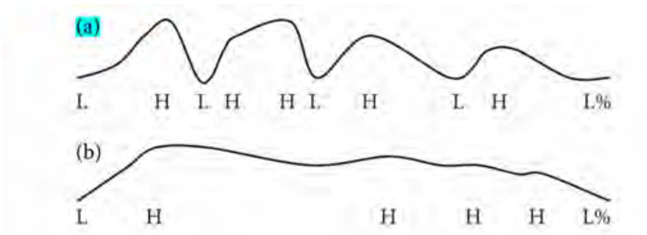


Figure 2.

(a) is more macro-rhythmic than (b) because the shape of each rise-fall in (a) is more similar than that in (b) (Jun, 2014:525).

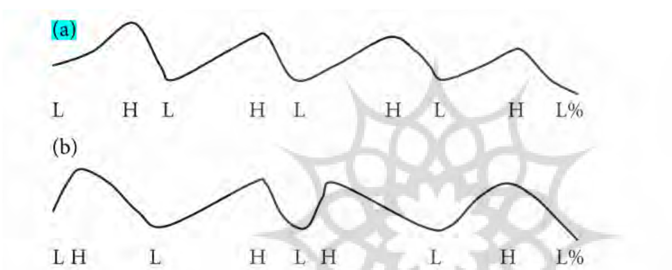
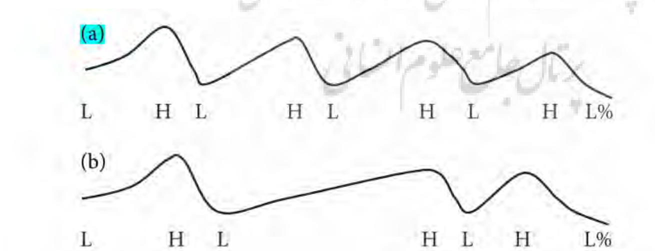


Figure 3.

(a) is more macro-rhythmic than (b) because the interval of each rise-fall is more regular in (a) than in (b) (Jun, 2014:525).



These three rules align with specific phonological criteria: the predominant type of phrase-medial tone in a language's tonal inventory (Figure 1), the number of phrase level tones within the tonal inventory (Figure 2), and the frequency of F0 rises within each word of a phrase (Figure 3).

Using these three criteria, languages can be classified into three groups of macro-rhythm: strong (such as Italian, Spanish, and Korean), medium (such as English and German), and weak (such as European Portuguese). It is important to note that macro-rhythm degree exists on a continuum and can be quantified by measuring various aspects of the F0 contour (Jun, 2014).

Quantifying macro-rhythm serves as a reliable approach for studying speech rhythm, as it reveals patterns of F0 movement according to the prosodic phonology of each language (Prechtel, 2020). John (2014: 535) classifies 37 languages based on the information derived from their prosodic phonology, but notes that Persian is absent from this typology. The framework she presents is preliminary, indicating the need for further research into the prosodic phonology of various languages and the quantification of their macro-rhythm.

The current study aims to introduce a method for quantifying macro-rhythm using data from Persian (specifically the Tehrani and Kerman dialects) and to propose a hypothesis regarding the typology of these two varieties based on macro-rhythm (weak, medium, strong). This contribution seeks to enhance the growing literature that confirms the cross-linguistic differences predicted by Jun's prosodic typology model (2014) in terms of frequency, variability, and regularity of F0 changes. This research provides insights into how macro-rhythm operates within Persian dialects and contribute to the understanding of rhythmic characteristics across languages. In the context of the Tehrani and Kermani dialects, tonal rhythm explains how variations in F0 patterns influence listener comprehension and speech production. The hypothesis suggests that Kermani may exhibit a stronger tonal rhythm due to its distinct prosodic features, which could lead to clearer communication and better processing compared to Tehrani. The study's hypothesis aligns with Jun's (2014) framework by proposing that Kermani's macro-rhythm may be stronger than Tehrani's, due to differences in prosodic structure. By quantifying these differences through macro-rhythm indices, the research aims to provide a clearer understanding of how these dialects fit within Jun's typological model.

Few studies have investigated the macro-rhythm of different languages. Burdin et al. (2014) demonstrated that differences in the phonetic realization of prominence are due to variations in macro-rhythm. They analyzed the prosodic features of focus in four languages: American English, Guarani, Moroccan Arabic, and K’iche. Their findings suggested that languages with stronger macro-rhythm employ strategies for focus-marking that maintain the regularity of F0 alternations, such as increasing duration. In contrast, languages with weaker macro-rhythm (e.g., American English and Moroccan Arabic) utilize strategies that disrupt this regularity, such as changing phrasing and deaccenting, along with varying pitch accents on focused and non-focused phrases. These findings provided an explanation for the differences in the prosodic characteristics of focus. Burdin et al. (2014) concluded that the acoustic differences in focus reflect cross-linguistic differences in macro-rhythm. However, they did not quantify the strength of macro-rhythm for each language.

Polyanskaya et al. (2019) quantified the macro-rhythm of English and Italian, aiming to provide quantitative data on how to measure macro-rhythm and acoustic evidence for it as a parameter in prosodic typology. Both Italian and English fall into the category of head-prominent languages with lexical stress; however, it was found that the macro-rhythm in Italian is stronger than in English. Specifically, the F0 alternations in Italian are more regular compared to those in English. The researchers noted that studies on the role of rhythm in speech processing and language acquisition should not be limited to syllabic rhythm but should also consider the impact of cross-linguistic differences in tonal rhythm.

Prechtel (2020) quantified the macro-rhythm of English and Spanish in two varieties of these languages (read speech and newscaster speech). Both languages are classified as Head-prominent languages with lexical stress. It was found that Spanish has a stronger macro-rhythm compared to English. The study utilized three metrics: normalized pairwise variability index (nPVI), macro-rhythm variation index (MacR_Var), and the frequency of F0 alternations within a phrase (MacR_Freq) to quantify different aspects of macro-rhythm strength.

Several studies have been conducted on the rhythm of the Persian language, focusing primarily on micro-rhythm regarding the theoretical framework proposed by Jun (2014). Buban (2007) compared quantitative rhythm measures in Persian with those in English and French, specifically examining vowel and inter-vowel intervals. She concluded that Persian is closer to syllable-timed languages rather than stress-timed ones, aligning it more with French and distinguishing it from English.

Asiaee and Nourbakhsh (2020) investigated the rhythmic structure of standard Persian based on durational parameters associated with c-interval, v-interval, and syllable to assess the potential for identifying the Persian speakers' use of Azeri as a form of voice disguise. They argued that Persian cannot be classified strictly as either syllable-timed or stress-timed but should be placed within the rhythmic category of rhythmically mixed languages.

Abolhasani Zadeh et al. (2013) and Abolhasani Zadeh, and Taghva (2019) introduced Persian as a syllable-timed language by examining the variability of vocalic and consonantal intervals. Taghva et al. (2020) further analyzed durational correlates of rhythm in Persian, positioning Persian along the continuum of syllable-timed and stress-timed languages, closer to the syllable-timed end.

Asadi and Alinezhad (2022) studied the acoustic parameters of speech rhythm based on intensity-based characteristics in Persian speakers' speech, investigating the influence of speaker factors on these parameters. They found that intensity-based parameters can effectively distinguish between Persian speakers, suggesting their potential use in speaker identification.

Masoumi and Modarresi Ghavami (2023) have determined the rhythmic features of the Persian language and, for this purpose, compared it with English, which is a stress-timed language. This study examines the impact of the phonological characteristics of Persian on the rhythm of the language. The research has shown that from a phonological perspective, Persian exhibits the features of a syllable-timed language.

The goal of the present paper is to introduce a method for quantifying macro-rhythm using Persian data and provide a hypothesis about the degree of macro-rhythm in Kermani and Tehrani dialects that despite their clear prosodic differences are categorized as stress-accent languages based on their phonetic features. It is the contention of this research that macro-rhythm indices can explicitly determine difference between the prosodic features of the two varieties of Persian. The question is: How do the macro-rhythmic characteristics of Kermani and Tehrani dialects differ? This question leads us to hypothesize that macro-rhythm indices can effectively reveal the distinctions in prosodic features between the two varieties of Persian. By quantifying these differences, we aim to enhance our understanding of how macro-rhythm influences linguistic phenomena within these dialects.

This paper is structured as follows: In section 2, we delve into the methodology. Section 3 focuses on the results, and analysis of the data using the measures introduced in section 2. Then in section 4, we discuss the findings and explore their theoretical implications. Finally, section 5 wraps up with a summary and conclusion of our key findings.

2. Methodology

We followed Prechtel (2020) for data collection, labeling, and the calculation of macro-rhythm indices.

2.1. Participants

Ten Persian speakers who were born and raised in Tehran (5 women and 5 men aged between 19 and 35 years, with an average age of 25.6 years) and ten Persian speakers who were born and raised in Kerman (5 women and 5 men aged between 18 and 42 years, with an average age of 27.4 years) volunteered to participate in this study. These individuals had not acquired any language other than Persian during their childhood.

2.2. Material

Twenty-two declarative sentences, each containing five lexical words, served as the experimental material. In designing the data, we took into account that factors such as information structure and IP length would influence the macro-rhythm. Therefore, all declarative sentences had five CWords and were produced without any focus and they were unmarked. This approach allowed us to examine the effects of various variables more accurately and to better identify the expected prosodic patterns. Since each sentence included five lexical words, it was anticipated that a maximum of five pitch accents and consequently five peaks would be observable in each utterance. Participants were prompted to read these sentences with natural intonation in a question-and-answer written dialogue. The researcher posed a question to which the target sentence was provided as the response by the participant (an example is provided below). To reduce errors, the questions and answers were repeated twice, and the more naturally produced sentence was used in the analyses. when a sentence was not produced in the neutral focus condition, it was excluded. In total, 440 sentences (22 sentences * 10 speakers * 2 varieties) were analyzed. The sentences were recorded using a Samsung headset model OG-HNF-EHS61ASFWE29, connected to a Lenovo ideapad 310 laptop in WAV format.

1. Question: leilɑ calɑfe bud mɑmɑn tʃi ɟoft Leila frustrated be.PST.3SG mom what say.PST.3SG
   "Leila was frustrated, what did mom said?"
   Answer: mɑmɑn ɟoft ɟarmɑ-je emruz leilɑ-ro divune carde
   mom say.PST.3SG hot-EZ today Leila-RA crazy make.PST.3SG
   "Mom said: Today's hot weather made Lila crazy"

2.3. Annotation

Each sentence produced as an Intonational Phrase (IP) was labeled using Praat software (version 6.2.12) (Boersma & Weenink, 2022) in four tiers: words, syllables (syll), F0 turning points (f0), and the number of peaks in the IP (freq), as shown in Figure 4. To prevent the influence of the IP boundary, sentences were labeled up to the last lexical word. In the final tier, the presence or absence of a lexical word was indicated by the numbers 1 and 0, respectively.

In the next stage, pitch contours were schematized using the procedure in Mennen, Schaeffler, and Docherty (2012). This step simplified the representation of the F0 contour and reduced the effects of fluctuations caused by consonants. The process was carried out as follows: first, the audio file was selected in Praat and a manipulation object was created. In the next step, all F0 points were removed, and the beginning and end of the pitch contour of the sentence are specified. For each Intonational Phrase (IP), the minimum and maximum F0 points were identified while disregarding fluctuations caused by micro-level variations. Finally, the previously labeled data were compared with this simplified representation and necessary adjustments were made. Examples of the schematized pitch contours are provided in Figures 4 and 5. The meanings of the labels are presented in Table 1, where each label corresponds to its order of occurrence; thus, each H label number corresponds to the preceding L or R label. For example, the first valley is labeled L1 and the first peak is labeled H1.

Figure 4.

Example of a schematized pitch contour of a Tehrani utterance, shown on top of the original contour. Labels are shown on the "f0" tier (H=peak, L= valley, R=rise). The labels are numbered in the order in which each f0 point occurs in the utterance. The "freq" tier marks the number of peaks per PWord per IP, with '1' indicating the presence of a peak.

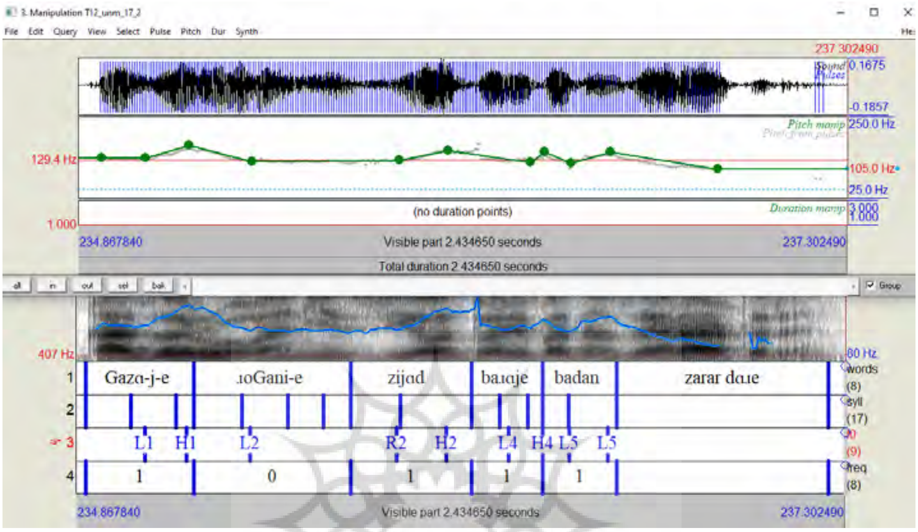


Figure 5.

Example of a schematized pitch contour of a Kermani utterance, shown on top of the original pitch track. Labels are shown on the "f0" tier (H=peak, L= valley, R=rise). The labels are numbered in the order in which each f0 point occurs in the utterance. The "freq" tier marks the number of peaks per PWord per IP, with '1' indicating the presence of a peak.

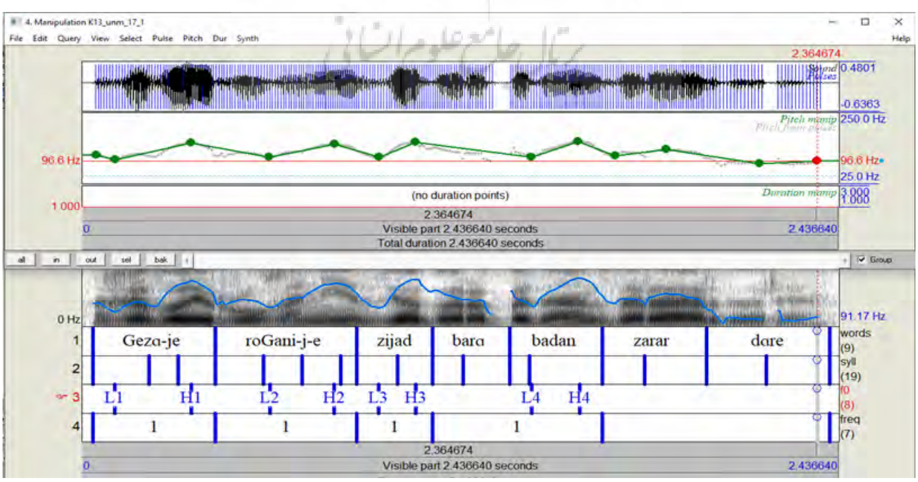


Table 1.

F0 labeling conventions

Label	Description
L	indicates where the F0 value reaches its lowest point before the next F0 rise. This is often referred to as a valley in the pitch contour.
H	indicates the point that marks the highest F0 value in a rise. This is often referred to as a peak
R	indicates the beginning of an F0 rise after a low plateau.

2.4. MacR Measures

Jun (2014) has proposed two measurements for quantifying macro-rhythm. The first one is the calculation of the Macro-Rhythm Variability Index (MacR_Var). Since macro-rhythm is defined as the phrase-medial tonal rhythm, where the unit is equal to or slightly larger than a word, its strength can be quantified by measuring the variability of intervals and the shape of tonal units within a phrase. The intervals of tonal units can be measured through the distances between peaks and valleys of F0, while their shape can be assessed through the slope of F0 rise and fall. Therefore, the Macro-Rhythm Variability Index (MacR_Var) is derived from the sum of the standard deviations of peak-to-peak distance, valley-to-valley distance, rising slope, and falling slope. This index serves as a quantitative measure for Rule 3 in the first section of this article. A pitch contour with a higher variability index indicates a weaker macro-rhythm:

$$MacR\_Var = SDp+SDv+SDr+SDf$$

SDp: standard deviation of peak-to-peak distance; SDv: standard deviation of valley-to-valley distance; SDr: standard deviation of rising slope; SDf: standard deviation of falling slope.

A script[1] on the target sentence files was run and the time and height values of the F0 labels were obtained. To calculate the slope of the rise, the difference in height between the peak (H) and the preceding valley (L or R) was computed. Similarly, the slope of the fall was obtained by calculating the difference in height between the valley and the preceding peak. The distance from peak to peak was calculated by measuring the time difference between two consecutive peaks in milliseconds (ms). The distance from valley to valley was also determined based on

1. https://www.ub.edu/phoneticslaboratory/praatscripts/extracts_f0_from_points.praat

the time difference between adjacent peaks.

The second method for measuring the degree of macro-rhythm is the frequency of L/H alternations within a phrase (MacR_Freq). Since it is expected that the tonal unit is equal to or slightly larger than a word, the MacR_Freq index is calculated as the average number of H/L alternations per word within a phrase, according to the following formula:

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

If this number is close to 1, the macro-rhythm is strong. This index provides a quantitative measure for Rule 2 in the first section of this article. In this definition, the term "word" corresponds to a lexical word rather than a morpho-syntactic word, and the phrase corresponds to an Intonational Phrase (IP). A lexical word includes the content word along with function words and unstressed clitics surrounding it.

The third index used for quantifying the number of L and H alternations in a language is known as the distribution of F0 targets Index. This index is derived using the nPVI (Normalized Pairwise Variability Index) formula (Grabe & Low, 2002), which calculates the number of F0 changes within an utterance. This formula has been previously employed in studies to quantify speech rhythm in terms of consonant and vowel durations. Polyanskaya et al. (2019) utilized this index to compute the variability of distances between peaks and valleys of F0, calculating its values for valley-to-valley distances (nPVI-H), peak-to-peak distances (nPVI-L), and distances from valleys to peaks (nPVI-LH).

$$n\,PVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right]$$

In this context, m represents the number of adjacent tonal intervals in an utterance, and d is the score of the k^{th} measurement. These values and scores are explained in section 3.3.

3. Results

In this section, the results of calculating the quantitative indices introduced in the previous section are described using a sample of data from each of the Tehrani and Kermani varieties.

3.1. MacR_Var

The Macro-Rhythm Variability Index (MacR_Var) is calculated as the sum of the standard deviations of the rising slope, falling slope, peak-to-peak distance, and valley-to-valley distance. The obtained values for a sentence per language variety are presented in Table 3. Using SPSS, we computed the standard deviation for each of these values.

Table 2.

The values of the standard deviations of the rising slope (SDr), falling slope (SDf), peak-to-peak distance (SDp), and valley to-valley distance (SDv) for Tehrani and Kermani sentences.

Kermani

	N	Mean	Std. Deviation
rising slope	4	46.50	4.655
falling slope	3	43.33	2.082
peak-to-peak distance	3	.4162	.13660
valley to-valley distance	3	.4468	.08320

Tehrani

	N	Mean	Std. Deviation
rising slope	3	26.00	5.292
falling slope	2	32.50	3.536
peak-to-peak distance	2	.6661	.21990
valley to-valley distance	2	.5745	.34351

The MacR_Var for both varieties is calculated as follows:

	MacR_Var	Kermani	Tehrani
MacR_Var =	SDr + SDf + SDp + SDv	6.9568	9.39141

To determine if Tehrani had greater variability than Kermani, a linear mixed effects model was run with MacR_Var index values as the dependent variable, language variety as the predictor, and speaker as the random intercept (table 4). The results indicated the variability in macro-rhythm between the two dialects, with the Tehrani showing a higher MacR_Var value compared to the Kermani ($\beta = 0.99$, SE =0.067, F= 654.22, $p < 0.01$).

3.2. MacR_Freq

To calculate this index, the number of peaks in each intonational phrase is divided by the number of lexical words in that phrase. As shown in Figure 4, the number of lexical words in the Tehrani utterance is 4, and the number of peaks is 3; in contrast, these numbers for the Kermani utterance are 4 and 4, respectively. Therefore, MacR_Freq is calculated as follows:

MacR_Freq	Tehrani	Kermani
$\frac{\text{Number of Peaks}}{\text{Number of Lexical Words}}$	$\frac{3}{4} = 0.75$	$\frac{4}{4} = 1$

The index for the Tehrani variety is less than 1, while for the Kermani variety, it is equal to 1.

A linear mixed effects model was run with language variety as the predictor and speaker as a random intercept (table 4). The results showed that language group was a significant predictor (β= 0.94, SE= 0.06, F = 7.29, $p < 0.01$), indicating that Kermani speakers had higher MacR_Freq values than Tehrani speakers.

3.3. nPVI

The variability in the distribution of f0 targets in an utterance was obtained using the nPVI formula. This value was calculated for the distances between valleys (nPVI-L), distances between peaks (nPVI-H), and distances from valleys to peaks (nPVI-LH) in the sample sentences.

For example, to compute nPVI-H, each peak-to-peak distance is assigned a number (from 1 to m), represented by k, where each d_k indicates the time of peak k. The difference between peak distance d_k and the subsequent peak distance d_{k+1} is calculated and divided by the mean of these two distances. This process yields $m-1$ values, which are averaged and multiplied by 100 to account for the decimal places introduced by normalization. In Kermani utterance, the times for each of the peaks (d_k) were obtained for k=1 to k=4, corresponding to H1, H2, H3, and H4 in seconds:

- d_1=0.35,
- d_2=0.81,

- d_3=1.07,
- d_4=1.6.

The normalized differences for each of the pairs were calculated as follows: (d_1 and d_2), (d_2 and d_3), and (d_3 and d_4). For each of these three consecutive pairs, the differences were computed, and then the average of these three values was calculated and multiplied by 100.

	Kermani	Tehrani
nPVI-L	76	89.5
nPVI-H	47	73.5
nPVI-LH	42	41.4

The values of nPV for valleys and peaks are higher in Tehrani than Kermani, indicating Kermani has more regular L/H intervals than Tehrani. Linear mixed effects models were run on each measure with variety as the predictor and speaker as random intercept, and the results showed that nPVI-H and nPV-l were significant (Table 4).

Table 3.

Linear mixed effects model results for MacR_Var, MacR_Freq, nPVI-L, nPVI-H, and nPVI-LH. Cells with a significant p-value are highlighted.

Fixed Effect	β	*SE*	F	*p*
MacR_Var	0.999340	0.067529	654.226	0.000
MacR_Freq	0.943030	0.063724	7.290328	0.007
nPVI-L	0.784664	0.055537	20359.738999	0.000
nPVI-H	8.214589	0.555090	9334.702	0.000
nPVI-LH	8.565186	0.579443	4.446176	0.03

4. Discussion

As outlined in Section 2, macro-rhythm can be quantified using three indices, each corresponding to specific rules that determine the strength of macro-rhythm. In Section 3, we introduced the method for calculating the values of these indices using samples from both the Tehrani and Kermani varieties of Persian. Based on these results, the variability index of macro-rhythm (MacR_Var) in Tehrani data is greater than that in Kermani's. In other words, according to Rule 3, the pitch contour of Kermani exhibits more regular intervals between sub-tonal units, indicating a stronger macro-rhythm.

The data also revealed that the frequency of F0 rise per word in a phrase for the Kermani data is equal to 1, while this number is less than 1 for the Tehrani data. According to Rule 2, these figures indicate that the pitch contour of the Kermani data is composed of sequences of similar sub-tonal units. Therefore, Kermani is predicted to accent CWords with greater regularity and thus have stronger macro-rhythm than Tehrani.

The third index used to quantify the number of L and H alternations, known as the distribution of F0 targets index (nPVI), showed that both nPVI-L and nPVI-H values were higher for the Tehrani data compared to the Kermani data. In other words, Kermani has less inter-peak and inter-valley variability than Tehrani. It can be concluded that the pitch contour of the Tehrani variety is generally smoother than that of the Kermani variety, indicating a weaker macro-rhythm. Comparing our nPV results with Polyanskaya et al. (2019) and Prechtel (2020), It seems that which of these three indices is significant depends on which languages we compare, and it cannot be said, for example, that in order for a language to have a weaker macro-rhythm, nPV-L, or nPVI-H, or nPVI-LH must necessarily be higher in that language. In Polyanskaya et al.'s (2019) study found nPVI-L and nPVI-LH were significant while nPVI-H was not significant while in Prechtel (2020) the nPV-H was significant.

Each of these three indices has successfully quantified an aspect of macro-rhythm strength in the two varieties of Persian. By aligning the findings of this study with the criteria established by Jun (2014), we can suggest that the strength of macro-rhythm in the Kermani variety is greater than that in the Tehrani variety. Jun's (2014) Prosodic typology is based on three parameters: prominence type, wordprosody, and macro-rhythm, and the typology of the two dialects of Persian in this framework is shown in table 5. The application of the method introduced in the present paper to other varieties of the Persian language could further enrich this typology.

Table 4.

Prosodic typology of 2 Persian varieties based on three parameters (Prominence Type, Word Prosody, and Macro-rhythm). The dashes indicate that macro-rhythm is a continuous feature.

Prominence Type	Word Prosody	Macro-rhythm		
		Strong	Medium	Weak
Head-edge*	stress	Kermani	Tehrani	
	Tone/ lexical pitch accent			
	Both			
	none			

* Sadat-Tehrani (2008).

The number of speakers for each dialect and the number of sentences analyzed in this study were limited. Therefore, we can only propose a hypothesis regarding the macro-rhythm of these two varieties of Persian. Future studies should involve a larger number of speakers, different types of sentences, and more extensive data from both language varieties to test the hypothesis presented in this study more thoroughly. Limitations in sample size and data collection are common challenges in linguistic research, particularly in studies focusing on prosody and intonation. The complexity of capturing natural speech patterns often necessitates a careful balance between the breadth of data collected and the depth of analysis. As noted by researchers such as Ladd (2008), a comprehensive understanding of intonational patterns requires sufficient data reflecting the variability inherent in natural speech.

By proposing a hypothesis based on the limited data available, this study opens avenues for further exploration into the macro-rhythmic characteristics of Persian dialects. The findings suggest that while preliminary conclusions can be drawn, they must be validated through larger-scale studies that encompass a wider range of speakers and contexts. This is particularly important given that prosodic features can vary significantly based on social factors, regional differences, and individual speaker characteristics (Gussenhoven, 2002).

It has been shown that regularity in pitch patterns can enhance listener processing and reduce cognitive load during speech (Gussenhoven, 2002); while languages or dialects with higher variability in pitch patterns may require more effort from listeners to decode meaning (Dilley et al., 2010). The implications of these findings extend beyond theoretical considerations; they also have practical applications in language learning and teaching. For instance, learners exposed to a dialect with a more consistent macro-rhythm, such as Kermani, might find it easier to grasp and replicate intonational patterns. In contrast, learners engaging with Tehrani may need to develop specific strategies to navigate its greater variability in pitch changes.

Research indicates that macro-rhythm plays a crucial role in second language (L2) acquisition. For instance, learners who practice imitating native speakers' pitch contours can improve their pronunciation and fluency. In a study involving Japanese learners of English, it was found that while they could effectively imitate pitch variations (macroR_Var), they struggled with synchronizing these variations with prosodic words (macroR_Freq), highlighting the challenges in mastering macro-rhythmic patterns (Ortega-Llebaria, Silva, & Nagao; 2023). This suggests that effective language learning requires not only understanding the individual sounds but also grasping the broader rhythmic structures.

It has also been suggested that macro-rhythm influences how listeners segment speech into meaningful units. The presence of strong tonal contrasts can aid in identifying boundaries between words and phrases, facilitating comprehension. For example, languages with clear macro-rhythmic patterns allow listeners to anticipate upcoming words based on tonal cues, leading to more efficient processing (Polyanskaya et al., 2019). Future research should explore how these rhythmic differences impact not only speaker production but also listener perception and processing. For instance, experiments could be designed to assess how varying levels of pitch change frequency influence comprehension in real-time conversations.

5. Conclusion

In this article, we introduced a new framework for studying the prosodic typology of the Persian language. The aim of this research was to present a method for quantifying macro-rhythm using data from Persian and to propose a hypothesis regarding the prosodic characteristics of the two varieties: Tehrani and Kermani, based on their macro-rhythm strength. This method allows for the comparison of different varieties of Persian with one another, and with other languages. For instance, in Masoumi's (2021) study, it was observed that despite the clear intonational differences between Standard Persian and Kermani, both varieties fall within the category of stress-timed languages in terms of phonetic features. However, this classification does not adequately account for the noticeable prosodic differences between these two varieties, which are quite apparent to Persian speakers. The present study demonstrated that by employing the macro-rhythm parameter and comparing it across both Standard Persian and Kermani, we can elucidate these clear intonational differences.

Phonological theory guides empirical research, while acoustic data inform phonological theory, either confirming or refining it. Using the method introduced in this study, we tested macro-rhythm, a typological parameter based on the phonological analysis of over 37 languages (Jun, 2014), using acoustic data. The application of this method for linguistic comparisons in future studies represents an important step in validating the typology of tonal rhythm and enhancing our understanding of cross-linguistic differences in language structure. It also sheds light on how processing strategies (such as segmentation) are adjusted based on different macro-rhythms across languages.

In summary, this study contributes to the growing body of research on prosodic typology by providing a quantitative framework for analyzing macro-rhythm in Persian dialects. The findings not only enhance our understanding of the Tehrani and Kermani varieties but also offer insights into broader linguistic principles governing rhythm and intonation. Future research should continue to explore these dimensions across various languages and dialects to deepen our understanding of linguistic variation, and how rhythm functions in human communication.

Author Contributions

All authors contributed significantly to the conception and design of the study. A.M. under the supervision of G.M.G., led the data collection, analysis, developed the methodology, and performed the statistical analyses. Both authors reviewed, edited, and approved the final version of the manuscript.

Data Availability Statement

The datasets analyzed during the current study are available from the first author upon reasonable request. Data sharing complies with ethical and privacy considerations related to the study participants.

Ethical Considerations

The study was approved by the Ethics Committee of the Allame Tabataba'i University[1] (Ethical code: IR.ATU.REC.1401.020). Informed consent was obtained from all individual participants included in the study. The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

1. https://ethics.research.ac.ir/EthicsProposalView.php?&code=IR.ATU.REC.1401.020

References

Abercrombie, D. (1967). *Elements of general phonetics*. Aldine Publishing Corporation. https://doi.org/10.1515/9781474463775

Abolhasani Zadeh, V., Taghva, N., Sharifi Moghadam, A., & Golshan, A. (2013). The effect of Farsi language rhythm on Iranian traditional non metric instrumental music. *Aion-linguistica*, 2, 11-19. https://www.researchgate.net/publication/336872021_Effect_of_Farsi_language_on_Iranian_traditional_music

Abolhasani Zadeh, V., & Taghva, N. (2019). Study and comparison of Persian language and English language rhythm. *Bialingual Interpretive Comparative Journal of Language and Literature*, *4*(6), 1-10. [In Persian] https://www.researchgate.net/publication/336871581_brrsy_w_mqaysh_rytm_zban_farsy_w_zban_anglysy?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InByb2ZpbGUifX0

Andreeva, B., Barry, W. J., & Steiner, I. (2007). Producing phrasal prominence in German. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1209-1212. https://scidok.sulb.uni-saarland.de/handle/20.500.11880/23564?locale=en

Arvaniti, A. (2009). Rhythm, timing, and the timing of rhythm. *Phonetica*, *66*(1-2):46-63. https://www.degruyter.com/document/doi/10.1159/000208930/html

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3):351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Asadi, H., & Alinezhad, B. (2022). Speech rhythm measures: Acoustic cues for speaker identification. *Journal of Language Researches*, *12*(2),24-49. [In Persian]. https://doi.org/10.22059/jolr.2021.304539.666624

Asiaee, M., & Nourbakhsh, M. (2020). Durational parameters of speech rhythm, a measure for detection of the Persian speakers' use of disguise? *Journal of Language Research*, *11*(1):1-23. [In Persian]. https://doi.org/10.22059/jolr.2020.298600.666587

Barry, W. J. (1981). Prosodic functions revisited again. *Phonetica* 38, 320-340. https://doi.org/10.1159/000260036

Barry, W. J., Andreeva, B., & Koreman, J. (2009). Do rhythm measures reflect perceived rhythm? *Phonetica*, *66*(1-2), 78-94. https://www.degruyter.com/document/doi/10.1159/000208932/html

Buban, N. (2007). Quantitative indices of Persian rhythm in comparison with those of English and French. *Selected Papers of the 7th Iranian Linguistics Conference* (387-403). Tehran: Allameh Tabataba'i University. [In Persian].

Burdin, R., Phillips-Bourass, S., Turnbull, R., Yasavul, M., Clopper, C., & Tonhauser, J. (2014). Variation in the prosody of focus in head- and head/edge-prominence languages, *Lingua*, *165*, 254-276. https://doi.org/10.1016/j.lingua.2014.10.001

Cumming, R. (2011). The language-specific interdependence of tonal and durational cues in perceived rhythmicality. *Phonetica* 68, 1–25. https://doi.org/10.1159/000327223

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*(1), 51-62. https://doi.org/10.1016/S0095-4470(19)30776-4

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the 11th international congress of phonetic sciences, 5*, 447-450. https://www.semanticscholar.org/paper/Phonetic-and-Phonological-Components-of-Language-Dauer/01b4c9c031d212683782b4de2eaee2fa98915ebb

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for delta C. In P. Karnowski & I. Szigeti (Eds.), *Language and Language Processing: Proceedings of the 38th Linguistics Colloquium, Piliscsaba 2003* (pp. 231-241). Peter Lang Publishing Group. https://doi.org/10.5167/uzh-111789

Dilley, L. & Shattuck-Hufnagel, S. (1999). Effects of repeated intonation patterns on perceived word-level organization. *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, 1487-1490. https://www.internationalphoneticassociation.org/icphsproceedings/ICPhS1999/papers/p14_1487.pdf

Dilley, L. & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59, 294-311. https://doi.org/10.1016/j.jml.2008.06.006

Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language, 63*(3), 274–294. https://doi.org/10.1016/j.jml.2010.06.003

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 515-546). Cambridge University Press. https://doi.org/10.1515/9783110197105.2.515

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In *Speech Prosody 2002, International Conference.* https://doi.org/10.21437/SpeechProsody.2002-7

Handel, S. (1993). The effect of tempo and tone duration on rhythm discrimination. *Perception and Psychophysics, 54*(3), 370-382. https://doi.org/10.3758/BF03205273

Jun, S-A. (2005). Prosodic Typology. In S-A. Jun (Ed.), *Prosodic typology* (pp. 430-453). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

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. https://doi.org/10.1093/ACPROF%3AOSO%2F9780199567300.003.0017

Kohler, K. (2008). The perception of prominence patterns. *Phonetica*, 65, 257-269. https://doi.org/10.1159/000192795

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Ladefoged, P. (1975). *A Course in phonetics*. Harcourt Brace Jovanovich. https://doi.org/10.1177/0033688293024002

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, 225–254. https://doi.org/10.1016/j.cognition.2003.10.012

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press. https://doi.org/10.7551/mitpress/12513.001.0001

Masoumi, A. (2002). *Analyzing two Persian varieties from a prosodic typology point of view* [Doctoral dissertation]. Allameh Tabataba'i University. [In Persian]. https://ganj.irandoc.ac.ir/#/articles/5a867e6208bde83ab36e4d54ddb2272c

Masoumi, A., & Modarresi Ghavami, G. (2023). The rhythmic type of Persian: A phonological perspective. *Journal of Researches in Linguistics*, *15*(1), 85-100. [In Persian] https://jrl.ui.ac.ir/article_27689.html

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: a comparison of English and German. *J. Acoust. Soc. Am.* 131(3): 2249–2260. https://doi.org/10.1121/1.3681950

Niebuhr, O. (2009). F0-based rhythm effects on the perception of local syllable prominence. *Phonetica*, 66, 95–112. https://doi.org/10.1159/000208933

Ortega-Llebaria, M., Silva Jr, L., & Nagao, J. (2023). Macro and micro-rhythm in L2 English: Exploration and refinement of measures. In *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 1582-1586). https://www.internationalphoneticassociation.org/icphsproceedings/ICPhS2023/full_papers/830.pdf

Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Press. https://archive.org/details/intonationofamer0000pike

Polyanskaya, L., Busà, M.G., & Ordin, Mikhail. (2019). Capturing cross-linguistic differences in macro-rhythm: the case of Italian and English. *Language and Speech*, *63*(2), 242-263. https://doi.org/10.1177/0023830919835849

Prechtel, C. (2020). Quantifying macro-rhythm in English and Spanish: A comparison of tonal rhythm strength [Master's thesis]. University of California. https://escholarship.org/uc/item/3n58r8zh

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition 73*(3), 265-292. https://doi.org/10.1016/S0010-0277(00)00101-3

Sadat-Tehrani, N. (2007). *The Intonational grammar of Persian* [Doctoral dissertation]. University of Manitoba. https://linguistics.ucla.edu/people/jun/Workshop2007ICPhS/Papers/NimaTehrani-PaperIntonWokshop.pdf

Taghva, N., Moloodi, A., & Abolhasanizadeh, V. (2021). Acoustic correlations of speech rhythms in Persian based on variability of between-speakers characteristics. *Journal of Researches in Linguistics*, *12*(2), 27-50. [In Persian]. https://doi.org/10.22108/jrl.2021.126261.1535

Tilsen, S. & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *Journal* *of the Acoustical Society of America*, *124*(2), EL34-39. https://doi.org/10.1121/1.2947626

Tilsen, S. & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America*, *134*(1): 628-639. https://doi.org/10.1121/1.4807565

Thomassen, J. M. (1982). Melodic accent: Experiments and a tentative model. *Journal of the Acoustical Society of America*, *71*(6), 1596-1603. https://doi.org/10.1121/1.387814

White, L. & Mattys, S.L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35, 501-522. https://doi.org/10.1016/j.wocn.2007.02.003