University of Tehran Prees

Journal of Language Research

Online ISSN: 2676-3362

https://jolr.ut.ac.ir

Speaker Identification in Virtual Environments: Investigating the role of Voiceless Fricatives in WhatsApp Audio Data

Rahil Davoudi[1], **Homa Asadi**[2]✉

1. M.A. Student of Computational Linguistics, Linguistics Department, University of Isfahan, Isfahan, Iran. Email: rahil.otp@fgn.ui.ac.ir

2. Corresponding Author, Assistant Professor of Linguistics, Linguistics Department, University of Isfahan, Isfahan, Iran. Email: h.asadi@fgn.ui.ac.ir

Article Info

Article Type:
Research Article

Article History:

Received:
24, January, 2025

In Revised Form:
2, March, 2025

Accepted:
23, August, 2025

Published Online:
20, September, 2025

Keywords: Acoustic phonetics, Speaker identification, Fricative consonants, Mel-frequency cepstral coefficients, Support vector machine algorithm

Abstract

In the digital age, accurate speaker identification plays a crucial role in forensic and security investigations. However, the widespread use of internet-based communication platforms, such as WhatsApp, has introduced new challenges in this field. Factors such as variable microphone quality, background noise, network distortions, and audio compression can significantly affect a speaker's acoustic features and reduce the accuracy of speaker identification systems. Despite these limitations, evaluating the performance of acoustic features under such conditions is essential for advancing forensic phonetics and improving its practical applications in real-world settings. This study examines the role of voiceless fricatives in capturing between-speaker variability in audio recordings obtained through WhatsApp. The novelty of this research lies in investigating the ability of Persian voiceless fricatives to distinguish speakers under non-ideal recording conditions. To achieve this goal, speech data from 100 male Persian speakers were collected, and Mel-frequency cepstral coefficients (MFCCs) were extracted from the voiceless fricative segments. These features were then used as input to a support vector machine (SVM) model for speaker classification. The results showed that when all voiceless fricatives were considered together, the model achieved an overall speaker identification accuracy of 69%. However, analyzing each fricative separately led to an increase in model accuracy. Among the individual fricatives, the /s/ fricative had the highest accuracy at 77%, followed by /ʃ/, /f/, and /x/ with accuracies of 75%, 74%, and 73%, respectively. These findings suggest that even in non-ideal recording conditions, such as WhatsApp recordings, voiceless fricatives can provide valuable information for speaker differentiation. However, this study only focuses on one type of non-ideal recording condition, and further research is needed to explore other potential sources of degradation. The results highlight the potential of voiceless fricatives in speaker identification applications, particularly in informal, uncontrolled, and real-world scenarios where high-quality recording equipment is unavailable.

Cite this The Author(s): Davoudi, R., Asadi, H., (2025): Speaker Identification in Virtual Environments: Investigating the role of Voiceless Fricatives in WhatsApp Audio Data; Journal of Language Researches, No. 1, Vol.16 , Serial No. 30, Spring & Summer - (99-18). DOI: 10.22059/jolr.2025.389287.666909

Publisher: University of Tehran Press.

https://jolr.ut.ac.ir/article_104217.html?lang=en

1.Introducrion

In today's digital age, social media platforms like WhatsApp have become ubiquitous channels for interpersonal communication, generating vast quantities of speech data that have significant implications for forensic investigations. As criminal activities increasingly involve digital evidence, forensic phonetics has emerged as a crucial discipline for examining speech characteristics in legal contexts. This field extracts speaker-specific features from speech signals to aid in identifying or eliminating suspects, with applications extending across security, law enforcement, and technological domain. The intersection of forensic phonetics and social media presents unique opportunities and challenges.WhatsApp, widely used globally and relevant in criminal investigations, employs the Opus codec which significantly alters speech acoustics, potentially obscuring speaker-specific features. Within this challenging framework, fricative sounds represent a promising yet underexplored avenue for speaker recognition. These sounds produced by forcing air through narrow constrictions in the vocal tract, generate complex acoustic patterns influenced by individual anatomical and physiological traits that may persist despite compression. Prior research has demonstrated that acoustic properties of fricatives contribute to speaker identification, but a critical gap exists in understanding their behavior in codec-compressed speech signals typical of social media platforms. While previous studies have examined fricatives in laboratory settings or telephone-quality recordings, WhatsApp voice messages remain largely unexplored despite their forensic relevance. This study addresses this gap by investigating the potential of voiceless fricatives for speaker identification in WhatsApp voice messages from Persian speakers. We extract Mel-frequency cepstral coefficients (MFCCs) from voiceless fricative segments and apply Support Vector Machine (SVM) for classification, determining whether speaker-specific information in voiceless fricatives survives the compression process to contribute valuable insights to forensic practice and speech acoustics research.

2. Literature Review

Previous research has shown that fricatives, despite receiving less attention than vowels, play a crucial role in speaker identification due to their complex articulatory mechanisms and sensitivity to individual variation (Gold & French, 2011; Shindler &bDraxler, 2013). Studies on various languages have highlighted the acoustic distinctiveness of voiceless fricatives, particularly /s/, in differentiating speakers (Gordon et al., 2002; Kavanagh, 2012). Schindler and Drexler (2013) found that fricatives could outperform vowel formant frequencies in speaker recognition in German. In Persian, Asadi et al. (2019) demonstrated that COG in /s/ was highly speaker-specific for both male and female speakers, while COG in /ʃ/ was particularly distinctive for female speakers.More recent studies have reinforced the utility of spectral moments and cepstral features, with Smorenburg & Heeren (2020) demonstrating that fricatives retain speaker-discriminatory potential even in telephony speech in Dutch. Ulrich et al. (2023) further confirmed the superiority of MFCCs over spectral moments in speaker identification in Russian. The findings collectively suggest that fricatives, particularly /s/, provide valuable speaker-specific acoustic cues, warranting further investigation into their role in forensic and linguistic applications.

3. Materials and Methods

A total of 100 male Persian speakers (ages 25–45, M = 29.66, SD = 4.8) participated in this study. All were native speakers of Standard Persian with undergraduate to doctoral degrees. The dataset consisted of 20 sentences, each designed to elicit target fricatives in diverse phonetic contexts. Data collection was conducted online via WhatsApp, given

its widespread use in Iran and forensic relevance. Participants recorded their voices in quiet environments using mobile phones. WhatsApp voice messages, encoded with the Opus codec, were converted to .wav format for analysis in Praat (Boersma & Weenink, 2024). Segmentation was performed using the BAS WebService Maus tool (Baumann & Winkelmann, 2013). Acoustic analysis focused on MFCCs, widely recognized as effective features in automatic speaker and speech recognition (Lin, 2017). Speaker differentiation was assessed using SVM, a robust classification technique (Temko & Nadeu, 2005; Kinnunen & Li, 2010). Performance was evaluated using accuracy, precision, recall, and F-score. To assess the discriminative power of each fricative, we applied SVM models separately to each sound, comparing classification accuracy. All analyses were conducted in Python 3.13.0 using the pandas and scikit-learn libraries.

4. Discussions and conclusion

Results indicate that voiceless fricatives effectively classify speakers despite the presence of background noise and compression artifacts in WhatsApp recordings. This suggests that fricatives retain speaker-specific characteristics even under suboptimal recording conditions. The effectiveness of MFCCs in analyzing fricatives played a key role in this outcome. While MFCCs are commonly used for voiced speech segments like vowels, our findings confirm their high discriminative power for voiceless fricatives as well. Among the examined fricatives, /s/ exhibited the highest speaker specificity, consistent with findings from English, German, and Russian. However, other fricatives—/ʃ/, /f/, and /x/—also demonstrated notable speaker-discriminative potential, with relatively close classification accuracy. Interestingly, speaker classification improved when fricatives were analyzed separately, rather than in combination, likely due to reduced feature overlap and greater model focus on the distinct spectral properties of each fricative. Overall, these findings highlight voiceless fricatives as robust acoustic markers for speaker identification, even in real-world speech data recorded via WhatsApp. This underscores their applicability in forensic phonetics and linguistic research, particularly in digital communication contexts.

دانشگاه تهران

پژوهش‌های زبانی

شاپای الکترونیکی: ۳۳۶۲-۲۶۷۶

https://jolr.ut.ac.ir/

شناسایی گویندگان در فضای مجازی: بررسی نقش آواهای سایشی بی‌واک در داده‌های صوتی واتس‌اپ

راحیل داودی[۱]، هما اسدی[۲] ✉

۱. دانشجوی کارشناسی ارشد زبانشناسی رایانشی، گروه زبانشناسی، دانشگاه اصفهان، اصفهان، ایران. رایانامه: rahil.otp@fgn.ui.ac.ir

۲. نویسنده مسئول استادیار گروه زبانشناسی، دانشگاه اصفهان، اصفهان، ایران. رایانامه: h.asadi@fgn.ui.ac.ir

اطلاعات مقاله

نوع مقاله:
علمی – پژوهشی

تاریخ دریافت:
۱۴۰۳/۱۱/۰۵

تاریخ بازنگری:
۱۴۰۳/۱۲/۲۴

تاریخ پذیرش:
۱۴۰۴/۰۶/۰۲

تاریخ انتشار:
۱۴۰۴/۰۶/۳۰

واژه‌های کلیدی:
آواشناسی آکوستیکی، شناسایی گوینده، آواهای سایشی، ضرایب کپسترال فرکانسی مل، الگوریتم ماشین بردار پشتیبان

چکیده

در عصر دیجیتال، شناسایی دقیق گوینده در تحقیقات قضایی و امنیتی از اهمیت ویژه‌ای برخوردار است. با این حال، گسترش ارتباطات مبتنی بر اینترنت و استفادهٔ گسترده از پیام‌رسان‌هایی مانند واتس‌اپ، چالش‌های جدیدی را در این حوزه ایجاد کرده است. کیفیت متغیر میکروفون، نویز پس‌زمینه، اختلالات شبکه و فشرده‌سازی صوتی از جمله عواملی هستند که می‌توانند ویژگی‌های آکوستیکی گوینده را تحت تأثیر قرار دهند و دقت سیستم‌های شناسایی را کاهش دهند. علیرغم این محدودیت‌ها، بررسی عملکرد ویژگی‌های آکوستیکی در چنین شرایطی برای پیشبرد حوزهٔ آواشناسی قضایی و بهبود کاربردهای عملی آن در محیط‌های واقعی ضروری است. این پژوهش به بررسی نقش آواهای سایشی بی‌واک در نشان‌دادن تغییرات بین‌گوینده‌ای در داده‌های صوتی ضبط‌شده از طریق پیام‌رسان واتس‌اپ می‌پردازد. نوآوری این پژوهش در بررسی توانایی آواهای سایشی بی‌واک زبان فارسی برای شناسایی گویندگان در شرایط ضبط غیرایده‌آل است. برای این منظور، داده‌های صوتی از ۱۰۰ گویشور مرد فارسی‌زبان جمع‌آوری شد و ضرایب کپسترال فرکانسی مل (MFCC) از زنجیرهٔ آواهای سایشی بی‌واک استخراج شده و به‌عنوان ورودی به مدل ماشین بردار پشتیبان (SVM) وارد شدند. نتایج نشان داد که دقت مدل در تشخیص گوینده، زمانی که تمامی آواهای سایشی بی‌واک به‌طور هم‌زمان در نظر گرفته شدند، ۶۹ درصد بوده است. با این حال، بررسی جداگانهٔ هر یک از آواهای سایشی، افزایش دقت مدل را نشان داد. در این میان، آوای سایشی /s/ با دقت ۷۷ درصد، بیشترین تأثیر را داشت. پس از آن، آواهای /ʃ/ ، /f/ و /x/ به‌ترتیب با دقت‌های ۷۵ درصد، ۷۴ درصد و ۷۳ درصد قرار گرفتند. این نتایج نشان می‌دهد که حتی در شرایط ضبط غیرایده‌آل، مانند داده‌های ضبط‌شده از طریق واتس‌اپ، آواهای سایشی بی‌واک می‌توانند اطلاعات ارزشمندی برای تمایز میان گویندگان ارائه دهند. با این حال، این پژوهش تنها به یک نمونه از شرایط ضبط غیرایده‌آل پرداخته و بررسی سایر عوامل مخدوش‌کنندهٔ بالقوه، نیازمند تحقیقات بیشتری است. یافته‌های این مطالعه، پتانسیل بالای آواهای سایشی بی‌واک را در کاربردهای شناسایی گوینده، به‌ویژه در سناریوهای غیررسمی، غیرکنترل‌شده و واقعی که فاقد تجهیزات ضبط باکیفیت هستند، نشان می‌دهد.

استناد: داودی، راحیل، اسدی، هما ؛ (۱۴۰۴): شناسایی گویندگان در فضای مجازی: بررسی نقش آواهای سایشی بی‌واک در داده‌های صوتی واتس‌اپ: پژوهش‌های زبانی، سال ۱۶، شمارهٔ۱، بهار و تابستان، پیاپی ۳۰- (۲۴-۱)

DOI: 10.22059/jolr.2025.389287.666909.

ناشر: مؤسسهٔ انتشارات دانشگاه تهران

https://jolr.ut.ac.ir/article_104217.html?lang=en

۱. مقدمه

زبان، به‌ویژه در قالب گفتار، یکی از مهم‌ترین ابزارهای هویتی افراد است که حاوی اطلاعات گوناگون و منحصربه‌فرد می‌باشد. گفتار انسان به‌عنوان یکی از پیچیده‌ترین ابزارهای ارتباطی، نقشی کلیدی در تعاملات اجتماعی ایفا می‌کند. این ابزار نه‌تنها وسیله‌ای برای انتقال اطلاعات زبانی است، بلکه حامل ویژگی‌های فردی، فرهنگی و اجتماعی نیز هست (رز، ۲۰۰۲؛ دلوو و همکاران، ۲۰۰۷؛ لی و کرایمن، ۲۰۱۹). با گسترش فناوری‌های دیجیتال و رسانه‌های اجتماعی، نقش گفتار در ارتباطات میان‌فردی بیش از پیش برجسته شده است. پلتفرم‌هایی نظیر واتس‌اپ و تلگرام به‌طور گسترده‌ای برای تبادل پیام‌های صوتی مورد استفاده قرار می‌گیرند؛ امری که امکان مطالعهٔ ویژگی‌های گفتاری و شناسایی گوینده را در شرایط واقعی فراهم می‌کند.

شناسایی گوینده کاربردهای مهمی در حوزه‌های حقوقی، امنیتی و فناوری دارد. آواشناسی قضایی[1]، به‌عنوان شاخه‌ای از زبان‌شناسی قضایی، به مطالعه و تحلیل ویژگی‌های آوایی گفتار برای تعیین هویت گوینده می‌پردازد. این حوزه با به‌کارگیری دانش آواشناسی و پردازش سیگنال گفتاری، امکان استخراج ویژگی‌های فردی از نمونه‌های صوتی را فراهم می‌کند (یسن، ۲۰۰۸). در محیط‌های حقوقی و قضائی، این رویکرد می‌تواند به‌عنوان ابزاری قدرتمند برای شناسایی یا حذف مظنونان به کار رود. با این حال، شناسایی گوینده همواره با چالش‌هایی همراه بوده است. برخلاف داده‌های آزمایشگاهی که در شرایط کنترل‌شده تولید می‌شوند، داده‌های واقعی، به‌ویژه در رسانه‌های اجتماعی، تحت تأثیر عواملی همچون نویزهای محیطی، تغییرات کیفیت ضبط و تفاوت‌های کانالی قرار دارند (گوری و همکاران، ۲۰۲۴). این شرایط، پیچیدگی تحلیل ویژگی‌های آوایی را افزایش داده و بر دقت سیستم‌های شناسایی گوینده تأثیر می‌گذارد. از جمله چالش‌برانگیزترین ویژگی‌های آوایی در این زمینه، آواهای سایشی[2] هستند که به دلیل حساسیت بالا به کیفیت ضبط و تغییرات محیطی، کمتر مورد توجه پژوهشگران قرار گرفته‌اند.

ازآنجاکه آواهای سایشی به‌دلیل ساختار تولیدی خاص خود، ظرفیت بالایی در تمایز بین گویندگان دارند، مطالعهٔ آن‌ها در این زمینه حائز اهمیت است. این آواها با ایجاد شکافی باریک در مجرای گفتار تولید می‌شوند که جریان هوا از میان آن عبور کرده و طیفی پیچیده و پرانرژی ایجاد می‌کند (کتفورد، ۱۹۷۷؛ شیدل، ۱۹۹۰). توزیع انرژی در طیف این آواها به ویژگی‌های آناتومیکی و فیزیولوژیکی گوینده بستگی دارد و ازاین‌رو می‌تواند اطلاعات منحصربه‌فردی دربارهٔ هویت افراد ارائه دهد (یانگمن و همکاران، ۲۰۰۰؛ اشتوارت–اشمیت، ۲۰۰۷؛ ریتز و یانگمن، ۲۰۰۹).

1. forensic phonetics
2. fricative sounds

پژوهش‌های پیشین نشان داده‌اند که ویژگی‌های آکوستیکی آواهای سایشی، از جمله گشتاورهای طیفی[1]، در تمایز گویندگان مؤثر هستند (کاواناگ، ۲۰۱۲؛ شیندلر و دراکسلر، ۲۰۱۳؛ اسمورنبرگ و هیرن، ۲۰۲۰؛ الریچ و همکاران، ۲۰۲۳). با این حال، بیشتر این مطالعات در شرایط کنترل‌شدهٔ آزمایشگاهی یا بر روی داده‌های تلفنی انجام شده‌اند. در مقابل، داده‌های به‌دست‌آمده از پیام‌رسان‌هایی مانند واتس‌اپ و تلگرام، به‌دلیل تأثیر عواملی همچون پهنای باند شبکه، نوع دستگاه ضبط و الگوریتم‌های فشرده‌سازی، کیفیتی متفاوت داشته و کمتر در تحقیقات آواشناسی بررسی شده‌اند. فشرده‌سازی صوتی در واتس‌اپ، که از الگوریتم‌های پیشرفتهٔ کاهش حجم داده استفاده می‌کند، می‌تواند موجب تغییر در ویژگی‌های آکوستیکی سیگنال گفتاری شده و چالش‌هایی را در شناسایی گوینده ایجاد کند.

در این پژوهش، با تمرکز بر آواهای سایشی بی‌واک در داده‌های صوتی واتس‌اپ، نقش بالقوهٔ این آواها در شناسایی گوینده بررسی می‌شود. هدف اصلی تحقیق، ارائهٔ تحلیلی جامع از تفاوت‌های آوایی بین گویندگان در محیطی واقعی و پویا است. در راستای این هدف، پژوهش حاضر از رویکردی نوآورانه بهره می‌برد که جنبه‌های مختلفی دارد. نخست، این پژوهش از داده‌هایی استفاده می‌کند که شرایط طبیعی گفتار را بازتاب می‌دهند و بدین ترتیب، تحلیل‌هایی واقعی‌تر و کاربردی‌تر ارائه می‌دهد. دوم، تمرکز بر آواهای سایشی بی‌واک، که تاکنون کمتر در شناسایی گوینده مورد استفاده قرار گرفته‌اند، می‌تواند به شناسایی ویژگی‌های آکوستیکی متمایزکنندهٔ جدیدی منجر شود. سوم، این پژوهش با بهره‌گیری از الگوریتم‌های یادگیری ماشین، به‌ویژه ماشین بردار پشتیبان، به تحلیل دقیق داده‌ها می‌پردازد؛ رویکردی که در شناسایی گوینده در زبان فارسی کمتر مورد توجه قرار گرفته است. این روش، از طریق تحلیل ضرایب کپسترال فرکانسی مل[2] از زنجیرهٔ آواهای سایشی بی‌واک، می‌تواند توانایی بالایی در تمایز میان گویندگان فراهم آورد. پرسش‌های پژوهش حاضر عبارت‌اند از:

۱.آیا ویژگی‌های آکوستیکی آواهای سایشی بی‌واک زبان فارسی در داده‌های صوتی واتس‌اپ توانایی تمایز قابل‌توجه بین گویندگان مختلف را دارد؟

۲.در صورت مثبت بودن پاسخ پرسش اول، کدام یک از آواهای سایشی بی‌واک زبان فارسی بیشترین توانایی را در تمایز بین گویندگان دارند؟

هدف از انتخاب سایشی‌های بی‌واک در مقابل سایشی‌های واک‌دار، نقش برجسته‌تر این آواها در نمایاندن تغییرات آکوستیکی بین گویندگان است. سایشی‌های بی‌واک، به‌دلیل تفاوت‌های قابل‌توجه در توزیع انرژی طیفی، نقش مهم‌تری در تفکیک گوینده و تشخیص گفتار

1. spectral moments
2. Mel-frequency cepstral coefficients (MFCCs)

دارند و اطلاعات بیشتری برای تحلیل تفاوت‌های بین‌گوینده فراهم می‌آورند (گوردون، ۲۰۰۲؛ کاواناگ، ۲۰۱۲).

۲. پیشینهٔ پژوهش

اگرچه آواهای سایشی به اندازهٔ واکه‌ها در پژوهش‌های آواشناسی توجه نگرفته‌اند، اما نسبت به سایر همخوان‌ها، بیشتر در مطالعات شناسایی گوینده بررسی شده‌اند (گلد و فرنچ، ۲۰۱۱). رز (۲۰۰۲) و یسن (۲۰۰۸) نیز بر این نکته تأکید کرده‌اند که ویژگی‌های آکوستیکی آواهای سایشی می‌توانند الگوهای گفتاری فردی را آشکار سازند. به دلیل سازوکارهای تولید پیچیده و حساسیت بالا به تفاوت‌های فیزیولوژیکی، این دسته از همخوان‌ها می‌توانند سرنخ‌های آکوستیکی ارزشمندی برای تمایز میان گویندگان، به‌ویژه در حوزهٔ قضایی، فراهم آورند.

یکی از نخستین پژوهش‌ها در زمینهٔ بررسی آواهای سایشی، مطالعهٔ گوردون و همکاران (۲۰۰۲) بود. آنان آواهای سایشی بی‌واک در هفت زبان در معرض خطر را با رویکرد آکوستیکی تحلیل کردند. افزون بر مقایسهٔ بین‌زبانی، تفاوت‌های میان گویندگان هر زبان نیز بررسی شد. سه ویژگی آکوستیکی دیرش، مرکز تجمع انرژی و میانگین قلهٔ طیفی مورد اندازه‌گیری قرار گرفت. نتایج نشان داد آواهای سایشی کناری بیشترین تفاوت را بین زبان‌ها و آوای /s/ بیشترین تفاوت را بین گویندگان نشان می‌دهد. محققان دریافتند که /s/ در هر دو سطح بین‌زبانی و بین‌گوینده‌ای، بیشترین توان تمایز را دارد.

کاواناگ (۲۰۱۲) پارامترهای آکوستیکی مختلف را برای تشخیص هویت گویندگان در دو لهجهٔ انگلیسی بررسی کرد. او به‌طور خاص بر روی پنج همخوان شامل سه خیشومی، یک کناری و یک سایشی در موقعیت‌های مختلف در واژه تمرکز کرد. برای این پژوهش، از ضبط صدای ۱۲ مرد جوان انگلیسی که از پایگاه داد *آی‌وی‌ای* انتخاب شده بودند، استفاده شد. نتایج نشان داد که مرکز تجمع انرژی و انحراف معیار در آوای /s/، دو پارامتر آکوستیکی مفید برای تشخیص هویت گویندگان هستند. نتایج نشان داد که آوای /s/ نسبت به سایر همخوان‌های موردبررسی تفاوت‌های فردویژهٔ بیشتری را نشان می‌دهد.

شیندلر و دراکسلر (۲۰۱۳) به بررسی امکان استفاده از آواهای خیشومی و سایشی برای شناسایی گویندگان آلمانی زبان می‌پردازند و پیشنهاد دادند که این آواها ممکن است حتی نسبت به فرکانس سازه‌های واکه‌ها، که به‌طور سنتی برای این منظور استفاده می‌شوند، عملکرد بهتری داشته باشند. روش‌های مورد استفاده در این پژوهش، شامل استفاده از گشتاورهای طیفی خیشومی‌ها و سایشی‌ها به‌عنوان ویژگی‌های آکوستیکی برای شناسایی گوینده بود. گشتاورهای طیفی مورد استفاده در این مطالعه، شامل مرکز تجمع انرژی، انحراف معیار[1]، چولگی[2] و

1. standard deviation
2. skewness

کشیدگی[1] بود. نتایج نشان داد که آواهای خیشومی و سایشی تغییرات بین‌گویندهٔ بالاتری نسبت به فرکانس سازه‌های واکه‌ها داشتند، که نشان‌دهندهٔ قابلیت بهتر این آواها در شناسایی گوینده است.

در رابطه با زبان فارسی، اسدی و همکاران (۱۳۹۴) در مطالعه‌ای به بررسی تأثیر پوشش‌های صورت نظیر ماسک جراحی، ماسک پلاستیکی، کلاه اسکی، کلاه کاسکت موتورسواری و همچنین شرایط بدون ماسک بر ویژگی‌های آکوستیکی سایشی‌های بی‌واک زبان فارسی در صدای ۵ گویشور زن فارسی‌زبان پرداختند. پارامترهای مربوط به مشخصه‌های طیفی و شدت سایش، از جمله مرکز تجمع انرژی، قلهٔ طیفی و شدت[2]، برای تحلیل انتخاب شدند. نتایج نشان داد که پوشش‌های صورت به‌طور معناداری بر شدت، قلهٔ طیفی و مرکز تجمع انرژی سایشی‌های بی‌واک /f/، /s/ و /ʃ/ تأثیر می‌گذارند. علاوه بر این، نتایج حاکی از آن بود که بیشترین تغییرات در شدت سایشی‌های بی‌واک در حالت استفاده از کلاه کاسکت و بیشترین تغییرات در مرکز تجمع انرژی در حالت استفاده از کلاه کاسکت و ماسک پلاستیکی مشاهده شده است. همچنین، نتایج نشان داد که قلهٔ طیفی سایشی‌های بی‌واک /f/، /s/ و /ʃ/ نسبت به شدت و مرکز تجمع انرژی تغییرات کمتری داشته است. در پژوهشی دیگر، اسدی و همکاران (۱۳۹۸) به بررسی تفاوت‌های بین گویندگان در همخوان‌های سایشی بی‌واک پرداختند. در این تحقیق، ۲۴ گویشور فارسی (۱۲ مرد و ۱۲ زن) شرکت کردند. متغیرهای مورد بررسی شامل تعامل میان متغیرهای مستقل (گوینده، جنسیت، تکرار و نوع سایشی بی‌واک) و اثر دو متغیر وابسته (مرکز تجمع انرژی و دیرش) بود. این مطالعه نشان داد که مرکز تجمع انرژی آوای /s/ در گروه مردان و مرکز تجمع انرژی آواهای /s/ و /ʃ/ در گروه زنان، مناسب‌ترین پارامترهای آکوستیکی برای تفکیک گویندگان هستند. با این حال، دیرش آوای /s/، علی‌رغم معنادار بودن تفاوت آن میان گویندگان، پارامتر مؤثری برای شناسایی گوینده محسوب نشد. این نتایج، بر لزوم انتخاب دقیق‌تر پارامترهای آکوستیکی برای بررسی و تحلیل سایشی‌های بی‌واک زبان فارسی تأکید می‌کنند.

اسمورنبرگ و هیرن (۲۰۲۰) به بررسی اطلاعات فردویژه در همخوان‌های سایشی /s/ و /x/ زبان هلندی که از مکالمات تلفنی استخراج شده‌اند، می‌پردازند. آن‌ها داده‌های آکوستیکی مربوط به همخوان‌های سایشی بی‌واک /s/ و /x/ را از گفتگوهای تلفنی ۴۳ گویندهٔ مرد که به زبان هلندی معیار صحبت می‌کردند و سبک بداهه داشتند، استخراج کرده و مورد تحلیل قرار دادند. در پژوهش آن‌ها پارامترهای دیرش، مرکز تجمع انرژی، انحراف معیار، شیب طیفی و دامنه اندازه‌گیری شدند. نتایج نشان داد که سایشی‌های مورد بررسی حتی در داده‌های تلفنی هم می‌توانند تمایزدهنده افراد باشند. همینطور نتایج نشان داد که سایشی‌های مورد بررسی اگر در پایانهٔ هجا قرار بگیرند و یا در مجاورت همسایه‌های دولبی قرار بگیرند ویژگی‌های تمایزدهندهٔ

1. kurtosis
2. intensity

بیشتری دارند نسبت به اینکه اگر در آغازهٔ هجا یا در مجاورت آواهای غیرلبی قرار گیرند.

یکی از جدیدترین پژوهش‌ها در زمینهٔ بررسی آواهای سایشی، پژوهش الریچ و همکاران (۲۰۲۳) بوده است که به بررسی تنوع آکوستیکی درون‌گوینده و بین‌گوینده در همخوان‌های سایشی زبان روسی می‌پردازد و پتانسیل آن‌ها را برای شناسایی گوینده و پیش‌بینی جنسیت بررسی می‌کند. این مطالعه بر روی صدای ۵۹ گویندهٔ روسی (۳۰ زن، ۲۹ مرد) انجام شد که هشت همخوان سایشی در بافت‌های آوایی مختلف را تولید می‌کردند. سپس با استفاده از دو روش یادگیری ماشین، یعنی درخت تصمیم‌گیری[1] و جنگل تصادفی[2]، داده‌ها را تحلیل کردند. نتایج نشان داد ضرایب کپسترال فرکانسی مل در پیش‌بینی جنسیت و هویت گوینده از مجموعه پارامترهای آکوستیکی گشتاورهای طیفی بهتر عمل کردند. دقت شناسایی افراد با استفاده از ضرایب کپسترال فرکانسی مل حدود ۶۴ درصد بود، در حالی که این میزان برای گشتاورهای طیفی تنها حدود ۲۲ درصد بود. در مجموع نتایج این پژوهش نشان داد که ضرایب کپسترال فرکانسی مل برای شناسایی گوینده مؤثرتر از پارامترهای آکوستیکی سنتی مانند گشتاورهای طیفی هستند. در زمینهٔ بررسی داده‌های صوتی در محیط رسانه‌های اجتماعی با رویکرد آکوستیکی تنها یک مطالعه یافت شد که اخیرا توسط گوری و همکاران (۲۰۲۴) انجام گرفته است. این مطالعه به بررسی تأثیر چهار پلتفرم شبکه‌های اجتماعی (واتس‌اپ، اینستاگرام، اسنپ‌چت و تلگرام) بر ویژگی‌های آکوستیکی واکه‌ها پرداخته است. داده‌های صوتی از ۴۰ گویشور انگلیسی‌زبان، ۲۰ مرد و ۲۰ زن، جمع‌آوری شد. پژوهشگران از تکنیک‌های یادگیری ماشین برای ارزیابی تأثیر این شبکه‌های اجتماعی بر فرکانس سازه‌های اول تا چهارم واکه‌ها استفاده کردند. یافته‌های مطالعه نشان داد که داده‌های صوتی تلگرام بهترین عملکرد را در نشان‌دادن تغییرات میان گویندگان و نیز پیش‌بینی جنسیت گویندگان داشته‌اند. پس از آن، داده‌های صوتی مستخرج از واتس‌اپ عملکرد بهتری نشان داده‌اند. نتایج برای داده‌های صوتی اینستاگرام متغیر بوده است و بدترین عملکرد مربوط به داده‌های اسنپ‌چت بوده است.

۳. روش‌شناسی

بخش‌های زیر به ارائهٔ اطلاعات در مورد شرکت‌کنندگان، روش گردآوری داده‌های صوتی، مراحل پردازش داده‌ها و در نهایت پارامترهای مورد مطالعه اختصاص یافته است.

۳-۱. شرکت‌کنندگان و داده‌های آوایی

در این پژوهش، ۱۰۰ گویشور مرد فارسی‌زبان با دامنه سنی ۲۵ تا ۴۵ سال (میانگین = ۲۹٫۶۶، انحراف معیار = ۴٫۸) مشارکت داشتند. همگی شرکت‌کنندگان به فارسی با لهجهٔ معیار سخن می‌گفتند و با توجه به سن، جنسیت و گویش، به‌منظور تشکیل یک گروه یکنواخت انتخاب شدند. این افراد همگی از اقشار تحصیل‌کرده جامعه و دارای مدارک کارشناسی، کارشناسی ارشد یا

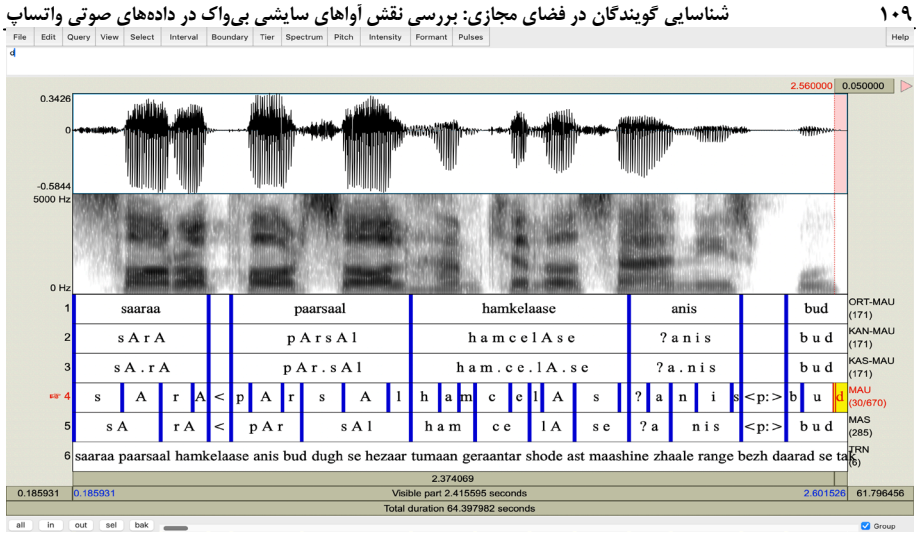
1. decision tree
2. random forest

دکتری بودند. برای ثبت داده‌های گفتاری، ۲۰ جمله طراحی شد که در آن‌ها آواهای سایشی هدف در موقعیت‌های آوایی متنوعی قرار می‌گرفتند. فرایند گردآوری داده‌ها از طریق برنامهٔ پیام‌رسان واتس‌اپ انجام شد؛ انتخاب این بستر به دلیل کاربرد گسترده آن در ایران و اهمیت آن در حوزهٔ آواشناسی قضایی صورت گرفت. بنا بر آمار وب‌سایت استاتیستا، واتس‌اپ در زمان گردآوری داده‌ها (۲۰۲۲–۲۰۲۳) بیش از دو میلیارد کاربر فعال ماهانه داشت و این تعداد تا ژوئن ۲۰۲۳ به حدود ۲٫۷ میلیارد نفر رسید. واتس‌اپ در سال‌های اخیر به پرکاربردترین پیام‌رسان جهان تبدیل شده است؛ این عوامل، استفاده از پیام‌های صوتی در این پلتفرم را برای پژوهش حاضر توجیه‌پذیر ساخت. تمامی ضبط‌ها توسط شرکت‌کنندگان و با استفاده از گوشی‌های همراه شخصی آن‌ها انجام شد. از آن‌ها خواسته شد در محیطی آرام قرار گیرند و براساس یک فایل راهنمای گفتاری که از پیش برایشان ارسال شده بود، جملات را ضبط کنند. برای اطمینان از اجرای صحیح دستورالعمل‌ها، تماس تصویری نیز با شرکت‌کنندگان برقرار می‌شد. گردآوری این پیکرهٔ گفتاری در زمان همه‌گیری کرونا انجام گرفت؛ بنابراین، شرکت‌کنندگان در خانه و بدون استفاده از ماسک حضور داشتند و به دلیل نبود تماس مستقیم و ضبط در فضای شخصی، خطری از نظر انتقال ویروس وجود نداشت.

۲-۳. تقطیع داده‌ها

جریان‌های صوتی واتس‌اپ معمولاً با کدک Opus فشرده می‌شوند (کارپیسک، ۲۰۱۵). برای تسهیل تحلیل‌های بعدی در نرم‌افزار پرات (بورسما و وینیک، ۲۰۲۴)، فایل‌های صوتی واتس‌اپ ابتدا با استفاده از وبسایت Online Audio Converter[1] به فرمت wav. تبدیل شدند. برای تحلیل سیگنال‌های صوتی در پرات، به یک شبکهٔ متنی (TextGrid) برای هر فایل صوتی نیاز است. برای ایجاد این شبکه‌های متنی، از درگاه BAS WebService Maus (کیسلر و همکاران، ۲۰۱۷) استفاده شد. در مرحلهٔ اول، رونویسی صوتی فایل‌ها به‌صورت دستی توسط پژوهشگران انجام شد و در فایل‌های متنی با پسوند txt. ذخیره شد. سپس، این فایل‌های متنی به همراه فایل‌های صوتی به‌عنوان ورودی به سرویس وب Maus داده شدند. سرویس وب Maus با استفاده از مسیر pipeline without ASR و با تنظیم زبان فارسی و خروجی TextGrid.، شبکه‌های متنی موردنیاز برای پرات را تولید کرد. پس از بارگذاری فایل‌ها و پذیرش شرایط، سرویس وب یک فایل فشرده حاوی شبکهٔ متنی را ایجاد می‌کند که برای تحلیل‌های بعدی در پرات مورد استفاده قرار گرفت. در شکل ۱ نمونه‌ای از تقطیع داده‌های صوتی به همراه موج صوتی و طیف‌نگاشت آن را مشاهده می‌کنید.

1. www.online-convert.com



شکل ۱: نمونه‌ای از تقطیع و برچسب‌گذاری داده‌های آوایی پژوهش

۳-۳. نحوهٔ استخراج پارامترهای آکوستیکی

در این پژوهش، برای تحلیل آواهای سایشی، از ضرایب کپسترال فرکانسی مل به‌عنوان ویژگی اصلی استفاده شده است. لین (۲۰۱۷) این ضرایب را یکی از رایج‌ترین و مؤثرترین ویژگی‌ها در سیستم‌های خودکار شناسایی صدا معرفی می‌کند. تولید بردارهای ویژگی ضرایب کپسترال فرکانسی مل طی چند مرحله انجام می‌شود. ابتدا طیف توان سیگنال محاسبه می‌شود که معمولاً با استفاده از تبدیل فوریهٔ گسسته انجام می‌گیرد. تبدیل فوریه، سیگنال را به فرکانس‌های تشکیل‌دهندهٔ آن تجزیه کرده و طیف فرکانسی حاصل را به دست می‌دهد. این تبدیل روی بازه‌های زمانی کوتاه (فریم‌های صوتی) انجام می‌شود که معمولاً اندازهٔ هر فریم حدود ۱۰ میلی‌ثانیه است. در مرحلهٔ بعد، طیف فرکانسی به مقیاس مل نگاشت می‌شود. این کار با کمک یک بانک فیلتر انجام می‌شود که فرکانس‌ها را بر اساس مقیاس مل هموار می‌کند. سپس لگاریتم مقادیر به‌دست‌آمده گرفته می‌شود تا مقیاس‌های بزرگ و کوچک طیف متعادل شوند. در مرحلهٔ پایانی، داده‌های حاصل از فیلتر مقیاس مل با استفاده از تبدیل کسینوسی گسسته به حوزهٔ کپستروم منتقل می‌شوند. این انتقال، داده‌ها را به ضرایب کپسترال تبدیل می‌کند که نمایانگر ویژگی‌های اصلی سیگنال صوتی در حوزه‌ای به نام کِفرِنسی هستند. در این حوزه، فرکانس‌های سازه‌ای به ضرایب پایین‌تر و فرکانس پایه به ضرایب بالاتر مربوط می‌شوند (لین، ۲۰۱۷: ۵۰). استخراج ضرایب کپسترال فرکانسی مل در این پژوهش با کمک دو اسکریپت از پیش‌نوشته‌شده که در نرم‌افزار پرات اجرا شده‌اند، انجام گرفت. اسکریپت اول وظیفهٔ استخراج ویژگی‌های آکوستیکی از هر فایل صوتی را بر عهده دارد. این اسکریپت فایل‌های ورودی را پردازش کرده و ویژگی‌هایی مانند زیروبمی، هارمونیک، فرکانس سازه‌ها و ضرایب کپسترال فرکانسی مل را از سیگنال استخراج می‌کند و سپس داده‌های حاصل را برای تحلیل‌های بعدی ذخیره می‌کند.

اسکریپت دوم، داده‌های ضرایب کپسترال را از فایل‌های خروجی دریافت کرده، آن‌ها را به ماتریس تبدیل و پردازش‌های تکمیلی را انجام می‌دهد.

تمامی این مراحل برای زنجیرهٔ آواهای سایشی بی‌واک انجام شد. ابتدا یک اسکریپت جداگانه توسط نویسندهٔ اول طراحی شد تا آواهای سایشی بی‌واک از سیگنال استخراج شوند و سپس اسکریپت‌های مرتبط با ضرایب کپسترال فرکانسی مل بر روی این آواهای سایشی اعمال گردید.

۴-۳. تحلیل آماری

به‌منظور بررسی تفاوت‌های میان گویندگان در داده‌های آوایی مستخرج از محیط واتس‌اپ از مدل ماشین بردار پشتیبان استفاده کردیم. این مدل یک تکنیک طبقه‌بندی تبعیضی است که عمدتاً بر دو فرضیه استوار است. نخست، تبدیل داده‌ها به فضایی با بعد بالا می‌تواند مسائل طبقه‌بندی پیچیده با سطوح تصمیم‌گیری پیچیده را به مسائل ساده‌تری تبدیل کند که قابلیت استفاده از توابع تبعیضی خطی را دارند. دوم، ماشین‌های بردار پشتیبان بر استفاده از تنها آن دسته از الگوهای آموزشی که در نزدیکی سطح تصمیم‌گیری قرار دارند متکی هستند، با فرض اینکه آن‌ها مفیدترین اطلاعات را برای طبقه‌بندی فراهم می‌کنند (تمکو و نادو، ۲۰۰۵). این روش در شناسایی گوینده و طبقه‌بندی واج‌ها به‌طور گسترده‌ای استفاده شده است (کینونن و لی، ۲۰۱۰).

برای ارزیابی عملکرد سیستم شناسایی گوینده، از معیارهای استاندارد دسته‌بندی شامل دقت[1]، صحت[2]، پوشش[3] و امتیاز اف[4] استفاده شد. دقت، معیاری برای اندازه‌گیری عملکرد کلی مدل است و نسبت تعداد پیش‌بینی‌های درست به تعداد کل پیش‌بینی‌ها را نشان می‌دهد. صحت، دقت پیش‌بینی‌های مثبت مدل را می‌سنجد و به‌صورت نسبت تعداد پیش‌بینی‌های مثبت درست به کل پیش‌بینی‌های مثبت تعریف می‌شود. پوشش، توانایی مدل در شناسایی موارد مثبت واقعی را نشان می‌دهد و بیان می‌کند چه درصدی از موارد مثبت واقعی توسط مدل شناسایی شده‌اند. امتیاز اف نیز، که میانگین هارمونیک صحت و یادآوری است، برای ارزیابی کلی مدل و ایجاد تعادل میان این دو معیار استفاده می‌شود. علاوه بر این، به‌منظور بررسی میزان تاثیرگذاری هر کدام از سایشی‌های بی‌واک در نشان‌دادن تمایزات میان گویندگان، ابتدا داده‌ها بر اساس نوع آوا تقسیم‌بندی شد و سپس مدل بردار پشتیبان در هر کدام از دسته های آوایی اجرا شد. در نهایت میزان دقت به دست‌آمده در هر دستهٔ آوایی با هم مقایسه شد. تمامی تحلیل‌های آماری داده‌های آوایی این پژوهش با استفاده از نرم‌افزار پایتون نسخهٔ ۳,۱۳,۰[5] انجام گرفته است.

1. accuracy
2. precision
3. recall
4. F1- Score
5. https://www.python.org/downloads/

تحلیل داده‌ها با استفاده از کتابخانهٔ pandas (مک کینی، ۲۰۱۰) در پایتون انجام شد. برای وظایف یادگیری ماشین، از کتابخانهٔ Scikit-learn (پدرگوسا و همکاران،۲۰۱۱) استفاده گردید که دسترسی به طیف گسترده‌ای از الگوریتم‌ها، از جمله ماشین بردار پشتیبان را فراهم می‌کند. تجسم داده‌ها و ترسیم نتایج با استفاده از کتابخانهٔ Matplotlib (هانتر، ۲۰۰۷) انجام شد.

۴. گزارش نتایج

در بخش‌های زیر به‌منظور پاسخگویی به پرسش‌های مطرح‌شده در مقدمه آزمون‌های آماری مناسب اجرا شده و نتایج حاصل از تحلیل داده‌ها گزارش خواهد شد.

۴-۱. تقسیم بندی داده‌های صوتی به داده‌های آموزش و آزمایش

همان‌طور که قبلا ذکر شد، مدل یادگیری ماشینی که برای تحلیل داده‌ها انتخاب کردیم، مدل ماشین بردار پشتیبان است. ضرایب کپسترال فرکانسی مل به‌عنوان ویژگی‌های ورودی و گوینده به‌عنوان هدف وارد مدل شدند. از میان داده‌ها ۸۰ درصد برای آموزش و ۲۰ درصد برای آزمایش انتخاب شدند. لازم به ذکر است تمامی داده‌ها در حالتی بررسی شدند که متعادل[1] شده بودند.

۴-۲. بررسی تغییرات بین‌گوینده آواهای سایشی بیواک بر اساس مدل بردار پشتیبان

پس از انجام تقسیم‌بندی، داده‌ها بر روی مدل ماشین بردار پشتبان آموزش داده شدند و نتایج زیر که شامل دقت، صحت، پوشش و امتیاز اف می‌باشند به دست آمد.

جدول ۱: نتایج مقادیر دقت، صحت، پوشش و امتیاز اف در داده‌های آموزش.

معیار ارزیابی	دقت	صحت	پوشش	امتیاز اف
ماشین بردار پشتیبان	۰,۷۲	۰,۷۱	۰,۷۱	۰,۷۰

جدول ۱ نتایج مربوط به داده‌های آموزش و جدول ۲ نتایج مربوط به داده‌های آزمایش را نشان می‌دهد. تمامی نتایج با استفاده از نمودار مشخصهٔ عملکرد[2] مقایسه شدند و الگوریتم‌های دسته‌بندی آن‌ها ارزیابی گردید تا از بیش‌برازش[3] احتمالی جلوگیری گردد.

جدول ۲: نتایج مقادیر دقت، صحت، پوشش و امتیاز اف در داده‌های آزمایش.

معیار ارزیابی	دقت	صحت	پوشش	امتیاز اف
ماشین بردار پشتیبان	۰.۶۹	۰.۶۸	۰.۶۹	۰.۶۸

نتایج جدول ۱ نشان می‌دهد که مدل عملکرد قابل قبولی بر روی داده‌های آموزش داشته است. مقادیر نزدیک به هم دقت، صحت و پوشش نشان می‌دهد که مدل تعادل خوبی بین شناسایی صحیح نمونه‌های مثبت و منفی برقرار کرده است. با این حال، برای ارزیابی دقیق‌تر

1. balanced data
2. receiver operating characteristic
3. overfitting

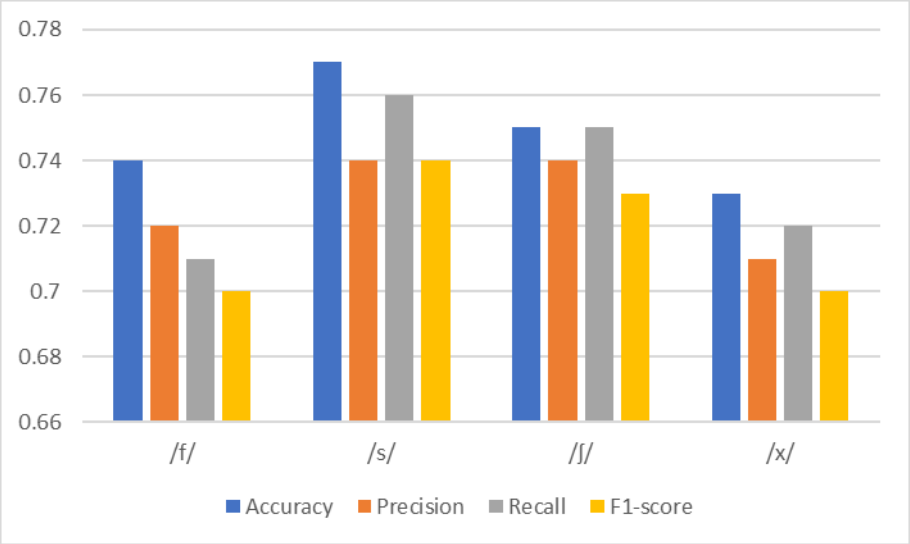
عملکرد مدل، لازم است که نتایج بر روی داده‌های آزمایش نیز بررسی شود. جدول ۲ نتایج مربوط به داده‌های آزمایش را نشان می‌دهد. این داده‌ها به مدل ارائه شده‌اند تا عملکرد آن در داده‌هایی که قبلاً ندیده است، ارزیابی شود. هدف از این کار، سنجش توانایی تعمیم‌پذیری مدل به داده‌های جدید و ناشناخته است. نتایج نشان می‌دهد که مدل عملکرد قابل قبولی بر روی داده‌های آزمایش نیز داشته است. اگرچه نسبت به داده‌های آموزش کمی افت عملکرد مشاهده شده است، اما این افت قابل چشم‌پوشی است. با توجه به اینکه داده‌های آزمایش داده‌هایی هستند که مدل قبلاً ندیده است، این نتایج نشان می‌دهد که مدل توانایی خوبی در تعمیم‌پذیری به داده‌های جدید دارد.

۳-۴. بررسی میزان توانایی آواهای سایشی بی‌واک در نشان‌دادن تغییرات آکوستیکی میان گویندگان

برای بررسی تأثیر هر یک از آواهای سایشی بی‌واک /f/، /s/، /ʃ/، /x/ بر عملکرد مدل ماشین بردار پشتیبان، داده‌ها به چهار دستهٔ جداگانه تقسیم شدند و مدل بر روی هر دسته به‌صورت مستقل آموزش داده شد. نتایج حاصل در جدول ۳ ارائه شده است. همان‌طور که مشاهده می‌شود، آوای /s/ بیشترین تأثیر را بر دقت مدل داشته و به دقت ۷۷ درصد دست یافته است. پس از آن، آواهای/ʃ/، /f/ و /x/ به ترتیب با دقت‌های ۷۵ درصد، ۷۴ درصد و ۷۳ درصد قرار گرفته‌اند. علاوه بر دقت، معیارهای دیگری مانند صحت، پوشش و امتیاز اف نیز برای ارزیابی عملکرد مدل در نظر گرفته شده‌اند که نتایج آن‌ها در جدول ۳ آورده شده است. شکل ۲ نمایش دیداری عملکرد هر کدام از آواها را به تفکیک نشان می‌دهد.

جدول ۳: نتایج مقادیر دقت، صحت، پوشش و امتیاز اف درآواهای سایشی بی‌واک بر اساس مدل ماشین بردار پشتیبان.

	/f/	/s/	/ʃ/	/x/
دقت	۰,۷۴	۰,۷۷	۰,۷۵	۰,۷۳
صحت	۰,۷۲	۰,۷۴	۰,۷۴	۰,۷۱
پوشش	۰,۷۱	۰,۷۶	۰,۷۵	۰,۷۲
امتیاز اف	۰,۷۰	۰,۷۴	۰,۷۳	۰,۷۰



شکل ۲: نمودار عملکرد هر آوا در نشان‌دادن تمایزات میان گویندگان بر اساس مدل ماشین بردار پشتیبان

۵. بحث و بررسی

در این پژوهش، توانایی ویژگی‌های آکوستیکی آواهای سایشی بی‌واک زبان فارسی در تمایز بین گویندگان مختلف در داده‌های صوتی واتس‌اپ بررسی شده است. هدف اصلی این مطالعه، پاسخ به این پرسش بود که آیا ویژگی‌های آکوستیکی آواهای سایشی بی‌واک زبان فارسی در داده‌های صوتی واتس‌اپ قادر به تشخیص مؤثر گویندگان مختلف هستند یا خیر. علاوه بر این، به دنبال شناسایی بهترین آوای سایشی بی‌واک برای دستیابی به بالاترین دقت در تمایز بین گویندگان نیز بوده‌ایم.

نتایج نشان می‌دهد که آواهای سایشی بی‌واک عملکرد قابل‌قبولی در دسته‌بندی گویندگان در داده‌های صوتی واتس‌اپ داشته‌اند. این یافته بیانگر آن است که علی‌رغم شرایط ضبط غیرایده‌آل در محیطی مانند واتس‌اپ، این آواها توانسته‌اند به‌خوبی در تمایز گویندگان مؤثر باشند. یکی از دلایل احتمالی این عملکرد مطلوب، روش تحلیل آکوستیکی مبتنی بر ضرایب کپسترال فرکانسی مل است. اگرچه در بسیاری از پژوهش‌های پیشین، ضرایب کپسترال فرکانسی مل عمدتاً برای تحلیل زنجیره‌های واک‌دار، مانند واکه‌ها یا سایر بخش‌های واک‌دار گفتار، مورد استفاده قرار گرفته‌اند، یافته‌های این پژوهش نشان می‌دهد که این ضرایب می‌توانند در تحلیل سیگنال‌های بی‌واک، مانند آواهای سایشی، نیز اثربخشی بالایی داشته باشند. سیگنال‌های بی‌واک، به‌ویژه آواهای سایشی، حاوی اطلاعات فرکانسی و انرژی پراکنده‌ای هستند که عمدتاً ناشی از نویزهای تولیدشده در جریان هوا و شکل‌گیری شکاف در دستگاه گفتار می‌باشند. این خصوصیات آکوستیکی، به‌ویژه در شرایطی که تولید آن‌ها وابسته به ویژگی‌های فیزیولوژیکی و ساختاری دستگاه گفتار است، می‌تواند به عنوان عاملی متمایزکننده بین گویندگان عمل کند. شایان ذکر است که ضرایب کپسترال فرکانسی مل به تغییرات نویز حساس هستند. در

این مطالعه، داده‌های صوتی توسط شرکت‌کنندگان در محیط‌های روزمره و با استفاده از گوشی‌های هوشمند ضبط شده‌اند. بنابراین، حضور نویزهای پس‌زمینه، مانند صداهای محیطی و تداخلات غیرکنترل‌شده، اجتناب‌ناپذیر بوده است. با این حال، عملکرد موفق ضرایب کپسترال فرکانسی مل در این شرایط می‌تواند نشان‌دهندهٔ قابلیت‌های بالقوهٔ این ویژگی‌ها در تحلیل داده‌های واقعی باشد.

در رابطه با میزان تاثیرگذاری آواهای سایشی نتایج حاصل از پژوهش حاضر نشان می‌دهد که در میان آواهای سایشی مورد بررسی، آوای لثوی /s/ بیشترین تأثیرگذاری را در نشان‌دادن تفاوت‌های آکوستیکی بین گویندگان داشته است. این یافته با نتایج پژوهش‌های پیشین از جمله گوردون وهمکاران (۲۰۰۲)، کاواناگ (۲۰۱۲)، شیندلر و دراکسلر (۲۰۱۳)، اسمورنبرگ و هیرن (۲۰۲۰) و الریچ و همکاران (۲۰۲۳) همخوانی دارد. این پژوهشگران نیز بر اهمیت آوای /s/ در تشخیص هویت گویندگان تأکید کرده‌اند و دلایل آن را به ویژگی‌های تولید این آوا، از جمله میزان سایش و جایگاه تولید، نسبت می‌دهند. لازم به ذکر است که پژوهش‌های پیشین مذکور در محیط‌های آزمایشگاهی کنترل‌شده و با استفاده از داده‌های استاندارد ضبط‌شده در شرایط ایده‌آل انجام شده‌اند و نه در بستر رسانه‌های اجتماعی مانند واتس‌اپ. با این حال، علیرغم تفاوت چشمگیر در منبع داده‌ها (داده‌های واقعی واتس‌اپ در مقابل داده‌های آزمایشگاهی)، نتایج حاضر از جهت تأثیرگذاری آوای /s/ در تمایز گویندگان، همسو با یافته‌های پیشین است. گفتنی است که سایر آواهای سایشی مورد مطالعه در این پژوهش نیز توانایی تمایز گویندگان را نشان دادند، هرچند توانایی آن‌ها به اندازهٔ آوای /s/ نبوده است. این یافته نیز با نتایج پژوهش الریچ و همکاران (۲۰۲۳) در خصوص آواهای سایشی زبان روسی مطابقت دارد که نشان داد افزون بر آوای/s/، سایر سایشی‌های زبان روسی مانند /f/ نیز توان تشخیص گویندگان را دارند. همچنین، شیندلر و دراکسلر (۲۰۱۳) نیز نشان دادند که افزون بر آوای /s/ که ظرفیت بالایی در تمایز گویندگان آلمانی زبان داشت، آوای /f/ نیز توانسته است تا حد زیادی در نشان‌دادن تغییرات بین-گوینده موفق باشد. نکتهٔ قابل تأمل این است که اگرچه داده‌های پژوهش حاضر از محیط واتس‌اپ با محدودیت‌های کیفیتی خاص خود (نظیر فشرده‌سازی سیگنال و نویزهای محیطی) جمع‌آوری شده، اما همگرایی نتایج با پژوهش‌های آزمایشگاهی پیشین، نشان می‌دهد که پیشرفت فناوری در حفظ کیفیت سیگنال‌های صوتی در پیام‌رسان‌های مدرن، امکان استخراج ویژگی‌های آکوستیکی پایدار را فراهم کرده است. این امر به‌ویژه در مورد آوای /s/ که انرژی آن در بازه‌های فرکانسی بالاتر متمرکز است، مشهودتر است، چرا که بهبود کدک‌های صوتی در سال‌های اخیر، انتقال اطلاعات فرکانس بالا را در بسترهایی مانند واتس‌اپ تسهیل کرده است.

پژوهش‌هایی که تاکنون آواهای سایشی را بررسی کرده‌اند، عمدتاً بر گشتاورهای طیفی متمرکز بوده‌اند (گوردون، ۲۰۰۲؛ کاواناگ، ۲۰۱۲؛ شیندلر و دراکسلر، ۲۰۱۳). یکی از دلایلی که آوای /s/ در این پژوهش‌ها موفق بوده، مربوط به گشتاور اول یعنی پارامتر آکوستیکی مرکز تجمع انرژی

بوده است. توزیع انرژی آوای /s/ در فرکانس‌های بالا متمرکز است و از این رو گشتاور اول به‌خوبی این تمرکز را که حساس به ویژگی‌های فردی و فیزیولوژیکی است، دریافت می‌کند. این موضوع نشان می‌دهد که استفاده از گشتاورهای طیفی ممکن است تنها در برخی از آواهای سایشی از جمله صفیری‌ها مثمرثمر باشد و در آواهایی که انرژی طیف فرکانسی‌شان پراکنده است چندان عملکرد مطلوبی نداشته باشند. با این وجود در تنها پژوهشی که از دیدگاه شناسایی گوینده به نقش تمایزدهندهٔ آواهای سایشی با استفاده از ضرایب کپسترال فرکانسی مل پرداخته بود، نشان داده شد که اکثر آواها قدرتی نزدیک به هم دارند (الریچ و همکاران، ۲۰۲۳). در این پژوهش نیز که از ضرایب فرکانسی کپسترال مل استفاده شده است، تمامی چهار آوای سایشی/f/، /s/، /ʃ/ و /x/ با اختلاف‌هایی اندک عملکرد خوبی داشتند و مقادیر دقت آن‌ها نزدیک به هم بوده است. همسویی این یافته با پژوهش الریچ و همکاران (۲۰۲۳) نشان می‌دهد که انتخاب مشخصه ورودی برای سیستم شناسایی گوینده می‌تواند تأثیر قابل‌ملاحظه‌ای بر عملکرد آواها داشته باشد. همچنین، در پژوهش الریچ و همکاران (۲۰۲۳)، تفاوت قابل‌توجهی بین عملکرد سیستم‌ها زمانی که گشتاورهای طیفی به‌عنوان ویژگی ورودی و زمانی که ضرایب کپسترال فرکانسی مل به‌عنوان ویژگی ورودی استفاده شدند، مشاهده شد؛ که ضرایب کپسترال عملکرد بسیار بالاتری داشتند.

نتایج پژوهش نشان داد که مدل، هنگامی که آواها جداگانه تحلیل شدند، در مقایسه با شرایطی که همهٔ آواها به‌طور هم‌زمان بررسی شدند، عملکرد بهتری داشت. این تفاوت می‌تواند به دلایلی همچون کاهش پیچیدگی داده‌ها، تمرکز بیشتر مدل بر ویژگی‌های خاص هر آوا و کاهش احتمال همپوشانی یا تداخل ویژگی‌ها بین آواهای مختلف باشد. ضرایب کپسترال فرکانسی مل به دلیل ماهیت ریاضی‌گونهٔ خود، ویژگی‌هایی تولید می‌کنند که حساس به تغییرات فیزیکی و آکوستیکی هستند. این ویژگی‌ها ممکن است در آواهای مختلف رفتار متفاوتی داشته باشند؛ به‌عنوان مثال، ضرایب کپسترال فرکانسی مل در آوای /s/ که انرژی آن در فرکانس‌های بالا متمرکز است، با آوای /f/ که توزیع انرژی گسترده‌تری دارد، تفاوت زیادی دارد. تحلیل همهٔ آواها به‌طور همزمان، مدل را ملزم می‌کند که مرزهای تصمیم‌گیری پیچیده‌تری در فضای چندبُعدی ایجاد کند، که این امر می‌تواند به کاهش دقت منجر شود. از سوی دیگر، تحلیل جداگانهٔ آواها با کاهش تعداد داده‌ها و ساده‌تر شدن فضای تصمیم‌گیری، امکان پردازش دقیق‌تر و بهینه‌سازی بهتر پارامترهای مدل را فراهم می‌آورد.

در مجموع نتایج نشان می‌دهد که آواهای سایشی بی‌واک می‌توانند به‌عنوان سرنخ‌های آکوستیکی در شناسایی گوینده، در داده‌های صوتی پیام‌رسان‌هایی مانند واتس‌اپ که کاربرد گسترده‌ای در میان کاربران دارند، مورد استفاده قرار گیرند. با اینکه سایشی‌ها پس از واکه‌ها دومین دسته از آواهای مورد توجه زبان‌شناسان و آواشناسان فعال در حوزه‌های قضایی هستند، اما به دلایلی نظیر حساسیت این آواها به کانال‌های ارتباطی و از دست رفتن اطلاعات مربوط به

فرکانس‌های بالا در خطوط تلفنی، تاکنون توجه کمتری به آن‌ها شده است. با این حال، نتایج این پژوهش نشان می‌دهد که با پیشرفت فناوری در عرصهٔ ارتباطات، کانال‌های ارتباطی تلاش کرده‌اند تا با بهبود کیفیت سیگنال‌های ورودی و خروجی، اطلاعات آوایی را بهتر حفظ کرده و فهم‌پذیری ارتباطات و رضایت کاربران خود را افزایش دهند. این پیشرفت‌ها باعث شده است تا آواهای سایشی نقش مؤثرتری در فرایندهای تشخیص گفتار و شناسایی گوینده ایفا کنند.

۶. نتیجه

این پژوهش با هدف بررسی توانایی ویژگی‌های آکوستیکی آواهای سایشی بی‌واک زبان فارسی در تمایز بین گویندگان مختلف در داده‌های صوتی ضبط‌شده در محیط واتس‌اپ انجام شد. نتایج نشان داد که این آواها، علی‌رغم شرایط ضبط غیرایده‌آل، می‌توانند به‌عنوان سرنخ‌های آکوستیکی مؤثر در فرآیند شناسایی گوینده عمل کنند. ضرایب کپسترال فرکانسی مل، به دلیل حساسیت بالای خود به تغییرات آکوستیکی، عملکرد قابل‌توجهی در تمایز بین گویندگان داشتند. به‌ویژه، آوای لثوی /s/ با دقت ۷۷ درصد بیشترین تأثیر را در شناسایی گویندگان داشت. تحلیل جداگانهٔ آواها در مقایسه با تحلیل ترکیبی آن‌ها، به دلیل کاهش پیچیدگی و تمرکز بیشتر بر ویژگی‌های خاص هر آوا، دقت بالاتری ارائه کرد. این یافته‌ها نشان می‌دهند که با پیشرفت فناوری‌های ارتباطی و بهبود کیفیت کانال‌های ارتباطی، استفاده از آواهای سایشی به‌عنوان ابزاری مؤثر در شناسایی گویندگان در محیط‌های کاربردی نظیر پیام‌رسان‌هایی مانند واتس‌اپ، قابلیت اجرایی بالایی دارد. این پژوهش محدودیت‌هایی نیز داشت. یکی از محدودیت‌های این پژوهش، تمرکز صرف بر گویشوران مرد بود که ممکن است نتایج را تنها به این گروه محدود کند. برای افزایش تعمیم‌پذیری نتایج، پیشنهاد می‌شود در پژوهش‌های آینده داده‌های گفتاری از گویشوران زن و گروه‌های متنوع‌تر سنی و اجتماعی جمع‌آوری شود. همچنین، بررسی تأثیر شرایط مختلف ضبط، مانند کیفیت‌های گوناگون میکروفون و محیط‌های ضبط، می‌تواند به درک پایداری این روش‌ها کمک کند. افزون بر این، پیشنهاد می‌شود ترکیب ضرایب کپسترال فرکانسی مل با سایر ویژگی‌های آکوستیکی، نظیر گشتاورهای طیفی بررسی شود تا دقت و قابلیت سیستم‌های شناسایی گوینده بهبود یابد.

منابع

اسدی، ه.، نوربخش م.، ساسانی ف.، تفاوت های بین–گوینده در سایشی‌های بی‌واک زبان فارسی. جستارهای زبانی. ۱۳۹۸؛ ۱۰ (۱) :۱۲۹–۱۴۷.

اسدی، ه.، حسینی کیونانی، ن.، و نوربخش، م. (۱۳۹۴). بررسیِ تأثیر فراخوانی صورت بر ویژگی‌های آکوستیکی سایشی‌های بی‌واک زبان فارسی: پژوهشی در چارچوب آواشناسی قضایی. زبان‌شناسی و گویش‌های خراسان، ۷(۱۳)، ۱–۱۵.

Boersma, P., & Weenink, D. (2025). *Praat: Doing phonetics by computer* (Version 6.4.26) [Computer software]. University of Amsterdam. http://www.praat.org

Catford, J. C. (1977). *Fundamental problems in phonetics*. Edinburgh University Press. https://doi.org/10.2307/412751

Dellwo, V., Huchvale, M., & Ashby, M. (2007). How is individuality expressed in voice? An introduction to speech production and description for speaker classification. In C. Müller (Ed.), *Speaker identification* (Vol. 1, pp. 1–20). Springer. https://doi.org/10.1007/978-3-540-74200-5_1

Gold, E., & French, P. (2011). International practices in forensic speaker comparison. *International Journal of Speech, Language and the Law*, *18*(2). https://doi.org/10.1558/ijsll.v18i2.293

Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. Journal of the International Phonetic Association, 32(2), 141-174. https://doi.org/10.1017/S0025100302001020

Gouri, G., Sharma, A., & Sharma, V. (2024). Forensic speaker and gender identification from voice samples recorded through mobile phones and social media applications: A statistical and machine learning approach. *Applied Acoustics*, *222*, 110074. https://doi.org/10.1016/j.apacoust.2024.110074

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass, 2*(4), 671–711. https://doi.org/10.1111/j.1749-818x.2008.00066.x

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252-1263. https://doi.org/10.1121/1.1288413

Karpisek, F., Baggili, I., & Breitinger, F. (2015). WhatsApp network forensics: Decrypting and understanding the WhatsApp call signaling messages. Digital Investigation, 15, 110-118. https://doi.org/10.1016/j.diin.2015.09.002

Kavanagh, C. (2012). *New consonantal acoustic parameters for forensic speaker comparison* (Doctoral dissertation, University of York). https://etheses.whiterose.ac.uk/id/eprint/3980/

Kinnunen, T., & Li, H. (2010). An overview of text-independent speaker recognition: From features to supervectors. *Speech Communication*, *52*(1), 12-40. https://doi.org/10.1016/j.specom.2009.08.009

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347. https://doi.org/10.1016/j.csl.2017.01.005

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

Lindh, J. (2017). Forensic comparison of voices, speech and speakers: Tools and methods in forensic phonetics**.** University of Gothenburg. Retrieved from https://gupea.ub.gu.se/handle/2077/52188

McKinney, W. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference* (pp. 51–56). https://pandas.pydata.org/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J.,

Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://scikit-learn.org/stable/

Reetz, H., & Jongman, A. (2009). *Phonetics: Transcription, production, acoustics, and perception* (1st ed.). Wiley-Blackwell.

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

Schindler, C., & Draxler, C. (2013). Using spectral moments as a speaker-specific feature in nasals and fricatives. In *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7849-7853). https://doi.org/10.21437/Interspeech.2013-639

Shadle, C. H. (1990). Articulatory-acoustic relationships in fricative consonants. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 187–209). Springer. http://eprints.soton.ac.uk/id/eprint/250178

Smorenburg, L., & Heeren, W. (2020). The distribution of speaker information in Dutch fricatives /s/ and /x/ from telephone dialogues. *The Journal of the Acoustical Society of America*, *147*(4), 2554-2567. https://doi.org/10.1121/10.0000674

Statista Research Department. (2021). Most popular global mobile messaging apps 2021. Retrieved October 16, 2023, from https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/

Statista Research Department. (2023). Number of unique WhatsApp mobile users worldwide from January 2020 to June 2023. Retrieved October 16, 2023, from https://www.statista.com/statistics/1306022/whatsapp-global-unique-users/

Stuart-Smith, J. (2007). Empirical evidence for gendered speech production: /s/ in Glaswegian. In J. Cole & J. Hualde (Eds.), *Change in phonology: Papers in laboratory phonology*. Mouton de Gruyter. https://eprints.gla.ac.uk/8985/

Temko, A., & Nadeu, C. (2005). Classification of acoustic events using SVM-based clustering schemes. *TALP Research Center, Universitat Politècnica de Catalunya*. https://upcommons.upc.edu/bitstream/handle/2117/2065/classification.pdf?sequence=3'

Ulrich, N., Pellegrino, F., & Allassonnière-Tang, M. (2023). Intra- and inter-speaker variation in eight Russian fricatives. *The Journal of the Acoustical Society of America*, *135*(4), 2098-2109 .https://doi.org/10.1121/10.0017827