Journal of Language Research
ZABANPAZHUHI

Journal of Language Research, *Alzahra University*
Vol. 16, No. 53, Winter 2025, pp.9-34
Research Article

Cross-linguistic rhythmic patterns in Persian-English bilingual speakers: Implications for speaker recognition*

Homa Asadi[1], Maral Asiaee[2]

Received: 2024/10/28 Accepted: 2024/12/02

Abstract

This study investigates rhythmic patterns in Persian-English bilingual speech, focusing on duration-based measures in a sample of late bilingual adult males. Using various rhythmic measures, including consonantal and vocalic duration, we explored cross-linguistic differences, individual consistency, and speaker identification potential. The results revealed significant differences between Persian (L1) and English (L2), particularly in vocalic measures, whereas consonantal measures exhibited greater consistency. Cross-linguistic correlations were stronger for consonantal measures than for vocalic measures, suggesting higher individual consistency in consonant timing. Speaker identification, conducted through linear discriminant analysis, achieved the highest accuracy with consonantal measures, with stronger performance in L1 than in L2. These findings indicate that while bilingual speakers adjust their rhythmic patterns to suit L2 demands, they retain individual characteristics, especially in consonant timing. This research has implications for understanding bilingual speech production and enhancing speaker recognition technology.

Keywords: Bilingualism, speech rhythm, speaker individuality, Persian, English

How to Cite:

Asadi, H; Asiaee, M (2025), Cross-linguistic rhythmic patterns in Persian-English bilingual speakers: Implications for speaker recognition, *Journal of Language Research*, 16 (53), 9-34.

https://doi.org/10.22051/jlr.2024.48741.2511 homepage: https://zabanpazhuhi.alzahra.ac.ir

* This research is part of project number 99029580, which has been financially supported by the Iran National Science Foundation (INSF).

1. Assistant Professor of Linguistics, University of Isfahan, Isfahan, Iran; (Corresponding author) h.asadi@fgn.ui.ac.ir

2. Postdoctoral Researcher, Adam Mickiewicz University, Poznań, Poland; marasi@amu.edu.pl

1. Introduction

The human voice is a highly complex acoustic signal, characterized by a unique combination of features that allow listeners to identify individual speakers with remarkable accuracy. This ability to recognize voices plays a fundamental role in social interactions, communication, and even forensic applications. While the exact mechanisms underlying voice recognition are still being explored, it is evident that numerous acoustic parameters contribute to shaping a speaker's distinctive vocal identity.

Previous studies have revealed substantial variations in the acoustic properties of speech, including both spectral and temporal parameters, among different speakers (Nolan, 1989; Rose, 2002; Jessen et al., 2005; Jessen & Becker, 2010; Gold et al., 2013; Dellwo et al., 2015; He & Dellwo; 2016). For instance, formant frequencies and fundamental frequency have been shown to be particularly influential in revealing speaker-specific characteristics among speakers (Goldstein, 1976; Braun, 1995; Debruyne et al., 2002; Gold et al., 2013; Jessen & Becker, 2010; Asadi et al., 2018). Additionally, spectral moments of consonants, especially fricatives, have been found to provide valuable speaker-specific information, further enhancing the distinctiveness of individual voices (Smorenburg & Heeren, 2020; Ulrich et al., 2023). Furthermore, studies have demonstrated that voice quality features, such as the balance between high-frequency harmonic and inharmonic energy, also contribute to the unique vocal signature of each speaker (Kreiman & Sidtis, 2011; Lee et al., 2019). The underlying rationale for this variability lies in the individual physiological characteristics of speakers' speech organs. However, it is crucial to acknowledge that social factors, including regional accents, dialects, and cultural influences, also play a non-trivial role in shaping the distinctive characteristics of a person's voice.

Beyond spectral features, temporal aspects of speech, particularly speech rhythm, are crucial in differentiating individual voices. Rhythmic patterns, characterized by the timing of morae, syllables, words, and phrases, vary considerably across different languages (Gibbon, 2022). These patterns convey not only the linguistic content of an utterance but also information

about the speaker's identity. Recent investigations have revealed that temporal aspects of speech, especially rhythmic patterns, contribute significantly to speaker identification. These patterns, reflecting the timing of consonantal and vocalic intervals, demonstrate substantial between-speaker variability (Dellwo et al., 2015; Asadi et al., 2018, Taghva et al., 2023). Individual differences in the timing of specific articulatory gestures have been identified as a major contributor to variability in the temporal patterns of speech production between speakers. Differences in the anatomical size of the articulatory apparatus can lead to distinct coordination strategies, resulting in variations in how speakers operate their speech organs, including the tongue, lips, and jaw, ultimately impacting the acoustic parameters of speech rhythm (Dellwo et al., 2015). Nevertheless, the connection between physiological factors and rhythmic variability is not straightforward, as acquired linguistic behaviors and language-specific prosodic patterns also contribute to the temporal characteristics of speech.

The influence of language-specific factors on speech rhythm measures is further exemplified by the phonotactic system of a given language. Languages traditionally fall into different rhythmic categories: stress-timed (e.g., English), syllable-timed (e.g., Persian), or mora-timed (e.g., Japanese) (Ladefoged, 1975, Lazard, 1999; Dellwo, 2010). These classifications reflect fundamental differences in temporal organization. For instance, syllable complexity has a significant impact on systematic variability in speech rhythm measurements (Prieto et al., 2012). It is proposed that stress-timed languages with more phonotactically complex structures exhibit higher levels of vocalic and consonantal intervals compared to languages with simpler structures (Ramus et al., 1999). Additionally, languages that allow vowel reduction often reflect this acoustically through highly variable vocalic intervals (Dellwo, 2010). Given the inherent differences in the rhythmic characteristics of speech across languages, it is plausible that bilingual speakers demonstrate variations when switching from one language to another. This is particularly evident when bilingual speakers are speaking in two typologically distinct languages.

This study aims to investigate the degree to which durational measures

of speech rhythm vary among bilingual speakers. By comparing Persian-English bilingual speakers, we explore the impact of language-specific rhythmic characteristics on individual speaker variability. Persian and English represent particularly interesting cases for comparison due to their distinct rhythmic properties. Persian exhibits a relatively simple syllable structure (CV(C)(C)) and minimal vowel reduction (Windfuhr, 1979; Sadeghi, 2015). In contrast, English features complex syllable structures, significant vowel reduction, and stress-based timing patterns (Dellwo, 2010). By analyzing how switching from Persian to English affects speech rhythm measures, we also aim to determine the extent to which this variability influences between-speaker rhythmic variability. Our findings will clarify the extent to which speech rhythm is influenced by language and individual speaker characteristics, providing insights into the complex relationship between these factors in bilingualism.

1.1. The role of speech rhythm in variability between languages and speakers

The rhythmic properties of speech have been the subject of extensive investigation in the fields of linguistics and phonetics. Numerous metrics have been devised to quantify speech rhythm from various phonetic units, particularly focusing on cross-linguistic attributes (Ramus et al., 1999; Dellwo, 2006; Grabe & Low, 2002; White & Mattys, 2007). While phoneticians have developed various measures to categorize languages rhythmically, the existence of such differences and the feasibility of language classification based on these measures remain debated (White & Mattys, 2007; Dellwo, 2010; Loukina et al., 2011). Loukina et al. (2011) demonstrated the diverse rhythmic variations across languages, highlighting that a single speech rhythm metric cannot optimally differentiate all language pairs. Consequently, multiple measures are essential for accurate identification of more than two languages. In light of this discovery, phoneticians embarked on a more comprehensive investigation into the factors contributing to variation in speech rhythm measures, resulting in the formulation of the hypothesis of speaker-specific rhythmic patterns.

Building on the understanding that speech rhythm is not solely a language-specific phenomenon, researchers investigated whether acoustic rhythm metrics can effectively distinguish between speakers of the same language. Numerous studies across diverse languages have demonstrated significant between-speaker variability in various rhythm measures, suggesting that these metrics can indeed capture individual speaker characteristics. Yoon (2010) found significant between-speaker variability in %V and VarcoV for English speakers. Wiget et al. (2010) and Leemann et al. (2014) also observed considerable individual differences in rhythmic measures for English and Zurich German speakers, respectively. Similarly, Dellwo et al. (2015) found substantial between-speaker variability in %V, ΔC(ln), ΔV(ln), and Δpeak(ln) among German and Swiss German speakers. Furthermore, Asadi et al. (2018) demonstrated the effectiveness of speech rhythm measures in distinguishing Persian speakers, identifying %V as a key factor. Taghva et al. (2023) extended this to Kalhori Kurdish, confirming the effectiveness of %V and syllable rate. Overall, duration-based metrics of speech rhythm have consistently proven their ability to capture speaker individuality across various languages. These studies mainly focused on speakers who spoke the same language.

Research on second language (L2) speech rhythm has shown that learners exhibit significant variability in their ability to acquire the temporal patterns of the target language, which is influenced by factors such as the rhythmic characteristics of their first language (L1), their level of proficiency in the L2, and the extent of their exposure to the target language. White and Mattys (2007) demonstrated that speakers may not implement the subtle adjustments required to accommodate the rhythmic differences between their L1 and a rhythmically similar L2, instead relying on their native language timing patterns, which may not be effective for rhythmically distinct languages. This finding was further supported by Li and Post (2014), who examined the development of English rhythm in L2 learners with typologically different L1s, specifically Mandarin and German. Their study revealed that while both groups followed similar developmental paths in acquiring vocalic variability and accentual lengthening, they diverged in the proportion of vocalic materials in

their L2 utterances, reflecting direct L1 transfer. The study highlights the role of proficiency in rhythm development and supports a multisystemic model of L2 rhythm acquisition, where different rhythmic properties are acquired at varying proficiency levels, influenced by both L1 transfer and universal acquisition processes. Ordin and Polyanskaya (2015) found that as French and German learners of L2 English became more proficient, their speech rhythm shifted from syllable-timed to stress-timed patterns typical of English. While German learners achieved native-like variability, French learners showed lower variability even at advanced levels, indicating persistent native language influences. Stockmal et al. (2005) provided additional insights by examining Latvian native speakers and Russian learners, demonstrating that high-proficiency learners could approximate native speech patterns, whereas low-proficiency individuals exhibited substantially more inconsistent rhythmic characteristics.

Bilingual speakers offer a unique perspective on speech rhythm variability, as they navigate between two linguistic systems with potentially distinct rhythmic properties. Lleó et al. (2011) studied German-Spanish bilingual children and found that their vocalic and consonantal interval timing patterns suggest an interaction between the rhythmic systems of both languages. Bunta and Ingram (2007) similarly demonstrated that bilingual children exhibit distinct speech rhythm patterns for their target languages, deviating from their monolingual peers. Henriksen (2016) conducted a study on adult bilinguals, further confirming these findings. He found that highly proficient adult Spanish-English bilinguals exhibited different rhythmic patterns in their L1 and L2 rhythms, suggesting distinct rhythmic production strategies in their two languages. Aldrich (2020) investigated the speech rhythm of adult early Spanish-English bilingual speakers and found language-specific rhythm production with more variability associated with English compared to Spanish. They concluded that this differentiation in terms of rhythm suggests a possibly unique abstract organization for each language at the prosodic level.

Previous studies on bilingualism have primarily focused on linguistic

differences between languages. However, the extent to which between-speaker variability is affected has been largely overlooked. Dellwo and Smith (2015) pioneered research in this area by examining speech rhythm characteristics in bilinguals. They tested the hypothesis that a speaker's temporal characteristics in one language would correlate with those in another. Using a variety of durational rhythmic features, they found significant speaker-specific effects for measures like %V, ΔV, and articulation rate. This suggests that speakers systematically vary in their suprasegmental temporal characteristics. The authors argue that these measures are likely influenced by individual differences in articulatory control.

1.2. The current study: Research questions

This research examines the speaker-specific temporal features of bilingual Persian-English speakers, with a particular emphasis on durational rhythmic variability. We aim to ascertain whether temporal features derived from consonantal, vocalic, and syllable intervals can also effectively discriminate bilingual speakers when they are engaged in speech production across languages with distinct phonological systems and rhythmic categories. This research seeks to investigate the extent to which rhythmic variability measures contribute to speaker individuality and to determine the degree to which such features are language-dependent or speaker-specific. We hypothesize that speakers may exhibit greater acoustic variability in their native language compared to their second language. The challenges inherent in second language acquisition may lead speakers to prioritize intelligibility over acoustic variation, potentially resulting in a reduction in variability within their speech patterns. The primary research questions addressed in this study are as follows:

1) How do rhythm metrics differ between Persian (L1) and English (L2) in bilingual speakers?
2) To what extent do speakers maintain consistent rhythmic patterns across their two languages?
3) Which rhythm metrics are most effective for speaker identification in

L1 and L2?

To address these research questions, an exploratory corpus analysis of a bilingual Persian-English speech corpus was conducted. To control for the influence of lexical variability on rhythm measures, participants produced an equivalent number of sentences in both languages. This ensured that rhythmic differences were not confounded by differences in lexical choices, which can affect rhythm in spontaneous speech (Wiget & et al., 2010, Dellwo et al., 2015). Extraneous factors such as age and accent were carefully controlled to minimize their influence on the analysis. The findings of this study are anticipated to be instrumental in advancing applied domains where knowledge of human individuality cues is crucial, such as automatic speaker recognition and forensic speaker identification.

2. Procedure

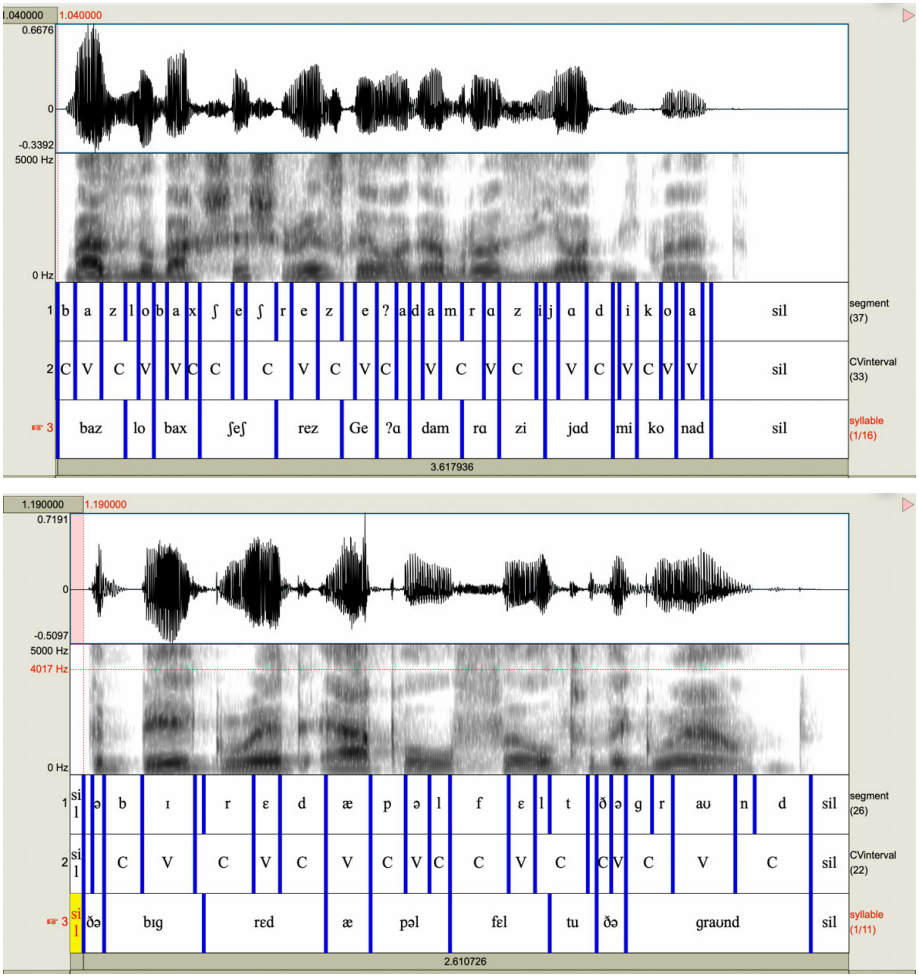
2.1. Participants and data segmentation

The study sample consisted of 20 male native Persian speakers, all of whom demonstrated proficiency in English. The participants were selected according to specific criteria, including that they were late bilinguals with Persian as their first language, exhibited minimal regional or social accent variation, and had no history of language or hearing impairment. The participants had a mean age of 27.6 years (SD = 3.1, range = 24-36). Oral English proficiency was assessed in an initial screening interview conducted by two experienced English teachers to ensure a consistent language level among participants. Audio recordings were made using a ZOOM H4n Pro handheld recorder in a controlled, noise-free environment to ensure the highest possible quality of data. The recordings were captured at a 44,100 Hz sampling rate with 16-bit quantization. The participants were instructed to read a predetermined list of sentences at a comfortable pitch and volume. The reading material comprised 50 sentences in Persian and 50 sentences in English, each selected for its phonetic richness and inclusion of a diverse range of lexical items, ensuring a broad representation of phonetic contexts. Each participant read the sentences in both languages naturally, with a three-second pause inserted

between successive utterances to allow for the clear separation of the speech samples. Subsequently, the speech data underwent comprehensive analysis using Praat software (version 5.2.34; Boersma and Weenink, 2024). The audio files were annotated at three distinct levels: segmental, consonantal-vocalic intervals (CV), and syllable. Vowels were segmented at the onset and offset of their steady-state formant patterns. The onset was marked when the formant trajectories stabilized, while the offset was identified at the point where the formants returned to baseline or transitioned into a neighboring consonant. Glides were segmented based on the formant transitions between two adjacent vowels. The onset of a glide was marked where the formant movement began, and the offset was placed where the formant stabilized into the adjacent vowel. For plosive consonants, the beginning was marked at the onset of the burst, characterized by a sudden increase in acoustic energy. The end of the plosive was defined as the point where the burst energy dissipated, and the following vowel or consonant began to emerge. In the case of voiced plosives, the vocalic transition also guided the offset placement. For other consonants, such as fricatives and affricates, segmentation relied on established acoustic landmarks, such as the onset of fricative noise or the abrupt energy changes characteristic of stops. Additionally, the authors manually annotated the syllable level to ensure an accurate reflection of the syllable structure. For quantitative analysis, a Praat script, designated as *DurationAnalyzer*, was employed to automatically calculate a range of rhythm metrics based on the duration of the annotated levels. This automated approach facilitated the extraction of relevant temporal features, which were then utilized to investigate cross-linguistic rhythm patterns in Persian and English. Figure 1 illustrates an annotated signal containing both Persian and English speech segments, as pronounced by speaker 1.

Figure 1.

Annotated speech signal illustrating Persian and English utterances produced by speaker 1



2.2. Acoustic rhythmic parameters

In this study, we examined duration-based metrics derived from consonantal and vocalic intervals, along with syllable units in speech signals. To minimize the influence of speech rate on the analysis and avoid potential artifacts, we used only rate-normalized measures. We selected eight duration-based metrics: one consonantal and vocalic proportion metric (%V), two variability measures for consonantal and vocalic durations (ΔV(ln), ΔC(ln)), two rate-normalized variability measures (n-PVI-V, n-PVI-C), two coefficients of variation measures (VarcoV and VarcoC), and a syllable rate-based measure

(articulation rate). These metrics are grounded in well-established temporal measures from previous research on speech rhythm (Ramus et al., 1999; Grabe and Low, 2002; Dellwo et al., 2015; Asadi et al., 2018; Aldrich, 2020) and were used to analyze the speech of bilingual speakers. The following duration-based measures were automatically calculated from the consonantal-vocalic (CV) interval tier, while the articulation rate was calculated from the syllable tier, expressed in terms of the number of syllables per second. Investigated measures are summarized below:

-%V: The percentage of overall speech duration accounted for by vocalic segments.

-ΔV(ln): The standard deviation of the natural logarithm of vocalic interval durations, representing variability in vocalic durations.

-ΔC(ln): The standard deviation of the natural logarithm of consonantal interval durations, indicating variability in consonantal durations.

-VarcoC: The variability in consonantal durations, determined by the standard deviation of consonant interval durations divided by the mean consonant interval duration, multiplied by 100 to express it as a percentage.

-VarcoV: The variability in vocalic durations, calculated as the standard deviation of vowel interval durations divided by the mean vowel interval duration, multiplied by 100 to express it as a percentage.

-n-PVI-V: The rate-normalized Pairwise Variability Index for vocalic intervals, calculated as the average absolute difference in duration between successive vocalic intervals, adjusted for speech rate.

-n-PVI-C: The rate-normalized Pairwise Variability Index for consonantal intervals, calculated similarly to n-PVI-V but for consecutive consonantal intervals.

-articulation rate: The number of syllables articulated per second.

2.3. Statistical analysis

All statistical analyses were carried out using the R statistical programming language, version 4.3.1 (R Core Team, 2023). To investigate the rhythmic differences between Persian-English bilinguals' production in their L1

(Persian) and L2 (English), we conducted a statistical analysis on the selected rhythm metrics: %V, ΔV(ln), ΔC(ln), VarcoC, VarcoV, n-PVI-V, and n-PVI-C. For each measure, we computed descriptive statistics, which included means, standard deviations, and coefficients of variation (CV). The CV, calculated as the ratio of standard deviation to mean, was particularly useful in assessing the relative variability of rhythm measures within each language context. To ensure a robust comparison between L1 and L2 production patterns, we employed paired-sample t-tests for each rhythm measure, as the same speakers produced both languages. Effect sizes were calculated using Cohen's d, with pooled standard deviations to account for the within-subjects design. The normality of the data for each rhythm measure was assessed using the Shapiro-Wilk test. The results indicated that all measures followed a normal distribution ($p > 0.05$), justifying the use of parametric tests for statistical analysis.

To examine the relationship between speakers' rhythmic patterns across their two languages, we conducted Pearson correlation analyses for each rhythm measure between L1 and L2 productions. These correlations helped determine whether speakers maintained consistent individual rhythmic characteristics across languages or demonstrated language-specific adaptations. Strong correlations ($r > 0.5$) were interpreted as evidence of cross-linguistic consistency in individual utterances, while weak correlations suggested more successful adaptation to language-specific rhythm patterns. The statistical significance of all correlations was assessed at $\alpha = 0.05$, with particular attention paid to the strength of correlations as indicated by the correlation coefficient (r).

To quantify the magnitude of cross-linguistic differences, we calculated mean differences and percentage changes between L1 and L2 for each rhythm measure. These calculations provided a clear indication of the direction and magnitude of change in rhythmic properties when speakers switched between languages. The percentage differences were calculated as $((L2-L1)/L1) \times 100$, offering a normalized measure of change that could be compared across different rhythm metrics. Results were rounded to three decimal places for

consistency.

Speaker identification analysis was conducted using Linear Discriminant Analysis (LDA) with a leave-one-out cross-validation procedure to evaluate the discriminative power of each rhythm metric. The analysis was performed separately for each language (L1 and L2) and for the combined dataset. Prior to LDA, the data was standardized to have a mean of zero and a standard deviation of one. For the combined analysis using all parameters, Principal Component Analysis (PCA) was applied as a dimensionality reduction technique, retaining components that explained 95% of the total variance. Classification accuracy was assessed through overall accuracy rates and confusion matrices, with statistical significance evaluated against chance-level performance using a chi-square test ($p < 0.05$). This framework allowed for a comprehensive evaluation of each rhythm metric's effectiveness in speaker identification across both languages.

3. Results

3.1. Preliminary data analysis

Table 1 provides descriptive statistics for speech rhythm metrics calculated for Persian-English bilingual speakers.

Table 1.

Descriptive data pertaining to speech rhythm metric grouped by language

Measure	Mean (SD)		CV (%)	
	Persian (L1)	English (L2)	Persian (L1)	English (L2)
ΔC(ln)	0.677 (0.087)	0.702 (0.0722)	12.85	10.26
n-PVI-C	64.69 (6.21)	68.59 (5.843)	9.60	8.59
VarcoC	0.990 (0.401)	0.851 (0.264)	40.51	31.02
ΔV(ln)	0.489 (0.034)	0.574 (0.045)	6.95	7.84
n-PVI-V	51.46 (3.47)	60.15 (5.89)	6.74	9.79
VarcoV	0.518 (0.063)	0.631 (0.087)	12.16	13.79
%V	37.07 (3.24)	33.65 (4.25)	8.74	12.63
Articulation rate	5.107 (0.557)	3.901 (0.483)	10.91	12.38

* CV = Coefficient of Variation (standard deviation/mean × 100%)

For consonantal-based measures, English has a slightly higher mean ΔC (ln) value (0.702) compared to Persian (0.677), while Persian shows a greater coefficient of variation (CV) for ΔC (ln), at 12.85% versus 10.26% in English. VarcoC, which measures variability in consonantal timing, is higher on average in Persian (0.990) than in English (0.851), with Persian displaying greater relative variability in this measure as well (CV of 40.51% compared to 31.02% in English). These differences suggest distinct rhythmic handling of consonant timing between the two languages.

For vocalic-based metrics, English shows higher mean values for ΔV (ln) and VarcoV, with means of 0.574 and 0.631, respectively, compared to 0.489 and 0.518 in Persian. English also has a higher mean n-PVI-V value (60.15) compared to Persian's 51.46, indicating greater variability in vowel durations. The %V measure differs considerably, with Persian showing a higher mean %V (37.07) than English (33.65). Finally, the articulation rate (rateSyl) is faster in Persian, averaging 5.107 syllables per second, compared to 3.901 in English.

3.2. Comparative statistics between between L1 (Persian) and L2 (English)

Table 2 presents the comparative statistics between L1 (Persian) and L2 (English) speech rhythm measures in Persian-English bilingual speakers, including t-tests, effect sizes (Cohen's d), and percentage differences (% Diff). The t-statistics and p-values reveal several significant differences between the two languages. Notably, n-PVI-C, ΔV (ln), VarcoV, n-PVI-V, %V, and articulation rate show statistically significant differences ($p < 0.05$), indicating that these rhythm metrics vary significantly when bilingual speakers switch between languages. For example, ΔV (ln) has a t-statistic of 7.83 and a highly significant p-value ($p < 0.001$), along with a large effect size (Cohen's $d = 2.12$), suggesting a substantial increase of 17.38% in English compared to Persian. Similarly, the articulation rate has a large negative effect size (Cohen's $d = -2.34$) and a significant difference ($p < 0.001$), showing a 23.61% decrease in English.

In terms of effect size, metrics such as VarcoV (Cohen's $d = 1.49$) and n-

PVI-V (Cohen's d = 1.81) show large positive differences in English, while %V and articulation rate display notable negative values, reflecting reductions in English.

Table 2.

Results of paired t-test along with effect sizes and Diff%

measure	t-stat	p-value	Cohen's d	%Diff
ΔC(ln)	1.12	0.276	0.31	3.69
n-PVI-C	2.31	0.032*	0.65	6.02
VarcoC	-1.43	0.168	-0.42	-14.04
ΔV(ln)	7.83	<0.001**	2.12	17.38
n-PVI-V	6.42	<0.001**	1.81	16.88
VarcoV	5.89	<0.001**	1.49	21.81
%V	-3.12	0.006*	-0.91	-9.22
Articulation rate	-8.45	<0.001**	-2.34	-23.61

* p < 0.05, ** p < 0.001; % Diff = ((L2-L1)/L1) × 100%; Cohen's d interpretation: small = 0.2, medium = 0.5, large = 0.8

3.3. Cross-linguistic correlations between L1 (Persian) and L2 (English)

Table 3 provides a summary of the correlation coefficients between Persian (L1) and English (L2) speech rhythm metrics, calculated for bilingual speakers. These correlations reveal the extent to which speech rhythm patterns are preserved across languages for individual bilingual speakers, offering insights into the potential transfer of rhythmic features between L1 and L2. The results show that ΔC(ln) and n-PVI-C exhibit strong cross-language correlations, with correlation coefficients of 0.62 (p = 0.004) and 0.58 (p = 0.007), respectively, suggesting that bilingual speakers maintain relatively consistent patterns in these measures across both languages. VarcoC (r = 0.45, p = 0.046) was significantly correlated across the languages albeit moderately. Other measures including ΔV (ln), n-PVI-V, %V and articulation rate displayed moderate correlations, indicating some degree of rhythmic similarity between L1 and L2, though to a lesser extent. Weak correlation was observed for VarcoV (r = 0.28, p = 0.232), suggesting minimal cross-language consistency for this measure. Generally, these findings suggest that while certain rhythm metrics

are strongly correlated across languages, others show only moderate or weak cross-linguistic consistency, reflecting a varied adaptation of rhythmic patterns in bilingual speech.

Table 3

Cross-Language Correlations Between L1 (Persian) and L2 (English)

parameter	r	p-value	correlation strenght
ΔC(ln)	0.62	0.004**	Strong
n-PVI-C	0.58	0.007**	Strong
VarcoC	0.45	0.046**	Moderate
ΔV(ln)	0.39	0.089	Moderate
n-PVI-V	0.31	0.184	Moderate
VarcoV	0.28	0.232	Weak
%V	0.42	0.065	Moderate
Articulation rate	0.35	0.13	Moderate

* Statistically significant correlations ($p < 0.05$, $p < 0.01$) were interpreted as weak ($r < 0.3$), moderate ($0.3 \leq r < 0.5$), or strong ($r \geq 0.5$) based on the magnitude of the correlation coefficient (r).

3.4. Suitability of rhythmic measures for speaker identification in L1, L2, and combined accuracy across both languages

To assess how effective rhythmic measures are in the speaker identification task, we performed linear discriminant analysis (LDA) once in each language and once across the whole data. To do so, a standard LDA algorithm with a leave-one-out cross-validation procedure was used. The results of this analysis are summarized in Table 4.

Table 4

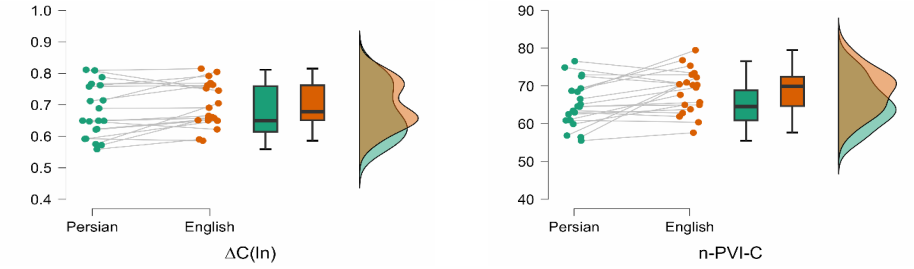
Identification accuracy in Persian (L1), English (L2) and across both languages

parameter	accuracy in Persian (L1)	accuracy in English (L2)	overall accuracy
ΔC(ln)	79.20%	71.80%	75.50%
n-PVI-C	74.50%	76.30%	75.40%
VarcoC	70.10%	68.40%	69.30%
ΔV(ln)	68.40%	65.20%	66.80%
n-PVI-V	64.00%	62.50%	63.70%
VarcoV	61.20%	60.10%	60.70%
%V	59.40%	58.70%	59.10%
Articulation rate	57.80%	55.30%	56.60%

While ΔC(ln), n-PVI-C, and VarcoC showed the highest accuracy in both Persian (L1) and English (L2), the order of their importance varied. In L1, ΔC(ln), n-PVI-C, and VarcoC achieved the highest accuracy rates, whereas in L2, n-PVI-C had the highest accuracy, followed by ΔC(ln) and VarcoC. Additionally, the higher accuracy rates observed for L1 speaker identification indicate that individual speakers exhibited more consistent and distinctive speech rhythm patterns in their native language compared to their second language. The raincloud chart of the two parameters that best identified speakers in their L1, L2, and overall is presented in Figure 2.

Figure 2

Raincloud chart of ΔC(ln) and n-PVI-C in Persian (L1) and English (L2)



4. Discussion

The current study investigated the influence of bilingualism on speech rhythm by examining duration-based rhythm metrics in Persian-English late bilingual speakers. The results reveal several significant patterns in how bilingual speakers handle rhythm across their L1 (Persian) and L2 (English), with implications for both language-specific rhythm characteristics and bilingual speech production.

Our findings revealed significant cross-linguistic differences between Persian and English in most rhythm metrics, particularly in vocalic measures. English exhibited higher values of ΔV(ln), VarcoV, and n-PVI-V compared to Persian, suggesting greater variability in vocalic intervals. These results are in accordance with the established notion of English as a stress-timed language, where vowel durations fluctuate significantly based on stress patterns. The large effect sizes for these measures (Cohen's d ranging from 1.49 to 2.12) highlight the substantial magnitude of these cross-linguistic differences. Additionally, English displayed a lower %V value (33.65%) compared to Persian (37.07%), indicating a proportionally smaller amount of vocalic regions in English speech. This finding aligns with the higher degree of syllabic complexity in English and its tendency to reduce vowels in unstressed positions.

In comparison to native English speakers, as examined in the study by White and Mattys (2007), Persian-English bilinguals in our study exhibited significantly lower values for vocalic rhythmic measures (%V, VarcoV, and n-PVI-V), as well as a slower articulation rate when speaking English. This indicates that native English speakers exhibit more pronounced stress-timed prosodic characteristics, marked by higher values for n-PVI-V and VarcoV, as well as a more consistent vocalic interval percentage. In contrast, the English speech of Persian-English bilinguals, as an L2, exhibited rhythmic properties more typical of syllable-timed languages like Persian, characterized by reduced vocalic duration variability and greater temporal uniformity. This supports the well-documented phonological transfer effect, where a speaker's first language (L1) influences the rhythmic and temporal characteristics of their second

language (L2) production (Henriksen, 2016; Aldrich, 2020).

Regarding bilingual studies, Aldrich (2020) found that both vocalic and consonantal rhythmic measures exhibited greater variability in English compared to Spanish among Spanish-English bilinguals. However, in our study, we observed this increased variability only for vocalic intervals. This suggests that Persian-English bilinguals primarily experience rhythmic influence in their vocalic patterns during language switching, while Spanish-English bilinguals exhibit influence in both vocalic and consonantal patterns. We assume this discrepancy may be due to the bilingual status of our participants, who are late bilinguals, compared to the early bilingual participants in Aldrich's study (2020). Furthermore, the articulation rates of the two languages were found to be markedly different, with Persian showing a faster articulation rate (5.107 syllables/second) compared to English (3.901 syllables/second). This outcome may be explained by the simpler syllable structure of Persian, which allows for a greater number of syllables to be produced within a given time frame compared to English. This finding corroborates the findings of Dellwo et al. (2015), who observed that German-Italian bilingual speakers demonstrated a higher articulation rate in Italian, characterized by its simpler syllable structure, relative to German. Aldrich (2020) also found similar findings that Spanish, which is simpler in syllable structure, had a higher articulation rate compared to English in early Spanish-English bilingual speakers.

The correlation analysis revealed interesting patterns regarding the consistency of rhythmic features across languages. Strong cross-language correlations were observed for consonantal measures (ΔC(ln): r = 0.62; n-PVI-C: r = 0.58), suggesting that speakers maintain relatively stable consonantal timing patterns across both languages. This finding suggests that certain aspects of temporal organization, especially those related to consonantal intervals, may be more resistant to cross-linguistic influence and may reflect individual characteristics rather than language-specific patterns. Conversely, vocalic-based measures showed weaker cross-language correlations (VarcoV: r = 0.28; n-PVI-V: r = 0.31), indicating that speakers are more flexible in adapting their vocalic timing patterns when switching between languages. This

adaptability in vocalic production might reflect speakers' conscious effort to accommodate the different rhythmic requirements of their L1 and L2.

Regarding the suitability of rhythm measures for showing between-speaker variability, the LDA analysis revealed that consonantal measures, particularly ΔC(ln) and n-PVI-C, emerged as the most reliable parameters for identifying speakers across both languages, achieving accuracy rates of 75.5% and 75.4% respectively. This finding suggests that consonantal rhythm measures may serve as more stable individual markers of speaker identity compared to vocalic measures. This finding is different from those of Asadi et al. (2018), who found vocalic measures like %V and articulation rate to be robust parameters in speaker identification tasks in Persian. We assume that this difference might result from two primary factors: the selection of rhythm measures and the statistical procedures employed. In the current study, the inclusion of additional consonantal measures, such as nPVI-C and VarcoC, allowed for a more detailed analysis of consonantal variability, which proved to be more effective in distinguishing speakers, particularly in a bilingual context. Moreover, the use of linear discriminant analysis, in contrast to the linear mixed effects models and multinomial logistic regression utilized by Asadi et al. (2018), may have highlighted distinct aspects of speaker variability, favoring consonantal over vocalic measures. These differences underscore the significance of tailoring measure selection and statistical modeling to the specific linguistic and speaker characteristics under investigation.

Results indicated that speaker identification was more effective in Persian (L1) than in English (L2). The generally higher identification accuracy in L1 suggests that speakers maintain more consistent individual patterns in their native language, likely due to well-established motor patterns and greater automaticity in L1 speech production. Additionally, the challenges associated with L2 acquisition may lead speakers to prioritize intelligibility over acoustic variation, thereby reducing the variability in their speech patterns. This is further supported by the lower accuracy of vocalic parameters in L2, which may stem from the difficulty of mastering vowel production in a second language. Foreign accents often arise from deviations in vowel articulation,

which differ from native speaker norms, contributing to the reduced effectiveness of vocalic measures in L2.

Several limitations should be considered. The use of read speech, while providing a controlled experimental setting, may not fully capture the dynamic nature of spontaneous speech. Furthermore, while the sample size of 20 male speakers is sufficient for statistical analysis, future research with a more diverse sample, including female speakers, is needed to confirm the generalizability of these findings

5. Conclusion

The current study examined rhythmic patterns in late Persian-English bilinguals, revealing significant cross-linguistic differences, particularly in vocalic measures, while consonantal timing showed greater stability across languages. The findings demonstrate that bilingual speakers adapt their rhythmic patterns when speaking English, as evidenced by significant changes in vocalic variability measures and articulation rate while maintaining certain individual characteristics, especially in consonantal timing. Results from the speaker identification analysis highlighted the potential of rhythm metrics, particularly consonantal measures, for speaker recognition applications, with higher accuracy rates in L1 compared to L2 production. These results expand insights into bilingual speech rhythm and have practical implications for bilingual speech production and speaker recognition technology, though future research with larger sample sizes and spontaneous speech would be valuable to confirm these patterns across different contexts and language pairs.

Appendix:

Ten sentences from our dataset are presented below.

1) The big red apple fell to the ground.
2) Seven seals were stamped on great sheets.
3) No doubt about the way the wind blows.
4) You cannot brew tea in a cold pot.
5) Help the woman get back to her feet.

6) Dill pickles are sour but taste fine.

7) The bark of the pine tree was shiny and dark.

8) Take the winding path to reach the lake.

9) The wrist was badly strained and hung limp.

10) Kick the ball straight and follow through.

۱) با روشن‌شدن هوا تظاهرکنندگان به‌سوی مجلس شورای اسلامی شروع به راهپیمایی کردند.

۲) بذل و بخشش رزق آدم را زیاد می‌کند.

۳) در مسابقه‌ی جام جهانی فوتبال، نروژ، ژاپن را شکست داد.

۴) هر روز صبح پنیر با کره می‌خورم.

۵) در کنفرانس ژنو صلح برقرار گردید.

۶) این ماهی طعم میگو دارد.

۷) اگر اسمت را عوض کنی به نفع تو است.

۸) این متن خیلی سنگین است.

۹) رعد و برق باعث رعب و وحشت چند نفر شد

۱۰) من طبل بلد نیستم.

References

Abercrombie, D. (1967). *Elements of General Phonetics.* Edinburgh University Press.

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., Nourbakhsh, M., He, L., Pellegrino, E. & Dellwo, V. (2018). Between-speaker rhythmic variability is not dependent on language rhythm, as evidence from Persian reveals. *International Journal of Speech, Language and the Law, 25*(2), 151- 174. https://doi.org/10.1558/ijsll.37110

Asadi, H., Nourbakhsh, M., Sasani, F., & Dellwo, V. (2018). Examining long-term formant frequency as a forensic cue for speaker identification: An experiment on Persian. In *Proceedings of the First International Conference on Laboratory Phonetics and Phonology*, Tehran, Iran (pp. 21–28). https://search.ricest.ac.ir/dl/search/defaultta.aspx?DTC=36&DC=306748

Boersma, P. & Weenink, D. (2024). Praat: Doing Phonetics by Computer. http://www.praat.org, Accessed 20 June 2024.

Braun, A. (1995). Fundamental frequency – How speaker-specific is it? In A. Braun & J.P. Köster (eds.), *Studies in Forensic Phonetics*, (pp. 9-23). Wissenschaftlicher Verlag.

Bunta, F., & Ingram, D. (2007). The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4-and 5-year-old children. *Journal of Speech, Language, and Hearing Research, 50*(4), 999-1014. https://doi.org/10.1044/1092-4388(2007/070)

Debruyne, F., Decoster, W., Van Gijsel, A., & Vercammen, J. (2002). Speaking fundamental frequency in monozygotic and dizygotic twins. *Journal of Voice*, *16*(4), 466-471. https://doi.org/10.1016/s0892-1997(02)00121-2

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for ΔC. In P. Karnowski & I. Szigeti (Eds.), Language and language-processing (pp. 231-241). Peter Lang. https://doi.org/10.5167/uzh-111789

Dellwo, V. (2010). *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence* [Unpublished doctoral dissertation]. Bonn University.

Dellwo, V., Leemann, A., & Kolly, M. J. (2015). Rhythmic variability between speakers: articulatory, prosodic, and linguistic factors. *The Journal of the Acoustical Society of America*, *137*(3), 1513–1528. https://doi.org/10.1121/1.4906837

Dellwo, V., Schmid, S., Leemann, A., & Kolly, M. J. (2015). Speaker-individual rhythmic characteristics in read speech of German-Italian bilinguals. *Trends in Phonetics and Phonology: Studies from German speaking Europe. Bern: Peter Lang*, 349-362. https://doi.org/10.5167/uzh-114508

Gibbon, D. (2022) Speech rhythms: learning to discriminate speech styles. In *Speech Prosody*, 302-306. doi: 10.21437/SpeechProsody.2022-62

Gold, E., French, J.P & Harrison, P. (2013). Examining long-term formant distributions as a discriminant in forensic speaker comparisons under a likelihood ratio framework. In *Proceedings of Meetings on Acoustics*, Montreal, Canada, (pp. 1-8). https://doi.org/10.1121/1.4800285

Goldstein U. G. (1976). Speaker-identifying features based on formant tracks. *The Journal of the Acoustical Society of America*, *59*(1), 176–182. https://doi.org/10.1121/1.380837

Gordon, M, Barthmaier, P, Sands.K. (2002). A Cross-linguistic study of voicelessfricatives. *Journal of the International Phonetic Association, 32*(2), 2-32. https://doi.org/10.1017/S0025100302001020[Opens in a new window]

Grabe, E. & Low, E. L. (2002). Durational variability in speech and rhythm class hypothesis. In N. Warner & C. Gussenhoven (Eds.), Papers in Laboratory *Phonology* 7 *(pp. 515-543), Mouton de Gruyter.* https://doi.org/10.1515/9783110197105.2.515

He, L. & Dellwo, V. (2016). The role of syllable intensity in between-speaker rhythmic variability. *The International Journal of Speech, Language and the Law*, 23, 243-273. https://doi.org/10.1558/ijsll.v23i2.30345

Henriksen, N. (2016). Convergence effects in Spanish-English bilingual rhythm. In *Speech Prosody* (pp.721-725). https://doi.org/10.21437/SpeechProsody.2016-148

Jessen, M. and Becker, T. (2010). Long-term formant distribution as a forensic phonetic feature. *Conference of the Acoustical Society of America*, Cancun, Mexico. https://doi.org/10.1121/1.3508452

Jessen, M., Köster, O., & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language, and the Law, 12*(2), 174-213. https://doi.org/10.1558/sll.2005.12.2.174

Kehoe, Margaret, Lleó, Conxita, & Rakow, Martin. (2011). Speech rhythm in the pronunciation of German and Spanish monolingual and German-Spanish bilingual 3- year-olds. *Linguistische Berichte*, 227, 323-351. 10.46771/2366077500227_3

Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies*. Wiley-Blackwell Publishing.

Lazard*, G. (1992).* Grammar of Contemporary Persian*. Mazda Publishers.*

Ladefoged, P. (2006). *A course in phonetics.* Thomson Wadsworth.

Lee, Y., Keating, P., & Kreiman, O. (2019). Acoustic voice variation within and between speakers. *The Journal of the Acoustical Society of America*, 146(3), 1568–1579. https://doi.org/10.1121/1.5125134

Leemann, A., Kolly, M.-J., & Dellwo, V. (2014). Speaker-individuality in suprasegmental temporal features: implications for forensic voice comparison. *Forensic Science International,* 238, 59-67. https://doi.org/10.1016/j.forsciint.2014.02.019

Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm:

Evidence from L1 Mandarin and German learners of English. *Studies in Second Language Acquisition, 36*(2), 223-255. https://www.jstor.org/stable/26328939.

Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *The Journal of the Acoustical Society of America,* 129(5), 3258-3270. https://doi.org/10.1121/1.3559709

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech, 43*(4), 377-401. https://doi.org/10.1177/00238309000430040301

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge University Press.

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America, 138*(2), 533-544. https://doi.org/10.1121/1.4923359

Pike, K. L. (1945). *The Intonation of American English*. University of Michigan Press.

Prieto, P., del Mar Vanrell, M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, 54, 681–702. https://doi.org/10.1016/j.specom.2011.12.001

R Core Team (2023) R: A Language and Environment for Statistical Computing (version 4.3.1). R Foundation for Statistical Computing. http://www.rproject.org/, Accessed 20 June 2023.

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

Rose, P. (2002). *Forensic speaker identification*, New York: Taylor & Francis.

Sadeghi, V. (2015). A phonetic study of vowel reduction in Persian. *Language Related Research,* 30, 165-187. http://lrr.modares.ac.ir/article-14-7916-en.html

Smorenburg, L., & Heeren, W. (2020). The distribution of speaker information in Dutch fricatives /s/ and /x/ from telephone dialogues. *The Journal of the Acoustical Society of America*, *147*(2), 949. https://doi.org/10.1121/10.0000674

Stockmal, V., Markus, D., & Bond, D. (2005). Measures of native and non-native rhythm in a quantity language. *Language and Speech*, 48(1), 55-63. https://doi.org/10.1177/00238309050480010301

Taghva, N., Moloodi, A., Abolhasanizadeh, V., & Tabei, R. (2023). A corpus study of durational rhythmic measures in the Kalhori variety of Kurdish. *Loquens*, *10*(1-2), e098. https://doi.org/10.3989/loquens.2023.e098

Ulrich, N., Pellegrino, F., & Allassonnière-Tang, M. (2023). Intra- and inter-speaker variation in eight Russian fricativesa). *The Journal of the Acoustical Society of America*, *153*(4), 2285. https://doi.org/10.1121/10.0017827

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501-522. https://doi.org/10.1016/j.wocn.2007.02.003

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., & Mattys, S. L. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America, 127*(3), 1559–1569. https://doi.org/10.1121/1.3293004

Windfuhr, G. L. (1979). *Persian grammar: History and state of its study*. De Gruyter Mouton.

Yoon, T.J. (2010). Capturing inter-speaker invariance using statistical measures of speech rhythm. *In Speech Prosody*, (pp. 1-4), Chicago/IL, USA. doi: 10.21437/SpeechProsody.2010-58