



Speech Rhythm Measures: Acoustic Cues for Speaker Identification

Homa Asadi¹ & Batool Alinezhad²
(29-49)

Abstract

Rhythmic characteristics of speech based on consonantal and vocalic intervals as well as syllabic intervals vary between speakers of the same language. Nonetheless, the rhythmicity of a speech signal is not solely dependent on the durational variability of phonetic intervals but it is also associated with the variability of the intensity patterns as well. Acoustic parameter of intensity is largely determined by the articulatory behaviors of the speech organs such as lip movement or mouth aperture. Therefore, it is plausible that speaker idiosyncrasy in movement of speech articulators and anatomical differences in individual's vocal tracts may influence the energy distribution across a speech signal which subsequently leads to the variability in the values of the intensity measures. Using experimental phonetics tools and from an explicitly speaker-specific perspective, the present research attempts to explore potential speaker-specific acoustic parameters of speech rhythm which are extracted from the intensity contours across Persian speakers. This research aims to discover whether intensity-based measures of speech rhythm are able to discriminate between speakers in Persian. Two types of acoustic rhythmic measures based on the mean syllable intensity (stdevM, varcoM, rPVIm, nPVIm) and peak syllable intensity (stdevP, varcoP, rPVIp, nPVIp) were selected for this study. Speech data from 12 Persian male speakers were recorded non-contemporaneously in laboratory environment on two different occasions separated by one to two weeks. Speech tokens were acoustically measured with PRAAT version 5.2.34 and statistical analyses were carried out with SPSS version 21 and R version 3.3.3. Results of the study indicated that speech rhythm measures based on intensity fluctuations play an important role in between-speaker rhythmic variability. In addition, discriminatory power of intensity-based measures is not affected by the language-dependent characteristics of Persian. The results also showed that the peak syllable intensity measures carry more speaker-specific information compared to the mean syllable intensity measures.

Keywords: Experimental phonetics, intensity-based measures, speaker identification, speech rhythm, between-speaker variability.

Received: 22, July, 2020 & Accepted: 2, January, 2021

doi
10.22059/jolr.2020.307695.666628
Print ISSN: 2288-1026-Online ISSN: 3362-2676
<https://jolr.ut.ac.ir>

1 .Email of the corresponding author: h.asadi@fgn.ui.ac.ir

Assistant Professor of Linguistics, University of Isfahan, Isfahan, Iran.

2 .Associate Professor of Linguistics, University of Isfahan, Isfahan, Iran.

1.Introduction

Speech is highly organized in time. In the present study, we study how suprasegmental temporal features can contribute to speaker-individuality in Persian. Temporal characteristic of speech is a newly developed method applied in forensic phonetics. The rationale behind this idea is that humans differ in terms of the anatomical dimensions of the articulators, which result in idiosyncratic temporal characteristic articulation, namely speaker-specific rhythm. This relationship, however, is very complex as numerous acquired and language-specific characteristics also influence the temporal structure of speech. Suprasegmental speech rhythm measures involve a tight interaction between speakers' anatomy and their learned behavior. It seems that a possible contribution of the individual kinematics of the articulators which might lead to individual temporal characteristics in the signal may influence aspects of the signal that stand in relation to its intensity. To date, there has only been research in languages like English and German whose speakers might vary considerably in the way they operate the articulators to produce complex phonotactics and vowel reductions. Therefore, it is conceivable that language-specific features of Persian influence on temporal characteristic of speech. This study therefore sets out to explore between-speaker rhythmic variability as well as within-speaker rhythmic variability across Persian speakers to see to what degree such temporal characteristics vary among Persian speakers. Certainly a comprehensive understanding of speaker-specific temporal characteristics can only be obtained via a contrastive analysis of such characteristics in numerous different languages with widely different phonological systems.

2.Literature Review

In this section, different numerous studies on speech rhythm measures were categorized based on the measures of speech interval duration and temporal characteristics of the amplitude envelope.

3.Methodology

To test between-and within-speaker variability, twelve male speakers of Standard Contemporary Persian were recorded on two different sessions, separated by a time-lapse of one to two weeks. Speakers were asked to read the 54 sentences one by one, with a pause, and in a natural way, without any marked intonation. The microphone was positioned approximately 20 cm away from the mouth of the speakers in a diagonal position. Speech tokens were analysed using Praat (version 5.2.34, Boersma and Weenink 2013). For this study, we selected suprasegmental intensity-based measures retrieved from mean and peak syllable units of speech signals. Statistical analysis of data was carried out using R (R core Team 2014) version 3.3.3 and the R package lme4 (Bates, Maechler, Bolker and Walker 2016) and SPSS (IBM Corp. 2012).

4.Results

In this section, we provide the results of different acoustical models i.e. two-way anova, multinomial logistic regression and principal component analysis that were employed on the collected speech data of Persian. In the present study, we explored potential speaker-specific acoustic parameters of speech rhythm in Persian from a forensic-acoustic perspective. Statistical analysis of speech data revealed that selected acoustic speech rhythm measures are able to discriminate between Persian speakers and among them the variation coefficient of syllable peak intensity levels (varcoP) could account better for variability across speakers.

5.Discussion

We discuss how temporal characteristics of speech signals based on intensity fluctuations vary between Persian speakers and which acoustic parameter of speech rhythm has more potential in showing between-speaker variability. Furthermore, we compared the results of the present study to the findings of previous studies to see in what way Persian has been similar or different from other investigated languages

6.Conclusions

Based on the results obtained we can conclude that speech rhythm measures can be potentially characterized as language-independent parameters which can be utilized as acoustic cues in cases where knowledge about speaker-specific rhythm is needed.

پارامترهای آکوستیکی ریتم گفتار: سرنخ‌های آکوستیکی برای تشخیص گوینده

هما اسدی^۱

استادیار گروه زبان‌شناسی، دانشکده زبان‌های خارجی، دانشگاه اصفهان، اصفهان، ایران.

بتول علی‌نژاد

دانشیار گروه زبان‌شناسی، دانشکده زبان‌های خارجی، دانشگاه اصفهان، اصفهان، ایران.

تاریخ دریافت مقاله: ۹۹/۰۵/۰۱؛ تاریخ پذیرش مقاله: ۹۹/۱۰/۱۳

علمی - پژوهشی

چکیده

مشخصه‌های ریتم گفتار بر اساس دیرش مناطق همخوانی، واکهای و نیز هجا می‌توانند گویشوران یک زبان خاص را از هم تمیز دهند. با وجود این، ریتم یک سیگنال آوایی تنها بر مبنای روابط دیرشی میان مناطق آوایی آن استوار نیست بلکه به تفاوت‌های مربوط به الگوی شدت در آن‌ها نیز بستگی دارد. پارامتر آکوستیکی شدت تا حد زیادی تحت تأثیر الگوهای حرکتی اندام‌های گفتار مانند حرکت لب‌ها و باز و بسته‌کردن دهان است. بنابراین انتظار می‌رود تفاوت‌های فردی در حرکت‌های اندام‌های گویایی و نیز تفاوت‌های آناتومیک مربوط به ساختار دستگاه آوایی افراد منجر به تغییر توزیع انرژی در سطح سیگنال و در نتیجه تغییر مقادیر شدت در سیگنال‌های آوایی افراد شود. در پژوهش حاضر قصد داریم با استفاده از ابزارهای آواشناسی آزمایشگاهی و با رویکرد تشخیص هویت گوینده پارامترهای آکوستیکی ریتم گفتار را بر مبنای مشخصه‌های شدت-بنیاد در زبان فارسی بررسی کنیم. هدف این پژوهش بررسی توانایی پارامترهای شدت-بنیاد در سطح هجا جهت تشخیص هویت گویندگان فارسی‌زبان است. دو دسته از پارامترهای شدت-بنیاد یکی بر مبنای میانگین شدت هجا (stdevM, varcoM, rPVIm, rPVIm) و دیگری بر مبنای شدت قله هجا (stdevP, (stdevP, varcoP, rPVIp, nPVIp) برای بررسی و تحلیل داده‌ها انتخاب شد. صدای دوازده گویشور مرد فارسی‌زبان در محیط آزمایشگاهی و در دو نوبت مجزا به فاصله یک تا دو هفته ضبط شد. نمونه‌های آوایی با استفاده از برنامه Praat ویرایش ۳، ۴، ۵، مورد تجزیه و تحلیل آکوستیکی قرار گرفت و تحلیل آماری داده‌ها و مقادیر به دست آمده از بررسی آکوستیکی نمونه‌های آوایی با استفاده از نرم‌افزار SPSS ویرایش ۲۱/۰ و نرم‌افزار R ویرایش ۳، ۳ صورت گرفت. نتایج نشان داد پارامترهای آکوستیکی ریتم گفتار بر مبنای نوسانات شدت نقش بسزایی در تشخیص هویت گویندگان فارسی‌زبان دارند. افزون بر آن، مشخصه‌های زبان-ویژه تأثیری در توانایی این پارامترها در تشخیص هویت گوینده ندارند. همچنین نتایج حاکی از آن بود که پارامترهای مبتنی بر شدت قله هجا ویژگی‌های فرد ویژه بیشتری را انعکاس می‌دهند.

واژه‌های کلیدی: آواشناسی آزمایشگاهی، ریتم گفتار، پارامترهای شدت-بنیاد، تشخیص گوینده.

۱- مقدمه

تعریف ریتم از ابتدای پیدایش این مفهوم تا به امروز همواره مبحثی چالش‌برانگیز بوده است و تاکنون پژوهشگران موفق به ارائه تعریفی دقیق و جامع از آن نشده‌اند. در میان

بحث‌های متداول در این باره، نسپور^۱ و همکاران (۲۰۱۱) تعریف افلاطون^۲ از ریتم را جامع‌ترین و در عین حال زیباترین تعریف می‌دانند. افلاطون در جلد دوم کتاب *قوانین*^۳ ریتم را «نظم در حرکت» (order in movement) تعریف می‌کند. زبان‌های مختلف دنیا با عنایت به تفاوت‌های واج‌شناختی که با هم دارند به لحاظ ریتم نیز با هم متفاوت‌اند و موضوع تقسیم‌بندی زبان‌ها به دسته‌های ریتمیک متفاوت (به‌ویژه زبان‌های تکیه-زمان^۴ و زبان‌های هجا-زمان^۵) همواره میان پژوهشگران محل بحث بوده است (پایک^۶: ۱۹۴۵، ابرکرومبی^۷: ۱۹۶۷، راموس: ۱۹۹۹، گریب و لو^۸: ۲۰۰۲، وایت^۹ و ماتیس^{۱۰}: ۲۰۰۷، دلوو^{۱۱}: ۲۰۱۰، آروانی‌تی^{۱۲}: ۲۰۱۲). زبان‌های تکیه-زمان نسبت به زبان‌های هجا-زمان دارای ساختار هجایی پیچیده‌تر هستند و کاهش واکه‌ای نیز در این زبان‌ها رخ می‌دهد. از این‌روی دیرش مناطق واکه‌ای^{۱۳} در این زبان‌ها کم‌تر، در حالی که دیرش مناطق همخوانی^{۱۴} در این زبان‌ها بلندتر است. منظور از منطقه همخوانی، منطقه‌ای است در گفتار که شامل یک یا چند همخوان بوده و یک واکه (و یا مکث) قبل و یا بعد از همخوان(ها) قرار می‌گیرد. منطقه واکه‌ای، بالعکس، منطقه‌ای است در گفتار که شامل یک یا چند واکه (یا بخش‌های واکه‌ای مانند واکه مرکب یا واکه سه‌گانه) بوده و یک همخوان (و یا مکث) قبل و یا بعد از واکه(ها) قرار می‌گیرد.

عمده پژوهش‌های انجام‌شده در مورد ریتم گفتار تمرکز خود را به مطالعه و بررسی تغییرات بین-زبانی دیرش در سطح منطقه همخوانی و واکه‌ای معطوف کرده بودند

-
- 1.M. Nespor
 2. Plato
 3. *The Laws*
 4. stressed-timed
 5. syllable-timed
 6. K. L Pike
 - 7.D. Abercrombie
 - 8.E.L. Low
 - 9.L. White
 10. S.L. Mattys
 - 11.V. Dellwo
 - 12.A. Arvaniti
 - 13.vocalic interval
 14. consonantal interval

(پایک: ۱۹۴۵، ابرکرومبی: ۱۹۶۷، گریب و لو: ۲۰۰۲، راموس^۱ و همکاران: ۱۹۹۹، دلوو: ۲۰۰۶، تیلسن^۲ و آروانی‌تی: ۲۰۱۳). تا به امروز شواهد زیادی نشان داده است که ویژگی‌های دیرشی منطقه همخوانی و منطقه واکه‌ای، همبسته‌های تغییرات ریتمیک بین-زبانی هستند (وایت^۳ و ماتیس^۴: ۲۰۰۷، دلوو: ۲۰۱۰، لوکینا^۵ و همکاران: ۲۰۱۱ آروانی‌تی: ۲۰۱۲). افزون بر نشان دادن تغییرات بین-زبانی، پارامترهای آکوستیکی ریتم گفتار بر اساس دیرش مناطق همخوانی و واکه‌ای و همچنین هجا می‌توانند افراد یک زبان خاص را از هم تمیز دهند. با وجود این، مطالعات اخیر تأکید کرده‌اند که برای درک بهتر این تغییرات باید جنبه‌های مختلفی که ریتم گفتار همراه با آن تغییر می‌کند نیز مدنظر قرار گرفته شود (فوکس^۶، ۲۰۱۶: ۷۰). هی^۷ و دلوو (۲۰۱۶) معتقدند اگرچه دیرش مناطق آوایی مانند آواهای منفرد یا هجا در ریتم گفتار قابل‌ملاحظه بوده‌اند، اما تاکنون توجه کمتری به این مسئله شده است که ریتم یک سیگنال آوایی تنها بر مبنای روابط دیرشی میان این مناطق نیست بلکه به تفاوت‌های مربوط به شدت^۸ در آن‌ها نیز بستگی دارد. مدل‌های ریتم گفتار که بر تفاوت‌های بین-گوینده بر اساس پارامترهای شدت-بنیاد ریتم گفتار تمرکز کرده باشند، بسیار اندک است. هی و دلوو (۲۰۱۴)، معتقدند پارامتر آکوستیکی شدت با سیکل‌های بازوبسته‌شدن دهان در ارتباط است. بنابراین انتظار می‌رود در صورتی که گویندگان در میزان باز و بسته‌کردن اندام گفتاری دهان متفاوت عمل کنند، این تفاوت‌ها در سیگنال‌های آوایی نیز رمزگذاری شود. آن‌ها معتقدند با توجه به همبستگی میان شدت و میزان باز و بسته‌کردن دهان، مشخصه‌های فرد ویژه بیشتری در پارامترهای آکوستیکی مربوط به شدت یافت می‌شود. همچنین حرکت دهان و حالت دهان نیز با میزان دامنه^۹ صدا مرتبط است مثلاً وقتی دهان بازتر است دامنه صدا معمولاً بلندتر است (گارنیر^{۱۰} و همکاران: ۲۰۰۸،

-
- 1 .F. Ramus
 - 2 .S. Tilsen
 - 3 .L. White
 - 4 .S. Mattys
 5. A. Loukina
 6. R. Fuchs
 7. L. He
 - 8 . intensity
 - 9 . amplitude
 - 10 .M. Garnier

چاندرااسکاران^۱ و همکاران: ۲۰۰۹). پس می‌توان ابراز داشت تفاوت‌های فردی در حرکت‌های اندام‌های گویایی مانند باز و بسته کردن دهان، حرکت لب‌ها، میزان جابه‌جایی آرواره و نیز تفاوت‌های آناتومیکی که در ساختار دستگاه آوایی افراد است می‌تواند منجر به تغییر مقادیر شدت در سیگنال‌های آوایی افراد بشود. اریکسون^۲ و همکاران (۲۰۱۵) نیز معتقدند که ساختار نوایی گفتار به لحاظ آوایی تا حد زیادی تحت تأثیر الگوهای حرکتی آرواره است. آن‌ها معتقدند که این الگوهای حرکتی آرواره از هر زبانی به زبان دیگر متفاوت است. این بدان معناست که گویندگان زبان‌های مختلف در نحوه حرکت دادن اندام‌های گویایی‌شان متفاوت عمل می‌کنند و این تفاوت در الگوهای حرکتی می‌تواند در سیگنال‌های گفتاری نیز انعکاس یابد. با این حال، برخی از این تفاوت‌ها از شکل و اندازه دستگاه گفتار سرچشمه می‌گیرد و لذا فرد کنترل کمتری روی آن‌ها دارد. به عبارتی، ساختار دستگاه گفتار به مثابه یک صافی عمل می‌کند که باعث ایجاد تغییراتی در مقادیر مربوط به فیزیک آواهای گفتار می‌شود. بنابراین می‌توانیم استدلال کنیم هر چه پارامترهای آکوستیکی بیشتر تحت تأثیر مشخصه‌های یک زبان باشند، آن پارامترها وابستگی بیشتری به ویژگی‌های اکتسابی فرد در نحوه حرکت دادن اندام‌های گویایی دارد و هر چه پارامترهای آکوستیکی نسبت به ویژگی‌های زبان-ویژه حساسیت کمتری داشته باشند، آن پارامترها بیشتر وابسته به آناتومی دستگاه گفتار هستند.

با توجه به مباحثی که تاکنون مطرح شد در پژوهش حاضر قصد داریم پارامترهای آکوستیکی ریتم گفتار بر مبنای مشخصه‌های شدت-بنیاد را در گفتار استخراج‌شده از گویندگان فارسی‌زبان بررسی کنیم. به این منظور دو دسته از پارامترهای شدت-بنیاد که بر مبنای میانگین شدت هجا^۳ و شدت قله هجا^۴ استخراج‌شده برای بررسی تفاوت-های بین-گوینده در فارسی‌زبانان انتخاب شده است. پژوهش حاضر به دنبال یافتن پاسخ برای پرسش‌های زیر است:

(۱) آیا مشخصه‌های شدت-بنیاد ریتم گفتار در زبان فارسی نیز فرد ویژه هستند؟

-
1. Ch. Chandrasekaran
 2. D. Erickson
 3. mean syllable intensity
 4. peak syllable intensity

۲) کدام پارامتر شدت-بنیاد بهتر می‌تواند تغییرات بین-گوینده را در زبان فارسی توضیح دهد؟

۳) اطلاعات فرد ویژه در پارامترهای شدت-بنیاد در چه سطوحی از هجا بهتر توزیع شده است؟

با توجه به این که پارامترهای شدت-بنیاد ریتم گفتار تحت تأثیر الگوهای واج‌آرایی و ساختار هجایی یک زبان هستند بنابراین تحلیل و بررسی این مشخصه‌ها در زبان‌های مختلف می‌تواند به درک بهتر ماهیت این پارامترها کمک کند و نیز این مسئله را روشن سازد که آیا این پارامترها می‌توانند فارغ از تأثیر عوامل زبان-ویژه به‌عنوان پارامترهای زبان-آزاد و جهانی در تشخیص هویت گوینده کاربرد داشته باشند یا خیر. دستاوردهای این پژوهش افزون بر آواشناسی آزمایشگاهی قابلیت به‌کارگیری در زمینه‌های مختلفی از جمله آواشناسی حقوقی و تشخیص خودکار گوینده را نیز دارد.

۲- پیشینه پژوهش

تفاوت‌های ریتم گفتار بر مبنای دیرش مناطق همخوانی و واکه‌ای با توجه به عامل گوینده در زبان‌های مختلف از جمله انگلیسی، آلمانی، آلمانی سوئیسی و فارسی بررسی شده است. ویگت^۱ و همکاران (۲۰۱۰) در پژوهش خود به بررسی پارامترهای آکوستیکی ریتم گفتار بر اساس دیرش در سطح مناطق همخوانی و واکه‌ای در زبان انگلیسی پرداختند. آن‌ها سه پارامتر آکوستیکی $V\%$ ، $varcoV$ و $nPVI-V$ را در شش گوینده انگلیسی‌زبان (۳ زن، ۳ مرد) با گویش معیار جنوبی بررسی کرده‌اند. نتایج یافته‌های ویگت و همکاران (۲۰۱۰) نشان داد که اثر پارامترهای آکوستیکی $V\%$ و $varcoV$ در ایجاد تمایز میان گویندگان معنی‌دار بوده است.

لیمن^۲ و همکاران (۲۰۱۴) در پژوهش خود به بررسی تأثیر ویژگی‌های زبرزنجیری زمانی در نشان دادن تفاوت‌های بین-گوینده در زبان آلمانی پرداخته‌اند. پارامترهای آکوستیکی $varcoC$ ، $varcoV$ ، $nPVI-V$ ، $nPVI-C$ ، $varcoPeak$ و $nPVI-Peak$ برای انجام این پژوهش انتخاب شد. شواهد به دست آمده از این پژوهش نشان داد که با وجود حضور دو عامل به‌عنوان منابع تغییرات درون-گوینده یعنی سبک گفتار و نیز کانال ارتباطی، گویندگان به لحاظ مشخصه‌های زبرزنجیری زمانی با هم تفاوت معنی‌دار داشته‌اند.

1. L. Wiget

2. A. Leeman

دلوو و همکاران (۲۰۱۵) نیز در پژوهش خود به بررسی تغییرات بین-گوینده در دو پیکره آوایی پرداختند که به لحاظ سرعت تولید و ویژگی‌های ساختاری زبان‌شناختی دارای تغییرات درون-گوینده بالا بودند. آن‌ها پارامترهای آکوستیکی دیرش-بنیاد $\Delta C(\ln)$ ، $\Delta V(\ln)$ و $\Delta peak(\ln)$ را برای تحلیل و بررسی‌های خود انتخاب کردند. نتایج نشان داد جمله‌ها به لحاظ پیچیدگی در ساختار مناطق همخوانی و واکه‌ای‌شان با هم تفاوت داشتند و این تفاوت‌ها بر مقادیر ریتمیک اثرگذار بوده است. در پایان این پژوهش نگارندگان استدلال می‌کنند که منبع اصلی تفاوت‌های ریتمیک بین-گوینده از تفاوت‌های مربوط به ساختار دستگاه آوایی افراد و نیز حرکت‌های اندام‌های گویایی سرچشمه می‌گیرد.

اسدی و همکاران (۲۰۱۸، ۲۰۱۷) نیز در پژوهش خود پارامترهای ریتم گفتار را بر اساس مناطق همخوانی و واکه‌ای در دو پیکره آوایی زبان فارسی با منبع تغییرات درون-گوینده متفاوت بررسی کردند. نتایج این پژوهش نشان داد که پارامترهای آکوستیکی دیرش-بنیاد قابلیت تمایز گویندگان فارسی‌زبان را دارند. آن‌ها همچنین به این نتیجه رسیدند که پارامتر آکوستیکی ۷٪ می‌تواند به‌عنوان یک پارامتر جهانی در تشخیص هویت گویندگان زبان‌های مختلف به کار گرفته شود.

هی و دلوو (۲۰۱۴، ۲۰۱۶) معتقدند ویژگی‌های منحصر به فرد افراد در نحوه حرکت دادن اندام‌های گویایی‌شان می‌تواند افزون بر تأثیرگذاری بر ویژگی‌های دیرشی سیگنال‌ها، بر جنبه‌های مربوط به شدت در سیگنال‌های آوایی نیز اثر بگذارد. هی و دلوو (۲۰۱۴) با مطرح کردن این ادعا توجه مطالعات بین-گوینده را به سوی مشخصه‌های شدت-بنیاد ریتم گفتار نیز کشاندند. بدین ترتیب آن‌ها پارامترهای آکوستیکی شدت-بنیاد در ریتم گفتار را مطابق با پارامترهای آکوستیکی دیرش-بنیاد طراحی کردند و به بررسی این پارامترها پرداختند. داده‌های آوایی پژوهش آنان از پیکره آوایی مربوط به گویندگان آلمانی استخراج شد. ویژگی‌های مربوط به تغییرات شدت (پارامترهای مربوط به میانگین و پارامترهای مربوط به قله هجا) در سطح هجا اندازه‌گیری شد. نتایج این پژوهش نشان داد تفاوت فراوانی میان گویندگان به لحاظ مشخصه ریتمیک شدت-بنیاد

وجود دارد و این مشخصه‌ها در مقایسه با پارامترهای دیرش- بنیاد حامل تغییرات فرد ویژه بیشتر هستند. حی و دلوو (۲۰۱۶) برای اثبات ادعای خود به مطالعه بیشتر پرداختند و در پژوهش دیگر خود از طریق اندازه‌گیری تغییرات شدت در سطح هجا در سیگنال‌های آوایی به بررسی ویژگی‌های فرد ویژه پرداختند. آن‌ها با استناد به پریر^۱ (۲۰۱۲) و دلوو و همکاران (۲۰۱۵) ابراز می‌دارند که تفاوت‌های فردی از دو منبع سرچشمه می‌گیرند؛ یکی تفاوت‌های ذاتی که در ساختار آناتومیک اندام‌های گویایی هر فرد وجود دارد و دیگری تفاوت‌های ناشی از ویژگی‌های اکتسابی. بنابراین این امکان وجود دارد که پاره‌ای از ویژگی‌های حرکتی منحصر به فرد افراد بر جنبه‌های مربوط به شدت نیز تأثیر بگذارند. بدین منظور، آن‌ها پارامترهای آکوستیکی شدت-بنیاد $stdevM$ ، $stdevP$ ، $varcoM$ ، $varcoP$ ، $nPVIm$ ، $rPVIm$ ، $rPVIp$ و $nPVIp$ را با در نظر گرفتن گوینده به‌عنوان متغیر مستقل پژوهش بررسی کردند. نتایج پژوهش نشان داد پارامترهای آکوستیکی شدت-بنیاد در هر دو پیکره آوایی گویندگان را به‌طور معنی‌داری از هم متمایز ساخت. همچنین بر اساس شواهد به دست‌آمده از این پژوهش، پارامترهای آکوستیکی شدت- بنیاد عملکرد بهتری نسبت به پارامترهای دیرش-بنیاد داشته‌اند.

۳- روش‌شناسی پژوهش

در بخش‌های زیر اطلاعات مربوط به شرکت‌کنندگان، نحوه ضبط داده‌های آوایی، تقطیع داده‌ها و نیز پارامترهای انتخاب‌شده برای پژوهش حاضر توضیح داده می‌شود.

۳-۱- شرکت‌کنندگان و داده‌های آوایی

به‌منظور کشف پارامترهای آکوستیکی تمایزدهنده میان گویشوران فارسی‌زبان، پیکره‌ای آوایی در محیط آزمایشگاهی ضبط گردید. در این پیکره آوایی صدای دوازده گویشور مرد فارسی‌زبان در دو جلسه مجزا ضبط شد. گویشوران به زبان فارسی معیار با لهجه تهرانی صحبت می‌کردند. متوسط سن شرکت‌کنندگان $\pm SD$ عبارت بود از $4/28 \pm 1/6$ سال و محدوده سنی آنان میان ۲۲ تا ۳۵ سال بود. علت ضبط داده‌های آوایی در دو جلسه مجزا اندازه‌گیری میزان تغییرات درون-گوینده پارامترهای آکوستیکی شدت-بنیاد ریتم گفتار با توجه به عامل گذر زمان بود. ضبط غیر هم‌زمان داده‌ها این امکان را میسر می‌سازد تا نمونه‌های آوایی افزون بر تغییرات بین-گوینده به لحاظ تغییرات درون-گوینده نیز بررسی و تحلیل شود. گویشوران هیچ‌گونه سابقه اختلال گفتاری و

شنوایی نداشتند. گویشوران تعداد ۵۴ جمله فارسی را در دو جلسه جداگانه به فاصله یک تا دو هفته تولید کردند. در مجموع تعداد ۱۲۹۶ (۱۲۹۶ = ۲ تکرار × ۵۴ جمله × ۱۲ گویشور) نمونه آوایی به دست آمد.

۲-۳- شیوه ضبط و تقطیع داده‌ها

صدای آزمودنی‌ها با استفاده از میکروفون رولاند ۴۴۱۰۰ هرتز ضبط شد. میکروفون به صورت مورب و به فاصله ۲۰ سانتی متری از دهان شرکت‌کنندگان قرار گرفت. از آن‌ها درخواست شد که جمله‌ها را بدون آهنگ نشان‌دار به صورت طبیعی و با مکثی حدود ۳ ثانیه میان هر جمله تولید کنند. نمونه‌های آوایی با استفاده از برنامه Praat ویرایش ۳۴، ۲، ۵ (بورسما^۱ و وینینک^۲: ۲۰۱۶) مورد تجزیه و تحلیل آکوستیکی قرار گرفت. متناسب با هر فایل صوتی، یک شبکه متنی ایجاد شد. کل پیکره آوایی به دو لایه تقسیم‌بندی شد که این کار به دو صورت دستی و خودکار صورت گرفت. لایه هجا به صورت دستی توسط نگارنده اول و لایه قله هجا نیز بر اساس قانون رسایی^۳ و با استفاده از یک برنامه خودکار به نام Peak Tier Creator که توسط لی هی^۴ نوشته شده است، ایجاد شد. در این پژوهش مقادیر به دست آمده از پارامترهای شدت-بنیاد ریتم گفتار با استفاده از دو برنامه اندازه‌گیری خودکار Intensity Variability و Intensity Analyzer نوشته لی هی اندازه‌گیری شده است. برنامه‌های فوق بر دو سطح هجا و قله هجا اعمال می‌شوند و مقادیر مربوط به پارامترهای شدت-بنیاد ریتم گفتار را به دست می‌دهند. تحلیل آماری داده‌ها و مقادیر به دست آمده از بررسی آکوستیکی نمونه‌های آوایی با استفاده از نرم‌افزار SPSS (IBM Corp. 2012) و نرم‌افزار R (R Core Team: 2016) صورت گرفت.

۳-۳- پارامترهای شدت-بنیاد ریتم گفتار

در این پژوهش تغییرات شدت بر اساس پراکنش سطوح شدت در هجا و نیز میانگین تفاوت‌های میان سطوح شدت در هجاهای متوالی اندازه‌گیری شده است (شکل ۱). از آنجایی که شدت یک پارامتر آکوستیکی است که به آسانی و با چرخش سر نیز تحت تأثیر قرار می‌گیرد از شاخص تغییرات دودویی (Pairwise Variability Index) برای

1. P. Boersma
2. D. Weenink
3. Sonority rule
4. Lei He

اندازه‌گیری میانگین (nPVI_m و rPVI_m) و تفاوت شدت در قله هجا (nPVI_p و rPVI_p) میان هجاهای متوالی استفاده شده است. با توجه به این که برخی از گویندگان ممکن است نسبت به برخی دیگر از گویندگان بلندتر صحبت کنند از پارامترهای نرمال شده نیز استفاده کردیم. پارامترهای آکوستیکی شدت-بنیاد ریتم گفتار که در این پژوهش انتخاب کردیم عبارت‌اند از پارامترهای شدت-بنیاد بر مبنای میانگین شدت هجا یعنی:

stdevM: انحراف معیار میانگین سطح شدت در هجا

varcoM: ضریب تغییر متوسط سطح شدت در هجا (انحراف معیار نرمال شده)

rPVI_m: شاخص خام تغییرات میانگین سطح شدت در هجاهای متوالی

nPVI_m: شاخص نرمال شده تغییرات میانگین سطح شدت در هجاهای متوالی

و پارامترهای شدت-بنیاد بر مبنای شدت قله هجا یعنی:

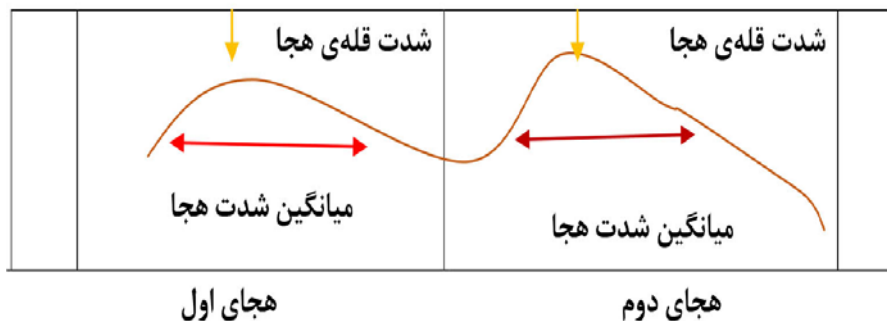
stdevP: انحراف معیار میانگین سطح شدت در قله هجا

varcoP: ضریب تغییر سطح شدت در قله هجا (انحراف معیار نرمال)

rPVI_p: شاخص خام تغییرات میانگین سطح شدت در قله‌های هجاهای متوالی

nPVI_p: شاخص نرمال شده تغییرات میانگین سطح شدت در قله‌های هجاهای

متوالی



شکل ۱: نمایش گرافیکی ترسیم‌کننده واحدهای اندازه‌گیری شدت-بنیاد بر مبنای میانگین شدت هجا و شدت قله هجا

۴- گزارش نتایج

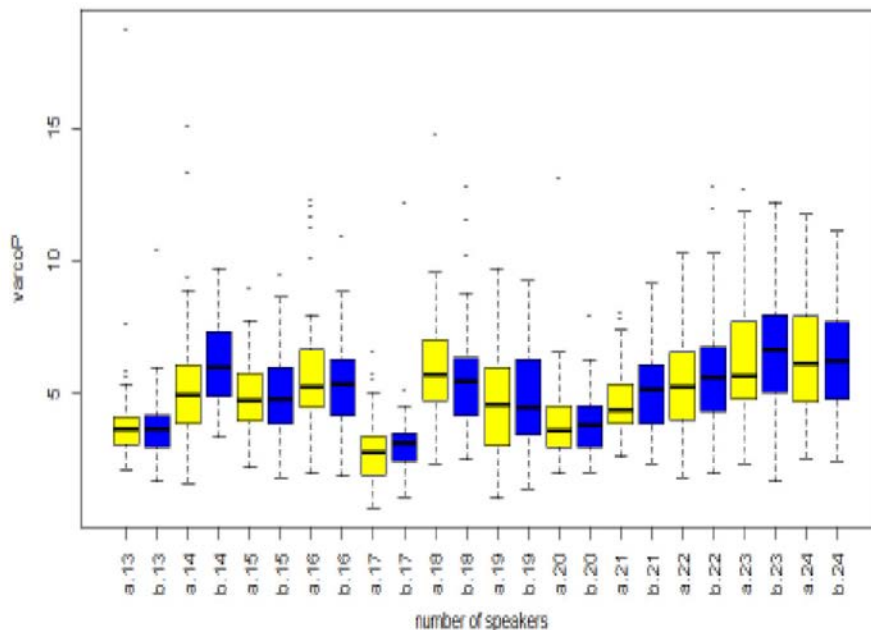
در بخش‌های زیر با توجه به پرسش‌های مطرح شده در بخش مقدمه به اجرای آزمون-های آماری مرتبط پرداخته و نتایج مربوط به تجزیه و تحلیل داده‌ها گزارش خواهد شد.

۴-۱- تغییرات بین-گوینده در پارامترهای شدت-بنیاد ریتم گفتار

به‌منظور بررسی تأثیر عامل گوینده بر متغیرهای آکوستیکی ریتم گفتار، آزمون تحلیل واریانس دوطرفه اجرا شد. گوینده و تکرار به‌عنوان متغیرهای مستقل و پارامترهای آکوستیکی شدت-بنیاد به‌عنوان متغیر وابسته پژوهش در نظر گرفته شدند. جدول ۱ نتایج حاصل از اجرای آزمون آماری تحلیل واریانس دوطرفه را بر پارامترهای شدت-بنیاد نشان می‌دهد. طبق جدول ۱، نتایج حاصل از اجرای آزمون تحلیل واریانس دوطرفه نشان می‌دهد که اثر گوینده بر پارامترهای شدت-بنیاد معنی‌دار بوده است. این بدان معناست که پارامترهای شدت-بنیاد ریتم گفتار میان گویندگان متفاوت بوده است و بنابراین توانایی ایجاد تمایز میان گویندگان فارسی‌زبان را دارند. اثر تکرار (دو جلسه مربوط به ضبط داده‌ها) بر پارامترهای شدت-بنیاد معنی‌دار نبود ($p > 0.05$) که نشان می‌دهد عامل گذر زمان به‌عنوان یک منبع تغییر درون-گوینده تأثیری بر این پارامترها ندارد. در شکل ۲ نمودار جعبه‌ای مربوط به تغییرات بین-گوینده را در پارامتر شدت-بنیاد varcoP مشاهده می‌کنید. همان‌گونه که نمودار نشان می‌دهد گوینده شماره ۱ بیشترین شباهت را به لحاظ پارامتر varcoP با گوینده شماره ۸ و بیشترین تفاوت‌ها را با گویندگان شماره ۳، ۴، ۶، ۱۱ و ۱۲ داشته است. به‌منظور بررسی مقایسه‌های معنی‌دار در پارامتر varcoP، آزمون تعقیبی بن‌فرونی^۱ را اجرا کردیم. طبق نتایج آزمون تعقیبی بن‌فرونی از میان ۱۳۲ مقایسه انجام‌شده در پارامتر varcoP ۸۲ مورد (۶۲ درصد) معنی‌دار گزارش شده است.

جدول ۱: نتایج مربوط تحلیل واریانس دوطرفه در بررسی تغییرات بین-گوینده پارامترهای شدت-بنیاد در گویندگان فارسی‌زبان

معنی‌داری	آماره F	میانگین مربعات	درجه آزادی	پارامتر آکوستیکی ریتم گفتار
$p < 0.001$	۱۳/۱۱	۲۳/۳۵	۱۱	stdevM
$p < 0.001$	۲۶/۳۳	۱۱۶/۴۹	۱۱	varcoM
$p < 0.001$	۷/۱۷	۱۴/۶۵	۱۱	rPVIm
$p < 0.001$	۱۸/۳۷	۹۶/۲۴	۱۱	nPVIm
$p < 0.001$	۲۴/۴۸	۴۱/۵۶	۱۱	stdevP
$p < 0.001$	۳۵/۱۳	۱۲۰/۲۵	۱۱	varcoP
$p < 0.001$	۱۳/۳۷	۱۰/۹۳	۱۱	rPVIp
$p < 0.001$	۲۰/۷۹	۳۲/۱۶	۱۱	nPVIp



شکل ۲: نمودار جعبه‌ای تغییرات بین-گوینده پارامتر شدت-بنیاد varcoP در گویندگان فارسی‌زبان (a) مربوط به جلسه اول ضبط داده‌ها و b) مربوط به جلسه دوم ضبط داده‌هاست.

۲-۴- سنجش توانایی پارامترهای شدت-بنیاد ریتم گفتار در توضیح تغییرات بین-گوینده در ادامه به‌منظور بررسی توانایی این پارامترها در توضیح تغییرات بین-گوینده، آزمون رگرسیون لجستیک^۱ چند اسمی را بر این پارامترها اجرا کردیم. گوینده به‌عنوان متغیر

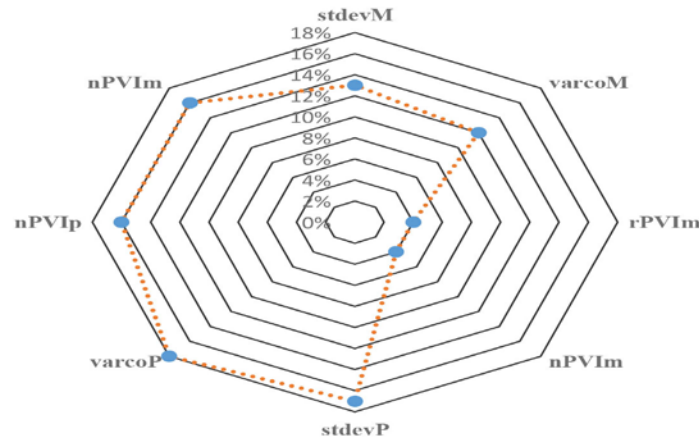
1. multinomial logistic regression

اسمی پاسخ^۱ و پارامترهای شدت-بنیاد به‌عنوان متغیرهای پیش‌بین^۲ وارد آزمون شدند. این آزمون نشان می‌دهد که هر پارامتر به چه میزانی می‌تواند به تشخیص‌گوینده کمک کند. میزان توانایی هر پارامتر در توضیح تغییرات بین-گوینده بر حسب درصد در جدول ۲ آورده شده است. میزان توانایی هر پارامتر با محاسبه مجموع مقادیر خی دو و آنگاه تقسیم مقدار خی دو هر پارامتر بر مجموع مقادیر خی دو و سپس تقسیم نتیجه حاصل بر ۱۰۰ به دست می‌آید.

پارامتر آکوستیکی	مقدار خی دو	درجه آزادی	معنی‌داری	درصد توضیح تغییرات بین-گوینده
stdevM	۶۵/۸۱	۱۱	$p < .0001$	۱۲/۶۵٪
varcoM	۶۳/۴۸	۱۱	$p < .0001$	۱۲/۲۰٪
rPVIm	۲۲/۵۵	۱۱	$p < .0001$	۴/۳۳٪
nPVIm	۲۳/۳۴	۱۱	$p < .0001$	۴/۴۸٪
stdevP	۸۷/۲۸	۱۱	$p < .0001$	۱۶/۷۸٪
varcoP	۹۳/۵۸	۱۱	$p < .0001$	۱۸٪
rPVIp	۸۱/۴۹	۱۱	$p < .0001$	۱۵/۶۶٪
nPVIp	۸۲/۶۰	۱۱	$p < .0001$	۱۵/۸۸٪

همچنین شکل ۲ نمایش گرافیکی توانایی نسبی هر پارامتر در نشان دادن تغییرات بین-گوینده است. نتایج نشان داد که پارامتر varcoP بیشترین قدرت را در توضیح تغییرات ریتمیک بین-گوینده داشته است و پس از آن پارامترهای stdevP و nPVIp بیشترین تغییرات بین گویندگان را نشان داده‌اند. نتایج نشان می‌دهد که پارامترهای شدت-بنیاد بر مبنای قله هجا نسبت به پارامترهای شدت-بنیاد بر مبنای میانگین حاوی اطلاعات فرد ویژه بیشتری هستند. به بیان دیگر، اطلاعات فردویژه در گفتار بیشتر در قله هجا رمزگذاری می‌شود. همچنین نتایج نشان می‌دهد پارامترهای مبتنی بر انحراف معیار مانند stdevP و stdevM و پارامترهای مبتنی بر ضریب تغییرات مانند varcoP و varcoM نسبت به پارامترهای مبتنی بر تغییرات دودویی (PVI) حاوی اطلاعات فردویژه بیشتری هستند.

1. nominal response variable
2. predicting variable



شکل ۳: نمودار رادار نشان‌دهنده قدرت پارامترهای شدت-بنیاد در توضیح تغییرات بین-گوینده

۳-۴- تحلیل مؤلفه‌های اصلی^۱

به‌منظور تحلیل همبستگی میان دو نوع پارامترهای شدت-بنیاد یعنی پارامترهای مبتنی بر میانگین شدت هجا و پارامترهای مبتنی بر شدت قله هجا، آزمون تحلیل عامل^۲ (تحلیل مؤلفه‌های اصلی) را اجرا کردیم. این آزمون ابتدا ماتریسی از ضرایب همبستگی را ایجاد می‌کند و سپس از طریق چرخش عامل‌ها حداکثر رابطه بین متغیرها و عامل‌ها را اندازه‌گیری و در نهایت بار عاملی^۳ هر پارامتر را محاسبه می‌کند. بار عاملی کوچک‌تر از ۰/۴ بیانگر این است که پارامتر مورد نظر تأثیر اندکی در توضیح واریانس عامل دارد و از این‌رو نمی‌تواند ذیل آن عامل طبقه‌بندی شود. عامل ۱ شامل پارامترهای مبتنی بر میانگین و عامل ۲ شامل پارامترهای مبتنی بر قله هجا هستند. خانه‌های سایه‌دار در جدول بزرگ‌ترین ارقام مربوط به هر پارامتر (بزرگ‌تر از حد آستانه ۰/۴) را نشان می‌دهند که نشان از همبستگی بالای میان پارامترهای هر عامل را دارد. نتایج نشان می‌دهد که پارامترهای مربوط به میانگین یعنی عامل اول و پارامترهای مربوط به قله یعنی عامل دوم مقوله‌های مستقلی را تشکیل می‌دهند و این نشان می‌دهد که این دو نوع پارامتر متعامد هستند و همبستگی کمی با هم دارند. از این‌رو پارامترهای شدت-بنیاد مبتنی بر میانگین و قله هجا اطلاعات متفاوتی در مورد گوینده رمزگذاری می‌کنند و ترکیب آن‌ها می‌تواند در تشخیص هویت گوینده مفید باشد.

1. principal component analysis
2. factor analysis
3. factor loadings

جدول ۳: نتایج آزمون تحلیل مؤلفه‌های اصلی در نشان دادن استقلال پارامترهای شدت-بنیاد

بار عاملی		پارامترهای شدت-بنیاد ریتم گفتار
عامل ۲	عامل ۱	
۰/۳۰۹	۰/۷۱۲	stdevM
۰/۳۲۷	۰/۷۲۵	varcoM
۰/۰۶۳	۰/۹۷۰	rPVI _M
۰/۱۴۳	۰/۹۵۸	nPVI _M
۰/۹۴۲	۰/۱۴۸	stdevP
۰/۹۳۵	۰/۱۶۹	varcoP
۰/۹۸۰	۰/۰۴۶	rPVI _P
۰/۹۷۵	۰/۰۶۳	nPVI _P

۵- بحث و بررسی

در پژوهش حاضر به بررسی تأثیر پارامترهای شدت-بنیاد ریتم گفتار در تشخیص هویت گویندگان فارسی‌زبان فارسی پرداخته‌ایم. نتایج حاصل از تجزیه و تحلیل داده‌های آوایی نشان داد که اثر گوینده در پارامترهای آکوستیکی شدت-بنیاد varcoM ، varcoP ، nPVI_P ، stdevM ، stdevP ، rPVI_P و nPVI_M معنی‌دار بوده است. همچنین گویندگان در دو جلسه جداگانه از ضبط داده‌ها به لحاظ پارامترهای شدت-بنیاد ثابت عمل کرده بودند که نشان می‌دهد این پارامترها نسبت به عامل گذر زمان به‌عنوان یک منبع تغییر درون-گوینده مقاوم بوده‌اند. از میان پارامترهای موردبررسی varcoP پارامتری بوده است که در نشان دادن تغییرات بین-گوینده بهترین عملکرد را داشته و در عین حال تغییرات درون-گوینده به لحاظ عامل گذر زمان در این پارامتر مشاهده نشده است. varcoP پارامتری نرمال شده است که از طریق اندازه‌گیری ضریب تغییر سطوح شدت در قلّه هجا بر سطح کل یک زنجیره آوایی استخراج می‌شود. نتایج حاکی از آن است که پارامترهای مربوط به قلّه هجا یعنی nPVI_P و varcoP نسبت به پارامترهای مربوط به میانگین عملکرد بهتری در تشخیص گویندگان داشته‌اند. این دو پارامتر عملکرد مشابهی در تشخیص تفاوت‌های بین-گوینده از خود نشان داده‌اند. همچنین در هنگام بررسی دقیق‌تر داده‌ها مشاهده شد تفاوت‌های موجود میان گویندگان لزوماً با یک

پارامتر مشخص نبود. بدان معنا که گاهی برخی گویندگان در پارامتری مانند varcoM مقادیر یکسانی داشته‌اند؛ درحالی‌که در پارامتر آکوستیکی دیگری مانند varcoP با هم متفاوت بوده‌اند. این موضوع نشان می‌دهد که بهتر است در پژوهش‌های مربوط به تشخیص گوینده ترکیب پارامترها نیز در نظر گرفته شود زیرا هر پارامتری می‌تواند حاوی اطلاعاتی باشد که ممکن است در سایر پارامترها وجود نداشته باشد. بنابراین هر چه پارامترها به لحاظ همبستگی از هم دورتر باشند ترکیب بهینه‌تری برای انعکاس ویژگی‌های فردویژه به وجود می‌آورند. بر اساس نتایج این پژوهش، ترکیب پارامترهای مربوط به میانگین هجا و پارامترهای مربوط به قله هجا می‌تواند سرخ مناسبی برای تشخیص هویت گوینده باشد زیرا پارامترهای مربوط به هر دسته اطلاعات متفاوتی را در مورد گویندگان حمل می‌کنند.

تفاوت‌هایی که پارامترهای شدت-بنیاد در سیگنال‌های آوایی گویندگان دارند با میزان باز کردن دهان و نیز حرکت آرواره‌های آن‌ها ارتباط دارد. افراد به هنگام تولید آوا دهان، لب‌ها و آرواره‌هایشان را مانند هم حرکت نمی‌دهند. تفاوت گویندگان در شیوه حرکت دادن اندام‌های گویایی‌شان بر مقادیر شدت در سیگنال‌های آوایی تأثیر می‌گذارد. این تفاوت‌ها در الگوهای مربوط به شدت رمزگذاری می‌شود و نتایجی که از تحلیل داده‌های این پژوهش به دست آمد نشان می‌دهد تفاوت‌هایی که گویندگان در حرکت اندام‌های گویایی خود مانند لب‌ها و یا نوک زبان داشته‌اند بیشتر در قله هجا رمزگذاری شده است. به سخن دیگر، همبسته‌های آکوستیکی حرکات تولیدی اندام‌های گفتار بیشتر در ویژگی‌های مربوط به قله هجا نمایان شده است. ارتباط میان الگوهای حرکتی آرواره و سایر اندام‌های گفتار بر مقادیر شدت در سیگنال‌های آوایی در پژوهش‌های گارنیر و همکاران (۲۰۰۸)، چاندراسکاران و همکاران (۲۰۰۹) و اریکسون و همکاران (۲۰۱۵) تأیید شده است. نتایج این پژوهش همراستا با یافته‌های هی و دلوو (۲۰۱۴، ۲۰۱۶) است. آن‌ها نیز در پژوهش‌های خود که گویندگان آلمانی‌زبان را مورد بررسی قرار داده بودند به این نتیجه رسیدند که پارامترهای ذکرشده قابلیت تشخیص گویندگان متفاوت را دارند. آن‌ها همچنین در پژوهش خود پارامترهای مربوط به قله هجا را دارای قدرت بیشتر در تمایز میان گویندگان عنوان کردند. نتیجه حاصل این موضوع را تأیید می‌کند که تفاوت‌های بین-گوینده فارغ از این‌که گوینده به چه زبانی صحبت می‌کند بیشتر در قله‌های هجا در سیگنال‌های آوایی رمزگذاری می‌شود.

در نهایت باید به این نکته اذعان داشت که زبان فارسی با توجه به این که نسبت به زبان‌های تکیه-زمان مانند انگلیسی و آلمانی ساختار هجایی ساده‌تر دارد و نیز کاهش واکه‌ای در آن به اندازه این زبان‌ها مشاهده نمی‌شود با این حال پارامترهای شدت-بنیاد ریتم گفتار توانسته‌اند تغییرات بین-گوینده را در این زبان نشان دهند. پارامترهای آکوستیکی ریتم گفتار تحت تأثیر ساختار واج‌آرایی زبان‌ها هستند و این انتظار می‌رفت که تفاوت‌های بین-گوینده ریتمیک در زبان‌های هجا-زمان کمتر مشاهده شود زیرا در این زبان‌ها خوشه‌های همخوانی پیچیده و نیز کاهش واکه‌ای وجود ندارد؛ از این روی گویندگان فضای کمتری برای تغییر در اختیار قرار دارند. با این حال نتایج این پژوهش نشان داده است که تغییرات بین-گوینده در زبان هجا-زمانی مانند فارسی نیز معنی‌دار بوده است. بنابراین می‌توانیم نتیجه بگیریم که پارامترهای شدت-بنیاد ریتم گفتار تحت تأثیر ویژگی‌های زبان-ویژه نیستند و می‌توانند به‌عنوان سرنخ‌های آکوستیکی زبان-آزاد در موارد مربوط به تشخیص هویت گوینده به کار گرفته شوند.

۶- نتیجه

پژوهش حاضر به بررسی نقش پارامترهای شدت-بنیاد ریتم گفتار در تشخیص هویت گویندگان فارسی‌زبان پرداخته است. هشت پارامتر آکوستیکی مربوط به میانگین هجا و قله هجا یعنی varcoP ، varcoM ، nPVIp ، nPVIIm ، stdevP ، stdevM ، rPVIp و nPVIIm در پیکره آوایی ضبط‌شده از دوازده گوینده مرد فارسی‌زبان مورد تجزیه و تحلیل آکوستیکی قرار گرفت. نتایج نشان داد که پارامترهای مربوط به قله هجا توانایی بیشتری در نشان دادن تغییرات بین-گوینده داشته‌اند. از این میان پارامتر varcoP نسبت به سایر پارامترها قدرت بیشتری در ایجاد تمایز میان گویندگان داشته است. همچنین نتایج پژوهش حاکی از آن بود که ترکیب پارامترهای مبتنی بر میانگین و قله هجا می‌تواند سرنخ سودمندی در تشخیص هویت گوینده باشد. همچنین بر اساس یافته‌های این پژوهش، پارامترهای شدت-بنیاد ریتم گفتار تحت تأثیر مشخصه‌های زبان-ویژه نیستند. از این یافته می‌توان چنین استنباط کرد که منبع اصلی تفاوت‌های بین-گوینده ریتمیک در پارامترهای شدت-بنیاد بیشتر ناشی از تفاوت‌های ساختاری دستگاه گفتار افراد است.

منابع

- Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Arvaniti, A. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3), 351–373.
- Asadi, H., Nourbakhsh, M., He, L., Pellegrino, E. and 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.
- Asadi, H., He, L., Pellegrino, E. and Dellwo, V. 2017. Between-speaker rhythmic variability in Persian. *The 26th annual conference of the International Association for Forensic Phonetics and Acoustics (IAFPA)*. Split, Croatia.
- Boersma, P. and Weenink, D. 2013. Praat: Doing Phonetics by Computer. <http://www.praat.org>, Accessed 13 July 2013.
- Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A. and Ghazanfar, A.A. 2009. The natural statistics of audiovisual speech. *PLoS Computational Biology*, 5(7), e1000436.
- 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., Leeman, A. and Kolly, M. 2015. Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *Journal of the Acoustical Society of America*, 137:1513-1528.
- Erickson, D., Kim, J., Kawahara, S., Wilson, I., Menezes, C., Suemitsu, A. and Moore, J. 2015. Bridging articulation and perception: The C/D model and contrastive emphasis. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, 1–5. Glasgow, UK.
- Fuchs, R. 2016. *Speech rhythm in varieties of English*. Singapore: Springer.
- Garnier, M., Wolfe, J., Henrich, N. and Smith, J. 2008. Interrelationship between vocal effort and vocal tract acoustics: a pilot study. In *Proceedings of INTERSPEECH*, 2302-2305. Brisbane, Australia.
- Grabe, E. and 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.
- He, L. and Dellwo, V. 2016. The role of syllable intensity in between-speaker rhythmic variability. *International Journal of Speech, Language and the Law*. Vol 23, 243-273.
- He, L., and Dellwo, V. 2014. Speaker idiosyncratic variability of intensity across syllables. In *Proceedings of INTERSPEECH*, 233-237. Singapore.

- IBM Corp. 2012. IBM SPSS Statistics for Windows (version 21.0). Armonk, NY: International Business Machines Corporation.
- Leemann, A., Kolly, M.-J., and Dellwo, V. 2014. Speaker-individuality insuprasegmental temporal features: implications for forensic voice comparison. *Forensic Science International*, 238, 59-67.
- Loukina, A., Kochanski, G., Rosner, B., Keane, E. and Shih, C. 2011. Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America*, 129(5), 3258–3270.
- Nespor, M., Shukla, M. Mehler, J. 2011. Stress-timed vs. syllable-timed languages. In M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice (eds.), *The Blackwell Companion to Phonology*, (pp. 1147–1159).
- Perrier, P. 2012. Gesture planning integrating knowledge of the motor plant's dynamics: a literature review from motor control and speech motor control. In S. Fuchs, M. Weirich, D. Pape and P. Perrier (eds) *Speech Planning and Dynamics* 191–238. Frankfurt am Main: Peter Lang.
- Pike, K. L. 1946. *Intonation of American English*. Ann Arbor: University of Michigan.
- R Core Team 2014. R: A Language and Environment for Statistical Computing (version 3.3.3). R Foundation for Statistical Computing. <http://www.Rproject.org>, Accessed 20 November 2016.
- Ramus, F., Nespor, M. and Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.
- Tilsen, S. and Arvaniti, A. 2013. Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America*, 134(1), 628–639.
- White, L., Mattys, S.L. 2007. Calibrating rhythm: First language and second language studies", *Journal of Phonetics*, 35(4), 501–522.
- Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., and Mattys, S. L. 2010. How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America*, 127(3), 1559–156.

