



<http://ui.ac.ir/en>

Journal of Researches in Linguistics

E-ISSN: 2322-3413

Document Type: Research Paper

12(2), 27-50

Received: 30.11.2020 Accepted: 27.01.2021

Acoustic Correlations of Speech Rhythms in Persian Based on Variability of Between-speakers Characteristics

Nafiseh Taghva

*PhD Candidate, Department of Foreign Languages and Linguistics, Faculty of Literature and Humanities, Shiraz University, Shiraz, Iran
taghvanafiseh@gmail.com*

Amirsaeid Moloodi *

*Assistant Professor, Department of Foreign Languages and Linguistics, Faculty of Literature and Humanities, Shiraz University, Shiraz, Iran
amirsaeid.moloodi@shirazu.ac.ir*

Vahideh Abolhasanizadeh

*Assistant Professor, Department of Foreign Languages, Faculty of Literature and Humanities, Shahid Bahonar University of Kerman, Kerman, Iran
vahidehabolhasani@yahoo.com*

Abstract

The durational variability of phonetic intervals is considered as one of the properties of speech rhythm. These intervals include segmental, vowel, consonantal, vocalic, intervocalic, voiced, unvoiced, syllable, and syllable peak intervals. The durational variability measure for some of these intervals, such as vowel, consonantal, vocalic, intervocalic intervals, determines the classification of languages based on their rhythm. Besides, in some cases, the speaker identification is only possible through the person's voice. The segmental and suprasegmental properties of a language can be used to identify the speaker. In this study, the acoustic correlations of Persian speech rhythm in a reading text are calculated by various durational measures. Also, the between-speaker rhythmic variability is considered to find the best rhythmic measures for Persian speaker identification. The results confirm that Persian is near to the syllable-based languages. Moreover, the results from segmental and supra-segmental consideration demonstrate significant between-speaker variability in Persian. Among phonetic intervals, nPVI-VC and V% (percentage of vocalic intervals) best discriminate between-speaker variability in Persian.

Keywords: Speech rhythm, Durational variability, Acoustic correlations, Between-speaker variability, Rhythmic measures

Introduction

The rhythmic properties of languages have been one of the controversial issues in linguistics in recent studies. Early studies on the classification of different rhythm types in language focused on the syllable and foot durations in which the speech rhythm was defined in terms of isochrony (Abercrombie, 1967; Lloyd James, 1940; Pike, 1945). They believed that Germanic languages had a simultaneous foot; that is why they were called "stress-timed" languages. It was also believed that Romance languages had similar syllables, so they were called "syllable-timed" languages.

However, such approaches can be easily violated in spontaneous speech (Dauer, 1987). Dauer argued that languages with different rhythms also differ in syllable weight and vowel reduction. Stress-timed languages usually have a complex syllable structure and a higher rate of vowel reduction. Ramus, Nesper, and Mahler (1999) examined this hypothesis by measuring the standard deviation of vocalic (ΔV) and consonantal intervals (ΔC) as well as the percentage of vocalic intervals (%V) for each sentence. Then Grabe and Low



This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Doi: [10.22108/jrl.2021.126261.1535](https://doi.org/10.22108/jrl.2021.126261.1535)

(2002) introduced the pairwise variability index (PVI) to measure durational variability between sequences of vocalic and consonantal intervals (nPVI-V and rPVI-C). Besides, Dellwo (2010) proposed other normalization methods for the speech rate, including the coefficient of variation (Varco) and the natural logarithm. Arvantini (2012) introduced amplitude envelope-based rhythm measure based on which she investigated the repetition of acoustic information rather than segmental units.

Another application of rhythm measures is in forensic sciences. As the speakers of a typical language have different voices, one of the aspects of forensic sciences is considering different voices between different speakers (Rose, 2004). Dellwo, Leeman, and Kolly (2015) cited three reasons for this diversity: the nature of the articulatory system, linguistic factors, and prosodic factors. Thus, we are faced with a variety of speakers' voices, which is called between-speaker variability. Recently, evidence from various datasets suggested that measuring rhythm based on different phonetic intervals could vary significantly in a language as a function of speakers (Leeman, Kolly and Dellwo, 2014; Wiget et al., 2010; Yoon, 2010).

Materials & Methods

Ten native speakers of contemporary standard Persian (5 men and 5 women) read a Persian text from the book "North Wind and the Sun" in the acoustic room at Shiraz University. The Persian version of this story contains seven complex sentences. Therefore, the dataset of this test comprised 70 tokens (10 speakers × 7 sentences).

This research corpus was acoustically analyzed in Praat (v 6.1.09, in which six tiers of TextGrids were created. In the first tier, the offset and onset of each segment were determined manually and transcribed according to IPA. Then in the second tier, the vowels and consonants were tagged. In the third tier, the vowel and consonants intervals were labeled based on the number of consonants and vowels. In the fourth tier, the vocalic and consonantal intervals were determined. In the fifth layer, the boundary between the existing syllables was tagged manually. Finally, in the sixth tier, the peak of each syllable was automatically identified according to the principle of sonority by a script written by Dellwo¹. Then, speech rhythm measures from previous works were used. All measures were automatically calculated using the existing script written by Dellwo.

The mean and standard deviation of the results obtained from the scripts was calculated in SPSS (v 23) to classify the Persian language rhythm. Moreover, Pearson correlation and one-way ANOVA test were used to distinguish the most robust between-speaker measure.

Discussion of Results and Conclusions

The results confirm that Persian is near to the syllable-based languages. Besides, it was revealed that seven metrics are statistically significant (Speech rate (syl/s), VarcoC, %V, nPVI-V, nPVI-VC, $\Delta C(\ln)$, $\Delta \text{Peak}(\ln)$). Based on the present study results, nPVI-VC and V% are the most powerful measures to show the between-speakers variability in Persian.

References

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh, UK: Edinburgh University Press.
- Abolhasani Zadeh, V., Taghva, N., Sharifi Moghadam, A., & Golshan, A. (2013). The effect of Farsi language rhythm on Iranian traditional non metric instrumental music. *Aion-linguistica*, 2, 11-19.
- Abolhasani Zadeh, V., & Taghva, N. (2019). Study and comparison of Persian language and English language rhythm. *Bilingual Interpretive-Comparative Journal of Language and Literature*, 4(6), 1-10. [In Persian]
- Adams, C. (1979). *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton.
- Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40, 351-373.
- Asadi, H., Nourbakhsh, M., He, L., Pelligrino, E., & Dellwo, V. (2018). Between-speaker rhythmic variability is not dependent on language rhythm, as evidence from Persian reveals. *The International Journal of Speech, Language and the Law*, 25(2), 151-172.


¹ <https://www.cl.uzh.ch/de/people/team/phonetics/vdellw.html>

- Barbosa, P., & Bailly, G. (1994). Characterisation of rhythmic patterns for text-to-speech synthesis. *Speech Communication*, 15, 127-137.
- Bertrán, A. P. (1999). Prosodic typology: On the dichotomy between stress-timed and syllable-timed languages. *Language Design*, 2, 103-131.
- Bosch, L., & Sebastián-Gallés, N. (1997). The role of prosody in infants' native language discrimination abilities: The case of two phonologically close languages. Paper presented in *EUROSPEECH-1997*, Rhodes, Greece, 231-234.
- Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51-62.
- Dauer, R. (1987). *Phonetic and phonological components of language rhythm*. Paper presented at International Congress of Phonetic Science (ICPhS) XI, Tallinn, Estonia.
- Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for delta C. In P. Karnowski & I. Sziget (Eds.), *Language and language-processing* (pp. 231-241). Frankfurt am Main: Peter Lang.
- Dellwo, V., Huckvale, M., & Ashby, M. (2007). How is individuality expressed in voice? An introduction to speech production & description for speaker classification. In C. Müller (Ed.), *Speaker Classification I* (pp. 1-20). Berlin: Springer.
- Dellwo, V. (2009). Choosing the right rate normalization method for measurements of speech rhythm. In S. Schmid, M. Schwarzenbach, and D. Studer (Eds.), *La Dimensione Temporale del Parlato: Atti del 5° Convegno Nazionale AISV 2009* (pp. 13-32). Torriana, Italy: EDK Editore.
- 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. (2012). Speaker idiosyncratic rhythmic features in the speech signal. *Proceedings of Interspeech*, Portland (USA).
- Dellwo, V. & Fourcin, A. (2013). Rhythmic characteristics of voice between and within languages. *Revue Tranel (Travaux neuchâtelois de linguistique)*, 59, 87-107.
- Dellwo, V., Leemann, A., & Kolly, M-J. (2015). Rhythmic variability between speakers: articulatory, prosodic, and linguistic factors. *Journal of the Acoustical Society of America*, 137, 1513-1528.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and rhythm class hypothesis. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology 7* (pp. 514-546). Berlin, Germany: Mouton de Gruyter.
- Haghshenas, A. (2004). *Phonetics*. Tehran: Agah. [In Persian]
- Liss, J. M., White, L., Mattys, S., Lansford, K., Lotto, K. A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech Language and Hearing Research*, 52, 1334-1352.
- He, L. (2017). Speaker idiosyncratic intensity variability in the speech signal. [Doctoral dissertation]. Zurich University.
- He, L., & Dellwo, V. (2016). The role of syllable intensity in between speaker rhythmic variability. *The International Journal of Speech*, 23(2), 243-273.
- Jessen, M. (2008). Forensic Phonetics. *Language and Linguistics Compass*, 2(4), 671-711.
- Lee, C. S., & Todd, N. P. M. (2004). Towards an auditory account of speech rhythm: application of a model of the auditory "primal sketch" to two multilanguage corpora. *Cognition*, 93, 225-254.
- 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.
- Leong, V., Stone, M. A., Turner, R. E., & Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *Journal of the Acoustical Society of America*, 136, 366-381.
- Leong, V., & Goswami, U. (2014). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience*, 8(96), 1-14.
- Lloyd James, A. (1940). *Speech Signals in Telephony*. London: Sir Isaac Pitman & Sons.
- Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *Journal of Acoustic Society of America*, 129, 3258-3270.
- McDougall, K. (2004). Speaker-Specific Formant Dynamics: An Experiment on Australian English /aI/. *International Journal of Speech, Language and the Law*, 11, 103-130.

- McDougall, K. (2006). Dynamic features of speech and the characterisation of speakers: Towards a new approach using formant frequencies. *International Journal of Speech, Language and the Law*, 13, 89–126.
- Milne, B. L. (1955). *Modern speech rhythm exercises*. London: Macmillan.
- Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, 83, 405–408.
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756–766.
- Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effect of rhythm and familiarity. *Journal of Memory and Language*, 43, 1–19.
- Nespor, I. (1990). On the rhythm parameter in phonology. In I. Roca (Ed.), *Logical Issues in Language Acquisition* (pp. 157–195). Dordrecht: Foris.
- Nolan, F. (2002). Intonation in speaker identification: an experiment on pitch alignment Features. *Forensic Linguistics*, 9, 1–21.
- Nolan, F., & Jeon, H. S. (2015). Speech rhythm: a metaphor? *Philosophical Transactions of the Royal Society B*, 369(1658), 1–11.
- O'Dell, M. L., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)* 1075–1078. San Francisco, USA.
- Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. (2012). Measuring child rhythm. *Language and Speech*, 55(2), 203–229.
- Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Pointon, G. E. (1980). Is Spanish really syllable-timed? *Journal of Phonetics*, 8, 293–304.
- Polyanskaya, L., & Ordin, M. (2015) Acquisition of speech rhythm in first language. *Journal of the Acoustical Society of America*, 138, 199–204.
- Ramus, F., & Mehler, J. (1999). Language identification with supra-segmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*. 105, 512–521.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top Tamarin monkeys. *Science*, 288, 349–351.
- Roach, P. (1982). On the distinction between “stress-timed” and “syllable-timed” languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 73–79). London: Edwards Arnold.
- Rose, P. (1996). Speaker verification under realistic forensic conditions. In P. McCormack and A. Russell (Eds), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology, Australian Speech Science and Technology Association, Canberra* (pp. 109–114).
- Rose, P. (2004). *Forensic Speaker Identification*. Boca Raton, FL: CRC Press.
- Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *Journal of Acoustic Society of America*, 124, 34–39.
- Tilsen, S. & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of Acoustic Society of America*, 134, 628–639.
- 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 Acoustic Society of America*, 127, 1559–1569.
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35, 501–522.
- White, L., Liss, J., & Dellwo, V. (2010) Assessment of rhythm. In A. Lowit & R. D. Kent (Eds.) *Assessment of motor speech disorders* (pp. 312–352). San Diego: Plural Publishing.
- Yoon, T. J. (2010). Capturing inter-speaker invariance using statistical measures of speech rhythm. *Proceedings of Speech Prosody 5, Chicago/Ill.* 1–4.

بررسی همبسته‌های آکوستیکی ریتم گفتار زبان فارسی با تمرکز بر تمایزات میان‌گوینده

* نفیسه تقوی

** امیرسعید مولودی 

*** وحیده ابوالحسنی زاده

چکیده

یکی از ویژگی‌های ریتم گفتار تغییرپذیری دیرش سطوح مختلف فواصل آوایی، از جمله فواصل واحدهای زبانی، فواصل همخوان و واکه، فواصل همخوانی و واکه‌ای، فواصل واکه‌داری و بی‌واکی، فواصل هجایی و فواصل قلّه هجا است. محاسبه تغییرپذیری دیرش برخی از این فواصل، از جمله فواصل همخوان و واکه و فواصل همخوانی و واکه‌ای، طبقه‌بندی زبان‌ها را براساس ریتم آنها مشخص می‌کند. علاوه بر این، تجارب روزمره ما نشان می‌دهد که در مواقعی شناسایی گوینده تنها از طریق صدای آن فرد امکان‌پذیر است. ویژگی‌های زنجیری و زبرزنجیری ریتم گفتار از جمله مواردی هستند که در شناسایی گوینده می‌توان از آنها استفاده کرد. در این مطالعه، همبسته‌های آکوستیکی ریتم گفتار زبان فارسی در یک متن خوانشی با محاسبه سنجه‌های مختلف دیرش فواصل ذکر شده بررسی می‌شوند. همچنین، به منظور یافتن بهترین سنجه ریتم شناسایی گوینده، تغییرپذیری ریتم میان‌گوینده داده‌های این پژوهش بررسی می‌شوند. نتایج به دست آمده از داده‌های این پژوهش، نظریه هجازمانی بودن زبان فارسی را تأیید می‌کند. علاوه بر این، نتایج بررسی ویژگی‌های زنجیری و زبرزنجیری ریتم داده‌های پژوهش حاضر حاکی از تغییرپذیری معنی‌دار میان‌گوینده گویشوران زبان فارسی است که از میان آنها سنجه مربوط به تغییرپذیری دوتایی فواصل همخوانی-واکه‌ای (nPVI_CV) و بعد از آن، سرعت هجا و درصد واکه‌ای بودن متغیرهای قوی‌تری برای نشان‌دادن تمایزات میان‌گوینده هستند.

کلیدواژه‌ها

ریتم گفتار، تغییرپذیری دیرش، همبسته‌های آکوستیک، میان‌گوینده، سنجه‌های ریتم



۱. مقدمه

بررسی ویژگی‌های ریتم زبان‌های مختلف از جمله موضوعات بحث‌برانگیز زبان‌شناسی در پژوهش‌های اخیر است. زمان دقیق وارد شدن بحث ریتم در حوزه گفتار به‌طور دقیق مشخص نیست؛ اما می‌توان تصور کرد که ابتدا به حوزه موسیقی وارد شده است. پژوهش‌های اولیه درباره ریتم گفتار تنها به بررسی ریتم در شعر می‌پرداختند؛ با این فرض که ریتم در گفتار عادی تظاهر پیدا نمی‌کند؛ اما اکنون مطالعات ریتم به تمام شاخه‌های زبان‌شناسی راه پیدا کرده است و در آواشناسی نیز نتایج جالبی را به خود اختصاص داده است. اکنون بالغ بر هشت دهه است که زبان‌ها را براساس ریتم آنها تقسیم‌بندی می‌کنند. در سال‌های اخیر نیز تلاش‌های مهمی در راستای بررسی ریتم گفتار در ارتباطات زبانی به‌خصوص در آواشناسی قضایی^۱ انجام شده است.

در اولین برخورد ممکن است موضوع ریتم زبان به‌مثابه ویژگی زبانی کم‌اهمیتی تلقی شود؛ زیرا ریتم اثر مستقیمی در واج‌شناسی و در فرایند ارتباط زبانی ندارد. به‌علاوه، برخلاف دیگر ویژگی‌های نوایی مانند آهنگ، نمی‌توانیم معنای دو عبارت، دو جمله یا دیگر واحدهای گفتار را براساس ریتم آنها متمایز کنیم. همچنین، ریتم از جنبه فرازبانی نیز ویژگی بارزی ندارد؛ برای مثال، نمی‌توان تنها براساس ریتم، دو احساس متفاوت را بیان کرد (Dellwo, 2010). بنا به دلایل ذکر شده، این پرسش مطرح می‌شود که اهمیت ریتم گفتار^۲ در چیست؟

تاکنون مطالعات ریتم زبان در زمینه آکوستیک حداقل در چهار حوزه اصلی نتایج مفیدی را عرضه کرده‌اند:

۱. گفتار کودک و یادگیری زبان دوم (Adams, 1979; Polyanskaya & Ordin, 2015; Milne, 1955;)

۲. فناوری گفتار (Barbosa and Bailly, 1994)

۳. آسیب‌شناسی گفتار^۳ (Leong & Goswami, 2014; White, Liss & Dellwo, 2010; Liss et al., 2009)

۴. شناسایی گوینده (Dellwo, Leeman & Kolly, 2012; Leeman, Kolly & Dellwo, 2014;)

در این پژوهش‌ها، هنوز توافقی در تعریف ریتم گفتار حاصل نشده است؛ اما دو جنبه زمان‌بندی و برجستگی طیفی، مکانیسم‌های قابل قبولی هستند که اساس این پدیده را تشکیل می‌دهند (Nolan & Jeon, 2015). با وجود این، جالب است که تقریباً تمام مدل‌های ریتم به‌نوعی براساس ویژگی‌های دیرشی گفتار بنا شده‌اند.

در پژوهش حاضر برآنیم تا به بررسی همبسته‌های دیرشی ریتم زبان فارسی در سطوح مختلف فواصل آوایی، از جمله فواصل واحدهای زبانی، فواصل همخوانی و واکه‌ای، فواصل واکه‌داری و بی‌واکی، فواصل هجایی و فواصل قله هجا پردازیم و سنجه‌های مناسب تشخیص میان‌گوینده^۴ را در زبان فارسی معرفی کنیم؛ با این هدف که درک نظری ما درباره ویژگی‌های ریتمیک فردی گوینده ارتقا یابد. انتظار می‌رود یافته‌های پژوهش حاضر در آینده در محیط تشخیص گوینده اعمال شود. پرسش‌های این پژوهش عبارت‌اند از:

الف. همبسته‌های آکوستیک^۵ زبان فارسی با توجه به سنجه‌های موجود چگونه عمل می‌کنند؟

1. Forensic phonetics

2. speech rhythm

3. speech pathology

4. between-speakers

5. acoustic correlations

- ب. آیا تفاوت‌های میان‌گوینده در ویژگی‌های زمانی زنجیری و زیرزنجیری زبان فارسی وجود دارد؟
 ج. کدام سنجه‌ی زمانی زنجیری و زیرزنجیری تفاوت‌های میان‌گوینده را بیشتر منعکس می‌کند؟

۲. پیشینه پژوهش

گمانه‌زنی‌های اولیه درباره‌ی طبقه‌بندی ریتم زبان‌های مختلف متمرکز بر مدت‌زمان هجاها و پایه‌ها^۱ هستند که در آنها ریتم گفتار از نظر جنبه‌های هم‌زمانی^۲ در نظر گرفته می‌شود. [ابر کرومبای](#)^۳ (1967)، [پایک](#)^۴ (1945) و [لیودجیمز](#)^۵ (1940) معتقدند زبان‌های ژرمنی^۶ از قبیل انگلیسی، آلمانی و هلندی دارای پایه‌های هم‌زمان هستند؛ یعنی دیرش فواصل بین تکیه‌ها در گفتار آنها باهم برابر است؛ از این رو، به آنها «تکیه‌زمانی»^۷ گفته می‌شود. همچنین، آنها بر این باورند که زبان‌های رومی تبار^۸ مانند فرانسه، ایتالیایی و اسپانیایی دارای هجاهای همسان هستند؛ یعنی مدت زمان هجایی برابر دارند؛ از این رو، زبان‌های «هجازمانی»^۹ خوانده می‌شوند. به‌علاوه در زبان‌های تکیه‌زمانی مانند انگلیسی، می‌توان تعدادی از هجاهای بدون تکیه را بین هجاهای تکیه‌دار قرارداد، بدون اینکه دیرش پایه‌ها به‌طور چشمگیری تغییر کند؛ اما به نظر می‌رسد برای زبان‌های هجازمانی مانند زبان فرانسه هجاهای تکیه‌دار یا بی‌تکیه، مدت‌زمان مشابهی دارند.

با وجود این، چنین رویکردی را می‌توان به راحتی در گفتار بداهه^{۱۰} نقض کرد؛ زیرا سنجه‌های اولیه‌ی اندازه‌گیری دیرش پایه و هجا نتوانسته چنین الگوهای یکسانی را در زمان‌بندی نشان دهند ([Dauer, 1983; Pointon, 1980; Bertrán, 1999](#)); تا آنجا که [نسپر](#)^{۱۱} (1990) استفاده از اصطلاحات تکیه‌زمانی و هجازمانی را نیز نمی‌پذیرد. در هر صورت، نتایج تعداد بی‌شماری از پژوهش‌ها تأیید می‌کنند که زبان‌هایی که از نظر ریتم باهم فرق دارند، در بزرگ‌سالان، نوزادان و حتی حیوانات از نظر ادراک نیز دارای تفاوت درخور توجه هستند ([Ramus et al., 2000; Nazi, Jusczyk & Johnson, 2000](#)); [Ramus & Mehler, 1999; Ramus, Nasper & Mehler, 1999; Nazzi, Bertoncini & Mehler, 1998; Bosch & Sebastián-Gallés, 1997](#).

در ادامه، پژوهشگران شروع به بررسی ویژگی‌های دیرش سایر واحدهای گفتار می‌کنند و معیارهای دیگری را برای تعیین کمیت تفاوت‌های ریتم بین زبان‌ها پیشنهاد می‌دهند. [دائر](#)^{۱۲} (1983, 1987) استدلال می‌کند زبان‌هایی که ریتم متفاوتی دارند از نظر وزن هجا و کاهش واکه‌ای نیز متفاوت هستند؛ به این صورت که زبان‌های تکیه‌زمانی معمولاً دارای ساختار هجایی پیچیده‌تر

1. feet

2. isocrony

3. D. Abercrombie

4. K. Pike

5. A. Lloyd James

6. Germanic

7. stress-based

8. Romance

9. syllable-based

10. spontaneous

11. I. Nespor

12. R. Dauer

و میزان بالاتری از کاهش واکه‌ای هستند؛ در صورتی که زبان‌های هجازمانی دارای ساختار هجایی ساده‌تر و میزان کمتر کاهش واکه‌ای هستند.

راموس، نسپر و مهلر^۱ (1999) به بررسی این فرضیه با اندازه‌گیری انحراف معیار دیرش فواصل همخوانی و واکه‌ای (ΔC) و ΔV) و همچنین، درصد واکه‌ای بودن ($\%V$) هر جمله می‌پردازند. سپس، گریب و لو^۲ (2002) شاخص تغییرپذیری دوتایی (PVI) را برای اندازه‌گیری تغییرپذیری دیرش فواصل واکه‌ای و همخوانی متوالی (nPVI-V و rPVI-C) معرفی می‌کنند. علاوه بر این، دلوو^۳ (2001, 2009) چند روش نرمال‌سازی مانند ضریب تغییر وارکو (varco) و لگاریتم طبیعی (تبدیل لگاریتمی)،^۴ که به اختصار ln نامیده می‌شود، را به منظور خنثی‌سازی سرعت گفتار ارائه می‌دهد.

در همین حین، رویکردهای دیگری در مطالعه ریتیم گفتار ظهور می‌کنند؛ مانند: مدل نوسانگر^۵ همراه (O'Dell & Nieminen, 1999)، مدل فاز مدولاسیون دامنه^۶ (Leong et al., 2014)، مدل طرح اولیه شنوایی^۷ (Lee & Todd, 2004)، و مدل طیفی از دامنه^۸ (Tilsen & Johnson, 2008). از میان این پژوهش‌ها، تنها مدل طیفی از دامنه (Tilsen & Johnson, 2008) در حوزه مورد بررسی پژوهش حاضر است که به توضیح آن خواهیم پرداخت. آروینینی^۹ (2012) روش دیگری برای اندازه‌گیری ریتیم گفتار معرفی می‌کند که سنج‌های ریتیم بسته دامنه^{۱۰} نام دارد. در این روش کمتر به خصوصیت‌های مربوط به واحدهای زبانی^{۱۱} پرداخته می‌شود و بیشتر به تکرار اطلاعات صوت‌شناختی مانند دامنه ضربه‌های حاصل از دامنه گفتار می‌پردازد. این رویکرد بر این نظریه استوار است که هجا دارای یک مرکز ادراکی^{۱۲} است که همبسته‌های صوتی آن شامل ترکیب پیچیده‌ای از قله‌های^{۱۳} بسته دامنه، حرکات فرکانس پایه^{۱۴} و کیفیت‌های مربوط به واحدهای زبانی هستند (Morton et al., 1976). همچنین، ویژگی‌های زمانی ضرب‌های هجا از نظر خصوصیت‌های درک ریتیم گفتار برجسته‌تر از مرزهای هجایی یا مربوط به واحدهای زبانی هستند (Tilsen & Arvintini, 2013; Tilsen & Johnson, 2008).

از دیگر حوزه‌هایی که سنج‌های ریتیم در آن بررسی می‌شود مطالعات علوم قضایی است. علوم قضایی در گفتار یک حوزه میان‌رشته‌ای است که پیوندی را میان نظریه‌ها و روش‌های آواشناسی اجتماعی، زبان‌شناسی اجتماعی، فیزیک گفتار، فناوری و پردازش گفتار و فعالیت‌های عملی در بافت‌های دادگاهی برقرار می‌کند (Jessen, 2008). یکی از جنبه‌های آواشناسی قضایی

1. F. Ramus, M. Nespor, & J. Mehler

2. E. Grabe & E. L. Low

3. V. Dellwo

4. logarithmic transform

5. the coupled-oscillator model

6. amplitude modulation phase model

7. auditory primal sketch model

8. amplitude modulation phase model

9. A. Arvaniti

10. amplitude envelope-based rhythm measure

11. segmental

12. p-center

13. peaks

14. F0

بررسی تغییرات میان‌گوینده و درون‌گوینده^۱ در گفتار و آواهای تولیدشده توسط انسان است؛ زیرا گویشوران مختلف یک زبان صداهای متفاوتی دارند (Rose, 2004). بنابراین، ما با تنوع میان‌گویشوران مواجه هستیم که آن را تنوعات میان‌گوینده می‌نامند. اگرچه یک فرض کلی وجود دارد که افراد مختلف دارای صداهای مختلفی هستند، درک این نکته ضروری است که حتی صدای یک گویشور نیز در موقعیت‌های مختلف متفاوت خواهد بود. رز (1996) بیان می‌کند که این یک حقیقت آوایی است که هیچ‌کس تا به حال یک گفتار را دقیقاً یکسان تولید نکرده است. رز (2004) چنین تنوعاتی را درون‌گوینده می‌نامد.

دلوو، لیمن و کولی^۲ (2015) سه دلیل برای این تنوعات بیان می‌کنند، که عبارت‌اند از: طبیعت اندام‌های تولید گفتار، عوامل زبان‌شناختی و عوامل نوایی. صدای هر گویشوری منحصر به فرد^۳ است؛ به همین دلیل است که معمولاً شنوندگان می‌توانند گویندگان را براساس صدای آنها تا حد نسبتاً دقیقی تشخیص دهند. ویژگی‌های متغیر زمان گفتار مانند کیفیت صدا یا ویژگی‌های دامنه فرکانس مانند فرکانس پایه و هارمونیک‌های آن، فرکانس سازه‌های^۴ واکه‌ها و ویژگی‌های کلی بسته طیفی^۵ نقش مهمی در شناسایی گویندگان انسانی و خودکار دارند (McDougall, 2004 Dellwo, Huckvale & Ashby, 2007; Nolan, 2002; Nolan, 2006;). دلیل منطقی آن این است که پارامترهای دامنه فرکانس تا حد زیادی نتیجه ویژگی‌های فیزیولوژیکی فردی اندام‌های تولید گفتار هر گوینده است؛ برای مثال، اندازه چاکنای و حنجره به ترتیب با میزان رزنانس چاکنای و میزان لرزش تارآواها ارتباط دارد (Dellwo et al., 2015).

در حال حاضر شواهد موجود از مجموعه داده‌های مختلف نشان می‌دهد که اندازه‌گیری ریتم براساس فواصل واکه‌ای و همخوانی می‌تواند در یک زبان به‌عنوان تابعی از گویشور به‌طوردرخور توجهی متفاوت باشد (Leeman, Kolly & Dellwo, 2010; Loukina et al., 2012; Arvintini, 2012; Wiget, et al., 2010; Yoon, 2010 2014;). ویگت و همکاران (2010) با بررسی ریتم گفتار زبان انگلیسی نشان می‌دهند که در زبان انگلیسی %V و VarcoV از تغییرپذیری میان‌گوینده بیشتری نسبت به nPVI برخوردار هستند. یوون (2010) همین نتایج را برای گفتار بداهه انگلیسی‌زبانان شمال آمریکا به دست می‌آورد. همچنین، اطلاعات گوینده‌ویژه در دیرش فواصل واکه‌داری و بی‌واکی در پژوهش‌های لیمن، کولی و دلوو^۶ (2014) و دلوو و فورسین^۷ (2013) در زبان آلمانی سوییس گزارش می‌شود.

در سال‌های اخیر پژوهش‌هایی در زمینه ویژگی‌های آکوستیک ریتم زبان فارسی انجام شده است. از جمله ابوالحسنی‌زاده و همکاران (2013) و ابوالحسنی‌زاده و تقوی (۱۳۹۸) که با بررسی تغییرپذیری دیرش^۸ فواصل واکه‌ای و میان‌واکه‌ای (nPVI-V و rPVI-C) زبان فارسی را جزو زبان‌های هجازمانی قرار می‌دهند. پژوهش دیگری که درباره همبسته‌های آکوستیک ریتم زبان فارسی انجام شده است، اثر اسدی و همکاران (2018) است. در این پژوهش تفاوت‌های درون‌گوینده و بین‌گوینده فارسی در

1. within-speaker

2. V.Dellwo, A. Leemann & M. J. Kolly

3. speaker-idiosyncratic

4. formants

5. spectral envelope

6. A. Leemann, M.-J. Kolly & V. Dellwo

7. V. Dellwo & A. Fourcin

8. durational variability

یک متن خوانشی با ۵ سرعت متفاوت در ده گویشور زبان فارسی معیار بررسی می‌شود. یافته‌های این پژوهش بیانگر این است که در ΔV [ln], ΔC [ln], nPVI-V، تفاوت معنی‌دار میان گوینده وجود دارد؛ همچنین، تأکید می‌کنند که $V\%$ به بهترین شکل تمایزهای میان گوینده را منعکس می‌کند.

در پژوهش [ابوالحسنی‌زاده و همکاران \(2013\)](#) و [ابوالحسنی‌زاده و تقوی \(۱۳۹۸\)](#) با استفاده از داده‌هایی که از خوانش جمله‌ها به دست آمده است، فقط به بررسی تغییرپذیری دیرش فواصل واکه‌ای و میان واکه‌ای (nPVI-V و rPVI-C) زبان فارسی پرداخته شده است. در پژوهش [اسدی و همکاران \(2018\)](#) نیز تمایزات میان گوینده در سه سطح واحدهای زبانی، همخوانی و واکه‌ای و هجایی در سرعت‌های مختلف گفتار بررسی شده است.

۳. مبانی نظری

سنجه‌های ریتم گفتار به دو دسته تقسیم‌بندی می‌شوند:

الف. سنجه‌های وابسته به دیرش فواصل گفتار (فاصله‌مبنا)^۱، که خود به سه دسته تقسیم می‌شود:

- سنجه‌های ضریب دیرش همخوانی و واکه‌ای، مانند درصد واکه‌ای بودن گفتار ([Ramus, Nesper & Mehler, 1999](#)). این سنجه‌ها به بررسی درصد واکه‌ای بودن گفتار ($V\%$) و درصد واکدار بودن گفتار ($VO\%$) می‌پردازند.
- سنجه‌های تغییرپذیری دیرش همخوانی و واکه‌ای که شامل ΔC و ΔV می‌شوند و به بررسی انحراف معیار دیرش فواصل واکه‌ای و همخوانی ([Ramus, Nesper & Mehler, 1999](#)) می‌پردازند. همچنین، میانگین تفاوت دیرشی بین فواصل همخوانی و واکه‌ای متوالی ([Grabe & Low, 2002](#)) که شاخص تغییرپذیری دوتایی (PVI) نام دارد و برای اندازه‌گیری تغییرپذیری دیرش فواصل واکه‌ای و همخوانی متوالی استفاده می‌شوند.
- سنجه‌های تغییرپذیری واکه‌ای و همخوانی نرمال‌شده سرعت، مانند ضریب تغییر دیرش فواصل واکه‌ای و همخوانی ([Dellwo, 2009; Dellwo, 2006](#)) که ضریب تغییر وارکو (varco) و لگاریتم طبیعی (تبدیل لگاریتمی) نامیده می‌شود و برای خنثی‌سازی سرعت گفتار به کار می‌روند. دیگر سنجه نرمال‌شده شاخص تغییرپذیری دوتایی نرمال‌شده (nPVI) است که به بررسی میانگین تفاوت میان فواصل واکه‌ای و همخوانی متناسب با دیرش فواصل دوتایی نرمال‌شده می‌پردازد ([Grabe & Low, 2002](#)).

ب. سنجه‌های ریتم بسته دامنه که در آن به خصوصیت‌های زمانی هجا پرداخته می‌شود ([Arvintini, 2012](#)) و به تکرار اطلاعات صوت‌شناختی مانند دامنه ضربه‌های حاصل از دامنه گفتار می‌پردازد.

در پژوهش حاضر همبسته‌های دیرشی ریتم زبان فارسی در سطوح مختلف فواصل آوایی، از جمله فواصل واحدهای زبانی، فواصل همخوانی و واکه‌ای، فواصل واک‌داری و بی‌واکی، فواصل هجایی و فواصل قلّه هجا بررسی می‌شوند و سنجه‌های مناسب تشخیص میان گوینده در زبان فارسی معرفی می‌شوند.

¹. interval-based

در پژوهش حاضر همبسته‌های دیرشی ریتم زبان فارسی در سطوح مختلف فواصل آوایی با توجه به سنجه‌های ذکر شده بررسی می‌شوند.

۴. روش پژوهش

با توجه به اینکه روش پژوهش حاضر آماری و آزمایشگاهی است، در این قسمت به معرفی گویشوران این پژوهش، چگونگی جمع‌آوری داده‌ها، ویرایش داده‌ها، سنجه‌های زمانی مورد استفاده و تحلیل‌های آماری پژوهش پرداخته می‌شود.

۴-۱. گویشوران

شرکت کنندگان این پژوهش شامل ده گویشور (۵ مرد و ۵ زن) زبان فارسی هستند که همگی با لهجه معیار صحبت می‌کنند و هیچ‌یک از آنها لهجه محلی و نارسایی گفتاری و شنوایی ندارند. بازه سنی گویشوران بین ۲۳ تا ۴۰ سال با میانگین سنی ۳۲/۱۸ و SD: ۷۹.۴ است. از نظر گروه اجتماعی تلاش شده است که گروه همسانی از گویشوران انتخاب شوند. به همین دلیل تمامی گویشوران از استادان و دانشجویان دانشگاه شیراز هستند.

۴-۲. داده‌ها

داده‌های پژوهش حاضر در اتاق آکوستیک دانشگاه شیراز با استفاده از ریکوردر zoom h4 با مشخصات ۹۶-bit/۲۴ kHz و قابلیت خشه‌گیری ۱۲۰-dBu EIN، ضبط شد. به صورتی که میکروفون به وسیله یک پایه به فاصله بیست سانتی‌متری از دهان شرکت کنندگان به صورت مورب قرار گرفت. از آنجا که این پژوهش بر تفاوت میان‌گوینده تمرکز دارد برای به حداقل رساندن تنوعات گفتاری سعی شد شرایط یکسانی برای تمامی شرکت کنندگان ایجاد شود و همه شرکت کنندگان یک متن واحد را با سرعت گفتار معمولی قرائت کنند. در همین راستا، متن فارسی داستان «باد شمال و خورشید» برای این پژوهش انتخاب شد. نمونه ضبط‌شده این متن به زبان‌های مختلف در سایت IPA^۱ نیز موجود است و تاکنون کارهای متعددی در سراسر دنیا با استفاده از این متن انجام شده است^۲ از جمله پژوهش آروینتینی (2012) و گریب و لو (2002). نسخه فارسی این متن شامل هفت جمله پیچیده است. تنها تفاوتی که متن این پژوهش با صدای ضبط‌شده زبان فارسی موجود در سایت IPA دارد، وجود واژه «لباده» است که با توجه به نام‌نوس بودن این واژه برای گویشوران معاصر، به «پالتو» تغییر پیدا کرد. متن داستان پیش از مصاحبه در اختیار گویشوران قرار می‌گرفت تا آن را مطالعه کنند. در زمان مصاحبه نیز به آنها گفته می‌شد که هر چند مرتبه که لازم بدانند می‌توانند از روی متن بخوانند تا به وضعیت دلخواه خود برسند؛ اگر در خواندن جمله‌ای دچار اشتباه می‌شدند از آنها خواسته می‌شد آن جمله را تکرار کنند.

۴-۲-۱. ویرایش داده‌ها

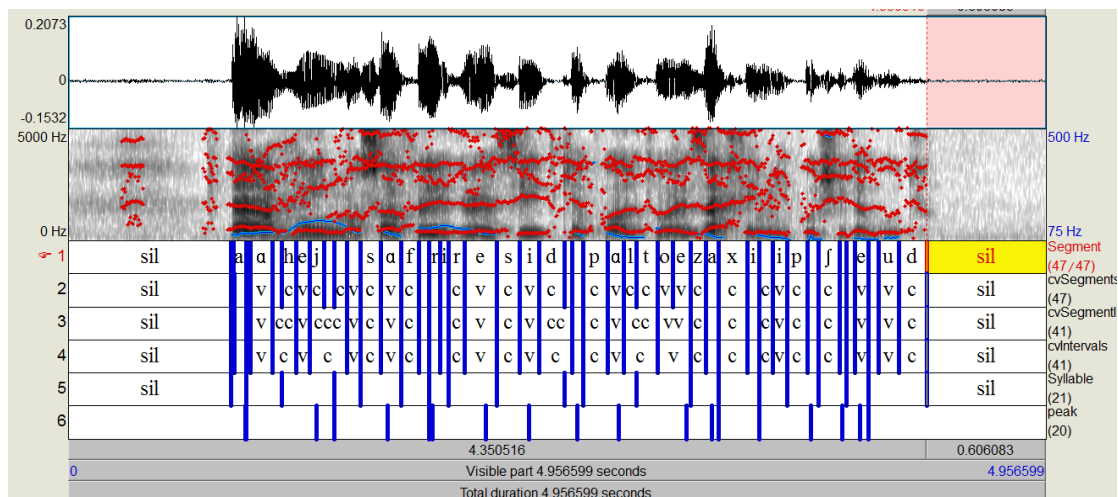
¹. International Phonetics Association

². <https://www.internationalphoneticassociation.org/content/ipa-handbook-downloads>

داده‌های پژوهش در نرم‌افزار پرت، نسخه ۰۹. ۱۰. ۰۶ مورد تجزیه و تحلیل آکوستیکی قرار می‌گیرد. به این صورت که شش لایه شبکه متنی^۱ ساخته می‌شود. در لایه اول آغاز و پایانه هر واحد زبانی به دقت مشخص می‌شود و با توجه به آوانگاری IPA آوانویسی می‌شوند؛^۲ مکث‌های موجود در اول، درون و آخر جمله‌ها مشخص می‌شوند. سپس، در لایه دوم، واژه‌ها و همخوان‌ها برحسب گذاری می‌شوند. در لایه سوم، فواصل همخوانی و واژه‌ای براساس تعداد همخوان‌ها و واژه‌ها تعیین می‌شوند؛ در لایه چهارم، فواصل همخوانی و واژه‌ای بدون نگارش تعداد واژه‌ها و همخوان‌های موجود در هر یک از فاصله‌ها مشخص می‌شوند. در لایه پنجم، مرز بین هجاهای موجود تعیین می‌شوند؛ در نهایت در لایه ششم، قله هر هجا برحسب اصل رسایی به صورت خودکار توسط برنامه رایانه‌ای،^۳ نوشته شده توسط دلوو^۴، مشخص می‌شوند. نمونه‌ای از شبکه متنی ساخته شده در شکل (۱) آورده می‌شود.

۳-۴. سنج‌های زمانی مورد استفاده

در پژوهش حاضر، از تعدادی از سنج‌های زمانی پژوهش‌های پیشین ریتم گفتار (Dellwo et al., 2015; Ramus et al., 1999; Tilsen & Arvintini, 2013; Dellwo et al., 2012; Dellwo, 2010 and 2009; White & Mattis, 2007; Grabe & Low, 2002.) استفاده می‌شود. تمامی سنج‌ها به صورت خودکار با استفاده از برنامه رایانه‌ای موجود^۵ نگارش شده توسط دلوو، محاسبه می‌شوند. این سنج‌ها برحسب لایه‌های شبکه متنی در جدول (۱) آورده می‌شوند.



شکل ۱. نمونه شبکه متنی ساخته شده برای این پژوهش (جمله دوم متن داستان باد شمال و خورشید)

Fig 1- An example of a TextGrid for this study (the second sentence of "The north wind and the sun" story)

1. TextGrid

2. <https://www.internationalphoneticassociation.org/content/ipa-chart>

3. script

4. Volker Dellwo, Associate Professor of Phonetics, Zurich University, Switzerland.

(<https://www.cl.uzh.ch/de/people/team/phonetics/vdellw.html>)

5. <https://www.cl.uzh.ch/de/people/team/phonetics/vdellw/software.html>

جدول ۱- سنج‌های محاسبه‌شده این مقاله براساس لایه‌های شبکه متنی

Table 1- The measures of this study based on the tiers of the TextGrids

سنجه	نام لایه	لایه
meanSeg rateSeg	segment	۱
meanCon, meanConLn, ΔCon, ΔConLn, varcoCon, rateCon nPVI_Con, ,rateVow, meanVow, meanVowLn, rPVI_Con varcoVow,rPVI_Vow,nPVI_Vow ΔVow, ΔVowLn	cv segment	۲
rateC, meanC, meanCLn, ΔC, ΔCLn, varcoC, rPVI_C, nPVI_C meanVLn, ΔV, ΔVLn, varcoV, meanV, rateV, rPVI_V, nPVI_V, % V,	cv interval	۴
rateSyl, meanSyl, meanSylLn, ΔSyl, ΔSylLn, varcoSyl, rPVI-Syl, nPVI-Syl	syllable	۵
meanPeak, ratePeak, meanPeakLn, ΔPeak, ΔPeakLn, varcoPeak rPVI_peak, nPVI_peak	peak tier	۶
rateCV, meanCV, meanCVLn, ΔCV, ΔCVLn, VarcoCV, nPVI-CV, rPVI-CV, % VO, nVoiced, meanVoiced, meanVoicedLn, ΔVoiced, ΔVoicedLn, varcoVoiced, rPVI_Voiced, nPVI_Voiced, meanUnvoicedLn, ΔUnvoiced, nUnvoiced, meanUnvoiced ,rPVI_Unvoiced ΔUnvoicedLn, varcoUnvoiced nPVI_Unvoiced	cv interval	۴

در این قسمت، از هر نمونه سنجه یک مورد توضیح داده می‌شود.

- درصد واکه‌ای بودن گفتار (%V) و درصد واکدار بودن گفتار (%VO):

$$(1) \%V = \frac{\sum_{i=1}^{n_v} v_i}{\sum_{i=1}^{n_v} v_i + \sum_{i=1}^{n_c} c_i} \times 100\%$$

که در این فرمول n_v تعداد فواصل واکه‌ای، n_c تعداد فواصل همخوانی، v_i دیرش فواصل واکه‌ای و c_i دیرش فواصل همخوانی است.

- سرعت فواصل مختلف براساس معادله زیر محاسبه می‌شود. فرمول (۲) سرعت فواصل همخوانی و واکه‌ای را نشان می‌دهد؛ مانند فرمول (۲):

$$(2) rateCV = \frac{N_{cv}}{d}$$

در این فرمول N_{cv} تعداد فواصل واکه‌ای یا همخوانی در جمله است و d دیرش جمله بدون در نظر گرفتن مکث‌ها است.

- دیرش فواصل مختلف انحراف معیار سرعت نرمال‌شده (Varco)، انحراف معیار تقسیم بر میانگین که varco نام دارد؛ مانند فرمول (۳):

$$(3) VarcoV = 100 \times \frac{\Delta V}{\bar{V}}$$

در این فرمول ΔV انحراف معیار دیرش فواصل واکه‌ای و \bar{V} میانگین دیرش فواصل واکه‌ای است.

- تغییرپذیری دوتایی دیرش فواصل مختلف نرمال شده؛ مانند فرمول (۴):

$$(4) nPVI_V = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right| -$$

که در آن m تعداد فواصل واکه‌ای و d_k دیرش فواصل k th

- سنجه‌هایی که پسوند Ln دارند نسخه‌های نرمال نشده همتای بدون Ln خود هستند؛ برای مثال فرمول (۵):

$$(5) \Delta Invl(ln) = \sqrt{\frac{n_{Invl} \sum_{i=1}^{Invl} (ln Invl_i)^2 - [\sum_{i=1}^{Invl} (ln Invl_i)]^2}{N_{Invl} \cdot (N_{Invl} - 1)}}$$

در این معادله $Invl$ (فواصل مورد نظر هستند، واکه‌ای، همخوانی یا قله‌ای) و N تعداد این فواصل مورد نظر است.

۴-۴. تحلیل‌های آماری

تمام محاسبه‌های مربوط به اندازه‌گیری سنجه‌های ریتم به وسیله برنامه رایانه‌ای موجود در نرم‌افزار پرت^۱ که توسط دلوو^۱ نوشته شده است، محاسبه می‌شود. سپس، برای بررسی پرسش اول این پژوهش مبنی بر عملکرد همبسته‌های آکوستیکی براساس سنجه‌های موجود، میانگین و انحراف معیار نتایج به دست آمده از برنامه رایانه‌ای، در برنامه SPSS نسخه ۲۳ بررسی می‌شود و با نتایج پژوهش‌های پیشین مقایسه می‌شود. لی و دلوو^۲ (2016) نشان می‌دهند که نمره^۳ Z در سنجه‌های ریتم برای پایگاه داده‌های بزرگ مفید است؛ به همین دلیل، در پژوهش حاضر از این تحلیل آماری استفاده نمی‌شود. در پاسخ به پرسش‌های دوم و سوم پژوهش درباره وجود تفاوت در ویژگی‌های زمانی زنجیری و زیرزنجیری میان گوینده زبان فارسی در مرحله اول با استفاده از تحلیل همبستگی پیرسون^۴ سنجه‌هایی که با یکدیگر همبستگی کم‌تری دارند، انتخاب می‌شوند؛ زیرا بررسی دو یا چند پارامتر با همبستگی بالا در پژوهش‌های آواشناسی حقوقی توجیه منطقی ندارد. سپس، با استفاده از آزمون آنوا یک طرفه^۵ که متغیر مستقل آن گویندگان و متغیر وابسته سنجه‌های ریتم هستند، قوی‌ترین سنجه‌ای که نشان‌دهنده تمایزات میان گوینده در داده‌های این پژوهش است مشخص می‌شود و در نمودارهای boxplot نمایش داده می‌شود.

¹ <https://www.cl.uzh.ch/de/people/team/phonetics/vdellw.html>

² L. He & V. Dellwo,

³ z-score

⁴ Pearson Correlation

⁵ Anova one-way

۵. یافته‌ها

همان‌طور که در بخش‌های قبل اشاره شد، تمرکز این پژوهش بر عملکرد سنجه‌های ریتم به کار گرفته شده در پژوهش‌های پیشین درباره زبان فارسی و شناسایی سنجه‌هایی است که تفاوت‌های میان‌گوینده را بیشتر منعکس می‌کند. در این پژوهش، از ۷۰ جمله زبان فارسی معیار (۷ جمله \times ۱۰ گویشور) استفاده می‌شود. تعداد دیرش فواصل بررسی شده به شرح زیر است:

فواصل مربوط به واحدهای زبانی: ۴۰۸۱، فواصل هجایی: ۲۰۱۶، فواصل میان‌واکه‌ای: ۱۹۱۸، فواصل واکه‌ای: ۱۸۰۲^۱، فواصل همخوانی-واکه‌ای: ۳۹۲۰، فواصل همخوان: ۱۱۴۸، فواصل واکه: ۱۷۶۶، فواصل بین قله‌ای: ۱۹۰۵، فواصل واکه‌داری: ۱۸۰۲، فواصل بی‌واکی: ۱۹۱۸

۵-۱. همبسته‌های ریتم زبان فارسی

میانگین و انحراف معیار تمامی سنجه‌های ذکر شده در جدول (۱)، برای داده‌های این پژوهش در جدول (۲) نشان داده می‌شود. در این جدول سنجه‌های مربوط به واحدهای زبانی با «Seg»، سنجه‌های مربوط به هجا با «Syl»، سنجه‌های مربوط به فواصل میان‌واکه‌ای با «C»، سنجه‌های مربوط به فواصل واکه‌ای «V»، سنجه‌های مربوط به فواصل همخوانی-واکه‌ای با «CV»، سنجه‌های مربوط به دیرش همخوان‌ها با «Con»، سنجه‌های مربوط به دیرش واکه‌ها با «Vow»، سنجه‌های مربوط به فواصل واکه‌داری با «Voiced»، سنجه‌های مربوط به فواصل بی‌واکی با «Unvoiced»، و در نهایت سنجه‌های مربوط به فواصل بین-قله‌ها با «Peak» مشخص شده‌اند.

راموس و همکاران (1999) با محاسبه ΔC ، ΔV ، $V\%$ پیوستاری را برای ریتم زبان‌های مختلف در نظر گرفته‌اند که در یک سر آن زبان‌های تکیه‌زمانی و در سر دیگر آن زبان‌های هجازمانی قرار دارند. آنها زبان انگلیسی را جزو زبان‌های تکیه‌زمانی و زبان فرانسه را نماینده مناسبی برای زبان‌های هجازمانی می‌دانند و کاتالان را در میانه این پیوستار قرار می‌دهند. درباره نتایج این پژوهش، همان‌طور که در جدول (۲) مشاهده می‌شود، انحراف معیار $V\%$ داده‌های پژوهش حاضر برابر $3/85$ است، در حالی که انحراف معیار $V\%$ زبان انگلیسی براساس یافته‌های راموس و همکاران (1999) برابر $5/4$ و زبان فرانسه $4/5$ و زبان ژاپنی $3/4$ است. علاوه بر این، میانگین ΔC و ΔV این پژوهش به ترتیب $0/03$ و $0/57$ است و ΔC و ΔV زبان انگلیسی، براساس یافته‌های راموس و همکاران (1999)، $1,25$ و $1,63$ و برای زبان فرانسه $1,21$ و $0,74$ است. در نتیجه، یافته‌های این پژوهش، زبان فارسی را در میانه پیوستار هجازمانی-تکیه‌زمانی نزدیک به هجازمانی قرار می‌دهد.

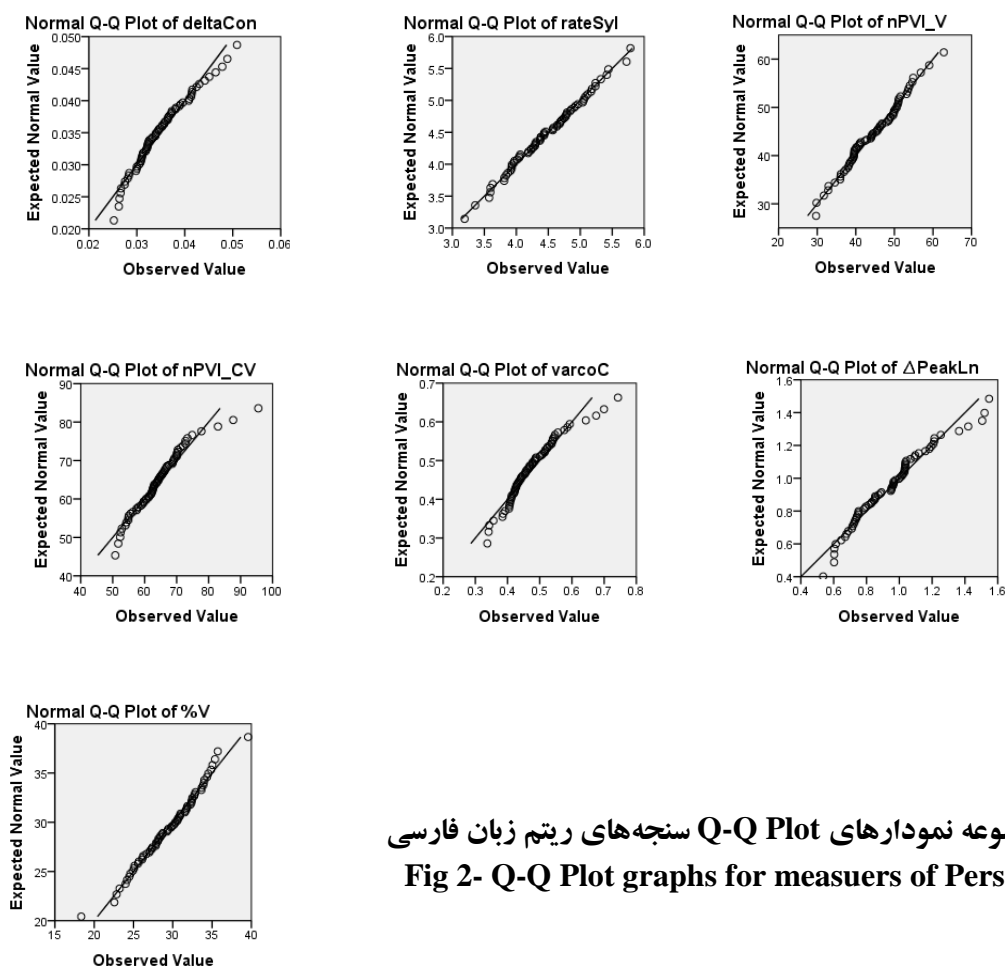
¹. vocalic

جدول ۲- میانگین و انحراف معیار سنج‌های ریتم زبان فارسی معیار

Table 2- Mean and standard deviation for measurers of Standard Persian rhythm

انحراف معیار	میانگین	سنجه	انحراف معیار	میانگین	سنجه
۰۶۶۵.	۴۰۹.	varcoCon	۰۴۳۴.	۱۲۶.	Δ Syl
۷۷۵.	۲۴۷.۴	rPVI_Con	۰۸۳.	۴۹۰.	Δ SylLn
۳۸۸.۹	۳۶۳.۵۰	nPVI_Con	۱۳۴.	۵۴۵.	varcoSyl
۱۰۰.۲	۹۰۳.۱۳	rateVow	۲۵۱.۳	۴۹۵.۱۱	rPVI-Syl
۰۱۸.	۰۷۴.	meanVow	۳۴۳.۹	۰۳.۵۱	nPVI-Syl
۱۴۷.	۶۹۲.۲-	meanVowLn	۰۳۹.۱	۴۵۳.۸	rateC
۰۰۸.	۰۲۸.	Δ Vow	۰۱۵.	۱۲۰.	meanC
۰۶۴.	۳۹۲.	Δ VowLn	۱۱۴.	۲۵۵.۲-	meanCLn
۰۶۷.	۳۷۸.	varcoVow	۰۱۳۵.	۰۵۷.	Δ C
۹۰۰.	۱۷۲.۳	rPVI_Vow	۰۹۴.	۵۴۳.	Δ CLn
۷۳۶.۷	۰۷۰.۴۳	nPVI_Vow	۰۷۹.	۴۷۴.	varcoC
۰۳۵.	۲۴۱.	meanPeak	۵۲۶.۱	۹۶۵.۵	rPVI_C
۵۹۰.	۲۲۶.۴	ratePeak	۰۴۵.۹	۸۱۴.۵۱	nPVI_C
۱۶۹.	۷۵۸.۱-	meanPeakLn	۹۹۸.۱	۶۴۸.۱۳	rateV
۰۴۹.	۱۸۵.	Δ Peak	۰۱۱۶.	۰۷۵.	meanV
۲۲۸.	۹۴۳.	Δ PeakLn	۱۴۴.	۶۷۹.۲-	meanVLn
۱۳۱.	۷۵۹.	varcoPeak	۰۰۷.	۰۳.	Δ V
۳۰۰.۴	۰۰۷.۱۸	rPVI_peak	۰۵۷.	۴۰۲.	Δ VLn
۶۰۶.۱۵	۱۳۸.۸۲	nPVI_peak	۰۶۰.	۳۹۴.	varcoV
۸۴۷.۳	۵۳۷.۲۹	Voiced%	۸۵۴.	۳۶.۳	rPVI_V
۰۱۱.	۰۷۵.	meanVoiced	۱۵۴.۷	۴۶۲.۴۴	nPVI_V
۱۴۴۹.	۶۷۹.۲-	meanVoicedLn	۸۴۷.۳	۵۳۷.۲۹	V%
۰۰۷.	۰۳.	Δ Voiced	۱۲۹.۱	۷۱۹.۸	rateCV
۰۵۷.	۴۰۲.	Δ VoicedLn	۱۰۴.	۳۸۱.۲-	meanCVLn
۰۶۰.	۳۹۴.	varcoVoiced	۰۳۲.	۱۰۳.	Δ CV
۸۵۴.	۳۶.۳	rPVI_Voiced	۰۷۳.	۶۳.	Δ CVLn
۱۵۴.۷	۴۶۲.۴۴	nPVI_Voiced	۱۸۸.	۸۶۸.	varcoCV
۰۱۵.	۱۲۰.	meanUnvoiced	۴۴۰.۱	۸۷۴.۷	rPVI_CV
۱۱۴.	۲۵۵.۲-	meanUnvoicedLn	۰۶۷.۸	۴۶۳.۶۴	nPVI_CV
۰۱۳.	۰۵۷.	Δ Unvoiced	۲۶۰.۱	۷۶۸.۱۱	rateCon
۰۹۴.	۵۴۳.	Δ UnvoicedLn	۰۰۹.	۰۸۶.	meanCon
۰۷۹.	۴۷۴.	varcoUnvoiced	۱۱۷.	۵۵۴.۲-	meanConLn
۵۲۶.۱	۹۶۴.۵	rPVI_Unvoiced	۰۰۵.	۰۳۵۰.۱.	Δ Con
۰۴۵.۹	۸۱۳.۵۱	nPVI_Unvoiced	۱۰۲.	۴۷۲۵۳.	Δ ConLn

مقایسه nPVI_V به دست آمده از این پژوهش (nPVI_V: ۴۴,۴۶) با نتایج کار گریب و لو (2002) نیز زبان فارسی را بین زبان فرانسه (nPVI_V: ۴۳,۰۵) و زبان کاتالان (nPVI_V: ۴۵) قرار می‌دهد. البته باید در نظر داشت که در این دو پژوهش ذکر شده جمله‌هایی انتخاب می‌شدند که مکث درون جمله‌ای نداشتند به همین دلیل تغییرپذیری دیرش آنها کمی متفاوت از این پژوهش است. همچنین، به همین دلیل است که در کار ابوالحسنی و همکاران (2013) و ابوالحسنی زاده و تقوی (۱۳۹۸) نیز، تاحدودی nPVI_V با یافته‌های این پژوهش متفاوت است و زبان فارسی را نزدیک‌تر به یافته‌های دو پژوهش یادشده و نزدیک‌تر به هجزامانی‌ها نشان داده است. در پژوهش حاضر، به دلیل نزدیک‌تر شدن داده‌ها به گفتار عادی گویشوران، جمله‌های دارای مکث نیز لحاظ می‌شوند؛ در هر صورت، برنامه رایانه‌ای استفاده شده در این پژوهش قادر به حذف مکث‌ها است. در راستای اثبات توزیع مناسب داده‌ها درباره متغیرهای این پژوهش با استفاده از نرم‌افزار spss و تحلیل همبستگی پیرسون سنجه‌هایی که همبستگی کم‌تری باهم داشتند، انتخاب می‌شوند و سپس، با استفاده از نمودار Q-Q plot بررسی می‌شوند. در زیر این نمودارها ارائه می‌شود.



شکل ۲- مجموعه نمودارهای Q-Q Plot سنجه‌های ریتم زبان فارسی
 Fig 2- Q-Q Plot graphs for measurers of Persain rhythm

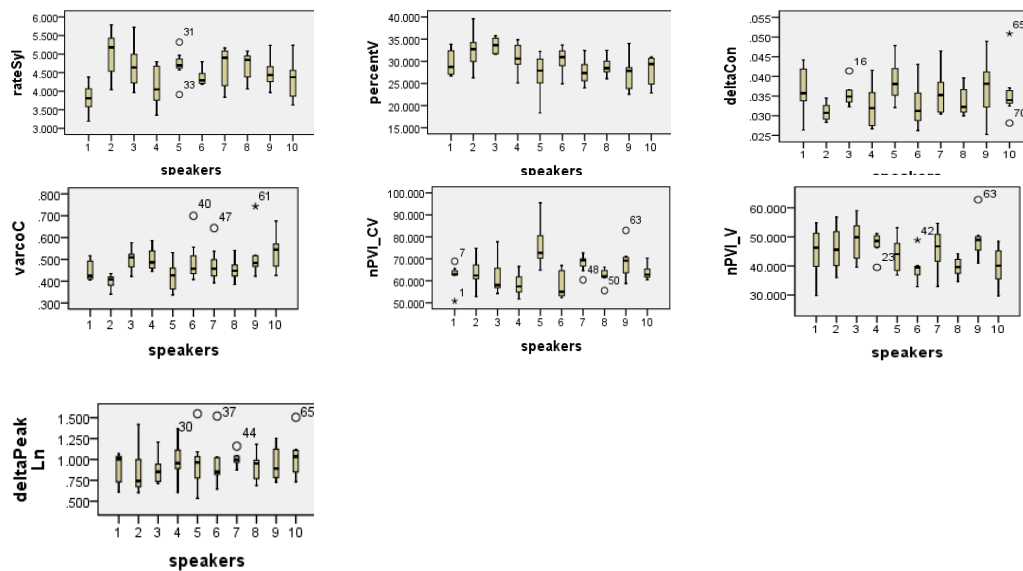
همان طور که ملاحظه می‌شود، متغیرهای انتخاب‌شده با توزیع داده‌ها مطابقت دارند؛ زیرا تا حد درخور توجهی خوشه‌های نقطه‌ها در خط مستقیم قرار گرفته‌اند. همچنین، خطی بودن نقطه‌ها نشان می‌دهد که داده‌ها به صورت معمولی توزیع یافته‌اند.

۲-۵. سنجه‌های نشانگر تغییرات میان‌گوینده در زبان فارسی

در این پژوهش، به منظور یافتن سنجه‌هایی که تفاوت‌های میان‌گوینده زبان فارسی را منعکس می‌کنند، از آزمون آنوا یک‌طرفه استفاده می‌شود. به این صورت که برای سنجه‌هایی که در قسمت (۱-۵) با استفاده از تحلیل همبستگی پیرسون انتخاب شدند، گویندگان متغیر مستقل و همبسته‌های ریتم، متغیرهای وابسته در نظر گرفته شدند. براساس عدد معنی‌داری که $0.05 \leq$ است، تعداد پنج سنجه از آنها معنادار بودند که به تفصیل در جدول (۳) بررسی می‌شوند. در این آزمون آماره F نشان‌دهنده نسبت تغییرات میان‌گوینده به تغییرات درون‌گوینده است. بنابراین، هرچه این عدد بالاتر باشد بیانگر این است که تغییرات میان‌گوینده این پارامتر نسبت به تغییرات درون‌گوینده آن بیشتر است و آن پارامتر تأثیر بیشتری را در نشان‌دادن تغییرات میان‌گوینده دارد. همان طور که ملاحظه می‌شود، از میان متغیرهایی که معنی‌دار هستند متغیر $nPVI_CV$ به دلیل داشتن آماره F بالاتر، سنجه قوی‌تری برای نشان‌دادن تمایزات میان‌گوینده براساس داده‌های این پژوهش است؛ دومین سنجه نشانگر تمایزات میان‌گوینده $rateSyl$ و سومین آنها $V\%$ است. در شکل (۳) تغییرات این سنجه‌ها برای ده گویشور این پژوهش با استفاده از نمودار $BoxPlot$ نشان داده شده‌است.

جدول ۳- نتایج آزمون آنوا یک‌طرفه برای نشان‌دادن تأثیر گوینده بر سنجه‌های ریتم گفتار
Table 3- ANOVA one-way to show the effects of speaker on the rhythm measures

سنجه	مجموع مجذورات	آماره F	عدد معناداری
$V\%$	۱۰۲۱,۳۱۹	۳,۰۹۵	۰,۰۰۴
$rateSyl$	۷,۰۱۸	۳,۱۳۱	۰,۰۰۴
$nPVI_CV$	۴۴۹۰,۱۲۸	۴,۴۴۸	۰,۰۰۰
$nPVI_V$	۸۶۹,۶۱۷	۲,۱۷۸	۰,۰۳۶
$VarcoC$	۰,۱۱۷	۲,۴۴۴	۰,۰۱۹
ΔCon	۰,۰۰۰	۱,۲۹۰	۰,۲۶۱
$\Delta PeakLn$	۰,۱۹۱	۰,۳۷۴	۰,۹۴۳



شکل ۳- مجموعه نمودارهای BoxPlot داده‌ها براساس گویندگان
 Fig 3- BoxPlot diagrams based on the speakers

۶. بحث و نتیجه‌گیری

در این پژوهش، به منظور بررسی ویژگی‌های ریتم زبان فارسی معیار از رویکردهای پژوهش‌های پیشین ریتم استفاده شد و همبسته‌های ریتم زبان فارسی با استفاده از سنج‌های مختلف ریتم زبان تعیین شد. این سنج‌ها به بررسی دیرش فواصل مختلف مربوط به واحدهای زبانی، واکه‌ای و همخوانی، واکه‌داری و بی‌واکی، هجایی و بین‌قله‌ای می‌پردازند. در این قسمت به پرسش‌های ذکر شده در بخش مقدمه، پاسخ داده می‌شود.

در زمینه پاسخ به پرسش اول، نتایج این پژوهش نشان می‌دهد جایگاه ریتم زبان فارسی معیار نزدیک به زبان‌های هجازمانی قرار داد. بررسی دیگر لهجه‌های زبان فارسی می‌تواند تفاوت ریتم دیگر لهجه‌های این زبان را مشخص کند. همچنین، بررسی همبسته‌های دیگر زبان‌های به کار گرفته شده در ایران می‌تواند ویژگی‌های ریتم بین‌زبانی و تأثیر آنها بر یکدیگر را مشخص کند. در خصوص ویژگی‌های ریتم بین‌زبانی شواهد رفتاری قوی وجود دارد که شنوندگان تفاوت بین زبان‌ها را براساس نوع تغییرپذیری دیرشی که در این پژوهش استفاده شد، درک می‌کنند. آزمایش‌ها نشان داده است که شنوندگانی که بالغ هستند (Ramus & Mehler, 1999) و همچنین، نوزادان تازه (Nazzi, Jusczyk & Johnson, 2000) می‌توانند زبان‌ها را با توجه به دسته‌بندی‌های مختلف ریتم تشخیص دهند؛ چنین ویژگی‌هایی در مرحله پیش‌زبانی^۱ به دست می‌آیند و ممکن است شنوندگان؛ به‌عنوان مثال نوزادانی که در یک محیط دوزبانه رشد می‌کنند، را قادر سازد که بین زبان‌های مختلف تفکیک قائل

^۱. pre-linguistic stage

شوند (Ramus et al., 1999). از آنجا که ویژگی‌های دیرشی فواصل همخوانی و واکه‌ای بین زبان‌ها به‌طور محسوسی درخور توجه است، به نظر می‌رسد تنوع بین‌گوینده نیز دارای اهمیت است.

در پاسخ به پرسش دوم، یافته‌های پژوهش نشان داد که سنجه‌های ریتم مبتنی بر مدت‌زمان فاصله همخوانی و واکه‌ای و همچنین ویژگی‌های بسته دامنه، مانند فواصل هجایی و قله هجا به‌شدت در میان‌گوینده متفاوت است. به نظر می‌رسد مکانیسم‌های کنترل اندام‌های تولید گفتار هر گویشور مسئول ایجاد چنین تفاوتی است. در اینجا این پرسش ممکن است مطرح شود: «چگونه حرکت اندام‌های گفتار هر فرد منجر به الگوهای فردی سنجه‌های ریتم می‌شود؟» دلوو، لیمن و کولی (۲۰۱۵) برای این منظور از دو نوع سنجه استفاده می‌کنند: ۱. سنجه میزان واکه‌ای ($V\%$)، بی‌واکی Voiced %، سنجه تغییرپذیری دیرشی (ΔV و $\Delta V[\ln]$) و ۲. تغییرپذیری قله دامنه درون هجا ($\Delta Peak$ و $\Delta Peak[\ln]$). نتایج معناداری از تفاوت میان‌گوینده این مطالعه به‌وسیله این سنجه‌ها به‌دست آمد ($\Delta V: 0.009$, $\Delta VLn: .013$, $\Delta Peak: .000$, $V\%: 0.004$, $Voiced\%: .004$). نتایج چنین مطالعاتی نشان‌دهنده این است که حرکت منحصر به فرد اندام‌های گفتار می‌تواند به‌دلیل زیستی، اکتسابی یا هر دو باشد.

در پاسخ به پرسش سوم این پژوهش، با استفاده از تحلیل همبستگی پیرسون و همچنین آزمون آنوا مشخص شد که متغیر $nPVI_CV$ بالاترین نسبت تغییرات میان‌گوینده به درون‌گوینده را دارد و نشان‌دهنده این است که این سنجه، توانایی بیشتری را در انعکاس تمایزات میان‌گوینده دارد. این متغیر در لایه چهارم شبکه‌های متنی بررسی می‌شود. در این لایه، مرز دقیق فواصل همخوانی و واکه‌ای مشخص شده است و $nPVI_CV$ تغییرپذیری این فواصل را نشان می‌دهد. بنابراین، می‌توان نتیجه گرفت که تغییرپذیری فواصل همخوانی-واکه‌ای متغیر مناسبی برای تمایزات میان‌گوینده در این سبک از زبان فارسی است. نکته‌ای که در اینجا درخور ذکر است این است که نتایج پژوهش حاضر، تفاوت معنی‌داری را در سنجه‌های فاصله‌ای هجایی نشان دادند ($rateSyl: 0.003$, $meanSyl: 0.002$, $meanSylLn: 0.015$, $\Delta Syl: .001$, $\Delta SylLn: .037$, $varcoSyl: .007$). بنابراین، سنجه‌های مربوط به فواصل هجایی نیز نشان‌دهنده تمایزات میان‌گوینده در زبان فارسی هستند. دلیل آن می‌تواند وجود انسدادی چاکنایی $/\text{?}/$ باشد. از آنجا که ساختار هجایی زبان فارسی مجوز شروع هجا با واکه را نمی‌دهد، در آغاز هجاهایی مانند «است» $/\text{?}ast/$ یک انسدادی چاکنایی قرار می‌گیرد (حق‌شناس، ۱۳۸۴). داده‌های این پژوهش نشان می‌دهند که در گفتار به‌هم‌پیوسته، این انسدادی چاکنایی اگر در اول هجاهایی که با واکه شروع می‌شوند مانند $/\text{?}ast/$ قرار گیرد دو حالت در تلفظ این هجا به وجود می‌آید. در حالت اول، گوینده $/\text{?}/$ را ادا می‌کند، که در این صورت ساختار هجایی دو واژه متوالی مانند «بیشتر است» به‌صورت $CVC\ CVC\ CVCC/$ است. در حالت دوم، ممکن است گوینده $/\text{?}/$ را تولید نکند، در چنین مواردی عدم تولید این انسدادی باعث تغییر ساختار هجا می‌شود؛ بنابراین، ساختار هجایی دو واژه متوالی ذکر شده به‌صورت $CVC\ CV / CVCC$ در می‌آید. این امر باعث می‌شود تمایزات معنی‌دار میان‌گوینده در ویژگی‌های مربوط به هجا به وجود آیند. می‌توان در پژوهش‌های آتی با استفاده از پیکره‌های بزرگ‌تر به بررسی دقیق‌تر این پدیده پرداخت.

همان‌طور که در جدول (۳) ملاحظه شد، $V\%$ که درصد واکه‌ای بودن گفتار را نشان می‌دهد نیز جزو قوی‌ترین سنجه‌های نشان‌دهنده تمایزات میان‌گوینده برای داده‌های این پژوهش است. این نتیجه، هم‌راستا با یافته‌های اسدی و همکاران (2018) و نیز

دلوو، لیمن و کولی (2015) است. بنابراین، از درصد واکه‌ای بودن گفتار نیز می‌توان به‌عنوان یک سنجۀ متمایزکننده گوینده استفاده کرد.

پرسش دیگری که در اینجا مطرح می‌شود این است که تنوع ریتم میان‌گوینده چه کاربردهایی می‌تواند داشته باشد؟ نتایج پژوهش حاضر می‌تواند برای هر نوع برنامه کاربردی که اطلاعات گوینده‌ویژه^۱ در آن نقش ایفا می‌کند، مورد استفاده قرار گیرد. برنامه‌های شناسایی گوینده از متغیرهای دامنه فرکانس مانند فرکانس‌های پایه و سازه یا کل ویژگی‌های بسته طیفی استفاده می‌کنند؛ زیرا توسط ویژگی‌های آناتومیک چاکنای شکل می‌گیرند. با این حال، این متغیرها از کانال‌های دیگر برای انتقال اطلاعات کاربردی زبانی و فرازبانی استفاده می‌کنند. در حالی که گفتار در زمان بسیار سازمان‌یافته، عمل می‌کند. واضح به نظر می‌رسد که سازمان‌دهی‌های زمانی زبرنجیری، عملکرد واضحی را برای انتقال اطلاعات زبان‌شناختی و فرازبان‌شناختی به کار نمی‌گیرند. در مواردی که گویشوران از متغیرها برای ایجاد تضادهای عملکردی در گفتار استفاده می‌کنند، برای تعدیل آنها باید بر این ویژگی‌ها کنترل داشته باشند و باید سیستم ادراکی گویشوران روی آنها تنظیم شود. با این توصیفات، گویشور توانایی کنترل کم‌تری بر پارامترهای ریتمیک به نسبت آهنگ و تکیه دارد (Dellwo, Leeman & Kolly, 2015). این ویژگی در شناسایی گویشور در موارد قضایی دارای اهمیت است.

قدردانی

در این قسمت وظیفه خود می‌دانیم که یاد استاد گران‌قدر دانشگاه شیراز، مرحوم دکتر مهرزاد منصوری، را گرامی بداریم که با همت و کوشش ایشان اتاق آواشناسی دانشگاه شیراز راه‌اندازی شد. روحشان شاد، نامشان جاودان. همچنین، از ده شرکت‌کننده این پژوهش که همگی از استادان و دانشجویان دانشگاه شیراز بودند و با ما در انجام این پژوهش همکاری کردند نهایت تشکر را داریم.

منابع

ابوالحسنی‌زاده، وحیده و تقوی، نقیسه. (۱۳۹۸). بررسی و مقایسه ریتم زبان فارسی و زبان انگلیسی. *دوفصلنامه زبان و ادبیات تفسیری-تطبیقی*، ۴(۱)، ۱-۱۰.

حق‌شناس، علی محمد. (۱۳۸۴). *آواشناسی (فوننتیک)*. تهران: نشر آگه.

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

Abolhasani Zadeh, V., Taghva, N., Sharifi Moghadam, A., & Golshan, A. (2013). The effect of Farsi language rhythm on Iranian traditional non metric instrumental music. *Aion-linguistica*, 2, 11-19.

¹. speaker-specific information

- Abolhasani Zadeh, V., & Taghva, N. (2019). Study and comparison of Persian language and English language rhythm. *Bilingual Interpretive-Comparative Journal of Language and Literature*, 4(6), 1-10. [In Persian]
- Adams, C. (1979). *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton.
- Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40, 351–373.
- Asadi, H., Nourbakhsh, M., He, L., Pelligrino, E., & Dellwo, V. (2018). Between-speaker rhythmic variability is not dependent on language rhythm, as evidence from Persian reveals. *The International Journal of Speech, Language and the Law*, 25(2), 151-172.
- Barbosa, P., & Bailly, G. (1994). Characterisation of rhythmic patterns for text-to-speech synthesis. *Speech Communication*, 15, 127-137.
- Bertrán, A. P. (1999). Prosodic typology: On the dichotomy between stress-timed and syllable-timed languages. *Language Design*, 2, 103-131.
- Bosch, L., & Sebastián-Gallés, N. (1997). The role of prosody in infants' native language discrimination abilities: The case of two phonologically close languages. Paper presented in *EUROSPEECH-1997*, Rhodes, Greece, 231-234.
- Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51-62.
- Dauer, R. (1987). *Phonetic and phonological components of language rhythm*. Paper presented at International Congress of Phonetic Science (ICPhS) XI, Tallinn, Estonia.
- Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for delta C. In P. Karnowski & I. Sziget (Eds.), *Language and language-processing* (pp. 231-241). Frankfurt am Main: Peter Lang.
- Dellwo, V., Huckvale, M., & Ashby, M. (2007). How is individuality expressed in voice? An introduction to speech production & description for speaker classification. In C. Müller (Ed.). *Speaker Classification I* (pp. 1-20). Berlin: Springer.
- Dellwo, V. (2009). Choosing the right rate normalization method for measurements of speech rhythm. In S. Schmid, M. Schwarzenbach, and D. Studer (Eds.), *La Dimensione Temporale del Parlato: Atti del 5° Convegno Nazionale AISV 2009* (pp. 13-32). Torriana, Italy: EDK Editore.
- 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. (2012). Speaker idiosyncratic rhythmic features in the speech signal. *Proceedings of Interspeech*, Portland (USA).
- Dellwo, V. & Fourcin, A. (2013). Rhythmic characteristics of voice between and within languages. *Revue Tranel (Travaux neuchâtelois de linguistique)*, 59, 87-107.
- Dellwo, V., Leemann, A., & Kolly, M-J. (2015). Rhythmic variability between speakers: articulatory, prosodic, and linguistic factors. *Journal of the Acoustical Society of America*, 137, 1513-1528.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and rhythm class hypothesis. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology 7* (pp. 514-546). Berlin, Germany: Mouton de Gruyter.
- Haghshenas, A. (2004). *Phonetics*. Tehran: Agah. [In Persian]

- Liss, J. M., White, L., Mattys, S., Lansford, K., Lotto, K. A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech Language and Hearing Research*, 52, 1334–1352.
- He, L. (2017). Speaker idiosyncratic intensity variability in the speech signal. [Doctoral dissertation]. Zurich University.
- He, L., & Dellwo, V. (2016). The role of syllable intensity in between speaker rhythmic variability. *The International Journal of Speech*, 23(2), 243-273.
- Jessen, M. (2008). Forensic Phonetics. *Language and Linguistics Compass*, 2(4), 671–711.
- Lee, C. S., & Todd, N. P. M. (2004). Towards an auditory account of speech rhythm: application of a model of the auditory “primal sketch” to two multilanguage corpora. *Cognition*, 93, 225-254.
- 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.
- Leong, V., Stone, M. A., Turner, R. E., & Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *Journal of the Acoustical Society of America*, 136, 366–381.
- Leong, V., & Goswami, U. (2014). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience*, 8(96), 1-14.
- Lloyd James, A. (1940). *Speech Signals in Telephony*. London: Sir Isaac Pitman & Sons.
- Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *Journal of Acoustic Society of America*, 129, 3258-3270.
- McDougall, K. (2004). Speaker-Specific Formant Dynamics: An Experiment on Australian English /aI/. *International Journal of Speech, Language and the Law*, 11, 103–130.
- McDougall, K. (2006). Dynamic features of speech and the characterisation of speakers: Towards a new approach using formant frequencies. *International Journal of Speech, Language and the Law*, 13, 89–126.
- Milne, B. L. (1955). *Modern speech rhythm exercises*. London: Macmillan.
- Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, 83, 405–408.
- Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756-766.
- Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effect of rhythm and familiarity. *Journal of Memory and Language*, 43, 1-19.
- Nespor, I. (1990). On the rhythm parameter in phonology. In I. Roca (Ed.), *Logical Issues in Language Acquisition* (pp. 157-195). Dordrecht: Foris.
- Nolan, F. (2002). Intonation in speaker identification: an experiment on pitch alignment Features. *Forensic Linguistics*, 9, 1–21.
- Nolan, F., & Jeon, H. S. (2015). Speech rhythm: a metaphor? *Philosophical Transactions of the Royal Society B*, 369(1658), 1-11.
- O’Dell, M. L., & Nieminen, T. (1999). Coupled oscillator model of speech rhythm. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)* 1075–1078. San Francisco, USA.

- Payne, E., Post, B., Astruc, L., Prieto, P., & Vanrell, M. (2012). Measuring child rhythm. *Language and Speech*, 55(2), 203–229.
- Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Pointon, G. E. (1980). Is Spanish really syllable-timed? *Journal of Phonetics*, 8, 293-304.
- Polyanskaya, L., & Ordin, M. (2015) Acquisition of speech rhythm in first language. *Journal of the Acoustical Society of America*, 138, 199–204.
- Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustical Society of America*. 105, 512-521.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top Tamarin monkeys. *Science*, 288, 349-351.
- Roach, P. (1982). On the distinction between “stress-timed” and “syllable-timed” languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 73-79). London: Edwards Arnold.
- Rose, P. (1996). Speaker verification under realistic forensic conditions. In P. McCormak and A. Russell (Eds), *Proceedings of the Sixth Australian International Conference on Speech Science and Technology, Australian Speech Science and Technology Association, Canberra* (pp. 109-114).
- Rose, P. (2004). *Forensic Speaker Identification*. Boca Raton, FL: CRC Press.
- Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *Journal of Acoustic Society of America*, 124, 34-39.
- Tilsen, S. & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of Acoustic Society of America*, 134, 628-639.
- 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 Acoustic Society of America*, 127, 1559– 1569.
- White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35, 501–522.
- White, L., Liss, J., & Dellwo, V. (2010) Assessment of rhythm. In A. Lowit & R. D. Kent (Eds.) *Assessment of motor speech disorders* (pp. 312–352). San Diego: Plural Publishing.
- Yoon, T. J. (2010). Capturing inter-speaker invariance using statistical measures of speech rhythm. *Proceedings of Speech Prosody 5, Chicago/Ill.* 1-4.