Journal of Language Research
ZABANPAZHUHI
Quarterly Scientific Journal of Language Research, Vol. 15, No. 49, Winter 2024
Alzahra University, http://jlr.alzahra.ac.ir
Article Type: Research
pp.61-82

Comparative Analysis of Speech Rhythm Measures for Persian Speaker Identification: Duration vs. Intensity[1]

Homa Asadi[2]

Received: 02/11/2023
Accepted: 03/12/2023

Abstract:

Previous studies have demonstrated the efficacy of speech rhythm measures in speaker identification across various languages with different phonotactic structures. In Persian language, in particular, two categories of speech rhythm metrics were examined: duration-based and intensity-based metrics. Building upon these prior works, the current study delves deeper into the discrimination capabilities of the mentioned measurement types—duration-based versus intensity-based—in the context of Persian speakers. To achieve this, a multinomial logistic regression model was employed on a dataset comprising 20 male Persian speakers, each reciting 100 sentences at a normal speaking pace. Findings revealed that, when distinguishing between Persian speakers, duration-based measures outperform intensity-based ones, however, this excellence is very slight. This observation is significant, as it sheds light on the suitability of specific rhythm metrics for Persian speaker identification. I postulate that this discrepancy in performance may be attributed to the simple syllable structure of Persian and the lesser reliance on intensity as a primary indicator of lexical stress. This research contributes valuable insights into the choice of rhythm metrics for optimal Persian speaker identification and underscores the importance of considering linguistic features when developing speaker recognition systems.

Keywords: forensic phonetics, speaker identification, speech rhythm measures, Persian language

---

[1] DOI: 10.22051/jlr.2023.45448.2370
[1]Assistant Professor of Linguistics, University of Isfahan, Isfahan, Iran;
h.asadi@fgn.ui.ac.ir, ORCID: https://orcid.org/0000-0003-1655-1336

1. Introduction

Acoustic parameters of speech rhythm, computed over the temporal characteristics of vocalic and consonantal intervals (hereafter referred to as C- and V-intervals), as well as syllabic intensity characteristics have consistently revealed noticeable between-speaker variability in multiple studies (Asadi et al, 2018; Dellwo et al., 2015; He & Dellwo, 2016; Leemann et al., 2014; Moez et al., 2018; Taghva et al., 2023; Weingartova, 2014). These investigations have collectively provided substantial empirical evidence supporting the idea that rhythmic parameters can be influential in distinguishing speakers across different languages, even when these languages are characterized by different phonotactic structures. The initial inquiries into this area primarily emerged from German and Swiss-German speech corpora, the two languages which are traditionally classified as stress-timed. Stress-timed languages are known for their intricate syllable structures and complex consonant clusters, setting them apart from syllable-timed languages, which generally exhibit less complex syllable structures and simpler consonant cluster patterns (Dellwo, 2010). Moreover, stress-timed languages exhibit relatively less variability in the temporal features of both consonantal and vocalic intervals (Dellwo, 2010). Studies focusing on stress-timed languages have consistently demonstrated the capability of speech rhythm metrics to distinguish speakers based on both temporal and intensity-related characteristics of speech (Dellwo et al., 2012; Dellwo et al., 2015; He & Dellwo, 2016).

Subsequent investigations ventured into the realm of syllable-timed languages, typified by Persian, where vocalic intervals tend to be prolonged while consonantal intervals favor brevity. This exploratory line of inquiry was underpinned by the hypothesis that, given the simple syllable structure of Persian and the absence of vowel reduction, there might exist variations in the release of consonant clusters, and consequently, speakers would have less freedom for differentiation (Asadi et al., 2018). However, contrary to initial expectations, these studies yielded consistent results, confirming the effectiveness of both duration-based and intensity-based metrics on discriminating speakers, regardless of the linguistic medium employed (Asadi

et al., 2018; Asadi & Alinezhad, 2022).

Nevertheless, it is important to note that the efficacy of rhythmic speech metrics for speaker identification is not uniform across languages. Comparative analyses of speech rhythm metrics in stress-timed languages, such as German and Swiss German, have demonstrated that intensity measures outperform duration-based metrics (He & Dellwo, 2016). This discrepancy is postulated to arise from the stronger association of intensity measures with underlying anatomical factors. Previous research has indicated that parameters such as the extent of oral aperture and intensity oscillations over time are closely linked to speech production (Chandrasekaran et al., 2009). Consequently, it is plausible that metrics assessing time-integrated intensity variability encompass a richer source of information regarding the idiosyncratic articulatory patterns of individual speakers than interval durations alone. Therefore, intensity-based rhythm metrics may expand the feature space, providing additional orthogonal dimensions for distinguishing different speakers (He & Dellwo, 2014; He & Dellwo, 2016).

Building upon these insights, the question arises as to whether similar outcomes can be expected in a linguistic context that substantially differs from the stress-timed languages of German and Swiss German, such as Persian. Persian exhibits distinct temporal attributes, characterized by longer vocalic intervals and shorter consonantal intervals. These different phonotactic features of Persian compared to the stressed-timed languages might lead one to question whether the efficacy of intensity-based metrics, closely linked to articulatory movements of speech organs, would still hold in this linguistic context. To address this conjecture and to further our understanding of the applicability of rhythm metrics across languages, I propose a comprehensive comparative analysis of both duration-based and intensity-based metrics within the context of Persian. This research endeavor aims to answer the two following research questions:

1) Which set of rhythm metrics, namely intensity-based or duration-based, exhibits superior performance in delineating between-speaker variability in Persian?

2) Which domain of metrics (intensity measures, duration measures, or a composite thereof) yields elevated recognition outcomes?

These questions are of paramount importance as they have the potential to shed light on the generalizability of speech rhythm metrics across diverse linguistic backgrounds and offer valuable insights into the intricate relationship between speech production and rhythm metrics in different languages. Consequently, this comparative analysis contributes significantly to the broader understanding of the role of acoustic metrics in the realm of linguistics and speaker identification.

2. Past research on between-speaker variability in speech rhythm measures

Speech rhythm metrics can be categorized into two distinct domains (Tilsen & Arvaniti, 2013). These domains encompass metrics which are rooted in the temporal duration of speech intervals and metrics which are reliant on the temporal characteristics of the amplitude envelope. These metrics are typically derived from various phonetic units, including CV-intervals (Grabe & Low, 2002; Ramus et al. 2009), syllables or foot (Nolan & Asu, 2009), consecutive CV-intervals (Grabe & Low, 2002), voiced and unvoiced intervals (Dellwo & Fourcin, 2013), and amplitude peak intervals (Marcus, 1981). Earlier studies on speech rhythm measures mainly focused on cross-linguistic rhythmic attributes, revealing that the durational features of C- and V-intervals serve as correlates of language-specific auditory rhythmic characteristics (Dellwo, 2006; Grabe & Low, 2002; Ramus et al., 1999; White & Mattys, 2007). Following this line of research, phoneticians proposed that speech rhythm measurements based on different phonetic regions like CV intervals and syllable units, may also hold the potential to capture between-speaker rhythmic variation within a given language. However, the explanation for the variability inherent in acoustic correlates of speech rhythms has been a subject of debate.

Yoon (2010) conducted one of the earliest studies on between-speaker rhythmic variability, applying rhythm metrics to a group of 10 native English speakers. His findings revealed significant differences in %V and VarcoV among

these speakers. In a similar vein, Wiget et al. (2010) investigated the reliability of various rhythm metrics in a dataset comprising 6 British English speakers. While their results indicated considerable variability across speakers, they emphasized that the sentence structure had the most essential impact on rhythm scores. With a similar goal in mind, Dellwo et al. (2012) examined the acoustic characteristics of speech rhythm in a dataset consisting of spontaneous speech from 8 German speakers. Their findings revealed that despite the variable nature of spontaneous speech, rhythmic measures effectively captured the variations between speakers. Furthermore, their study demonstrated that vocalic durations were more effective than consonantal intervals in showing between-speaker variability. In another study, Leemann et al. (2014) explored between-speaker variability in suprasegmental temporal features using a corpus containing 16 speakers of Zurich German. Their research uncovered a high degree of variability between speakers in both read and spontaneous speech. Additionally, they found that suprasegmental temporal features remained consistent across different speaking styles and channel variations, which enhanced their applicability in forensic casework. In another study, Dellwo et al. (2015) identified strong and consistent between-speaker variability in durational metrics, such as %V, ΔC(ln), ΔV(ln), and ΔPeak(ln), across two German speech corpora with different sources of within-speaker acoustic variability. This study underscored the stability of the investigated rhythmic measures in the presence of within-speaker variability stemming from articulation rate and linguistic structural characteristics.

Following extensive investigations into stressed-timed languages, Asadi et al. (2018) employed specific acoustic metrics primarily derived from CV intervals and syllables in two sets of Persian speech corpora. Their research postulated that the phonotactic structure of Persian, characterized by a simple syllable structure and reduced vowel reduction, may limit the degree of speaker differentiation. Nonetheless, findings in Persian supported earlier observations in stressed-timed languages, indicating that the temporal characteristics of speech rhythm measures effectively expose acoustic variations between speakers, regardless of the language spoken. A significant

outcome of this study was the recognition of %V as a highly influential factor, with the potential to serve as a universal acoustic parameter suitable for forensic voice comparison. Moez et al. (2018) conducted research on another syllable-timed language, namely French, and reported that durational parameters of voiced and unvoiced intervals are significantly influenced by the speaker. Subsequent studies in Persian further confirmed the results originally found by Asadi et al. (2018). Taghva et al. (2021) discovered that metrics like syllable rate VarcoC, %V, nPVI-V, nPVI-VC, ΔC(ln), and ΔPeak(ln) performed exceptionally well in Persian speaker identification. They also highlighted %V as the most crucial parameter for distinguishing between speakers. In the most recent study, Tahgva et al. (2023) applied the same metrics to Kalhori Kurdish, leading to similar findings. They identified %V and syllable rate as the most effective metrics for capturing between-speaker variability in both read and spontaneous speech corpora of Kalhori Kurdish. In general, research utilizing duration-based metrics has consistently demonstrated their ability to reveal substantial variance among speakers in diverse languages, such as English, German, Swiss German, Persian and French.

Following studies on between-speaker rhythmic variability, researchers explored an alternative dimension of speech rhythm measures, suggesting that the aspect related to intensity might also exhibit speaker-related variability. This quantification of speech rhythm in terms of intensity variability was prompted by the well-established connection between mouth aperture size and signal intensity (Chandrasekaran et al., 2009; Garnier, 2008). For example, a wider mouth aperture results in higher amplitude, indicating that idiosyncratic variations in mouth and articulatory organ movements encode a wealth of speaker-specific characteristics related to intensity. He & Dellwo (2014, 2016) proposed the idea that the part of the signal capturing intensity contours may show more speaker individuality because the alternations in mouth opening and closing coincide with the intensity curve of the speech signal. To validate their hypothesis, they applied two sets of intensity variability measures to speech corpora containing speakers from Zurich German and Northern German. The results demonstrated a significant

amount of between-speaker variability across both datasets. Importantly, they found that intensity measures contained more speaker-specific information than measures based on the temporal characteristics of speech, attributing this superiority to their association with anatomical factors. Persian was also investigated in terms of intensity-based measures (Asadi & Alinezhad, 2023). Initially, it was anticipated that given Persian's emphasis on duration as the primary indicator of lexical stress (Sadeghi, 2011), between-speaker intensity variability might be less observable in Persian. However, the results emerging from the analysis of Persian corpora demonstrated that speech rhythm measures derived from the amplitude envelope of the signal still retain the capacity to capture between-speaker variability in Persian.

In summary, previous research underscores the efficacy of speech rhythmic metrics in delineating acoustic variations among speakers. These studies posit that speech rhythm measures are closely intertwined with anatomical factors that are less subject to conscious control by speakers. Nevertheless, they do not discount the relevance of language-specific attributes. Specifically, this study aims to juxtapose the efficacy of two categories of speech rhythm metrics - namely, duration-based and intensity-based metrics - with the dual objectives of determining which domain exhibits superior performance and extending prior findings to ascertain whether the contribution of speech rhythmic metrics is solely a product of biological predispositions or if linguistic elements also wield influence over these acoustic features, thereby broadening our understanding of the intricate dynamics underpinning speech variability.

3. Materials and Methods

3.1. Data collection and segmentation

A corpus of 20 native male Persian speakers, specifically of the Tehrani variety, with an age range spanning from 22 to 35 years, were instructed to articulate a set of 100 sentences at a normal speaking pace, punctuated by 3-second intervals between each sentence. Subsequently, the speech samples underwent detailed analysis using Praat software (version 5.2.34, Boersma & Weenink, 2013) and were annotated across three distinct tiers: segments,

syllables, and peak tiers. In the initial step, the onsets and offsets of speech segments were meticulously marked through manual annotation, utilizing Praat's built-in annotation function. The syllable tier, on the other hand, was carefully delineated through manual annotation performed by the author. A Praat script *DurtaionAnalyzer*[1] was used to calculate automatically all duration-based speech rhythm measures. Computational calculations for intensity-based measures were carried out within the Praat environment, facilitated by a script developed by Lei He from the University of Zurich.

3.2. Acoustic rhythmic parameters

Building upon the influential speech rhythm metrics that exhibited reasonable success in forensic voice comparison studies, two sets of speech rhythm metrics were devised. For this study, duration-based measures derived from CV intervals and syllable units of speech signals were selected. To mitigate potential distortions introduced by speech rate variations, rate-normalized measures were exclusively applied. Drawing upon established temporal measures within the realm of speech rhythm research (Asadi et al., 2018; Asadi & Alinezhad, 2023; Dellwo et al., 2012; Grabe & Low, 2002; Leeman et al., 2014; Ramus et al., 1999), I employed acoustic rhythmic metrics on the collected dataset. In total, I selected five duration-based measures as follows: one vocalic duration ratio (%V), two consonantal and vocalic duration variability measures (ΔV(ln), ΔC(ln)), one rate-normalized vocalic variability measure (n-PVI-V). Additionally, the articulation rate (the number of syllables per second) was calculated from the syllable tier and the following duration-based measures were obtained from the CV interval tier (Dellwo et al., 2015; Leemann et al., 2014):

- %V: proportion over which speech is vocalic;
- ΔV(ln): standard deviation of the neutral log normalized durations of vocalic intervals;
- ΔC(ln): standard deviation of the neutral log normalized durations of

[1] This script is available on http://www.pholab.uzh.ch/static/volker/software/plugin_durationAnalyzer.zip.

consonantal vocalic intervals;

- n-PVI-V: rate-normalized averaged durational differences between consecutive vocalic intervals.

From the intensity contour, I quantified both the mean (M) and peak (P) intensity corresponding to a speech syllable. To obtain the amplitude envelope of the signal, a full-wave rectification was performed on the signal, followed by low-pass filtering at 10 Hz. The mean syllable intensity was determined by summing the intensity values within the contour delineated by syllable onset and offset and then dividing this sum by the duration of the syllable. Likewise, the peak intensity was computed at the syllable peak point using cubic function interpolation. The following intensity measures (He & Dellwo, 2016) were calculated:

- stdevM: the standard deviation of average syllable intensity levels;
- stdevP: the standard deviation of syllable peak intensity levels;
- varcoM: the variation coefficient of average syllable intensity levels (normalized stdevM);
- varcoP: the variation coefficient of syllable peak intensity levels.

3.3. Statistical Analysis

The statistical analysis of the data was conducted using R (R Core Team, 2021) version 4.2.2,. Initially, a Pearson pairwise correlation was employed to examine the extent of correlation between the duration-based and intensity-based measures. This preliminary analysis aimed to ascertain whether these two distinct sets of metrics fall within different categories. Highly correlated measures indicate a substantial overlap in the information they convey, while measures with low or negligible correlations do not share similar information. Next, a multinominal logistic regression model using the assembled speech data was constructed to address the question of which type of rhythmic metrics better accounted for between-speaker variability. In this model, the speaker was designated as the nominal response variable, and the chosen acoustic parameters were treated as predictors. To quantify the extent to which each intensity-based and duration-based measure contributed to

explaining between-speaker variability, I computed the likelihood ratio χ^2 for each acoustic parameter. This value was then divided by the sum of likelihood ratio χ^2 values for all parameters, providing a measure of the proportion of between-speaker variability attributable to each specific acoustic parameter. To express the variability explained by each measure as a percentage, I calculated the percentage of χ^2 values for each measure over the sum of all χ^2 values for all measures. The χ^2 value of the final model was calculated by taking the difference between the -2 log-likelihood ratios (-2LL) of the null model and the final model. Likewise, the χ^2 value of each tested measure was computed by taking the difference between the -2LLs of the final model and each reduced model.

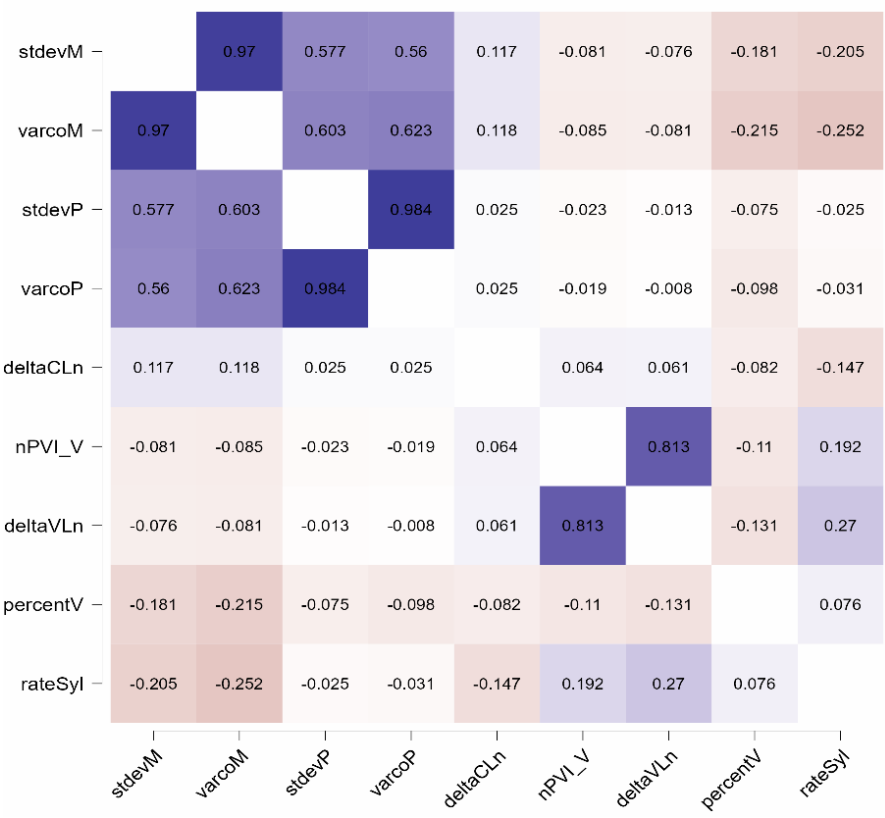
4. Results

4.1. Correlations between speech rhythm measures

I conducted pairwise correlation analyses among the speech rhythm measures to assess their predictive relationships with one another. Figure 1 illustrates the pairwise correlations between the duration-based and intensity-based measures in the collected dataset. The findings depicted in Figure 1 highlight a substantial interdependence among intensity-based measures. Furthermore, correlations among all duration-based measures, with the exception of %V and syllable rate, are substantially high. These results indicate that %V and rate of syllable form distinct categories that exhibit minimal correlation with both other duration-based measures and intensity-based measures. The robust correlations within each measurement type, whether duration-based or intensity-based, suggest that these metrics convey unique information and belong to separate categories. As indicated by Figure 1, these two sets of measures encapsulate distinct insights into speech rhythm variability, and, therefore, a combined analysis of both sets can provide a more comprehensive understanding of the speaker-specific characteristics within speech signals.

Figure 1

Correlation matrices showing Pearson's correlation coefficient r of the intensity and durational speech rhythm measures in our corpus. All correlations were highly significant (p < 0.005).



4.2. Comparative analysis of speech rhythm measures using multinomial logistic regression

In order to ascertain which types of rhythmic measures most effectively explained between-speaker variability, a multinomial logistic regression model was developed using the collected dataset. Results demonstrated that the effect of the speaker was highly significant for all the tested rhythmic measures. As revealed in Table 1, the most substantial effects were observed for the measures %V and syllable rate, closely followed by varcoP and stdevP. This outcome suggests that the most salient parameters indicative of speaker individuality primarily fall within the domain of duration-

based metrics. Nevertheless, it is worth noting that intensity-based measures also demonstrated commendable performance. The metrics obtained from peak intensity contours collectively explained roughly 29% of the between-speaker variability, while intensity metrics derived from the mean intensity contours accounted for approximately 19% of the acoustic rhythmic variability among speakers. The individual impact of each of the 9 predictors in the regression model is visually depicted in the radar chart presented in Figure 2. The length of the radius associated with each measure correlates with its weight in the created model, signifying the relative importance of each intensity measure in explaining variability between speakers.

Findings concerning the most effective parameters for capturing between-speaker variability in Persian highlight %V and syllable rate as the top parameters within the duration-based category. %V and syllable rate accounted for approximately 23% and 21% of the acoustic rhythmic variability in Persian speakers, respectively. Among intensity-based measures, stdevP and varcoP emerged as the most robust parameters for characterizing speaker identity. StdevP and VarcoP, derived from peak syllable characteristics, collectively contributed to approximately 28% of the acoustic rhythmic variability among speakers. Figure 3 presents boxplots illustrating the between-speaker variability observed in %V and varcoP, respectively. Upon exaining the correlation heatmap (Figure 1), it is apparent that there is a relatively low correlation between these top parameters. Even though %V and syllable rate both fall within the same category, they exhibit minimal correlation. This is a significant finding, as it suggests that these parameters convey distinct information, thus making them suitable for combination in speaker identification.

Table 1

Summary of the results of multinomial logistic regression on speech rhythm measures

-2 Log Likelihood of Reduced Model		χ2 [df]		P	Variability explained
Model fitting information					
null model	12941.563				
full model	7287.641	5653.924[171]		<0.0001	
Likelihood ratio test of each speech rhythm measure					
stdevM	7582.546	294.906 [19]		<0.0001	9.50%
varcoM	7590.376	302.736 [19]		<0.0001	9.81%
stdevP	7742.691	455.051 [19]		<0.0001	14.66%
varcoP	7728.237	440.597 [19]	∑χ2 = 3076.677	<0.0001	14.19%
%V	8000.653	713.014 [19]		<0.0001	22.97%
ΔVln	7368.771	81.131 [19]		<0.0001	2.61%
ΔCln	7437.624	149.984 [19]		<0.0001	4.83%
n-PVI-V	7314.323	26.684 [19]		<0.0001	0.85%
Syllable rate	7926.898	639.258 [19]		<0.0001	20.59%

Figure 2

Radar chart depicting the relative contributions of each analyzed parameter within the multinomial logistic regression model for speaker variation

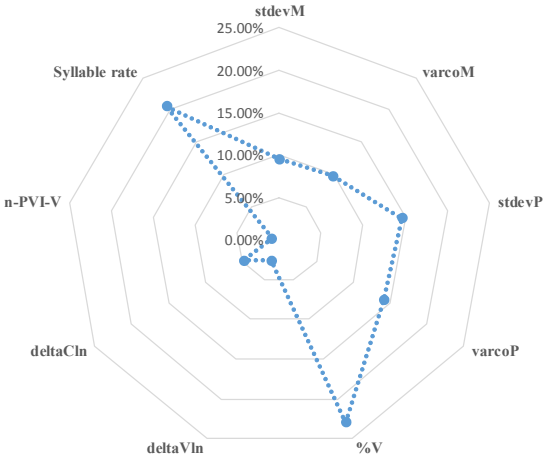
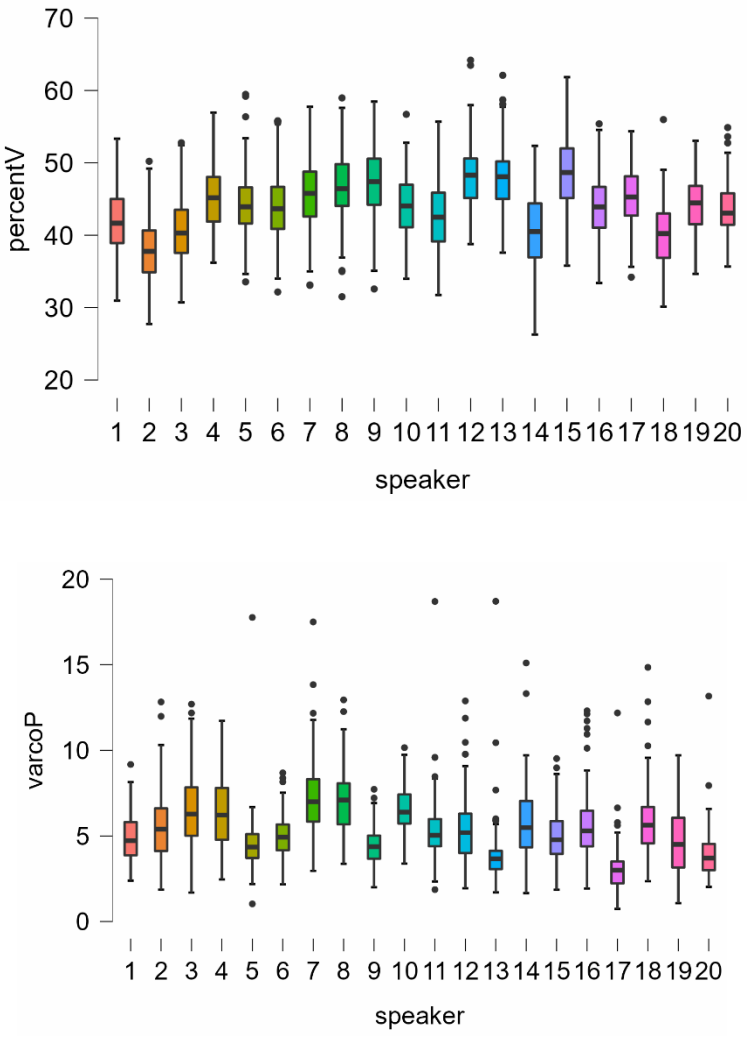


Figure 3

Boxplots showing speech rhythmic variability for %V (top) and VarcoP (bottom)



5. Discussion

In this study, I selected a diverse set of speech rhythm measures from different categories, encompassing both duration-based and intensity-based metrics. The primary objective was to assess which category of these metrics is better for capturing between-speaker variability in Persian. Initially, a pairwise Pearson correlations was performed to evaluate the predictive relationships among these parameters. This finding suggests that these two disparate

categories of speech rhythm measures offer distinct and complementary information about speech signals produced by individual speakers. Consequently, it becomes apparent that the fusion of these categories has the potential to yield more robust and comprehensive results in the realm of speaker identification. A closer examination of the correlation heatmap (see Figure 1) highlights that intensity-based measures display a strong interdependence among themselves, with a high level of correlation. In contrast, the duration-based measures exhibit more varied correlation patterns. As is shown in Figure 1, a high correlation was observed between ΔV(ln) and n-PVI-V, indicating that these two measures, both derived from vocalic intervals, convey similar information. However, the remaining duration-based parameters demonstrated relatively weak correlations.

Based on the data analysis, duration-based measures emerged as the more effective set of metrics when compared to their intensity-based counterparts, despite a relatively modest difference in their performance. This superior performance can be examined from two key perspectives. Firstly, the results of the multinomial logistic regression analysis unequivocally highlighted that duration-based measures, particularly %V and syllable rate, wielded the most significant influence in distinguishing between speakers when compared to the other duration-based parameters. %V and syllable rate, in combination, accounted for approximately 43% of the acoustic rhythmic variability among speakers. This finding underscores the significant influence of these two specific measures on excellence in duration-based metrics. A second perspective that significantly contributes to our understanding of the data stems from the examination of the pairwise correlation analysis. Intensity-based measures exhibited noticeable correlations among themselves, indicating redundancy in the information they extracted from the speech signal. In contrast, when we turn our attention to the duration-based measures, particularly highlighting the prominent parameters of %V and syllable rate, we observe weak correlations with the remaining metrics. This observation holds profound significance as it aligns with the insights of Rose (2002), who noted that an assemblage of acoustic parameters with low inter-correlations often

leads to enhanced discrimination performance. Hence, it can be reasonably inferred from the results that combining the most effective duration-based parameters, specifically %V and syllable rate, with intensity-based measures derived from the peak contours of intensity, such as varcoP and stdevP, may yield a better performance in the context of Persian speaker identification. This synthesis of parameters bridges the inherent strengths of both duration and intensity measures, culminating in a holistic approach that leverages the linguistic features of Persian and the subtleties of acoustic rhythm to advance the accuracy and reliability of speaker differentiation.

Upon scrutinizing the performance of the two domains of the analyzed rhythmic measures, these findings revealed that duration-based metrics exhibited a greater performance in discerning Persian speakers. However, it is noteworthy that intensity-based measures also yielded commendable results, albeit to a slightly lesser extent. One plausible rationale for these observed patterns can be traced back to the phonotactic structure of the Persian language. It is well-documented that syllable-timed languages, such as Persian, exhibit a higher percentage of vocalic intervals (%V) and shorter ΔC (the standard deviation of C-intervals), resulting in more compact and evenly distributed C-intervals (Dellwo & Wagner, 2003; Ramus et al., 1999). This phenomenon could provide speakers of syllable-timed languages with a greater degree of flexibility in modulating the vocalic aspects of their speech. The rich diversity of vocalic intervals in Persian allows for a more nuanced expression in speech, contributing to the effectiveness of duration-based measures, particularly %V, in capturing the rhythmic variability among speakers. Furthermore, this study highlights the significance of syllable rate as a crucial factor in showing variations among Persian speakers. At the syllable level, syllable-timed languages are often characterized by reduced complexity and typically lack vowel reduction. This linguistic feature affords their speakers increased flexibility in both syllable production and rate. In the case of Persian, the relatively simple syllable structure and the absence of vowel reduction may have been instrumental in generating variations in syllable rates among speakers.

Findings of this study, showcasing the superiority of duration-based measures over intensity-based ones, are in contrast with the conclusions drawn by He & Dellwo (2016), who found that intensity-based measures were more effective in distinguishing between German speakers. Their postulation centered on the notion that intensity-based measures were more closely linked to the idiosyncratic movements of speech articulatory organs, with a particular emphasis on those influencing the area of mouth opening. This contrast in findings underscores the importance of the relationship between speech rhythm measures and the linguistic and phonetic characteristics specific to each language. When comparing the results of this study to those of He and Dellwo (2016), it becomes clear that, although intensity-based measures perform adequately in the context of Persian, their performance still is not on par with %V and syllable rate, two key durational measures. The differences in intensity variability across languages can be attributed to a multitude of factors. One contributing factor is the complexity of syllables, which significantly impacts the systematic variability in speech rhythm measurements (Prieto et al., 2012). It is generally assumed that languages with more phonotactically complex structures exhibit higher levels of intensity variability than languages with simpler structures. Another important factor influencing the variability of intensity-based rhythmic measures is vowel reduction. Languages that allow vowel reduction tend to exhibit higher syllabic intensity because reduced or centralized vowels produce lower amplitude envelope levels, resulting in lower intensity levels in terms of mean intensity or peak intensity (He, 2017). Persian, categorized as a syllable-timed language (Lazard, 1992; Windfuhr, 1979), adheres to a simple syllable structure (CV(C)(C)), and it lacks the vowel reduction patterns found in stress-timed languages (Bijankhan, 2018; Lazard, 1992, 2010; Sadeghi, 2015; Windfuhr, 1979). Taking into account these considerations and exploring the multifaceted interplay between language-specific attributes and speaker variability, it becomes evident that the linguistic characteristics of Persian wield a substantial influence over the observed variations in intensity measures among speakers.

An alternative explanation for the superior performance of duration-based measures in Persian, as compared to the less effective performance of intensity-based measures, can be attributed to the way in which Persian signals lexical stress. The identification of lexical stress in languages is a complex process influenced by a multitude of interconnected perceptual factors (Fry, 1958). Languages that rely on intensity as a primary acoustic cue for indicating lexical stress or prominence typically exhibit higher levels of intensity variability when compared to languages that do not heavily rely on intensity to signal stress placement (Wang, 2008). In the case of Persian, duration is regarded as the most reliable indicator of stress, whereas intensity serves as a less reliable indicator of stress position in Persian (Sadeghi, 2011). Given this linguistic feature, it is reasonable to infer that the intensity levels of speech signals in Persian would be less affected by speaker individuality when compared to languages where intensity plays a more prominent role in stress placement. In Persian, the emphasis on duration as the primary marker of lexical stress may result in a relatively consistent intensity pattern across speakers, making intensity-based measures less effective for distinguishing among them.

In light of the various factors that have been discussed, it becomes increasingly evident that the linguistic characteristics of Persian play an important role in shaping the observed variations in speech rhythm measures among speakers. Our findings shed light on the fact that speech rhythm metrics are not solely a product of biological variability but are equally molded by linguistic factors. The intricate interplay between these biological and linguistic dimensions plays an integral role in defining the unique acoustic profiles of individual speakers. These insights underscore the importance of considering linguistic characteristics when delineating the most appropriate acoustic parameters for forensic voice comparison. Such considerations extend beyond the mere acoustic variability within speech and encompass the underlying linguistic structures and patterns that help shape these acoustic characteristics. Recognizing the profound impact of language-specific features on speaker differentiation is important in advancing the precision and reliability of voice

analysis, particularly in the realm of speaker identification and related applications.

In sum, this study emphasizes the multifaceted nature of speech rhythm measures, where both the biological attributes of the speakers and the linguistic traits of the language they speak converge to define the acoustic landscape of speech. Acknowledging the intricate relationship between these factors is fundamental to enhancing the accuracy and effectiveness of forensic voice analysis, ultimately contributing to the advancement of the forensic voice comparison field and its applications.

6. Conclusion

This study extends the previous works on Persian speaker identification using various speech rhythm measures. Previous works showed that speech rhythm measures extracted from CV intervals and syllable intensity have the capability to show speaker variability in Persian. In this study, I have taken a step further by conducting a comparative analysis of rhythmic metrics stemming from different categories of speech rhythm measures. Through this investigation, I sought to discern the most effective domain of rhythmic parameters for capturing between-speaker variability. This study has revealed that speech rhythm metrics computed from the durational aspects of speech, especially %V and syllable rate, exhibit slightly superior performance when compared to those derived from the amplitude envelope of the signal. However, it is essential to acknowledge that intensity-based measures, especially those extracted from syllable peaks, have also demonstrated commendable performance in capturing the variations among Persian speakers. The relatively low correlation observed between duration-based measures and intensity-based measures suggest that a combination of these two categories of measures might lead to more robust results in speaker identification tasks. Interestingly, these findings differ from previous studies in which intensity-based measures were found to perform better. I hypothesize that this discrepancy may be attributed to the phonotactic structure of Persian, which plays a significant role in shaping the between-speaker variability observed in this study. Findings

clearly showed that the importance of considering the linguistic characteristics of the target language in the interpretation of acoustic measures for speaker identification cannot be understated.

References

Asadi, H. & Alinezhad, B. (2023). Between-speaker syllable intensity variability in Persian. In *20th International Congress of the Phonetic Sciences (ICPhS)*, 3804-3808, Prague, Czech Republic.

Asadi, H., & Alinezhad, B. (2022). Speech Rhythm Measures: Acoustic Cues for Speaker Identification. *Language Research*, 12(2), 29-49. https://doi.org/10.22059/jolr.2021.304539.666624

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

Bijankhan, M. (2018) Phonology. In A. Sadeghi & P. Shabani-Jadidi (Eds.), *The Oxford Handbook of Persian Linguistics,* 111–141. Oxford: Oxford University Press.

Boersma, P. & Weenink, D. (2013). Praat: Doing Phonetics by Computer. http://www.praat.org, Accessed 13 July 2013.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A. & Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, 5(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

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

Dellwo, V. & Fourcin, A. (2013). Rhythmic characteristics of voice between and within languages. *Travaux Neuchâtelois de Linguistique*, 59: 87–107. https://www.zora.uzh.ch/id/eprint/91230/

Dellwo, V., Leemann, A. & Kolly, M. (2012). Speaker idiosyncratic rhythm features in the speech signal. In *Proceedings of INTERSPEECH*, Portland, USA. https://doi.org/10.5167/uzh-68554

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. & Wagner, P. (2003). Relations between language rhythm and speech rate. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, 471-474. Barcelona, Spain. https://doi.org/10.5167/uzh-111779

Fry, D.B. (1958). Experiments in the perception of stress. *Language and Speech*, 1(2), 126-152. https://doi.org/10.1177/002383095800100207

Garnier, M., Wolfe, J., Henrich, N. & Smith, J. (2008). Interrelationship between vocal effort and vocal tract acoustics: a pilot study. In *Proceedings of INTERSPEECH*, 2302-2305. Brisbane, Australia. http://dx.doi.org/10.21437/Interspeech.2008-588

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, 515-543, Berlin and New York: 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*. Vol 23, 243-273. https://doi.org/10.1558/ijsll.v23i2.30345

He, L., & Dellwo, V. (2014). Speaker idiosyncratic variability of intensity across syllables. In *Proceedings of INTERSPEECH*, 233-237, Singapore. https://doi.org/10.5167/uzh-103024

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

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

Marcus, S. (1981). Acoustic determinants of perceptual center (p-center) location. *Perception and Psychophysics*, 30, 247–256. https://doi.org/10.3758/bf03214280

Moez, Ajili., Bonastre, Jean- François., Rossato, Solange. (2018). Voice comparison and rhythm: Behavioral differences between target and non-target comparisons. In *Proceedings of INTERSPEECH*, 1061-1065. Hyderabad, India. https://doi.org/10.21437/Interspeech.2018-61

Nolan, F. & Asu, E. L. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica*, 66(1–2), 64–77. https://doi.org/10.1159/000208931

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 (2021) R: A Language and Environment for Statistical Computing (version 3.3.3). R Foundation for Statistical Computing. http://www.Rproject.org, Accessed 20 November 2021.

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

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

Sadeghi, V. (2011). Acoustic correlates of lexical stress in Persian. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, 1738-1741. Hong Kong.

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

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. https://doi.org/10.22108/jrl.2021.126261.1535

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

Tilsen, S. & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of theamplitude 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

Wang, Q. (2008). L2 stress perception: The reliance on different acoustic cues. In *Speech Prosody*, 635-638. Campinas, Brazil.

Weingartova, Lenka. (2014). Rhythm metrics for speaker identification in Czeck. *ActaUniversitatis Carolinae Philologica*, 1(10), 33-42.

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? *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*. New York: De Gruyter Mouton.

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