https://jrl.ui.ac.ir/?lang=en

Journal of Research in Linguistics
E-ISSN: 2322-3413
18(1), 55-70
Received: 22.12.2024 Accepted: 11.01.2025

Research Paper

The influence of speaking rate on the dynamics of vowel formant frequencies in Persian

Homa Asadi*
Assistant professor, Department of Linguistics, Faculty of Foreign Languages, University of Isfahan, Isfahan, Iran
h.asadi@fgn.ui.ac.ir

Abstract
This study investigates the effect of speaking rate variations on the formant frequencies of Persian vowels using acoustic analysis. Initially, the first to fourth formant frequencies were extracted using long-term analysis techniques. Subsequently, a comparative assessment was conducted across three distinct speaking rates: normal, slow, and fast. The statistical significance of differences between these speaking rates was evaluated using multivariate analysis of variance (MANOVA). A random forest algorithm was utilized to assess the capability of formant frequencies in capturing speaker-specific variations across the three speaking rates. Based on the results, the second formant frequency demonstrated the most substantial sensitivity to speaking rate modifications. Regarding the role of speaking rate on speaker-specific characteristics of formant frequencies, it was found that at slow speaking rate, first and third formant frequencies exhibited superior performance in showing speaker-specific characteristics, whereas in normal and fast speaking rates, third and fourth formant frequencies emerged as more significant indicators of individual variations. These findings indicate that acoustic parameters are affected by speaking rate variations and highlight the necessity of considering this factor in acoustic analyses and speaker identification studies.
Keywords: Acoustic phonetics, speech production, speaking rate, vowels, formant frequency

Introduction:

Speech production is a complex, multidimensional process involving precise coordination of dynamic articulatory movements and speech organs to achieve specific targets in the geometry of the vocal tract (Guenther & Hickok, 2016; Kearney & Guenther, 2019). This process, driven by physiological and neurological mechanisms, generates idiosyncratic acoustic features that convey not only linguistic messages but also speaker-specific information such as the speaker's gender, age, emotional state, and health condition. Such speaker-specific details are crucial for applications like speaker identification and forensic analysis. However, speaker identification faces significant challenges due to variability in speech patterns, particularly speaking rates, which can change depending on context, emotional state, or physical condition. For instance, spontaneous conversations often involve faster speaking rates, while slower rates are more common in reading prepared texts. Moreover, individual differences in speaking rates add another layer of complexity. These variations affect key acoustic parameters, such as formant frequencies, which are critical in distinguishing speakers. Despite their importance, the influence of speaking rate on these frequencies, especially in Persian speakers, remains underexplored. This research addresses this gap by systematically investigating how speaking rate (slow, normal, and fast) impacts the formant frequencies of Persian vowels and their effectiveness in speaker identification. Formant frequencies, chosen for their proven ability to differentiate speakers, exhibit high inter-speaker variability and low intra-speaker variability, making them ideal parameters for such analysis. The study also aims to identify which formants (F1–F4) undergo the most significant changes across varying speaking rates and assesses their reliability in encoding speaker-specific information. The findings of this study could enhance the understanding of the effects of

*Corresponding author

https://doi.org/10.22108/jrl.2025.143702.1884

speaking rate on the acoustic properties of voice, offering valuable insights for fields such as acoustic phonetics and speaker identification.

Materials and Methods:

This study employed a controlled experimental design to investigate the effects of speaking rate on formant frequencies and their role in speaker identification. Eighteen male Persian speakers with a Tehrani accent, aged 25-36 years, participated. All participants reported normal hearing and speaking abilities with no history of speech or hearing disorders. Speech data were collected using a standardized method inspired by the BonnTempo Corpus (Dellwo, 2010). Participants read the short, standardized text "*The North Wind and the Sun*" at three distinct speaking rates: slow, normal, and fast. Prior to recording, participants rehearsed the text multiple times to ensure fluency and familiarity. Recordings were conducted in a controlled acoustic chamber using a Shure SM7B microphone with a sampling rate of 44.1 kHz and 16-bit quantization. Audio data were processed using Praat software and the WebMAUS automatic segmentation tool. The first four formant frequencies (F1-F4) were extracted using long-term analysis methods, capturing dynamic changes in formant frequencies over time. Vowels were extracted and concatenated for each speaker to optimize the dataset for analysis. Formant frequencies were calculated using the LPC-Burg algorithm in Praat with 5 ms frame intervals, supported by a custom script. Statistical analyses were conducted using R (version 2.2.4). A MANOVA test was performed to examine differences in mean formant frequencies across the three speaking rates. Shapiro-Wilk tests were conducted to ensure data normality. Significant findings were further analyzed using one-way ANOVA and post-hoc tests. A random forest model was used to assess the ability of formant frequencies to differentiate between speakers. Feature importance analysis within the random forest model identified the formants that most significantly contributed to speaker differentiation at each speaking rate. Parameters with higher mean accuracy in decrease scores were deemed most effective for speaker identification.

Discussion of Results and Conclusions:

Results indicated that formant frequencies exhibited significant variability across different speaking rates, with varying degrees of sensitivity among formants. Consistent with Wismer and Barry (2003), the second formant (F2) demonstrated the highest sensitivity to speaking rate. As F2 is closely associated with the front-back dimension of vowels (Madresi-Qavami, 2014), these findings suggest that speakers may produce vowels with increased frontness at slower rates, potentially due to a greater emphasis on precise articulation. Conversely, the first formant (F1), primarily linked to tongue height and mouth opening (Nourbakhsh, 2013; Madresi-Qavami, 2014), showed the least sensitivity to speaking rate changes. This suggests that speakers maintain greater consistency in tongue height and mouth opening across different speaking rates while adjusting the front-back articulation more significantly.

Regarding speaker-specific characteristics, formant contributions varied across speaking rates. F1 and F3 exhibited the highest speaker-specific characteristics in slow speech. In fast speech, F1 measures were higher, indicating less tongue elevation and greater mouth opening, while slower speech revealed greater individual differences in tongue height and mouth opening. In contrast, F3 and F4 demonstrated better speaker discriminatory potential at normal and fast speaking rates. Notably, F3 and F4 consistently exhibited high speaker-specific characteristics across all speaking rates, corroborating previous findings (Goldstein, 1976; McDougall, 2004, 2006). These results highlight the robustness of F3 and F4 for speaker identification across varying speaking rates.

This study has limitations. The data was collected from only 18 male Persian speakers, limiting the generalizability of the findings. Including female voices and expanding the dataset would provide more comprehensive and robust results. Despite these limitations, the findings have significant implications for speech-processing technologies, speaker identification, and linguistic studies. By considering the acoustic properties of vowels and consonants across different speaking rates, researchers can improve speech and speaker recognition systems and gain a deeper understanding of inter- and intra-speaker variability.

https://doi.org/10.22108/jrl.2025.143702.1884

نشریه پژوهش‌های زبان‌شناسی
سال هجدهم، شماره اول، شماره ترتیبی۳۴، ۱۴۰۵، ص ۵۵-۷۰
تاریخ وصول: ۱۴۰۳/۱۰/۰۲تاریخ اصلاحات: ۱۴۰۳/۱۰/۱۸ تاریخ پذیرش: ۱۴۰۳/۱۰/۲۲

مقاله پژوهشی

تأثیر سرعت گفتار بر دینامیک فرکانس سازه‌های واکه‌های زبان فارسی

* **هما اسدی**

چکیده

پژوهش حاضر با استفاده از تحلیل آکوستیکی، تأثیر تغییرات سرعت گفتار را بر فرکانس سازه‌های واکه‌های زبان فارسی بررسی کرده است. در این پژوهش، ابتدا فرکانس سازه‌های اول تا چهارم واکه‌ها با استفاده از روش بلندمدت استخراج شد و مقایسه‌ای میان مقادیر این فرکانس‌ها در سه سرعت گفتاری (عادی، آهسته و سریع) صورت گرفت. سپس، تفاوت‌های معنادار میان سرعت‌های گفتاری با استفاده از آزمون تحلیل واریانس چندمتغیره مورد بررسی قرار گرفت. در ادامه، با بهره‌گیری از الگوریتم جنگل‌های تصادفی، توانایی فرکانس سازه‌ها در نشان دادن تغییرات فردویژه در سه سرعت گفتاری مختلف ارزیابی شد. نتایج نشان داد که فرکانس سازهٔ دوم بیشترین تأثیر را از سرعت گفتار می‌پذیرد. همچنین، در ارتباط با نقش سرعت گفتار در فردویژگی فرکانس سازه‌ها، مشخص شد که در سرعت آهسته، فرکانس سازه‌های اول و سوم عملکرد بهتری در نشان دادن ویژگی‌های فردویژه دارند؛ اما در سرعت‌های عادی و سریع، فرکانس سازه‌های سوم و چهارم اهمیت بیشتری پیدا می‌کنند. این یافته‌ها حاکی از آن است که پارامترهای آکوستیکی تحت تأثیر تغییرات سرعت گفتار دستخوش تغییر می‌شوند و بنابراین، در تحلیل‌های آکوستیکی و شناسایی گوینده باید به این عامل توجه ویژه‌ای داشت.

کلیدواژه‌ها: آواشناسی آکوستیکی، تولید گفتار، سرعت گفتار، واکه‌ها، فرکانس سازه

* استادیار گروه زبان‌شناسی، دانشکده زبان‌های خارجی، دانشگاه اصفهان، اصفهان، ایران(نویسنده مسئول) h.asadi@fgn.ui.ac.ir

https://doi.org/10.22108/jrl.2025.143702.1884

۱. مقدمه

تولید گفتار پدیده‌ای بسیار پیچیده و چندبعدی است که شامل هماهنگی دقیق حرکات مفصلی پویا و اندام‌های گفتاری برای دستیابی به اهداف مشخص در هندسهٔ مجرای صوتی می‌شود (Guenther & Hickok, 2016; Kearney & Guenther, 2019) . این فرایند، که توسط سازوکارهای فیزیولوژیکی و عصبی هدایت می‌شود، ویژگی‌های آکوستیکی منحصربه‌فردی ایجاد می‌کند که نقش مهمی در انتقال پیام‌های زبانی ایفا می‌کنند. این ویژگی‌های آکوستیکی نه تنها محتوای زبانی گفتار (آنچه گفته شده است) را به شنوندگان منتقل می‌کنند، بلکه اطلاعات غنی دیگری درباره گوینده (چه کسی سخن می‌گوید)، وضعیت عاطفی و حتی شرایط فیزیکی او مانند خستگی یا بیماری ارائه می‌دهند. این اطلاعات، که تحت عنوان «اطلاعات فردویژه»[1] نیز شناخته می‌شوند، نقش کلیدی در شناسایی و تحلیل هویت گوینده دارند.

صدا به‌عنوان یکی از بارزترین ویژگی‌های هویتی فردی، همواره در مرکز توجه محققان حوزه‌های زبان‌شناسی، روانشناسی و علوم کامپیوتر بوده است. ویژگی‌های آکوستیکی صدا، نظیر فرکانس، شدت و طیف زمانی، اطلاعات ارزشمندی درباره گوینده، احساسات او و حتی وضعیت سلامت وی ارائه می‌دهند. یکی از زیربخش‌های مرتبط با شناسایی گوینده، مقایسهٔ قضایی گوینده[2] است. در این فرایند، نمونه‌های صوتی شناخته‌شده و ناشناخته با هدف ارزیابی احتمال تعلق آن‌ها به یک گوینده یا گویندگان مختلف مقایسه می‌شوند. این مقایسه، در صورتی که به روش آکوستیکی انجام شود، از پارامترهایی استفاده می‌کند که به‌خوبی توانایی ارائه اطلاعات فردویژه را داشته باشند. پارامترهای مناسب باید علاوه بر نشان دادن تفاوت‌های میان گویندگان، تغییرپذیری بسیار اندکی نیز در یک گویندهٔ واحد داشته باشند (Nolan, 1983; Rose, 2002).

با این حال مقایسهٔ قضایی گویندگان به‌آسانی صورت نمی‌گیرد و همواره چالش‌هایی پیش روی این حیطه قرار دارد. یکی از این چالش‌های مهم، تغییرات درون‌گوینده‌ای ناشی از سرعت گفتار است. گویندگان بسته به شرایط ارتباطی و زمینه‌های مختلف، سرعت گفتار خود را تغییر می‌دهند؛ برای مثال، ممکن است در مکالمات بداهه سریع‌تر ولی در خواندن متون آماده، آهسته‌تر صحبت می‌کنند. البته گویندگان به‌طور ذاتی هم در سرعت گفتار با هم متفاوت هستند. مثلاً برخی گویندگان سریع‌تر صحبت می‌کنند؛ درحالیکه، برخی دیگر سرعت آهسته‌تری دارند. همچنین، عواملی مانند بیماری‌های حنجره، سن یا حالات عاطفی آنی نیز می‌توانند تغییرات بیشتری در سرعت گفتار ایجاد کنند. این تغییرات ممکن است تأثیر قابل‌توجهی بر فرکانس سازه‌ها و دیگر ویژگی‌های آکوستیکی صدا داشته باشند و در نتیجه بر دقت شناسایی گوینده نیز اثرگذار شوند. با این حال، مطالعات مقایسه‌ای منظم درباره اثرات سرعت گفتار بر فرکانس سازه‌ها و شناسایی گوینده، به‌ویژه در زبان فارسی، هنوز محدود باقی مانده‌اند.

پژوهش حاضردر همین راستا به بررسی تأثیر سرعت گفتار بر فرکانس سازه‌های واکه‌های زبان فارسی می‌پردازد. هدف این پژوهش، شناسایی تفاوت‌های فرکانس سازه‌های واکه‌ها در سرعت‌های مختلف (آهسته، عادی و سریع) و ارزیابی تأثیر این تفاوت‌ها بر توانایی این پارامترها در شناسایی گوینده است. پرسش‌های اصلی این تحقیق عبارتند از:

۱. آیا فرکانس سازه‌های واکه‌های زبان فارسی در سرعت‌های مختلف تفاوت معناداری دارند؟

۲. کدامیک از فرکانس سازه‌ها یعنی فرکانس سازهٔ اول، فرکانس سازهٔ دوم، فرکانس سازهٔ سوم و فرکانس سازهٔ چهارم بیشترین تغییرات را در سرعت‌های گوناگون نشان می‌دهند؟

۳. آیا تغییرات سرعت گفتار بر فرکانس سازه‌ها بر توانایی این پارامترها در شناسایی گوینده تأثیرگذار است؟

این پرسش‌ها در داده‌های آوایی طراحی‌شده‌ای که شامل نمونه‌هایی با سرعت‌های مختلف گفتاری است، مورد آزمایش قرار می‌گیرند. انتخاب فرکانس سازه‌ها به‌عنوان پارامترهای مورد بررسی در این پژوهش چند دلیل دارد. اول اینکه این پارامترها توانایی بالقوه‌ای در تمایز

[1] speaker-specific information (indexical information)
[2] forensic voice comparison

https://doi.org/10.22108/jrl.2025.143702.1884

گویندگان نشان داده‌اند (اسدی و علی‌نژاد ۱۳۹۹؛ Alderman, 2005 ; McDougall, 2006; Jessen & Becker, 2010; Gold et al., 2013; Goldstein, 1976) و از سوی دیگر، تغییرات میان‌گوینده‌ای آن‌ها بالا و تغییرات درون‌گوینده‌ای آن‌ها کم بوده است (Rose, 2002; Gold et al., 2013). همچنین، فرکانس سازه‌ها ارتباط مستقیمی با عوامل تولیدی گفتار دارند. برای مثال، فرکانس سازهٔ اول با افراشتگی زبان رابطهٔ عکس دارد و فرکانس سازهٔ دوم با میزان پیشین و پسین‌شدگی در ارتباط است (Ladefoged, 2006). فرکانس سازه‌های سوم و چهارم نیز بیشتر نشانگر اطلاعات فردویژه‌اند (Rose, 2002; Gold et al., 2013; Asadi et al., 2018). با توجه به اینکه فرکانس سازه‌ها بازتاب‌دهندهٔ بخشی از حرکات تولیدی دستگاه گفتار هستند و در عین حال اطلاعات فردویژه را نیز به‌خوبی نمایان می‌سازند، انتظار می‌رود سرعت گفتار، که به‌طور مستقیم با تغییرات در حرکات اندام‌های تولیدی مرتبط است، بر مقادیر فرکانس سازه‌ها تأثیر بگذارد و از این روی تصور می‌شود که شناسایی گوینده نیز تحت تاثیر این امر قرار بگیرد.

انتظار می‌رود که نتایج حاصل از این پژوهش بتوانند به درک عمیق‌تر از تأثیر سرعت گفتار بر ویژگی‌های آکوستیکی صدا کمک کرده و در زمینه‌هایی مانند آواشناسی آکوستیکی و شناسایی گوینده مورد استفاده قرار گیرند. این نتایج می‌توانند گامی مؤثر در بهبود دقت سیستم‌های شناسایی گوینده و توسعهٔ روش‌های تحلیل صدا در این سامانه‌ها نیز باشند.

۲. پیشینهٔ پژوهش

تحقیقات آزمایشگاهی مرتبط با اندازه‌گیری حرکات تولیدی اندام‌های گویایی در سرعت‌های مختلف، پیچیدگی و ظرافت سازوکارهای تولید گفتار انسان را بیش از پیش آشکار می‌کنند. این مطالعات نشان داده‌اند که تغییرات سرعت گفتار بر حرکات مفصلی اندام‌های گفتاری، از جمله لب‌ها، فک و زبان، تأثیر قابل‌ملاحظه‌ای دارد و این تأثیر در سطوح مختلف آوایی و حرکتی قابل مشاهده است. از جمله مطالعات پیشگام در این حوزه، می‌توان به پژوهش گی[1] و هیروز[2] (1973) اشاره کرد که به بررسی درک سازوکارهای تولید گفتار بر تولید همخوان‌های لبی پرداختند. آن‌ها در پژوهش خود از آزمایش االکترومایوگرافی[3] و تصویربرداری مستقیم و پرسرعت حرکتی، به صورت ترکیبی، برای توصیف تأثیر سرعت گفتار بر تولید همخوان‌های لبی استفاده کردند. یافته‌ها نشان داد افزایش سرعت گفتار با افزایش سطح فعالیت عضله و همچنین افزایش سرعت حرکت اندام‌های تولیدی گفتار همراه است. آن‌ها همچنین نشان دادند که افزایش سرعت گفتار با تغییرات پیچیده‌ای در فعالیت عضلانی همراه است. این تغییرات شامل بسته‌شدن بیشتر لب‌ها و بازشدن بیشتر دهان در همخوان‌های دولبی و واکه‌های گرد می‌باشد که نشان‌دهندهٔ تطبیق پیچیدهٔ دستگاه گفتار با سرعت‌های گفتاری متفاوت است. در پژوهشی دیگر، شایمن[4] و همکاران (1997) با بررسی دقیق پیکربندی لب‌ها به نکات جالب‌توجهی دست یافتند. آن‌ها دریافتند که نوع بیرون‌زدگی لب[5] در سرعت‌های گفتار مختلف ثابت نیست و تغییرات قابل‌توجهی دارد. به‌طور خاص، سرعت‌های گفتار کندتر عدم تقارن، بی‌نظمی و تغییرات بیشتری را در شکل هندسی لب نسبت به سرعت‌های عادی و سریع نشان دادند.

علاوه بر تفاوت‌های مشاهده‌شده در حرکات اندام‌های گفتاری، مطالعات آواشناختی متعددی نیز نشان داده‌اند که سرعت گفتار می‌تواند بر ویژگی‌های آکوستیکی گفتار نیز تأثیر بگذارد. به عنوان مثال، ویسمر[6] و بری[7] (2003) در مطالعهٔ خود به بررسی تغییرات مسیر فرکانس سازهٔ دوم پرداختند و نشان دادند که سرعت گفتار به طور نظام‌مندی بر مسیرهای سازه‌ای[8] تأثیر می‌گذارد. آن‌ها همچنین تغییر پایانهٔ فرکانس سازهٔ دوم[9] را به عنوان یک شاخص منفرد قوی معرفی کردند که می‌تواند نشانگر تغییرات سرعت گفتار در واکه‌های ساده

[1] Gay, T.
[2] Hirose, H.
[3] electromyography
[4] Shaiman. S
[5] lip protrusion
[6] Weismer. G
[7] Berry. J
[8] Formant trajectories
[9] F2 offset

https://doi.org/10.22108/jrl.2025.143702.1884

میان گویندگان باشد. در همین راستا، پیترمن[1] (2000) نیز با مطالعهٔ پاره‌گفتارهای تولیدی دو گویشور در ده سرعت گفتاری مختلف به این نتیجه رسید که فرکانس سازهٔ اول و دوم در سرعت‌های بالاتر تغییر می‌کنند. جادن[2] و ویلدینگ[3] (2004) نیز در مطالعهٔ خود که به بررسی تاثیر سرعت گفتار بر تولید آواها پرداخته بودند، نشان دادند که در سرعت گفتاری آهسته نرخ تولید هجا در ثانیه به‌طور قابل‌توجهی نسبت به سرعت عادی و معمول گفتاری کاهش می‌یابد. آن‌ها همچنین نشان دادند که مساحت فضای واکه‌ای در سرعت گفتاری آهسته گسترده‌تر می‌شود. آگویل[4] و همکاران (2009) نیز در مطالعهٔ خود به بررسی تأثیر سرعت گفتار بر هم‌تولیدی واکه‌ها و همخوان‌ها پرداختند. آن‌ها پاره‌گفتارهای تولیدشده توسط شش گویندهٔ آمریکایی-انگلیسی را در سه سرعت گفتاری عادی، سریع و بسیار سریع بررسی کردند. نتایج این پژوهش نشان داد که همخوان‌های لثوی و کامی نسبت به آواهای لبی در حالت گفتاری سریع، کاهش بیشتری در میزان آغازهٔ فرکانس سازهٔ دوم[5] داشته‌اند. در رابطه با واکه‌ها، آن‌ها نشان دادند مساحت فضای واکه‌ای در گفتار سریع‌تر نسبت به گفتار عادی کوچکتر است. مفرد[6] و گرین[7] (2010) نیز در پژوهش خود با هدف بررسی تعامل میان تولید گفتار و ویژگی‌های آکوستیکی گفتار به بررسی تأثیر سرعت گفتار بر حرکات تولیدی و ویژگی‌های آکوستیکی گفتار پرداختند. آن‌ها حرکت زبان را در تولید واکه‌ها در سرعت‌های گفتاری عادی، آهسته، سریع و بلند در ده گویندهٔ سالم بررسی کردند. نتایج نشان داد که جابه‌جایی زبان نقش مهمی در پیش‌بینی فاصلهٔ آکوستیکی واکه‌ها دارد. آن‌ها همچنین نشان دادند که فاصلهٔ فرکانس سازه‌ها در سرعت گفتاری آهسته مشخص‌تر و باثبات‌تر است؛ در حالیکه، این ثبات در سرعت‌های گفتار عادی، آهسته و سریع و نیز گفتار بلند مشاهده نشد.

به طور خلاصه، پژوهش‌های پیشین نشان داده‌اند که تغییرات سرعت گفتار تأثیر قابل توجهی بر جنبه‌های مختلف تولید گفتار دارد. این تأثیر هم بر حرکات فیزیکی اندام‌های گویایی (مانند لب‌ها و فک) و هم بر ویژگی‌های آکوستیکی گفتار (مانند فرکانس‌های سازه و فضای واکه‌ای) مشهود است. با این حال، علی‌رغم اهمیت این موضوع، هنوز پژوهشی جامع که به بررسی دقیق تأثیر سرعت گفتار بر تمام فرکانس‌های سازه در زبان فارسی بپردازد، به‌ویژه از منظر شناسایی گوینده، انجام نشده است. تنها مطالعه‌ای که تاکنون با دیدگاه شناسایی گفتار و گوینده در این زمینه انجام شده، توسط شهربابکی و همکاران (2018) بر روی زبان‌های آلمانی و فرانسوی بوده است. این مطالعه نشان داد که تغییرات سرعت گفتار می‌تواند بر تشخیص همخوان‌ها، واکه‌ها و حتی دقت مدل‌های خودکار تشخیص گوینده تأثیر بگذارد. بر این اساس، پژوهش حاضر به بررسی تأثیر مستقیم سرعت گفتار بر ویژگی‌های آکوستیکی واکه‌های زبان فارسی و همچنین بررسی تأثیر آن بر ویژگی‌های فردویژهٔ فرکانس سازه‌های واکه‌های زبان فارسی می‌پردازد.

۳. روش‌شناسی پژوهش

در این بخش، ابتدا به معرفی شرکت‌کنندگان و توضیح دربارهٔ داده‌های آوایی مورد استفاده در این پژوهش پرداخته خواهد شد. سپس، نحوهٔ تقطیع داده‌های آوایی و آماده‌سازی آن‌ها برای تحلیل آکوستیکی تشریح می‌شود. در ادامه نیز نحوهٔ استخراج فرکانس سازه‌های اول تا چهارم توضیح داده می‌شود و در پایان نیز مراحل تحلیل‌های آماری که برای به‌دست‌آوردن نتایج نهایی طی شدند، به‌تفصیل تشریح می‌گردد.

۳-۱. شرکت‌کنندگان و داده‌های آوایی

در این مطالعه، ۱۸ شرکت‌کنندهٔ مرد فارسی‌زبان با لهجهٔ تهرانی مشارکت داشتند. این افراد در محدودهٔ سنی ۲۵ تا ۳۶ سال قرار داشتند که میانگین سنی آن‌ها ۳۱.۳ سال و انحراف استاندارد سنی ۳.۷ محاسبه شد. تمامی شرکت‌کنندگان از سلامت کامل شنوایی و گفتاری

[1] Pitermann. M.
[2] Tjaden, K.
[3] Wilding, G.
[4] Agwuele, A.
[5] F2 Onset
[6] Mefferd, A. S.
[7] Green, J. R.

https://doi.org/10.22108/jrl.2025.143702.1884

برخوردار بودند و پیشینه‌ای از اختلالات شنوایی یا گفتاری در خود گزارش نکردند. شرکت‌کنندگان همگی دانشجوی مقاطع کارشناسی، کارشناسی ارشد یا دکترا در رشته‌های گوناگون تحقیقاتی بودند. این داده‌های آوایی با بهره‌گیری از روش‌های استاندارد و مشابه با روش پیکرهٔ آوایی بن‌تمپو[1] (Dellwo, 2010) که به زبان آلمانی جمع‌آوری شده بود، طراحی و اجرا گردید. برای جمع‌آوری داده‌ها، از گویندگان خواسته شد تا متن معروف «باد شمال و خورشید» را که یک متن کوتاه و استانداردشده برای مطالعات آوایی است، در سه سرعت گفتاری متفاوت بخوانند. پیش از شروع ضبط‌ها، به‌منظور آشنایی کامل شرکت‌کنندگان با متن و رفع هرگونه خطای احتمالی ناشی از ناآشنایی با واژگان یا ساختار جملات، از آن‌ها خواسته شد چندین بار متن را با دقت بخوانند. ابتدا از گویندگان خواسته شد که متن را با سرعت عادی خود بخوانند. در مرحلهٔ بعد از آن‌ها درخواست شد که سرعت گفتار خود را تا حد ممکن کاهش دهند. در نهایت، از گویندگان خواسته شد که متن را با حداکثر سرعت ممکن بخوانند این روند منجر به تغییرپذیری شدید نرخ هجا در سه نسخه خوانده‌شده از متن شد.

۳-۲. شیوهٔ ضبط و تقطیع داده‌ها

برای ضبط داده‌ها، شرکت‌کنندگان در یک محیط آزمایشگاهی استاندارد و کنترل‌شده قرار گرفتند. ضبط‌ها در یک اتاقک آکوستیک با استفاده از میکروفون مدل SHURE SM7B انجام شد. نرخ نمونه‌برداری روی ۴۴.۱ کیلوهرتز و کمی‌سازی با دقت ۱۶ بیت تنظیم شد. پس از انجام این مرحله، داده‌های آوایی وارد نرم‌افزار پرات[2] (Boersma and Weenink, 2023) شد تا مراحل بخش‌بندی و برچسب‌گذاری آن‌ها انجام گیرد. این فرایند بر اساس اطلاعات شروع و پایان هر بخش آوایی صورت گرفت. به‌منظور خودکارسازی این فرایند، از درگاه WebMAUS[3] استفاده شد که ابزاری مناسب برای پردازش خودکار داده‌های آوایی محسوب می‌شود. در اولین مرحله، برای هر پاره‌گفتار تولیدی توسط شرکت‌کنندگان، یک پروندهٔ متنی با فرمت txt تهیه شد. این پرونده‌های متنی توسط نویسنده براساس توصیه‌نامهٔ مختص زبان فارسی که در تارنمای رسمی این درگاه ارائه شده بود، آماده‌سازی شدند تا داده‌ها در قالب استاندارد و قابل پردازش قرار گیرند. در مرحلهٔ بعد، هر فایل صوتی به‌همراه فایل متنی متناظر خود به سیستم WebMAUS وارد شد. این کار از طریق بخش pipeline without ASR انجام گرفت و گزینهٔ زبان فارسی به‌عنوان زبان هدف انتخاب گردید. پس از اتمام پردازش، سیستم به‌طور خودکار فایل‌های TextGrid متناظر با هر فایل صوتی را تولید و به‌عنوان خروجی ارائه کرد. بدین ترتیب، به ازای هر فایل صوتی یک فایل TextGrid متناظر نیز به دست آمد. در مرحلهٔ نهایی، برای اطمینان از دقت و صحت پردازش داده‌ها، نویسنده به‌صورت دستی نیز فایل‌های TextGrid را بازبینی کرد. این کار با هدف شناسایی و اصلاح خطاهای احتمالی انجام شد؛ چراکه در فرایند خودکار ممکن است به‌دلیل پیچیدگی‌های آوایی یا تداخل‌های سیگنال صوتی، خطاهایی در بخش‌بندی و برچسب‌گذاری رخ دهد. علاوه بر این، به دلیل تغییرات سرعت گفتار (آهسته، عادی و سریع) و پیچیدگی‌های خاص هر سرعت گفتار، فرایند دستی بازبینی در این بخش اهمیت بیشتری پیدا کرد.

۳-۳. نحوهٔ استخراج پارامترهای آکوستیکی

پس از آماده‌سازی و پیش‌پردازش داده‌ها، پارامترهای آکوستیکی فرکانس سازه‌های اول تا چهارم از واکه‌ها استخراج گردید. در این مطالعه، استخراج پارامترها به روش بلندمدت صورت گرفت. این روش، که به‌طور گسترده در مطالعات شناسایی گوینده به کار می‌رود، از اندازه‌گیری دینامیکی (اندازه‌گیری تغییرات پیوسته فرکانس سازه‌ها در طول زمان) در سطح پاره‌گفتار برای نمایش ویژگی‌های آوایی گوینده استفاده می‌کند و دقت بیشتری نسبت به روش‌های کوتاه‌مدت دارد (Asadi et al., 2018; Gold et al., 2013; Moos, 2010). در تحلیل بلندمدت، فرکانس‌های سازه‌ها به‌صورت دینامیکی و با محاسبه میانگین مقادیر فرکانس در قاب‌های چندمیلی‌ثانیه‌ای طی یک بازهٔ زمانی

[1] BonnTempo
[2] Praat
[3] این درگاه، با ارائه امکان تقطیع خودکار داده‌های آوایی، به‌ویژه برای زبان فارسی، ابزاری کاربردی برای پژوهشگران و توسعه‌دهندگان در حوزهٔ پردازش زبان طبیعی فراهم کرده است. این قابلیت از طریق آموزش گسترده بر روی داده‌های صوتی فارسی در دو سال اخیر محقق شده است.

https://doi.org/10.22108/jrl.2025.143702.1884

طولانی از ضبط گفتار به دست می‌آید (Gold et al., 2013; Rose, 2002). این روش به‌دلیل توانایی در ثبت تغییرات پیوسته و طبیعی گفتار، برای نمایش دقیق‌تری از فردویژگی گویندگان و تحلیل رفتار آکوستیکی واکه‌ها بسیار مناسب است. برای اجرای این فرایند، در گام نخست تمامی واکه‌ها از سیگنال‌های صوتی استخراج و به‌صورت پشت‌سرهم مرتب شدند. این مرحله به‌منظور تجمیع کامل بخش‌های آوایی از هر گوینده صورت گرفت تا داده‌های آوایی برای تحلیل‌های بعدی بهینه‌سازی شوند. برای این منظور، از افزونهٔ رایگان Praat Vocal Toolkit (Corretge, 2022) که در محیط نرم‌افزار پرات توسعه یافته، استفاده شد. در گام دوم، مقادیر فرکانس سازه‌ها با استفاده از الگوریتم مبتنی بر LPC-Burg در نرم‌افزار پرات و با قاب‌بندی 5 میلی‌ثانیه به‌طور خودکار استخراج شدند. فرایند استخراج فرکانس‌ها توسط یک اسکریپت از پیش نوشته‌شده که به‌طور اختصاصی برای محیط پرات طراحی شده بود، انجام گرفت.

۳-٤. تحلیل آماری

تمامی تحلیل‌های آماری نمونه‌های آوایی این پژوهش با استفاده از نرم‌افزار R (R Core team, 2022) ویرایش 2.2. 4 صورت گرفت. برای آماده‌سازی داده‌ها برای تحلیل آماری، ابتدا داده‌های پرت با استفاده از معیارهای سه انحراف معیار از میانگین حذف گردیدند. سپس، برای بررسی تفاوت میانگین فرکانس سازه‌های اول تا چهارم در سه سرعت گفتاری (آهسته، عادی و سریع)، از تحلیل واریانس چندمتغیره[1] (MANOVA) استفاده شد. این روش به دلیل توانایی آن در بررسی همزمان چندین متغیر وابسته (فرکانس‌های سازه‌ها) و در نظر گرفتن همبستگی‌های بالقوهٔ میان آن‌ها، انتخاب گردید. متغیرهای وابسته شامل مقادیر فرکانس سازه‌های اول تا چهارم و متغیر مستقل سرعت گفتاری با سه سطح (آهسته، عادی و سریع) بودند. پیش از اجرای تحلیل واریانس چندمتغیره، نرمال بودن توزیع داده‌ها با آزمون شاپیرو-ویلک (Shapiro-wilk) بررسی شد. پس از اجرای تحلیل واریانس چندمتغیره، در صورت مشاهدهٔ تفاوت معنی‌دار (سطح معناداری کمتر از 0.05) در اثر عامل اصلی سرعت گفتار، از تحلیل‌های تکمیلی تحلیل واریانس یک‌راهه با استفاده از آزمون‌های تعقیبی استفاده گردید.

برای بررسی اینکه فرکانس سازه‌های اول تا چهارم در سه سرعت گفتاری مختلف تا چه حد می‌توانند به شناسایی گوینده‌ها کمک کنند، از آزمون جنگل‌های تصادفی[2] استفاده شد. در این تحلیل، متغیر وابستهٔ گوینده (با ۱۸ سطح، که هر سطح مربوط به یکی از گویندگان است) و متغیرهای مستقل شامل مقادیر فرکانس سازه‌های اول تا چهارم در نظر گرفته شدند. علاوه بر این، سرعت گفتاری (آهسته، عادی، سریع) به عنوان یک متغیر کمکی در مدل گنجانده شد. از آنجاییکه هدف این تحلیل بررسی میزان فردویژگی هر فرکانس سازه در هر سرعت گفتاری بوده است، از روش اهمیت ویژگی[3] استفاده شد تا مشخص گردد که کدام فرکانس سازه‌ها در هر سرعت بیشترین تاثیر را در تمایز گویندگان دارند. برای ارزیابی دقیق‌تر، میزان فردویژگی هر فرکانس سازه در سه سرعت گفتاری مختلف محاسبه شد و اهمیت آن‌ها در تمایز گویندگان برای هر سرعت به‌طور جداگانه بررسی گردید. میانگین کاهش دقت[4] در آزمون جنگل‌های تصادفی نشان می‌دهد که کدام پارامتر بیشتر فردویژه است. پارامترهایی که در تمایز گویندگان نقش بیشتری دارند، عدد کاهش دقت میانگین بالاتری دارند. این تحلیل کمک می‌کند تا مشخص شود کدام فرکانس سازه‌ها در هر سرعت گفتاری قدرت بیشتری در شناسایی گوینده دارند.

٤. گزارش نتایج

در این بخش ابتدا گزارش توصیفی مربوط به مقادیر میانگین و انحراف معیار فرکانس سازه‌های اول تا چهارم در سه سرعت گفتاری ارائه می‌شود. سپس، در ادامه با استفاده از آزمون تحلیل واریانس چندمتغیره تاثیر سرعت گفتار بر فرکانس سازه‌های اول تا چهارم بررسی می‌شود. پس از آن نیز، با استفاده از مدل جنگل‌های تصادفی توانایی هر پارامتر موردبررسی در سرعت‌های گفتاری مختلف تجزیه و تحلیل می‌شود.

[1] Multivariate Analysis of Variance
[2] Random Forest
[3] Feature Importance
[4] Mean decrease in accuracy

https://doi.org/10.22108/jrl.2025.143702.1884

٤-١. تحلیل توصیفی تغییرات آکوستیکی میان سرعت‌های مختلف

جدول (۱) آمار توصیفی سرعت‌های گفتاری مختلف را بر حسب دیرش هجا (میلی‌ثانیه) نشان می‌دهد. همچنین، در جدول (۲) آمار توصیفی مربوط به هر کدام از پارامترهای آکوستیکی موردبررسی یعنی فرکانس سازهٔ اول، فرکانس سازهٔ دوم، فرکانس سازهٔ سوم فرکانس سازهٔ چهارم به تفکیک هر کدام از سرعت‌های گفتاری مورد بررسی (آهسته، عادی و سریع) ارائه شده است.

جدول ۱-آمار توصیفی نشانگر میانگین و انحراف معیار دیرش هجا در سرعت‌های گفتاری مختلف (آهسته، عادی و سریع)

Table 1- Descriptive statistics showing the mean and standard deviation of syllable duration at different speaking rates (slow, normal and fast)

سرعت گفتاری	میانگین	انحراف معیار
آهسته	۰.۳۰۶	۰.۱۵۸
عادی	۰.۲۰۰	۰.۰۷۹
سریع	۰.۱۳۹	۰.۰۵۱

جدول ۲- آمار توصیفی نشانگر میانگین و انحراف معیار فرکانس سازه‌های اول تا چهارم بر حسب هرتز در سرعت‌های گفتاری مختلف (آهسته، عادی و سریع)

Table 2- Descriptive statistics showing the mean and standard deviation of the first to fourth formant frequencies in Hz at different speaking rates (slow, normal and fast)

سرعت گفتاری	پارامتر	میانگین	انحراف معیار
آهسته	فرکانس سازهٔ اول	۵۶۳.۲۲	۳۰۲.۶۲
	فرکانس سازهٔ دوم	۱۶۷۵.۴۷	۵۱۳۰۸
	فرکانس سازهٔ سوم	۲۷۶۹.۴۶	۴۷۵.۳۲
	فرکانس سازهٔ چهارم	۳۷۲۸.۵۲	۵۴۸.۰۶
عادی	فرکانس سازهٔ اول	۵۵۷.۰۸۵	۲۷۷.۰۷
	فرکانس سازهٔ دوم	۱۶۰۲.۶۹	۴۷۷.۶۵
	فرکانس سازهٔ سوم	۲۷۲۷.۱۰	۴۶۶.۸۲
	فرکانس سازهٔ چهارم	۳۶۹۶.۶۸	۵۰۴.۵۴
سریع	فرکانس سازهٔ اول	۵۷۶.۵۴	۲۹۸.۸۰
	فرکانس سازهٔ دوم	۱۵۹۷.۳۳	۴۷۱.۶۷
	فرکانس سازهٔ سوم	۲۷۲۷.۰۷	۴۶۰.۴۹
	فرکانس سازهٔ چهارم	۳۶۶۸.۴۲	۵۲۳.۷۰

همانگونه که جدول (۱) نشان می‌دهد دیرش هجا در سرعت‌های گفتاری مختلف متغیر بوده است. دیرش هجا در سرعت آهسته بیشترین و در حالت سریع کمترین مقدار را داشت. بر اساس نتایج جدول (۲)، میانگین فرکانس سازهٔ اول در هر سه سرعت تغییرات کمی دارد، به‌طوری‌که در سرعت آهسته کمی بالاتر از سرعت عادی است، اما در حالت سریع مجدداً اندکی افزایش می‌یابد. انحراف معیار این سازه نشان‌دهندهٔ پراکندگی قابل‌توجهی در داده‌ها، به‌ویژه در سرعت آهسته، است.

در مورد فرکانس سازهٔ دوم، میانگین در سرعت آهسته بالاترین مقدار را نشان می‌دهد و در سرعت عادی و سریع کاهش می‌یابد. پراکندگی داده‌ها در این سازه با توجه به انحراف معیار، در سرعت عادی و سریع نسبتاً مشابه است، اما در سرعت آهسته تفاوت بیشتری وجود دارد. فرکانس سازهٔ سوم میانگین بسیار نزدیکی در هر سه سرعت دارد و تغییرات کمی را نشان می‌دهد. انحراف معیار نیز در هر سه سرعت مقدار مشابهی دارد که نشان‌دهندهٔ یکنواختی نسبی در پراکندگی داده‌ها است. فرکانس سازهٔ چهارم نیز در سرعت‌های مختلف تغییر

https://doi.org/10.22108/jrl.2025.143702.1884

اندکی دارد، به‌طوری‌که در سرعت آهسته کمی بالاتر از سایر سرعت‌ها است. انحراف معیار این سازه نشان می‌دهد که پراکندگی داده‌ها در سرعت‌های آهسته و سریع نسبتاً بالاست، اما در سرعت عادی کمی کمتر است.

٤-٢. بررسی معناداری تغییرات آکوستیکی فرکانس سازه‌های اول تا چهارم میان سرعت‌های مختلف

در ادامه به منظور بررسی تاثیر سرعت گفتار بر مقدار فرکانس سازه‌های اول تا چهارم از آزمون تحلیل واریانس چندمتغیره استفاده شد. فرکانس سازه‌های اول تا چهارم به‌عنوان متغیر وابسته و سرعت گفتاری به‌عنوان متغیر مستقل وارد آزمون تحلیلی شد. نتیجهٔ این آزمون در جدول‌های شمارهٔ (۳) آورده شده است. این جدول چهار عامل اصلی شامل درجهٔ آزادی، میانگین مربعات، آمارهٔ F و سطح معنی‌داری را برای هر سازه نشان می‌دهد.

جدول ۳- نتایج مربوط به آزمون تحلیل واریانس چندمتغیره در بررسی تاثیر سرعت گفتار بر فرکانس سازه‌های اول تا چهارم

Table 3- The results of the multivariate analysis of variance in examining the effect of speaking rates on the first to fourth formant frequencies

پارامتر	درجهٔ آزادی	میانگین مربعات	آمارهٔ F	معنی‌داری
فرکانس سازهٔ اول	۲	۳۸۸۱۶.۵۹	۲۱.۲۸	P<0.0001
فرکانس سازهٔ دوم	۲	۳۶۵۶۷۲.۸۳	۱۹۵.۹۱	P<0.0001
فرکانس سازهٔ سوم	۲	۱۵۸۱۳۶.۹۵	۶۸.۷۷	P<0.0001
فرکانس سازهٔ چهارم	۲	۱۳۸۹۴۶.۴۳	۷۱.۸۲	P<0.0001

همان‌گونه که جدول شمارهٔ (۳) نشان می‌دهد سرعت گفتار تأثیر معناداری بر فرکانس سازه‌های اول تا چهارم داشته است. اما به‌منظور اینکه مشخص شود کدام سرعت‌های گفتاری تفاوت معناداری با هم داشته‌اند از آزمون تعقیبی توکی (Tukey) استفاده کردیم. نتایج این آزمون نشان داد که برای فرکانس سازهٔ اول میان سرعت گفتاری عادی و آهسته تفاوت معنادار نبوده است ($p=0.080$)؛ در حالیکه، در سایر مقایسه‌های دودویی (آهسته با عادی، آهسته با سریع) تفاوت معنادار بوده است ($p<0.001$). برای فرکانس سازهٔ دوم نتایج متفاوتی به‌دست آمد. فرکانس سازهٔ دوم میان سرعت عادی و سریع تفاوت معناداری نداشته است ($p=0.060$)؛ در حالی که، برای سایر مقایسه‌های دودویی تفاوت معنادار گزارش شده است. به همین ترتیب، فرکانس سازهٔ سوم نیز در مقایسهٔ دودویی میان سرعت‌های آهسته با عادی و همینطور سرعت آهسته با سریع تفاوتی معنادار نشان داده است؛ درحالیکه، تفاوت آن میان سرعت عادی با سریع معنادار نبوده است ($p=0.498$). در رابطه با فرکانس سازهٔ چهارم، همهٔ تفاوت‌ها (آهسته با عادی، آهسته با سریع و عادی با سریع) معنادار گزارش شده است ($p<0.001$). همچنین آمارهٔ F نشان می‌دهد کدام پارامتر آکوستیکی بیشترین تأثیرپذیری را در سرعت‌های گفتاری مختلف داشته است. هر چه مقدار این آمار بالاتر باشد بدان معناست که آن پارامتر بیشتر تحت‌تأثیر عامل سرعت گفتار بوده است. با نگاهی به مقادیر این آماره در جدول (۳)، مشخص است که مقدار آمارهٔ F برای فرکانس سازهٔ دوم نسبت به سایر سازه‌ها بیشتر است. این نتیجه نشان می‌دهد که فرکانس سازهٔ دوم بیشترین تأثیرپذیری را در سرعت‌های گفتاری مختلف داشته است.

۴-۳. اندازه‌گیری میزان فردویژگیِ فرکانس سازه‌های واکه‌های در سرعت‌های مختلف

برای اندازه‌گیری میزان توانایی پارامترهای آکوستیکی مورد بررسی در شناسایی گوینده در سرعت‌های مختلف گفتار، از روش اهمیت ویژگی مدل جنگل تصادفی استفاده شد. این روش به‌طور خاص بررسی می‌کند که کدام ویژگی‌ها (در این تحقیق، فرکانس‌های سازه‌های اول تا چهارم) بیشترین تأثیر را در تمایز بین گویندگان دارند. اهمیت ویژگی‌ها معمولاً بر اساس میانگین کاهش دقت اندازه‌گیری می‌شود. میانگین کاهش دقت معیاری است که در مدل‌های جنگل‌های تصادفی برای سنجش اهمیت ویژگی‌ها استفاده می‌شود. این معیار با اندازه‌گیری تأثیر حذف هر ویژگی بر دقت پیش‌بینی مدل تعیین می‌شود. به عبارت دیگر، برای هر ویژگی، دقت مدل پس از حذف آن

https://doi.org/10.22108/jrl.2025.143702.1884

ویژگی محاسبه شده و تفاوت آن با دقت مدل اصلی به‌عنوان مقدار میانگین کاهش دقت در نظر گرفته می‌شود. ابتدا این آزمون بدون در نظر گرفتن تعامل میان سرعت گفتار و فرکانس سازه‌ها برای مجموع سرعت‌ها اجرا شد. سپس به منظور بررسی دقیق‌تر، آزمون برای هر سرعت گفتاری به‌صورت جداگانه انجام گرفت تا میزان تأثیر هر یک هر یک از فرکانس سازه‌ها در تشخیص گوینده مشخص شود. جدول (۴)، درصد درستی تشخیص را برای هر پارامتر، هم در مجموع سرعت‌ها و هم به تفکیک سرعت گفتاری، نشان می‌دهد. این درصدها بر اساس مقدار میانگین کاهش دقت محاسبه شده‌اند. مجموع مقادیر میانگین کاهش دقت برای تمامی ویژگی‌ها (فرکانس سازه‌های اول تا چهارم) در هر گروه (آهسته، عادی، سریع و نیز مجموع سرعت‌ها) محاسبه شد. سپس درصد هر ویژگی با استفاده از فرمول ۱ که در آن $iMDA$ نمایانگر میانگین کاهش دقت برای ویژگی i و $totalA$ نمایانگر مجموع دقت‌های مدل در حالت اصلی است، تعیین گردید.

$$\text{درصد کاهش دقت برای ویژگی } i = 100 \times \left(\frac{iMDA}{totalA}\right)$$

فرمول ۱- محاسبهٔ درصد کاهش دقت برای ویژگی i بر اساس میانگین کاهش دقت

Formula 1- Calculation of the accuracy decrease percentage for feature i based on mean decrease in accuracy

این روش برای تمام گروه‌ها اعمال شد تا سهم نسبی هر ویژگی از میانگین کاهش دقت در هر گروه مشخص شود. این مقادیر نشان‌دهندهٔ اهمیت نسبی هر ویژگی در کاهش دقت میانگین مدل بوده و توضیحات لازم برای تکرارپذیری کامل ارائه شده است.

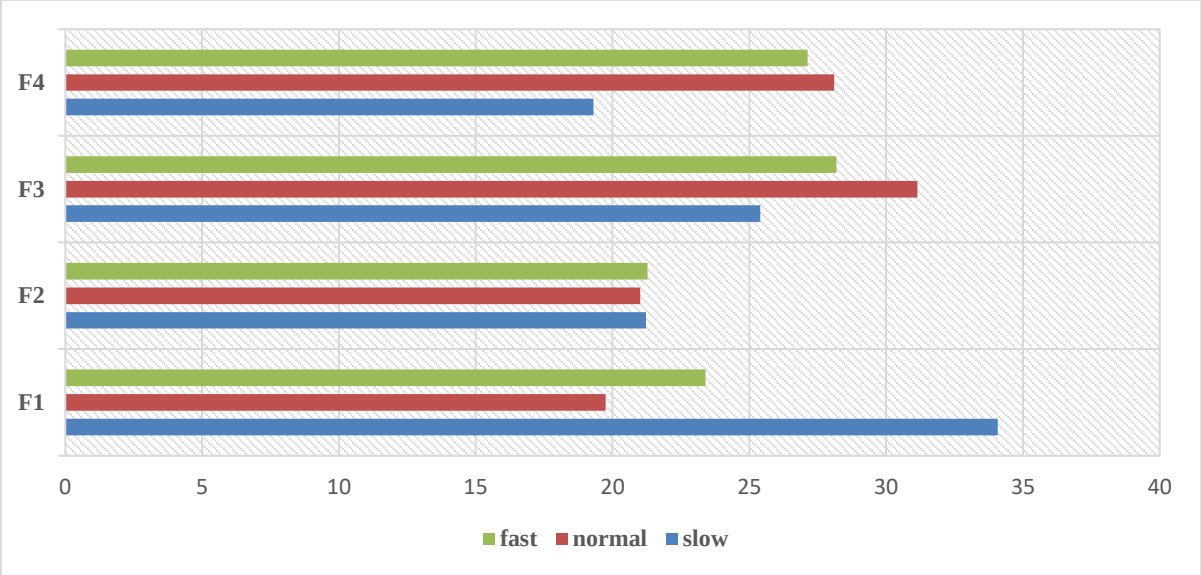
جدول ٤- نتایج حاصل از سنجش اهمیت پارامترها بر اساس میانگین کاهش دقت بر حسب درصد

Table 4- The results of measuring the feature importance based on the mean decrease in accuracy in terms of percentage

پارامتر	آهسته (٪)	عادی (٪)	سریع (٪)	مجموع سرعت‌ها (٪)
فرکانس سازهٔ اول	۳۴.۰۸	۱۹.۷۵	۲۳.۴۰	۲۴.۲۳
فرکانس سازهٔ دوم	۲۱.۲۲	۲۱.۰۱	۲۱.۲۸	۲۰.۴۸
فرکانس سازهٔ سوم	۲۵.۴۰	۳۱.۱۴	۲۸.۱۹	۲۷.۹۹
فرکانس سازهٔ چهارم	۱۹.۳۰	۲۸.۱۰	۲۷.۱۳	۲۷.۳۰

همان‌گونه که نتایج مندرج در جدول شمارهٔ (۴) نشان می‌دهد، میزان فردویژگی فرکانس سازه‌ها در سرعت‌های گفتاری مختلف تغییراتی داشته است. در سرعت گفتاری آهسته، فرکانس سازهٔ اول با ۳۴.۰۸ درصد بیشترین اهمیت را در کاهش میانگین دقت داشته‌اند. پس از آن فرکانس سازهٔ سوم با سهم ۲۵.۴۰ درصد بهترین عملکرد را از خود نشان داده است. در سرعت گفتاری عادی، فرکانس سازهٔ سوم و چهارم به ترتیب با ۳۱.۱۴ درصد و ۲۸.۱۰ درصد عملکرد مطلوبی داشته‌اند. در سرعت گفتاری سریع نیز، فرکانس سازهٔ سوم با با ۲۸.۱۹ درصد و پس از آن فرکانس سازهٔ چهارم با ۲۷.۱۳ درصد نسبت به سایر فرکانس سازه‌ها در نشان‌دادن تغییرات فردویژه بهتر عمل کرده‌اند. همچنین، در صورت در نظر گرفتن کل سرعت‌های مختلف بدون تفکیک آن‌ها، این فرکانس سازهٔ سوم بوده است که با ۲۷.۹۹ درصد نسبت به سایر پارامترها بهتر عمل کرده است و پس از آن نیز فرکانس سازهٔ چهارم با ۲۷.۳۰ عملکرد خوبی از خود نشان داده است. به‌طور خلاصه می‌توان ابراز داشت در سرعت گفتاری آهسته، سهم فرکانس سازهٔ اول و سوم بالاتر است؛ در حالیکه، در سرعت‌های عادی و سریع، سازه‌های سوم و چهارم اهمیت بیشتری پیدا می‌کنند. در حالت مجموع، فرکانس سازه‌های سوم و چهارم بیشترین سهم را دارند، که نشان‌دهندهٔ نقش مهم‌تر این سازه‌ها در ترکیب داده‌ها است. شکل (۱) نمایش گرافیکی توان پارامترهای موردبررسی در سرعت‌های گفتاری موردبررسی را به تصویر می‌کشد.

https://doi.org/10.22108/jrl.2025.143702.1884



شکل ۱- نمودار میله‌ای نشانگر توانایی پارامترهای آکوستیکی فرکانس سازه‌های اول تا چهارم در سه سرعت گفتاری مختلف (سرعت گفتاری عادی با رنگ نارنجی، آهسته با آبی و سریع با خاکستری نشان داده است)

Figure 1- Bar graph showing the ability of the acoustic parameters of the first to fourth formant frequencies at three different speaking rates (normal speaking rate is shown in orange, slow in blue and fast in gray)

۵. بحث و بررسی

در پژوهش حاضر، تأثیر سرعت گفتار بر فرکانس سازه‌های اول تا چهارم واکه‌های زبان فارسی در پیکره‌ای آوایی با تغییرات درون‌گوینده‌ای ناشی از سرعت گفتار مورد بررسی قرار گرفته است. هدف اصلی این مطالعه در دو گام دنبال شد: ابتدا، بررسی معناداری تفاوت فرکانس سازه‌های اول تا چهارم در سرعت‌های گفتاری آهسته، عادی و سریع بود. در این گام همچنین به دنبال یافتن سازه‌هایی بودیم که بیشترین تأثیر را از عامل سرعت گفتار می‌پذیرند. سپس، در گام دوم، هدف این بود که مشخص شود آیا میزان فردویژگی فرکانس سازه‌ها تحت تأثیر سرعت گفتار قرار می‌گیرد یا خیر؛ به عبارت دیگر، بررسی شد که آیا تأثیر سرعت گفتار بر فرکانس سازه‌ها در بین گویندگان مختلف متفاوت است یا خیر.

بر اساس نتایج این پژوهش، فرکانس سازه‌ها در سرعت‌های گفتاری مختلف تغییرات معنادار داشته‌اند، هرچند میزان این تغییر در سازه‌های مختلف یکسان نبوده است. فرکانس سازهٔ دوم بیشترین حساسیت را به عامل سرعت گفتاری نشان داده است. این نتیجه همراستا با یافته‌های ویسمر و بری (2003) است که نشان دادند فرکانس سازهٔ دوم به‌ویژه تغییرپایانهٔ این سازه یک شاخص قوی در نشان‌دادن تغییرات سرعت گفتاری میان گویندگان است. همچنین، پیترمن (2000) نیز در پژوهش خود نشان داد که فرکانس سازهٔ دوم در سرعت‌های بالاتر تغییر می‌کنند، اگرچه این پژوهش تنها بر روی دو گویشور انجام شده است، اما دامنهٔ سرعت‌های گفتاری بررسی‌شده تا حدود ده سرعت مختلف بود. فرکانس سازهٔ دوم با میزان پیشین‌بودن یا پسین‌بودن واکه در ارتباط است (مدرسی‌قوامی؛ ۱۳۹۳). به‌عبارتی، هرچه واکه پیشین‌تر باشد، میزان فرکانس سازهٔ دوم آن بیشتر، و هرچه واکه پسین‌تر باشد میزان فرکانس سازهٔ دوم آن کمتر است. بر اساس نتایج این پژوهش، میانگین فرکانس سازهٔ دوم در سرعت گفتاری آهسته نسبت به سرعت گفتاری عادی بیشتر و این میزان در گفتار عادی نیز نسبت به سریع بالاتر بوده است. این احتمال می‌رود که در حالت گفتار آهسته، گویندگان واکه‌ها را در مجموع پیشین‌تر تلفظ کرده‌اند، خصوصاً در رابطه با واکهٔ پسین /ɑ/ این احتمال وجود دارد، یا اینکه گویندگان تمرکز بیشتری در تولید واکه‌های پیشین برای انتقال مفاهیم داشته‌اند. با توجه به اینکه فرکانس سازهٔ دوم در انتقال پیام زبان‌شناختی و تفکیک واکه‌ها اهمیت دارد، گزینهٔ دوم محتمل‌تر به نظر می‌رسد؛ زیرا در حالت آهسته، کنترل گویندگان بر حرکات تولیدی اندام‌های گویایی بیشتر است و ازآنجاکه تمرکز بیشتر بر انتقال پیام زبانی است، گویندگان بر این دسته از واکه‌ها تأکید بیشتری داشته‌اند. بر اساس نتایج، فرکانس سازهٔ اول کمترین تأثیرپذیری را از سرعت گفتار داشته است. این سازه

https://doi.org/10.22108/jrl.2025.143702.1884

که با میزان افراشتگی زبان و بازبودن دهان در ارتباط است (نوربخش، ۱۳۹۲؛ مدرسی‌قوامی، ۱۳۹۳)، کمترین حساسیت را نسبت به سرعت گفتار از خود نشان داده است. این نتیجه نشان می‌دهد که گویندگان به‌هنگام صحبت در سرعت‌های گفتاری مختلف بیشتر بر مشخصهٔ پیشین و پسین‌شدگی واکه تمرکز کرده‌اند که اثرات آن در فرکانس سازهٔ دوم نمایان می‌شود. همچنین، با توجه به اینکه این دو سازه نقش مهمی در تمایز واکه‌ها دارند، به نظر می‌رسد که گویندگان در بافت‌هایی که فهم‌پذیری مطلب اهمیت دارد مانند سرعت‌های گفتاری متفاوت، الگوی پایدارتری در سازهٔ اول دارند؛ درحالی‌که، سازهٔ دوم بیشتر تحت تأثیر تغییرات سرعت گفتار قرار گرفته است.

در بررسی میزان فردویژگی فرکانس سازه‌ها، نتایج نشان داد پتانسیل این پارامترها در سرعت‌های گفتاری گوناگون دستخوش تغییراتی می‌گردد. در سرعت گفتاری آهسته، فرکانس سازهٔ اول و فرکانس سازهٔ سوم بهترین عملکرد را داشته‌اند. فرکانس سازهٔ اول همانگونه که پیشتر نیز ذکر شد با افراشتگی زبان و نیز میزان بازبودن دهان در ارتباط است. هرچه زبان افراشته‌تر باشد میزان این سازه پایین‌تر است و هرچه مجرای دهان در تولید واکه بازتر باشد میزان این سازه بیشتر است. با نگاهی به آمار توصیفی این پژوهش مشخص است که میانگین فرکانس سازهٔ اول در سرعت گفتاری آهسته ۵۶۳.۲۲، در سرعت گفتاری عادی ۵۵۷.۰۸۵ و در سرعت گفتاری سریع ۵۷۶.۵۴ است. به‌عبارتی، در حالت گفتار سریع، میزان فرکانس سازهٔ اول بیشتر از حالت عادی و آهسته است. همچنین در حالت آهسته نیز این میزان نسبت به حالت عادی بیشتر بوده است. بر این اساس می‌توان نتیجه گرفت که گویندگان در حالت‌های گفتاری آرام و سریع، درجهٔ افراشتگی زبان کمتری و در عین حال مجرای دهان بازتری داشته‌اند. با توجه به اینکه فرکانس سازهٔ اول در سرعت گفتاری آهسته حامل ویژگی‌های فردویژهٔ بیشتری بوده است این بدان معناست که گویندگان در حالت گفتاری آهسته رفتار متفاوتی در میزان افراشتگی زبان و نیز بازکردن دهان از خود نشان داده است که تجلی آن در فرکانس سازهٔ اول پدیدار گشته است و موجب شده است این پارامتر مشخصه‌های گوینده‌محور را بهتر نشان دهد.

بر اساس نتایج، در سرعت‌های گفتاری عادی و سریع، فرکانس سازهٔ سوم و چهارم نسبت به سایر سازه‌ها در نشان‌دادن تغییرات فردویژه عملکرد بهتری داشته‌اند. همچنین، در مجموع، زمانی‌که کل سرعت‌های گفتاری با هم در نظر گرفته شدند، باز هم فرکانس سازهٔ سوم و پس از آن فرکانس سازهٔ چهارم بهترین عملکرد را داشتند. پژوهش‌های پیشین نیز بر فردویژگی فرکانس سازه‌های سوم و چهارم و نقش ویژهٔ آن‌ها در شناسایی گوینده تأکید کرده‌اند (Goldstein, 1976; Alderman, 2005; Gold et al., 2013; McDougall, 2004; McDougall, 2006; Jessen & Becker, 2010). در پژوهشی هم که توسط اسدی و همکاران (2018) به روش بلندمدت انجام شد نشان داده شد که فرکانس سازهٔ سوم و چهارم دو پارامتر قدرتمند در نشان‌دادن تمایزات فردویژه میان گویندگان مرد و زن فارسی‌زبان هستند. بر این اساس، می‌توان اظهار داشت که این سازه‌ها در سرعت‌های گفتاری مختلف همچنان فردویژگی خود را حفظ کرده‌اند و این یافته بر بهینه‌بودن نقش این پارامترها در شناسایی گوینده صحه می‌گذارد.

این پژوهش محدودیت هایی نیز دارد. داده های آوایی تنها از صدای هجده مرد فارسی زبان استخراج شده است. بررسی تأثیر سرعت گفتار در پیکره‌های آوایی بزرگ‌تر و همچنین در صدای زنان می‌تواند به داده‌های جامع‌تر و نتایج تعمیم‌پذیرتری منجر شود. با این حال، یافته‌های این پژوهش کاربردهای مهمی در زمینه‌های مختلف از جمله فناوری‌های پردازش گفتار، شناسایی گوینده و مطالعات زبان‌شناختی دارد. توجه به ویژگی‌های آکوستیکی واکه‌ها و همینطور همخوان‌ها در سرعت‌های مختلف گفتاری می‌تواند هم به بهبود سیستم‌های شناسایی گفتار کمک کرده و هم در مطالعات بین‌گوینده‌ای و درون‌گوینده‌ای برای درک بهتر تفاوت‌های فردی مؤثر باشد.

۶. نتیجه

پژوهش حاضر به بررسی تأثیر سرعت گفتار بر فرکانس سازه‌های اول تا چهارم واکه‌های زبان فارسی و میزان فردویژگی این پارامترها در سرعت‌های گفتاری مختلف پرداخت. نتایج نشان داد که سرعت گفتار تأثیر معناداری بر فرکانس سازه‌ها دارد، و این تغییرات بر توانایی این پارامترها در نشان‌دادن مشخصه‌های فردویژه نیز تأثیرگذار هستند. نتایج نشان داد که فرکانس سازهٔ دوم بیشترین حساسیت را به تغییرات سرعت گفتار نشان داده است. در مقابل، فرکانس سازهٔ اول کمترین تأثیرپذیری را از سرعت گفتار داشته است؛ که این امر نشان‌دهندهٔ

https://doi.org/10.22108/jrl.2025.143702.1884

پایداری نسبی این سازه در سرعت‌های مختلف و تأکید گویندگان بر حفظ تمایز زبانی واکه‌ها از طریق مشخصه‌های دیگر مانند پیشین و پسین بودن است. در ارتباط با میزان فردویژگی فرکانس سازه‌ها، نتایج نشان داد که فرکانس سازۀ اول در سرعت آهسته بیشترین ویژگی‌های فردمحور را نمایان می‌کند. این امر به رفتار متفاوت گویندگان در میزان افراشتگی زبان و بازکردن دهان در این سرعت گفتاری نسبت داده شد. در مقابل، فرکانس سازه‌های سوم و چهارم در سرعت‌های گفتاری عادی و سریع عملکرد بهتری در نشان‌دادن ویژگی‌های فردی گویندگان داشتند. این نتیجه نشان می‌دهد که این پارامترها در سرعت‌های مختلف گفتاری همچنان ویژگی‌های فردمحور خود را حفظ می‌کنند. نتایج این مطالعه نه تنها از نظر آواشناسی آکوستیکی می‌تواند اهمیت داشته باشد، بلکه می‌تواند کاربردهای بالقوه‌ای در زمینه‌هایی مانند آموزش زبان، شناسایی گوینده و تکنولوژی‌های تبدیل متن به گفتار نیز داشته باشد. پژوهش‌های آتی می‌توانند با افزایش اندازه نمونه، بررسی عوامل بیشتر و استفاده از مدل‌های پیچیده‌تر، به بررسی این عامل تأثیرگذار بپردازند.

منابع فارسی

اسدی، هما؛ علی نژاد، بتول. (۱۳۹۹). بررسی ویژگی‌های فردویژۀ واکه‌های سادۀ زبان فارسی بر اساس نظریۀ منبع‌صافی. نشریه پژوهش‌های زبان شناسی (۲)۱۲، ۲۴۱-۲۶۲. https://doi.org/10.22108/jrl.2021.128697.1577

مدرسی قوامی، گلناز. (۱۳۹۳). آواشناسی: بررسی علمی گفتار. تهران: سمت.

نوربخش، ماندانا. (۱۳۹۲). آواشناسی فیزیکی با استفاده از رایانه. تهران: نشر علم.

References

Agwuele, A., Sussman, H. M., & Lindblom, B. (2009). The effect of speaking rate on consonant vowel coarticulation. *Phonetica 65*(4), 194–209. https://doi.org/10.1159/000192792

Alderman, T. (2005). *Forensic speaker identification: A likelihood ratio-based approach using vowel formants.* Munich: LINCOM.

Asadi, H., & Alinezhad, B. (2020). Speaker-specific features of simple vowels in Persian based on the source-filter theory. *Journal of Researches in Linguistics 12*(2), 241-262. [In Persian] https://doi.org/10.22108/jrl.2021.128697.1577

Asadi, H., Nourbakhsh, M., Sasani, F., and Dellwo, V. (2018). Examining long-term formant frequency as a forensic cue for speaker identification: An experiment on Persian. In M. Nourbakhsh, H. Asadi, and M. Asiaee (Eds), *Proceedings of the First International Conference on Laboratory Phonetics and Phonology* (21-28). Tehran: Neveesh Parsi Publications.

Boersma, P., & Weenink, D. (2023) Praat: Doing Phonetics by Computer. http://www.praat.org.

Corretge, R. (2022). Praat Vocal Toolkit. https://www.praatvocaltoolkit.com

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

Gay, T., & Hirose, H. (1973). Effect of speaking rate on labial consonant production. A combined electromyographic-high-speed motion picture study. *Phonetica 27*(1), 44–56. https://doi.org/10.1159/000259425

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

Guenther, F. H., & Hickok, G. (2016). Neural models of motor speech control. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (725-740). Academic Press.

Goldstein, U. (1976). Speaker-identifying features based on formant tracks. *The Journal of the Acoustical Society of America 59*(3), 176-182. https://doi.org/10.1121/1.380837

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

Kearney, E., & Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. *Language, Cognition and Neuroscience 34*(9), 1214–1229. https://doi.org/10.1080/23273798.2019.1589541

Ladefoged, P. (2006). *A course in phonetics*. Boston: Wadsworth Cengage Learning.

McDougall, K. (2004). Speaker-specific formant dynamics: an experiment on Australian English /aɪ/. *International Journal of Speech, Language and the Law 11*(1), 103-130. https://doi.org/10.1558/sll.2004.11.1.103

McDougall, K. (2006). Dynamic features of speech and the characterization of speakers: Toward a new approach using formant frequencies. *International Journal of Speech, Language and the Law 13*(1), 89-126. https://doi.org/10.1558/ijsll.v13i1.89

Mefferd, A. S., & Green, J. R. (2010). Articulatory-to-acoustic relations in response to speaking rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research: JSLHR 53*(5), 1206–1219. https://doi.org/10.1044/1092-4388(2010/09-0083)

Modarresi Ghavami, G. (2011). *Phonetics: The scientific study of speech*. Tehran, Samt. [In Persian]

Moos, A. (2010). Long-term formant distribution as a measure of speaker characteristics in read and spontaneous speech. *The Phonetician 101*(102), 7-24.

Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

Nourbakhsh, M. (2013). *Acoustic phonetics using computer*. Tehran: Nashre Elm. [In Persian]

Pitermann, M. (2000). Effect of speaking rate and contrastive stress on formant dynamics and vowel perception. *The Journal of the Acoustical Society of America 107*(6), 3425–3437. https://doi.org/10.1121/1.429413

R Core Team. (2022). R: A language and environment for statistical computing (version 4.2.2). R Foundation for Statistical Computing. http://www.Rproject.org.

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

Shahrebabaki, A. S., Imran, A. S., Olfati, N., & Svendsen, T. (2018). Acoustic feature comparison for different speaking rates. In *Human-Computer Interaction. Interaction Technologies: 20th International Conference, HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part III 20* (176-189). USA: Springer International Publishing. https://doi.org/10.1007/978-3-319-91250-9_14

Shaiman, S., Adams, S. G., & Kimelman, M. D. (1997). Velocity profiles of lip protrusion across changes in speaking rate. *Journal of Speech, Language, and Hearing Research: JSLHR 40*(1), 144–158. https://doi.org/10.1044/jslhr.4001.144

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research: JSLHR 47*(4), 766–783. https://doi.org/10.1044/1092-4388(2004/058)

Weismer, G., & Berry, J. (2003). Effects of speaking rate on second formant trajectories of selected vocalic nuclei. *The Journal of the Acoustical Society of America 113*(6), 3362–3378. https://doi.org/10.1121/1.1572142

https://doi.org/10.22108/jrl.2025.143702.1884