Journal of Linguistics and Khorasan Dialects
https://jlkd.um.ac.ir

https://doi.org/10.22067/jlkd.2025.91521.1295

Quarterly Journal of Linguistics & Khorasan Dialects
Vol. 17, No. 1 /Serial Number 38 (Spring 2025)

Analysis of Acoustic Voice Quality Parameters for Identifying Persian Speakers

Homa Asadi[1]
Assistant Professor of Linguistics, University of Isfahan, Isfahan, Iran
https://orcid.org/0000-0003-1655-1336

Received: January 5, 2025 ⁘ Revised: January 24, 2025
Accepted: January 26, 2025 ⁘ Published Online: June 07, 2025

How to cite this article:
Asadi, H. (2025). Analysis of Acoustic Voice Quality Parameters for Identifying Persian Speakers. Journal of Linguistics and Khorasan Dialects, 17 (1), 1-21. (in Persian with English abstract)
https://doi.org/10.22067/jlkd.2025.91521.1295

Abstract

This study acoustically examines voice quality parameters in two groups of Persian-speaking men and women. It aimed to assess the ability of voice quality parameters to differentiate Persian speakers and to evaluate the extent to which these parameters capture speaker-specific information. Additionally, this research sought to expand existing knowledge in the field of voice quality and address the limited scope of previous studies on Persian. Acoustic data were collected from 20 female and 20 male speakers in a laboratory setting. Multivariate Analysis of Variance (MANOVA) was used to analyze inter-speaker differences, and the Random Forest Algorithm was employed to assess feature importance. Six voice quality parameters were selected for analysis: jitter (frequency perturbation), shimmer (amplitude perturbation), harmonic-to-noise ratio (HNR), the ratio of the amplitudes of the first and second harmonics (H1-H2), cepstral peak prominence (CPP), and fundamental frequency (F0). The results demonstrated significant acoustic differences among Persian speakers based on voice quality features, though the discriminative power of these features was not uniform. For male speakers, CPP, HNR, and H1-H2 were identified as the most discriminative features, respectively. For female speakers, F0, CPP, and HNR emerged as the key features for speaker identification. The findings highlight the significant role of voice quality parameters in identifying Persian speakers. However, achieving higher accuracy in speaker recognition systems requires considering gender differences and the relative importance of various variables. Moreover, the limited number of participants may affect the generalizability of the results. Thus, future studies are recommended to include larger and more diverse speaker samples.

Keywords: Acoustic Phonetics, Speaker-Specific Information, Voice Quality, Persian Speech.

1. Email: h.asadi@fgn.ui.ac.ir

نشریه علمی
زبانشناسی و گویش های خراسان
https://jlkd.um.ac.ir

https://doi.org/10.22067/jlkd.2025.91521.1295https://doi.org/10.22067/jlkd.2024.89453.1267

مجلهٔ زبانشناسی و گویش‌های خراسان، شمارهٔ ۳۸، بهار ۱۴۰۴
مقالهٔ پژوهشی

تحلیل پارامترهای آکوستیکی کیفیت صدا برای شناسایی گویندگان فارسی‌زبان

هما اسدی

استادیار زبانشناسی، دانشگاه اصفهان، اصفهان، ایران [1]

https://orcid.org/0000-0003-1655-1336

صص ۱-۲۱

ارجاع به این مقاله:

اسدی، ه. (۱۴۰۴). «تحلیل پارامترهای آکوستیکی کیفیت صدا برای شناسایی گویندگان فارسی‌زبان»، زبانشناسی و گویش‌های خراسان، بهار، صص ۱-۲۱.

https://doi.org/10.22067/jlkd.2025.91521.1295

چکیده

پژوهش حاضر به بررسی آکوستیکی پارامترهای کیفیت صدا در دو گروه زنان و مردان فارسی‌زبان می‌پردازد. این مطالعه با هدف ارزیابی توانایی پارامترهای کیفیت صدا در تمایز گویندگان فارسی‌زبان و بررسی میزان اثرگذاری این پارامترها در شناسایی ویژگی‌های گوینده‌محور طراحی شده است. علاوه بر این، با هدف گسترش دانش موجود در حوزهٔ کیفیت صدا و پر کردن خلأ مطالعات محدود پیشین در زبان فارسی انجام شده است. داده‌های آوایی از ۲۰ گویشور زن و ۲۰ گویشور مرد در محیط آزمایشگاهی ضبط شدند. برای تحلیل تفاوت‌های میان گویندگان از آزمون تحلیل واریانس چندمتغیره و برای ارزیابی اهمیت ویژگی‌ها، از الگوریتم جنگل تصادفی بهره گرفته شد. شش پارامتر کیفیت صدا شامل فرکانس‌پریشی، دامنه‌پریشی، نسبت هارمونیک به نویز، نسبت دامنهٔ هارمونیک‌های اول و دوم، برجستگی قلهٔ طیفی و فرکانس پایه انتخاب شدند. نتایج نشان داد ویژگی‌های کیفیت صدا در نشان‌دادن تفاوت‌های آکوستیکی میان گویندگان فارسی‌زبان دارای تفاوت‌های معناداری بوده‌اند، اگرچه توانایی آن‌ها در تمایز گویندگان به‌طور یکسان نبوده است. برای گویندگان مرد، پارامترهای برجستگی قلهٔ طیفی، نسبت هارمونیک به نویز و نسبت دامنهٔ هارمونیک‌های اول و دوم، به ترتیب بیشترین توانایی را در تمایز آن‌ها از یکدیگر دارند. برای گویندگان زن، فرکانس پایه، برجستگی قلهٔ طیفی و نسبت هارمونیک به نویز به‌عنوان مهم‌ترین ویژگی‌ها برای تشخیص هویت شناخته شدند. نتایج این پژوهش نشان می‌دهد پارامترهای کیفیت صدا نقش قابل‌توجهی در شناسایی گویندگان فارسی‌زبان دارند. بااین‌حال، برای دستیابی به دقت بالاتر در سیستم‌های شناسایی گوینده، توجه به تفاوت‌های جنسیتی و اهمیت متغیرهای مختلف ضروری است. از طرف دیگر، محدودیت تعداد شرکت‌کنندگان ممکن است بر تعمیم‌پذیری نتایج تأثیر بگذارد؛ بنابراین، پیشنهاد می‌شود در پژوهش‌های آینده، از نمونه‌های بزرگ‌تر و تنوع بیشتر در گویندگان استفاده شود.

کلیدواژه‌ها: آواشناسی آکوستیکی، اطلاعات گوینده‌محور، کیفیت صدا، گفتار فارسی.

دریافت: ۱۴۰۳/۱۰/۱۶ ؛ بازنگری: ۱۴۰۳/۱۱/۰۵ ؛ پذیرش: ۱۴۰۳/۱۱/۰۷ ؛ انتشار برخط: ۱۴۰۴/۰۳/۱۷

پست الکترونیکی: 1. h.asadi@fgn.ui.ac.ir

۱. مقدمه

صدا، همانند اثرانگشت صوتی، هویت منحصربه‌فرد هر فرد را آشکار می‌سازد. بلین و همکاران (۲۰۰۴) صدا را "چهرهٔ صوتی" می‌نامند و معتقدند همان‌گونه که انسان از روی چهره قابل‌شناسایی است، صدایش نیز می‌تواند در شناسایی هویتش تأثیرگذار باشد. کیفیت صدا[1]، به‌عنوان یکی از بنیادی‌ترین ویژگی‌های گفتار، در این شناسایی نقشی محوری ایفا می‌کند. صدا نه‌تنها ابزار انتقال کلمات است، بلکه احساسات، نگرش‌ها و حتی هویت اجتماعی فرد را نیز منتقل می‌کند (لاور، ۱۹۶۸؛ کرایمن و سیتدیس، ۲۰۱۱). تحلیل‌های آکوستیکی، با بررسی دقیق ویژگی‌های صوتی، لایه‌های پنهان این ارتباط پیچیده را آشکار کرده و اطلاعات ارزشمندی دربارهٔ خصوصیات فیزیولوژیکی، عاطفی و حتی شخصیتی گوینده ارائه می‌دهند.

کیفیت صدا، به معنای گستردهٔ آن، به‌عنوان ویژگی‌های بلندمدتی تعریف می‌شود که دائماً در الگوی صدای گفتار یک گوینده وجود دارند (ابرکرومبی، ۱۹۶۷؛ اسلینگ، ۲۰۰۶). بااین‌حال، در معنای محدودتر، کیفیت صدا صرفاً به "انواع واک‌سازی[2] تولید شده در حنجره" اشاره دارد (اسلینگ ۲۰۱۳). در هر دو معنا، ویژگی‌های کیفیت صدا سیگنال‌های اطلاعاتی هستند که به شنوندگان اجازه می‌دهند نه‌تنها ویژگی‌های فیزیولوژیکی دائمی مانند جنسیت گوینده، بلکه ویژگی‌های اجتماعی یا روان‌شناختی یک گوینده معین را نیز درک کنند. بااین‌وجود، ماهیت این تغییرات در فضای صوتی گفتار گویندگان و نحوهٔ برخورد شنوندگان با این تغییرات برای تشخیص هویت یک گوینده هنوز نیاز به درک کامل دارد.

کیفیت صدا معمولاً از دو منبع اصلی ناشی می‌شود. اولاً، این ویژگی‌ها به‌شدت با ویژگی‌های آناتومیکی و فیزیولوژیکی دستگاه صوتی یک گوینده، یعنی مختصات حنجره و نیز ساختار مجرای صوتی فوق‌العاده مرتبط هستند؛ و ثانیاً، این ویژگی‌ها به "تنظیمات عضلانی بلندمدت"، که معمولاً یا از طریق عملکرد خاص اندام‌های گفتاری مفصلی یا از طریق رفتارهای آموخته‌شده به دست می‌آید نیز بستگی زیادی دارند (لاور، ۱۹۶۸). با وجود تفاوت‌های ذاتی میان عوامل زیست‌شناختی و اکتسابی، هر دو عامل اطلاعات خاص گوینده‌محور را به گفتار منتقل می‌کنند، که در ویژگی‌های آکوستیکی کلام گویندگان منعکس می‌شود و به‌عنوان سرنخ‌هایی آکوستیکی برای شنوندگان جهت شناسایی گوینده، حتی زمانی که گفتار بسیار متغیر است، عمل می‌کند.

1. voice quality
2. phonation types

کیفیت صدا در پارامترهای آکوستیکی متعددی نمود پیدا می‌کند. پارامترهایی نظیر فرکانس پایه[1]، فرکانس‌پریشی[2]، دامنه‌پریشی[3]، برجستگی قلهٔ طیفی[4]، نسبت نویز به هارمونیک[5] و نسبت دامنه هارمونیک‌های اول و دوم[6] به‌عنوان شاخص‌هایی کلیدی در رمزگذاری کیفیت صدا شناخته می‌شوند (کرایمن و سیتدیس،۲۰۱۱). فرکانس‌پریشی و دامنه‌پریشی، که نوسانات کوچک در فرکانس پایه و شدت سیگنال صوتی را منعکس می‌کنند، تحت تأثیر ویژگی‌های فیزیولوژیکی حنجره و تارهای صوتی هستند (تیخیرا و همکاران، ۲۰۱۳). این نوسانات، به دلیل تفاوت‌های فردی در ساختار حنجره و تارهای صوتی، میان گویندگان مختلف متفاوت است. به‌عنوان‌مثال، لرزش‌های کوچک در تارهای صوتی یا حتی تفاوت‌های جزئی در ساختار حنجره می‌توانند مقادیر این پارامترها را تغییر دهند. برجستگی قلهٔ طیفی به‌عنوان معیاری از کیفیت آکوستیکی صدا، نسبت انرژی هارمونیکی به انرژی نویز در طیف صوتی را اندازه‌گیری می‌کند. این پارامتر، که در تشخیص صداهای طبیعی از غیرطبیعی نیز کاربرد دارد، برای ارزیابی صدای پایدار و واضح اهمیت دارد (مورتون و همکاران، ۲۰۲۰). نسبت نویز به هارمونیک نیز معیاری برای سنجش کیفیت صدا است که نشان‌دهندهٔ شفافیت و نویز سیگنال صوتی است (تیخیرا و همکاران، ۲۰۱۳). همچنین، نسبت دامنهٔ هارمونیک‌های اول و دوم شاخصی از تنش و ارتعاش تارهای صوتی است و اطلاعاتی دربارهٔ نوع صدای تولیدشده ارائه می‌دهد (چای و همکاران، ۲۰۲۲). فرکانس پایه نیز به تعداد دفعات ارتعاش تارهای صوتی در ثانیه اشاره دارد و مستقیماً به عوامل فیزیکی حنجره مانند طول و حجم تارهای صوتی وابسته است (لدفوگد و جانسون، ۲۰۱۵). این پارامترها، به دلیل انعکاس ویژگی‌های فیزیولوژیکی و آکوستیکی فرد، برای مطالعه‌های فرد ویژگی و تفکیک گویندگان اهمیت ویژه‌ای دارند.

پژوهش‌های پیشین نشان داده‌اند که پارامترهای کیفیت صدا می‌توانند به‌طور معناداری در میان گویندگان مختلف تفاوت داشته باشند (لی و همکاران، ۲۰۱۹؛ لی و کرایمن، ۲۰۲۲a). به‌عنوان‌مثال، تحقیقاتی که بر روی گویندگان زبان‌های مختلف انجام شده، نشان داده‌اند که برخی از این پارامترها مانند فرکانس‌پریشی و دامنه‌پریشی، به دلیل تأثیرپذیری از ویژگی‌های آناتومیکی و فیزیولوژیکی، می‌توانند به‌عنوان شاخص‌های گوینده‌محور عمل می‌کنند (هیوز و همکاران، ۲۰۲۳). علاوه بر این، مطالعات دیگری نیز نشان داده‌اند که پارامترهایی مانند نسبت نویز به هارمونیک

1. fundamental frequency (*f0*)
2. jitter
3. shimmer
4. cepstral peak prominence (CPP)
5. harmoinc-to-noise ratio (HNR)
6. the difference between first and second harmonic amplitudes (H1-H2)

و ضریب انرژی صدا می‌توانند برای تمایز میان صدای طبیعی و غیرطبیعی مفید باشند (شاما و همکاران، ۲۰۰۶؛ مورتون و همکاران، ۲۰۲۰؛ فرناندز و همکاران، ۲۰۲۳).

با وجود نقش کلیدی پارامترهای کیفیت صدا در تمایز گویندگان، تأثیر این پارامترها در زبان فارسی به‌ویژه از دیدگاه شناسایی گوینده کمتر موردمطالعه قرار گرفته است. این پارامترها خاصیت زبان‌ویژه نیز دارند و در رمزگذاری ویژگی‌های زبانی نظیر تکیه و نواخت در برخی زبان‌ها نیز نقش دارند (واگنر و براون، ۲۰۰۳). از سوی دیگر، پارامترهای فردویژه ممکن است میان زبان‌ها متفاوت باشند. بدین معنا که پارامتری که در یک زبان مختصات گوینده را رمزگذاری می‌کند، ممکن است در زبان دیگر چنین نقشی نداشته باشد (کینوشیتا، ۲۰۰۱). ازاین‌رو، پژوهش حاضر در صدد بررسی میزان توانایی پارامترهای آکوستیکی کیفیت صدا در متمایز کردن افراد و میزان فردویژگی آن‌ها در زبان فارسی است. پرسش‌های این پژوهش عبارتند از:

۱) کدام‌یک از این پارامترهای کیفیت صدا تأثیر بیشتری برای نشان‌دادن تفاوت میان گویندگان دارند؟

۲) آیا تأثیر پارامترهای کیفیت صدا در تمایز بین گویندگان مختلف بسته به جنسیت گویندگان متفاوت است؟

این پژوهش بر روی پایگاه داده‌ای از گفتار طبیعی گویشوران فارسی‌زبان انجام خواهد شد. داده‌ها با استفاده از روش‌های آماری و الگوریتم‌های یادگیری ماشین تحلیل می‌شوند تا مشخص شود کدام پارامترها بیشترین نقش را در تمایز گویندگان ایفا می‌کنند. نتایج این پژوهش می‌تواند در حوزه‌هایی چون فناوری‌های شناسایی گفتار، آسیب‌شناسی صدا و مطالعات زبان‌شناسی کاربردهای متعددی داشته باشد. همچنین، تحلیل این پارامترها می‌تواند به درک بهتر فرآیند تولید صدا و شناسایی ویژگی‌های فردی کمک کند.

۲- پیشینهٔ پژوهش

آواشناسان حوزهٔ شناسایی گوینده همواره در پی یافتن بهترین ویژگی‌های صوتی بوده‌اند تا بتوانند تفاوت‌های صوتی میان افراد مختلف را به‌وضوح نشان دهند و درعین‌حال، تفاوت‌های صوتی درون یک فرد را به حداقل برسانند (نولان، ۱۹۸۳؛ رز، ۲۰۰۲؛ اسدی و همکاران، ۱۳۹۴). در سال‌های اخیر، پژوهشگران به ویژگی‌های کیفیت صدا توجه ویژه‌ای نشان داده‌اند و این ویژگی‌ها به‌طور گسترده در سیستم‌های خودکار شناسایی گوینده مورداستفاده قرار گرفته‌اند.

بیشتر پژوهش‌های انجام‌شده در زمینهٔ شناسایی گوینده با استفاده از مجموعه ویژگی‌های کیفیت صدا، به چند سال اخیر تعلق دارند. البته در میان پارامترهای کیفیت صدا، فرکانس پایه همواره به‌عنوان یک شاخص کلیدی در مطالعات شناسایی گوینده مطرح بوده است. پژوهش‌های متعددی در زبان‌های مختلف، بر نقش برجستهٔ فرکانس

پایه در تمایز بین گویندگان تأکید کرده‌اند (رز، ۲۰۰۲؛ مک دوگال،۲۰۰۴؛ یسن و همکاران، ۲۰۰۵؛ لابوتین و همکاران، ۲۰۰۷؛ اسدی و همکاران، ۲۰۱۸). بااین‌حال، به‌رغم اهمیت فرکانس پایه، سایر پارامترهای کیفیت صدا کمتر موردتوجه قرار گرفته‌اند.

پارک و همکاران (۲۰۱۶) در مطالعهٔ خود، به بررسی تأثیر ویژگی‌های کیفیت صدا بر دقت سیستم‌های تشخیص هویت گوینده پرداختند. آن‌ها از پایگاه دادهٔ صوتی UCLA که شامل صداهای گویندگان انگلیسی‌زبان است، استفاده کردند. در این پژوهش، دو دسته از ویژگی‌ها موردبررسی قرار گرفت: ضرایب طیفی فرکانسی مل[1] و ویژگی‌های کیفیت صدا. هدف اصلی، ارزیابی تأثیر ترکیب این دو نوع ویژگی بر عملکرد سیستم تشخیص هویت بود. نتایج نشان داد که ویژگی‌های کیفیت صدا به‌تنهایی و همچنین در ترکیب با ضرایب طیفی فرکانسی مل، به بهبود عملکرد سیستم کمک شایانی می‌کند. بر اساس نتایج، عملکرد سیستم پس از ترکیب این دو دسته از پارامترها حدود ۱۵ درصد بهبود نشان داده است. در مطالعه‌ای دیگر، پارک و همکاران (۲۰۱۷) به بررسی پارامترهای کیفیت صدا در پاره‌گفتارهای کوتاه استخراج‌شده از سبک‌های گفتاری متنوعی همچون خوانشی، بداهه و سبک گفتار مخصوص حیوانات خانگی پرداختند. هدف اصلی این پژوهش، بررسی پایداری ویژگی‌های کیفیت صدا در شرایط آکوستیکی بسیار متغیر و متنوع گفتاری بود؛ به‌عبارت‌دیگر، محققان می‌خواستند بدانند آیا ویژگی‌های صوتی که معمولاً به گویندهٔ خاص نسبت داده می‌شوند، در شرایط گفتاری مختلف و متغیر همچنان پایدار باقی می‌مانند یا خیر. نتایج این مطالعه نشان داد که افزودن ویژگی‌های کیفیت صدا به سیستم شناسایی گوینده، به‌طور قابل‌توجهی عملکرد این سیستم را بهبود بخشیده است.

لی و همکاران (۲۰۱۹) با هدف بررسی نقش پارامترهای کیفیت صدا در ایجاد تفاوت‌های صوتی بین افراد و همچنین تغییرات صوتی در صدای یک فرد، ۱۳ پارامتر مرتبط با کیفیت صدا و نوسانات این پارامترها را موردمطالعه قرار دادند. آن‌ها پژوهش خود را بر روی داده‌های صوتی ۵۰ زن و ۵۰ مرد انگلیسی‌زبان که با سبک خوانشی صحبت کرده بودند، انجام دادند. نتایج نشان داد که این پارامترهای کیفی صدا به‌خوبی می‌توانند تفاوت‌های صوتی بین افراد را نشان دهند و به‌ویژه، نسبت دامنه هارمونیک‌های اول و دوم در این تفاوت گذاری نقش مهمی دارد. لی و کرایمن (۲۰۲۲a) در ادامه پژوهش قبلی خود، به بررسی همان پارامترهای کیفیت صدا در سبک گفتاری بداهه پرداختند. هدف آن‌ها از این کار، بررسی این موضوع بود که آیا نتایج به‌دست‌آمده در سبک خوانشی، قابل‌تعمیم به سبک‌های

1. Mel-frequency cepstral coefficients

گفتاری دیگر نیز هست یا خیر. آن‌ها فرض کردند ازآنجایی‌که سبک خوانشی الگوهای صوتی منظم‌تری دارد، ممکن است نتایج حاصل از آن به‌طور کامل به سبک بداهه قابل‌تعمیم نباشد. ازاین‌روی، آن‌ها برای بررسی بیشتر، سبک گفتاری بداهه را انتخاب کردند و مجدداً همان پارامترهای کیفی صدا و نوسانات آن‌ها را در داده‌های این سبک، که از همان گروه گویندگان استخراج شده بود، مورد تحلیل قرار دادند. نتایج این پژوهش نیز نشان داد که پارامترهای کیفیت صدا در سبک گفتاری بداهه نیز می‌توانند تفاوت‌های بین افراد را نشان دهند. بااین‌حال، اهمیت نسبی این پارامترها در سبک بداهه متفاوت بود. به‌طور خاص، پارامتر فرکانس پایه که در سبک خوانشی اهمیت کمتری داشت، در سبک بداهه از اهمیت بیشتری برخوردار بود. لی و کرایمن (۲۰۲۲b) در پژوهشی دیگر، به بررسی تفاوت‌های بین‌زبانی پارامترهای کیفیت صدا در سه زبان انگلیسی، کره‌ای سئول و وایت همونگ[1] پرداختند. نتایج این پژوهش نشان داد که این پارامترها به‌طور مشترک، تفاوت‌های میان گویندگان این سه زبان را آشکار می‌سازند. بیشترین تغییرات آکوستیکی در پارامترهای مرتبط با هارمونیک‌ها و انرژی مشاهده شد که به‌طور مستقیم با ویژگی‌های فیزیکی دستگاه گفتار در ارتباط هستند. محققان دریافتند که با افزایش پیچیدگی واج‌شناختی یک زبان، ویژگی‌های زبان‌ویژه مانند نسبت دامنهٔ هارمونیک‌های اول و دوم، نقش پررنگ‌تری در شکل‌دهی فضای آکوستیکی گویندگان ایفا می‌کنند.

ووبی و همکاران (۲۰۲۱) با رویکردی رایانشی، پیشنهاد دادند که پارامترهای کیفیت صدا، به‌ویژه فرکانس‌پریشی و دامنه‌پریشی، به‌عنوان ویژگی‌های مکمل در کنار ضرایب طیفی فرکانسی مل در مدل‌های شبکه‌های عصبی عمیق[2] برای شناسایی گوینده به کار روند. پژوهش ووبی و همکاران (۲۰۲۱) نشان داد که ادغام ویژگی‌های کیفیت صدا با ویژگی‌های طیفی در مدل‌های شناسایی گوینده، منجر به ارتقای قابل‌توجهی در دقت سیستم شده است. نتایج حاصل از تحلیل آن‌ها بر روی یک پایگاه دادهٔ صوتی بزرگ زبان انگلیسی، مؤید بهبود حدود ۱۵ درصدی در عملکرد سیستم است. این امر نشان می‌دهد که پارامترهای کیفیت صدا نه‌تنها در مدل‌های ساده‌تر، بلکه در مدل‌های پیشرفته‌تر پردازش گفتار و تشخیص گوینده نیز به‌عنوان ویژگی‌های قدرتمندی عمل می‌کنند و به‌طور قابل‌توجهی به افزایش دقت این سیستم‌ها کمک می‌نمایند. این امر نشان‌دهندهٔ اهمیت در نظر گرفتن این پارامترها در طراحی و توسعهٔ الگوریتم‌های پیچیده‌تر پردازش سیگنال گفتار است. در همین راستا، هیوز و همکاران (۲۰۲۳) با تمرکز بر ویژگی‌های آکوستیکی، به بررسی توانایی پارامترهای کیفیت صدا و ویژگی‌های فیلتر در تمایز بین گویندگان پرداختند. آن‌ها با

1. White Hmong
2. deep neural networks

استخراج ویژگی‌های مختلف از نشانگر تردید "um" در داده‌های صوتی ۹۰ گویندهٔ مرد انگلیسی‌زبان با لهجهٔ استاندارد جنوب بریتانیا، به این نتیجه رسیدند که ترکیب پارامترهای کیفیت صدا (به‌عنوان ویژگی‌های منبع) با ویژگی‌های فیلتر (مانند فرکانس سازه و ضرایب طیفی فرکانس مل) می‌تواند منجر به بهبود قابل‌توجهی در دقت تشخیص گوینده شود. این یافته، اهمیت مکمل بودن اطلاعات موجود در ویژگی‌های منبع و فیلتر را در شناسایی هویت گوینده برجسته می‌سازد.

پژوهش‌های پیشین به‌طور گسترده‌ای به نقش کلیدی پارامترهای کیفیت صدا در شناسایی گویندگان پرداخته‌اند. بااین‌حال، بررسی این پارامترها در زبان فارسی و به‌ویژه در زمینهٔ شناسایی گوینده، کمتر موردتوجه قرار گرفته است. در این پژوهش، با بهره‌گیری از یافته‌های تحقیقات پیشین، به بررسی توانایی پارامترهای کیفیت صدا در تمایز گویندگان فارسی‌زبان پرداخته‌ایم.

۳- روش‌شناسی پژوهش

در این بخش، روش‌شناسی پژوهش به‌تفصیل شرح داده می‌شود. ابتدا جامعهٔ آماری و داده‌های صوتی مورداستفاده در مطالعه معرفی خواهند شد. سپس به تعریف و توضیح پارامترهای کیفیت صدا پرداخته خواهد شد. در ادامه، نحوهٔ استخراج این پارامترهای آکوستیکی از فایل‌های صوتی با استفاده از نرم‌افزار پرات (Praat) به‌صورت گام‌به‌گام توضیح داده خواهد شد. درنهایت، مراحل تحلیل آماری انجام‌شده بر روی داده‌های استخراج‌شده و روش‌های آماری به‌کاررفته برای پاسخ‌گویی به پرسش‌های پژوهش ارائه خواهد شد.

۳-۱- شرکت‌کنندگان و داده‌های آوایی

در این پژوهش، ۴۰ شرکت‌کننده (۲۰ مرد و ۲۰ زن) با میانگین سنی ۳۲.۶ سال، همگی دانشجو، شرکت داشتند. برای جمع‌آوری داده‌های صوتی، شرکت‌کنندگان در یک محیط آزمایشگاهی استاندارد قرار گرفتند و صحبت‌های آن‌ها با استفاده از یک میکروفونSM7B SHURE ضبط شد. میکروفون به فاصلهٔ ۲۰ سانتی‌متری از دهان شرکت‌کنندگان قرار داده شد. از شرکت‌کنندگان خواسته شد تا به‌صورت آزاد و دربارهٔ هر موضوع دلخواهی به مدت زمان ۱۰ دقیقه صحبت کنند. برای تسهیل انتخاب موضوع، فهرستی از موضوعات متنوع مانند رشتهٔ تحصیلی، سرگرمی و نحوهٔ گذراندن اوقات فراغت ارائه شد و پرسش‌هایی نیز در این زمینه مطرح گردید. پیش از آغاز ضبط، از تمامی شرکت‌کنندگان اطمینان حاصل شد که هیچ سابقه‌ای از اختلالات گفتاری یا شنوایی ندارند. روند ضبط

به‌صورت ترتیبی بود و هر شرکت‌کننده به‌تنهایی وارد محیط ضبط می‌شد و داده‌های خود را تولید می‌کرد. داده‌های صوتی با نرخ نمونه‌برداری ۴۴ کیلوهرتز و وضوح ۱۶ بیت ضبط شده و به‌عنوان مبنایی برای تحلیل‌های آکوستیکی بعدی مورداستفاده قرار گرفتند.

۲-۳- پارامترهای کیفیت صدا

در این پژوهش پارامترهای کیفیت صدا از داده‌های صوتی استخراج شدند. پارامترهای موردبررسی در قالب جدول ۱ ارائه شده است.

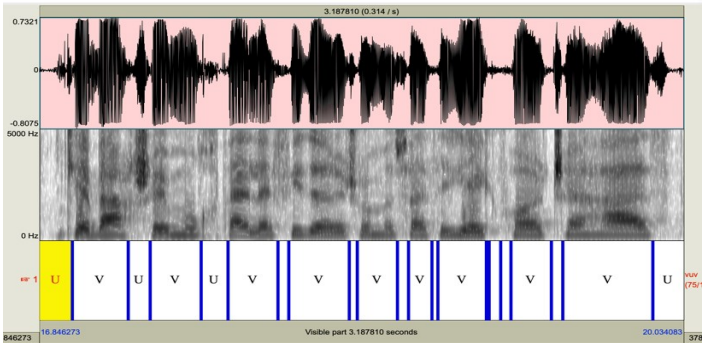
جدول ۱ : تعریف پارامترهای آکوستیکی کیفیت صدا

پارامتر	تعریف
فرکانس‌پریشی	میزان نوسانات تصادفی در دورهٔ تناوب ارتعاش تارهای صوتی
دامنه‌پریشی	اندازه‌گیری نوسانات تصادفی در دامنهٔ موج صوتی
نسبت هارمونیک به نویز	نسبت انرژی مؤلفه‌های هارمونیک به انرژی نویز موجود در سیگنال صوتی.
نسبت دامنهٔ هارمونیک‌های اول و دوم	نسبت دامنهٔ اولین هارمونیک (H1) به دومین هارمونیک (H2)
فرکانس پایه	پایین‌ترین فرکانس سیگنال صوتی که به‌عنوان فرکانس پایهٔ صدا شناخته می‌شود و با ادراک زیروبمی (Pitch) مرتبط است.
برجستگی قلهٔ طیفی	یک معیار برای ارزیابی برجستگی قله طیفی، که کیفیت صوت و وجود هارمونیک‌های قوی را اندازه‌گیری می‌کند.

۳-۳- نحوه استخراج پارامترهای آکوستیکی از نرم‌افزار پرات

برای استخراج پارامترهای آکوستیکی کیفیت صدا، ابتدا فایل صوتی مورد نظر در نرم‌افزار پرات (بورسما و وینینگ، ۲۰۲۴)، بارگذاری شد. سپس، به‌منظور تجزیه‌وتحلیل دقیق‌تر سیگنال صوتی، از افزونهٔ Praat Vocal Toolkit (کارتج، ۲۰۲۲) استفاده شد. این افزونه امکان استخراج خودکار سیگنال‌های صوتی واکدار و بی‌واک را فراهم

می‌کند. در ابتدا، با بهره‌گیری از دستور cut pauses کلیهٔ وقفه‌ها و سکوت‌های موجود در سیگنال حذف شدند. سپس، سیگنال‌های صوتی واک‌دار و بی‌واک به همراه شبکهٔ متنی متناظر آن‌ها با استفاده از دستور voiced and unvoiced extract استخراج گردید. شکل ۱ نمونه‌ای از این تحلیل و نحوه استخراج مناطق واک‌دار و بی‌واک از سیگنال‌های صوتی را نشان می‌دهد. در ادامه، با اجرای یک اسکریپت از پیش تعریف‌شده، تمامی سیگنال‌های صوتی واک‌دار به‌صورت زنجیره‌ای پیوسته به یکدیگر متصل شدند. درنهایت، محاسبات مورد نظر با گام زمانی ۵ میلی‌ثانیه بر روی این زنجیرهٔ سیگنال انجام شد. برای محاسبهٔ فرکانس پایه، از ابزار "Pitch" استفاده شد و تنظیمات مربوط به بازهٔ فرکانسی بهینه اعمال گردید. مقادیر فرکانس‌پریشی و دامنه‌پریشی با تبدیل فایل صوتی به یک سری رویداد زمانی (PointProcess)و سپس محاسبهٔ این پارامترها از طریق ابزارهای مربوطه استخراج شدند. برای اندازه‌گیری نسبت هارمونیک به نویز، سیگنال صوتی به یک شیء هارمونیکی (Harmonicity) تبدیل شد و میانگین این نسبت از طریق ابزار"Get mean" محاسبه گردید. جهت محاسبه نسبت دامنهٔ هارمونیک‌های اول و دوم، طیف فرکانسی صوت با ابزار"Spectrum" ایجاد شد و سپس، دامنهٔ هارمونیک‌های اول و دوم در طیف فرکانسی اندازه‌گیری و نسبت آن‌ها محاسبه گردید. برای توصیف شکل طیف فرکانسی به‌صورت دقیق‌تر، از ضرایب طیفی استفاده شد. با انجام پردازش‌های تکمیلی با استفاده از اسکریپت‌های تخصصی، برجستگی قلهٔ طیفی محاسبه شد. تمامی این مراحل با استفاده از اسکریپت‌های سفارشی‌شدهٔ مناسب محیط پرات انجام شد. درنهایت، تمام داده‌های استخراج‌شده در قالب یک فایل متنی ذخیره شدند. این فایل حاوی مقادیر عددی پارامترهای آکوستیکی برای هر یک از سیگنال‌های صوتی است. داده‌های حاصل، برای انجام تحلیل‌های آماری و مقایسهٔ گروه‌های مختلف در مراحل بعد مورد استفاده قرار گرفتند.



شکل ۱ : نمونه‌ای از نحوهٔ جداسازی مناطق واک‌دار و بی‌واک سیگنال‌های آوایی

۴-۳- تحلیل آماری

در این مطالعه، برای تحلیل ویژگی‌های آکوستیکی صدا از روش‌های آماری متنوعی استفاده شد. به‌منظور انجام تحلیل‌های آماری از نرم‌افزار R نسخهٔ 4.4.2 (R Core Team, 2024) استفاده شد. برای بررسی اینکه آیا پارامترهای کیفیت صدا تفاوت معناداری بین زنان و مردان وجود دارد، از آزمون t مستقل[1] استفاده گردید که امکان مقایسهٔ میانگین‌های دو گروه مستقل را فراهم می‌کند. سطح معناداری $p < 0.05$ به‌عنوان معیار تفاوت معنادار در نظر گرفته شد. همچنین به‌منظور بررسی تأثیر گوینده بر پارامترهای کیفیت صدا، از تحلیل واریانس چند متغیره[2] استفاده شد. در این تحلیل، پارامترهای آکوستیکی کیفیت صدا به‌عنوان متغیرهای وابسته در نظر گرفته شدند. گوینده نیز به‌عنوان متغیر مستقل وارد مدل شد تا بتوان تفاوت‌های بین گویندگان مختلف را در این پارامترها بررسی کرد. این آزمون به ما امکان می‌دهد تا به‌طور هم‌زمان تأثیر گوینده بر تمامی این متغیرهای وابسته را ارزیابی کنیم و از وجود تفاوت‌های معنادار بین گروه‌های گویندگان اطمینان حاصل نماییم. تحلیل واریانس چند متغیره به دلیل توانایی در کنترل خطای نوع اول در مقایسه با انجام آزمون‌های جداگانه بر روی هر متغیر وابسته، برای تحلیل داده‌های چندمتغیره مناسب است. پیش از انجام آزمون‌ها، نرمال‌بودن توزیع داده‌ها با استفاده از آزمون شاپیرو-ویلک[3] و همگنی واریانس‌ها با استفاده از آزمون لون[4] بررسی شد. نتایج نشان داد که داده‌ها در هر دو گروه دارای توزیع نرمال هستند ($p > ۰.۰۵$) و واریانس‌ها نیز همگن هستند ($p > ۰.۰۵$). همچنین، برای ارزیابی میزان فردویژگی پارامترهای صوتی از الگوریتم جنگل تصادفی[5] استفاده شد. الگوریتم جنگل تصادفی یکی از روش‌های یادگیری ماشین[6] است که از ترکیب چندین درخت تصمیم[7] تشکیل می‌شود. در این مطالعه، از این الگوریتم برای تحلیل اهمیت ویژگی‌های آکوستیکی در تمایز بین گویندگان استفاده شد. جنگل تصادفی با ایجاد مجموعه‌ای از درختان تصمیم که هرکدام بر روی زیرمجموعه‌ای تصادفی از داده‌ها و ویژگی‌ها آموزش می‌بینند، عمل می‌کند. این روش قابلیت محاسبهٔ اهمیت نسبی هر ویژگی را با استفاده از شاخص جینی[8] فراهم می‌کند. شاخص جینی میزان کاهش ناخالصی در گره‌های درخت را که توسط هر ویژگی ایجاد می‌شود، اندازه‌گیری می‌کند. هر چه یک ویژگی بیشتر به کاهش ناخالصی کمک

1. independent t-test
2. MANOVA
3. Shapiro-Wilk test
4. Levene's test
5. random forest
6. machine learning
7. decision tree
8. Gini index

کند، اهمیت بیشتری در تمایز بین کلاس‌ها (در اینجا گویندگان) دارد. این روش به‌طور خاص برای مطالعات فردویژگی مناسب است، زیرا علاوه بر طبقه‌بندی، می‌تواند سهم هر ویژگی آکوستیکی را در تمایز بین گویندگان به‌صورت کمی مشخص کند.

۴- گزارش نتایج

در این بخش، نتایج حاصل از تحلیل‌های آماری به‌تفصیل ارائه می‌شود. در ابتدا، تفاوت‌های جنسیتی میان پارامترهای کیفیت صدا در گروه‌های زنان و مردان بررسی خواهد شد. سپس، به تحلیل واریانس چندمتغیره پرداخته و تفاوت‌های فردی در هر یک از این دو گروه جنسیتی موردبررسی قرار می‌گیرد. درنهایت، قدرت تمایز هر یک از پارامترهای صوتی در تشخیص تفاوت‌های فردی و رتبه‌بندی آن‌ها ارائه خواهد شد.

۴-۱- تحلیل آماری مقایسه‌ای پارامترهای کیفیت صدا در دو گروه زنان و مردان

در این بخش آمار توصیفی مربوط به میانگین و انحراف معیار پارامترهای کیفیت صدا در هر دو گروه مردان و زنان ارائه شده است. این مقادیر در جدول ۲ گزارش شده است. معناداری تفاوت این پارامترها با استفاده از آزمون t مستقل در جدول ۳ آمده است.

جدول ۲: میانگین و انحراف معیار پارامترهای آکوستیکی کیفیت صدا در دو گروه زنان و مردان. (عدد اول میانگین و عدد درون پرانتز نشان‌دهندهٔ انحراف معیار است).

پارامتر	زنان	مردان
فرکانس‌پریشی	۰/۱۸۹ (۰.۰۴۳)	۰.۲۷۱ (۰.۹۵)
دامنه‌پریشی	۰.۹۸۷ (۰.۲۹۱)	۱.۲۰۳ (۰.۳۱۲)
نسبت هارمونیک به نویز	۱۵.۱۷۱ (۲.۸۹۱)	۱۱.۶۵۴ (۱.۶۱۲)
نسبت دامنه هارمونیک‌های اول و دوم	۴.۹۱۶ (۳.۶۵۷)	۰.۶۰۱ (۳.۲۴۶)
فرکانس پایه	۲۰۵.۶۹ (۲۳.۱۵۷)	۱۱۷.۲۱ (۰.۰۳۸۴)
برجستگی قلهٔ طیفی	۷۴.۴۸ (۶.۷۵)	۷۴.۸۵ (۵.۲۳)

طبق داده‌های جدول ۱، گویندگان زن به‌طور متوسط فرکانس پایه، نسبت هارمونیک به نویز و نسبت دامنهٔ هارمونیک‌های اول و دوم بالاتری نسبت به مردان دارند. بااین‌حال، برجستگی قلهٔ طیفی در هر دو گروه جنسیتی مشابه بوده و تفاوت قابل‌توجهی مشاهده نشده است. همچنین، نتایج نشان می‌دهد که فرکانس‌پریشی در صدای مردان نسبت به زنان بیشتر است؛ درحالی‌که، دامنه‌پریشی در صدای مردان اندکی بیشتر از زنان است.

جدول ۳: مقایسهٔ میانگین‌های پارامترهای کیفیت صدا بین گروه‌های زنان و مردان با استفاده از آزمون t مستقل

پارامتر	t-value	p-value	Effect Size (Cohen's d)
فرکانس‌پریشی	۳٫۵۹۴	<۰٫۰۰۱	۱٫۱۳۷
دامنه‌پریشی	۲٫۳۲۴	<۰٫۰۰۱	۰٫۷۳۵
نسبت هارمونیک به نویز	-۴٫۸۹۵	<۰٫۰۰۱	−۱٫۵۴۸
نسبت دامنهٔ هارمونیک‌های اول و دوم	−۴٫۰۸۶	<۰٫۰۰۱	−۱٫۲۹۲
فرکانس پایه	−۱۷.۰۰۶	<۰.۰۰۱	−۵.۳۷۸
برجستگی قلهٔ طیفی	۰٫۱۹۲	۰٫۸۴۹	۰٫۰۶۱

نتایج جدول ۲ نشان می‌دهد که برخی از ویژگی‌های آکوستیکی، مانند فرکانس‌پریشی، دامنه‌پریشی، نسبت دامنهٔ هارمونیک‌های اول و دوم، نسبت هارمونیک به نویز و فرکانس پایه، تفاوت آماری معناداری میان گروه زنان و مردان دارند و در شناسایی تفاوت‌ها مؤثر هستند. در این میان، فرکانس پایه بیشترین تفاوت را نشان داده است. از سوی دیگر، ویژگی‌هایی مانند برجستگی قلهٔ طیفی تفاوت معناداری نشان نداده‌اند و تأثیر کمتری در تحلیل داشته‌اند.

۲-۴- نتایج مربوط به تغییرپذیری پارامترهای کیفیت صدا میان گویندگان

برای بررسی معناداری تفاوت‌های بین گروه‌های جنسیتی در پارامترهای صوتی، از آزمون تحلیل واریانس چندمتغیره استفاده شد. پیش از انجام آزمون، نرمال‌بودن توزیع داده‌ها با آزمون شاپیرو–ویلک و همگنی واریانس‌ها با آزمون لوین تأیید گردید.

جدول ۴: نتایج آزمون تحلیل واریانس چندمتغیره در نشان دادن تفاوت‌های آکوستیکی میان گویندگان در گروه مردان

پارامتر	F-value	p-value	Effect Size
فرکانس‌پریشی	۱۸٫۴۵	<۰٫۰۰۱	۰٫۷۷
دامنه‌پریشی	۱۵٫۹۲	<۰٫۰۰۱	۰٫۷۳
نسبت هارمونیک به نویز	۱۹٫۸۳	<۰٫۰۰۱	۰٫۷۸
نسبت دامنهٔ هارمونیک‌های اول و دوم	۲۱٫۷۶	<۰٫۰۰۱	۰٫۸۰
فرکانس پایه	۱۲٫۵۴	<۰٫۰۰۱	۰٫۶۸
برجستگی قلهٔ طیفی	۲۴٫۳۱	<۰٫۰۰۱	۰٫۸۲

جدول ۵: نتایج آزمون تحلیل واریانس چندمتغیره در نشان دادن تفاوت‌های آکوستیکی میان گویندگان در گروه زنان

پارامتر	F-value	p-value	Effect Size
فرکانس‌پریشی	۱۶٫۹۲	<۰٫۰۰۱	۰٫۷۵
دامنه‌پریشی	۱۹٫۳۷	<۰٫۰۰۱	۰٫۷۸
نسبت هارمونیک به نویز	۲۵٫۶۶	<۰٫۰۰۱	۰٫۸۳
نسبت دامنهٔ هارمونیک‌های اول و دوم	۲۳٫۵۱	<۰٫۰۰۱	۰٫۸۱
فرکانس پایه	۳۱٫۲۹	<۰٫۰۰۱	۰٫۸۶
برجستگی قلهٔ طیفی	۲۷٫۸۴	<۰٫۰۰۱	۰٫۸۴

نتایج جدول‌های ۴ و ۵ نشان داد که تفاوت معناداری میان گویندگان در هر دو گروه زنان و مردان در همهٔ پارامترهای آکوستیکی موردبررسی وجود دارد. بااین‌حال اندازهٔ اثر این تفاوت‌ها با هم متفاوت است. اندازهٔ اثر این تفاوت‌ها بر اساس شاخص η^2 در سطح متوسط تا بزرگ ارزیابی شده است.

۳-۴- میزان فردویژگی پارامترهای کیفیت صدا

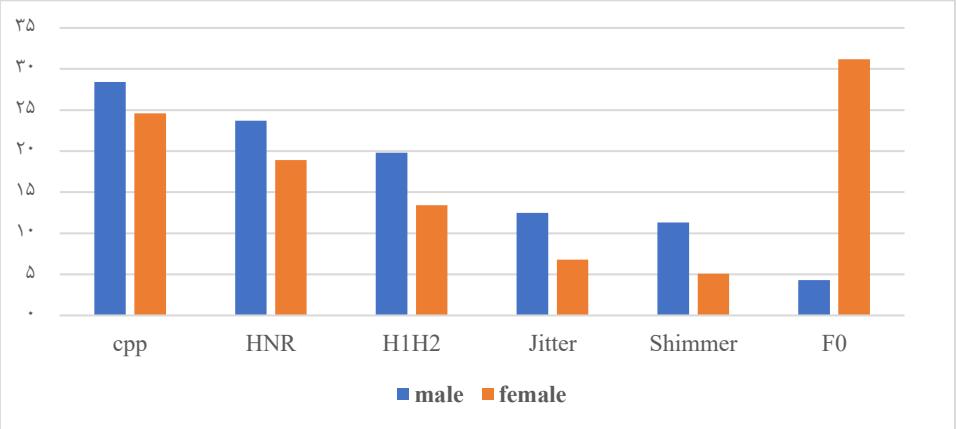
برای سنجش اهمیت ویژگی‌های کیفیت صدا، از الگوریتم جنگل تصادفی بهره بردیم. شاخص جینی که پیش‌تر توضیح داده شد، میزان خلوص هر گره در درخت تصمیم را نشان می‌دهد. هر چه ویژگی‌ای بتواند گره‌ها را خالص‌تر کند، اهمیت بیشتری در تشخیص گویندگان مختلف دارد. نتایج این آزمون در جدول ۶ آورده شده است. اهمیت ویژگی‌ها بر حسب درصد بیان شده است.

جدول ۶: اهمیت پارامترهای آکوستیکی کیفیت صدا برحسب درصد

پارامتر	زنان	مردان
فرکانس‌پریشی	۶.۸	۱۲.۵
دامنه‌پریشی	۵.۱	۱۱.۳
نسبت هارمونیک به نویز	۱۸.۹	۲۳.۷
نسبت دامنهٔ هارمونیک‌های اول و دوم	۱۳.۴	۱۹.۸
فرکانس پایه	۳۱.۲	۴.۳
برجستگی قلهٔ طیفی	۲۴.۶	۲۸.۴

بر اساس نتایج، پارامترهایی که در گروه مردان بیشترین اهمیت را دارند، برجستگی قلهٔ طیفی، نسبت هارمونیک به

نویز و نسبت دامنهٔ هارمونیک‌های اول و دوم هستند. در گروه زنان، پارامتر فرکانس پایه، برجستگی قلهٔ طیفی و نسبت هارمونیک به نویز به‌ترتیب بیشترین اهمیت را داشتند. پارامترهایی مانند برجستگی قلهٔ طیفی و نسبت هارمونیک به نویز در هر دو گروه اهمیت بالایی دارند، اما سهم آن‌ها در گروه مردان بیشتر است. فرکانس پایه اهمیت بالایی برای گروه زنان دارد، درحالی‌که در گروه مردان اهمیت بسیار پایینی دارد. پارامترهای فرکانس‌پریشی و دامنه‌پریشی در هر دو گروه کم‌اهمیت‌ترند، اما این اهمیت در گروه مردان کمی بیشتر از گروه زنان است. شکل ۲ نمایش دیداری اهمیت پارامترهای کیفیت صدا را در هر دو گروه زنان و مردان به شکل مقایسه‌ای نشان می‌دهد.



شکل ۲: نمودار میله‌ای نشان‌گر اهمیت پارامترهای کیفیت صدا در دو گروه زنان و مردان. اهمیت پارامترها در گروه زنان با میله‌های قرمز و در گروه مردان با میله‌های آبی نشان داده شده است.

۵. بحث و بررسی

در این پژوهش، پارامترهای آکوستیکی مرتبط با کیفیت صدا بررسی شدند تا نقش آن‌ها در نشان‌دادن شاخص‌های گوینده‌محور در میان گویندگان فارسی‌زبان مشخص شود. هدف پژوهش این بود که ارزیابی شود آیا این پارامترها توانایی تفکیک گویندگان فارسی‌زبان را دارند و تا چه اندازه می‌توانند تمایزهای بین گوینده‌ها را نشان دهند. نتایج در سه سطح تحلیل شدند: تفاوت‌های جنسیتی، ویژگی‌های گوینده‌محور، و اهمیت پارامترهای آکوستیکی در برجسته‌کردن تفاوت‌های بین گویندگان.

نتایج تحلیل نشان داد که بین پارامترهای آکوستیکی صدای زنان و مردان تفاوت‌های آماری معناداری وجود دارد.

بزرگ‌ترین تفاوت مشاهده‌شده در فرکانس پایه با اندازهٔ اثر ۵٫۳۷۸ بود که نشان‌دهندهٔ تفاوت بسیار بزرگ بین دو گروه است. این یافته البته قابل پیش‌بینی بود زیرا آناتومی دستگاه گفتار زنان و مردان به‌ویژه طول و اندازهٔ حنجره و تارهای صوتی در زنان و مردان متفاوت است. همچنین، تفاوت معناداری در نسبت هارمونیک به نویز با اندازهٔ اثر ۱٫۵۴۸ مشاهده شد که نشان‌دهندهٔ هارمونیک‌های منظم‌تر صدای زنان است. این تفاوت‌ها را می‌توان به تفاوت‌های آناتومیکی دستگاه صوتی بین زنان و مردان، به‌ویژه تفاوت در اندازه و کشش تارهای صوتی، نسبت داد.

در رابطه با تفاوت‌های آکوستیکی میان گویندگان، نتایج نشان داد که پارامترهای موردبررسی در هر دو گروه زنان و مردان تفاوت معناداری میان گویندگان داشتند. این بدان معناست که این پارامترها پتانسیل قابل‌توجهی برای به‌کارگیری در پروژه‌های شناسایی گوینده به‌ویژه سیستم‌های شناسایی گویندهٔ مربوط به گفتار فارسی دارند. این یافته با پژوهش‌های پیشین همسو است که نشان داده‌اند پارامترهای کیفیت صدا نقش بسیار مهمی در تشخیص هویت گوینده ایفا می‌کنند (ووبی و همکاران، ۲۰۲۱؛ لی و کرایمن، ۲۰۲۲a؛ هیوز و همکاران، ۲۰۲۳). ازآنجایی‌که این پارامترها به فیزیولوژی دستگاه گفتار مرتبط هستند و افراد دستگاه گفتار منحصربه‌فردی دارند که ساختار آن از فردی به فرد دیگر متفاوت است، این پارامترها توانایی بالایی در تفکیک گویندگان نشان می‌دهند. البته نباید از تأثیر ویژگی‌های اکتسابی و فرهنگی بر پارامترهای کیفیت صدا غافل شد. همان‌طور که پژوهش‌ها نشان داده‌اند (اسلینگ، ۲۰۰۰؛ کرایمن و سیدتیس، ۲۰۱۱)، برخی از پارامترهای کیفیت صدا به عوامل اجتماعی و فرهنگی نیز وابسته‌اند. این نکته اهمیت زیادی دارد، زیرا نشان می‌دهد که پارامترهای کیفیت صدا نه‌تنها میان افراد مختلف بلکه بین زبان‌ها و فرهنگ‌های مختلف نیز تفاوت دارند.

اگرچه نتایج نشان داد که تفاوت‌های میان گویندگان فارسی‌زبان در پارامترهای کیفیت صدا معنادار بوده است اما با این حال میزان فردویژگی شاخص‌های گوینده‌محور که این پارامترها رمزگذاری می‌کنند با هم یکسان نیست. بدان معنا که برخی پارامترها نسبت به برخی دیگر از اهمیت بالاتری برخوردارند. بر اساس نتایج به‌دست‌آمده از این پژوهش اهمیت پارامترها در دو گروه زنان و مردان متفاوت بوده است. این یافته نشان می‌دهد که انتخاب پارامترهای مناسب برای شناسایی گوینده، علاوه بر سایر عوامل، به جنسیت نیز وابسته است. تحلیل داده‌ها نشان داد که پارامترهای برجستگی قلهٔ طیفی، نسبت هارمونیک به نویز و نسبت دامنهٔ هارمونیک‌های اول و دوم به ترتیب بیشترین توانایی را در تفکیک گویندگان مرد داشته‌اند. در مقابل، پارامترهای فرکانس پایه، برجستگی قلهٔ طیفی و نسبت هارمونیک به نویز به‌ترتیب مهم‌ترین پارامترها برای تفکیک گویندگان زن بودند. این یافته نشان می‌دهد که دو پارامتر

برجستگی قلهٔ طیفی و نسبت هارمونیک به نویز فارغ از جنسیت گویندگان می‌توانند به‌عنوان شاخص‌های کلی گوینده‌محور در شناسایی گوینده به‌کار گرفته شوند. نسبت هارمونیک به نویز میزان تناوبی‌بودن سیگنال صدا را نشان می‌دهد، به‌طوری‌که مقادیر بالاتر نشان‌دهندهٔ صدایی تمیزتر و تناوبی‌تر است که می‌تواند برای تشخیص گویندگان مفید باشد. این ویژگی تغییرات ظریف در رنگ صدا[1] را که خاص افراد است، به همراه تفاوت‌های موجود در نحوهٔ تولید گفتار افراد، ثبت می‌کند. به‌طور مشابه، برجستگی قلهٔ طیفی، با کیفیت سیگنال صدا مرتبط است و به‌طور خاص بر برجستگی قله‌های طیفی در طیف صدا تمرکز دارد. مقدار بالاتر برجستگی قلهٔ طیفی عموماً نشان‌دهندهٔ صدایی متمایزتر و واضح‌تر است که می‌تواند در شناسایی گوینده ارزشمند باشد. هر دوی این ویژگی‌ها نسبت به ویژگی‌های منحصربه‌فرد دستگاه گفتار فرد و نحوهٔ تولید صدا توسط حنجره حساس هستند و آن‌ها را به ابزارهای قدرتمندی برای تمایز بین گویندگان، صرف‌نظر از جنسیت، تبدیل می‌کنند.

نتایج حاصل از این پژوهش، نکته جالبی را آشکار ساخت. نکتهٔ جالب‌توجه، کم‌اهمیت بودن فرکانس پایه در گروه مردان است. باوجوداینکه فرکانس پایه در گروه زنان از اهمیت بالایی برخوردار بوده، در گروه مردان کمتر توانسته است گویندگان را از هم متمایز کند. این نتیجه می‌تواند نشان دهد که درحالی‌که فرکانس پایه یک ویژگی قوی برای شناسایی گویندگان زن است، ممکن است در صدای مردان به اندازه کافی متغیر یا منحصربه‌فرد نباشد. همچنین این یافته نشان می‌دهد که زنان و مردان در به‌کارگیری پارامترهای اکوستیکی متمایزکنندهٔ هویت فرد به روش‌های متفاوتی عمل می‌کنند. اگرچه فرکانس پایه در گروه مردان نیز تفاوت معناداری نشان داده است، بااین‌حال گروه مردان از این پارامتر برای رمزگذاری شاخص‌های گوینده‌محور نسبت به سایر پارامترهای کیفیت صدا استفاده کمتری داشته‌اند. یافته‌ها حاکی از آن است که نسبت دامنهٔ هارمونیک اول به دوم نیز در شناسایی گویندگان، چه زن و چه مرد، بسیار مؤثر است. این در حالی است که پارامترهایی مانند فرکانس‌پریشی و دامنه‌پریشی کمترین تأثیر را در این زمینه نشان داده‌اند. نسبت دامنهٔ هارمونیک اول به دوم، به‌عنوان معیاری برای سنجش ساختار هارمونیک صدا، می‌تواند ویژگی‌های منحصربه‌فرد هر فرد را آشکار کند. حساسیت این ویژگی به پیکربندی دستگاه صوتی و تنظیمات حنجره (لاور، ۱۹۶۸؛ اسلینگ، ۲۰۰۶)، آن را به ابزاری قدرتمند برای تمایز بین گویندگان مختلف تبدیل کرده است. برعکس، فرکانس‌پریشی و دامنه‌پریشی که به‌ترتیب تغییرات در فرکانس و دامنهٔ سیگنال صوتی را اندازه می‌گیرند، بیشتر با ناپایداری صدا و اختلالات صوتی مرتبط هستند تا ویژگی‌های فردی گویندگان. حساسیت بالای این پارامترها به

1. timbre

نویز، نفس‌دار بودن و سایر ناهنجاری‌های تولید صدا، آن‌ها را برای شناسایی گوینده در شرایط ایده‌آل، مانند یک مجموعه دادهٔ گفتار تمیز، کمتر قابل‌اعتماد می‌سازد.

۶. نتیجه‌گیری

پژوهش حاضر با هدف بررسی تفاوت‌های آکوستیکی در پارامترهای کیفیت صدا بین گویندگان مرد و زن فارسی‌زبان انجام شده است. نتایج نشان داد که برخی ویژگی‌های آکوستیکی مانند فرکانس‌پریشی، دامنه‌پریشی، نسبت دامنهٔ هارمونیک‌های اول و دوم، نسبت هارمونیک به نویز و فرکانس پایه، تفاوت‌های آماری معناداری بین دو گروه جنسیتی دارند. این یافته حاکی از آن است که این پارامترها می‌توانند به‌عنوان سرنخ‌هایی آکوستیکی برای تشخیص جنسیت گوینده در زبان فارسی مورد استفاده قرار گیرند. علاوه بر تفاوت‌های جنسیتی، نتایج نشان داد که پارامترهای کیفیت صدا به‌طور قابل‌توجهی بین افراد یک جنس نیز متفاوت است؛ به‌عبارت‌دیگر، هر فرد از الگوی منحصربه‌فردی از پارامترهای آکوستیکی برخوردار است که می‌تواند برای شناسایی هویت وی مورد استفاده قرار گیرد. تحلیل‌ها نشان داد که برای گویندگان مرد، پارامترهای برجستگی قلهٔ طیفی، نسبت هارمونیک به نویز و نسبت دامنهٔ هارمونیک‌های اول و دوم به ترتیب بیشترین توانایی را در تفکیک آن‌ها از یکدیگر دارند. در مقابل، برای گویندگان زن، فرکانس پایه، برجستگی قلهٔ طیفی و نسبت هارمونیک به نویز به‌عنوان مهم‌ترین شاخص‌ها برای تشخیص هویت شناخته شدند. نتایج این پژوهش نشان می‌دهد که پارامترهای کیفیت صدا نقش مهمی در شناسایی گویندگان فارسی‌زبان دارند. بااین‌حال، برای دستیابی به دقت بالاتر در سیستم‌های شناسایی گوینده، لازم است به تفاوت‌های جنسیتی و همچنین به اهمیت متفاوت پارامترهای مختلف در هر گروه جنسیتی نیز توجه شود. این یافته‌ها می‌توانند در طراحی و توسعهٔ سیستم‌های تشخیص هویت مبتنی بر صدا، به‌ویژه برای زبان فارسی، کاربرد فراوانی داشته باشند. اگرچه نتایج این پژوهش گام مهمی در جهت شناسایی گویندگان فارسی‌زبان برداشته است، اما دارای محدودیت‌هایی نیز هست. یکی از مهم‌ترین این محدودیت‌ها، تعداد محدود گویندگان شرکت‌کننده در پژوهش است. اگرچه این تعداد برای انجام تحلیل‌های آماری و رسیدن به نتیجه‌ای مطلوب کافی است اما برای دستیابی به نتایج دقیق‌تر و جامع‌تر، پیشنهاد می‌شود در مطالعات آتی، پارامترهای آکوستیکی کیفیت صدا بر روی یک پیکرهٔ بزرگ‌تر صوتی و با استفاده از مدل‌های پیشرفته‌تر یادگیری ماشین مانند شبکه‌های عصبی عمیق تحلیل شوند.

منابع

اسدی، ه.، حسینی کیونانی، ن.، و نوربخش، م. (۱۳۹۴). بررسیِ تأثیر پوششِ صورت بر ویژگی‌های آکوستیکی سایشی‌های بی‌واک زبان فارسی: پژوهشی در چارچوب آواشناسی قضایی .زبانشناسی و گویش‌های خراسان. (۷) ۱۳، ۱۳۵–۱۴۸. https://doi.org/10.22067/lj.v7i13.56117

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Asadi, H., Hosseini-Kivanani, N., & Nourbakhsh, M. (2015). Effects of face covering on acoustic properties of voiceless fricatives in Farsi: A forensic approach. *Journal of Linguistics & Khorasan Dialects*, *7*(13), 135-148. [In Persian] https://doi.org/10.22067/lj.v7i13.56117

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

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer*. http://www.praat.org

Chai, Y., & Garellek, M. (2022). On H1–H2 as an acoustic measure of linguistic phonation type. *The Journal of the Acoustical Society of America*, *152*(3), 1856-1870. https://doi.org/10.1121/10.0014175

Corretge, R. (2022). *Praat vocal toolkit*. http://www.praatvocaltoolkit.com

Esling, J. H. (2000). Crosslinguistic aspects of voice quality. In R. D. Kent & M. J. Ball (Eds.), *Voice quality measurement* (pp. 25–35). Singular Publishing Group.

Esling, J. H. (2006). Voice quality. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (pp. 470–474). Elsevier. https://doi.org/10.1016/b0-08-044854-2/00032-8

Esling, J. H. (2013). Voice quality. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 6144–6150). Blackwell.

Fernandes, J. F., Freitas, D., Júnior, A.C., & Teixeira, J. P. (2023). Determination of harmonic parameters in pathological voices—Efficient algorithm. *Applied Sciences*, *13*(4), 2333. https://doi.org/10.3390/app13042333

Hughes, V., Cardoso, A., Foulkes, P., French, P., Gully, A., & Harrison, P. (2023). Speaker-specificity in speech production: The contribution of source and filter. *Journal of Phonetics*, *97*, 101224. https://doi.org/10.1016/j.wocn.2023.101224

Jessen, M., Koster, O., & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *The International Journal of Speech, Language and the Law*, *12*(2), 174-213. https://doi.org/10.1558/sll.2005.12.2.174

Kinoshita, K. (2001). Testing realistic forensic speaker identification in Japanese: A likelihood ratio based approach using formants (Unpublished doctoral dissertation). Australian National University.

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies*. Wiley-Blackwell.

Labutin, P., Koval, S., & Raev, A. (2007). Speaker identification based on the statistical analysis of f0. In *Proceedings of IAFPA*.

Ladefoged, P., & Johnson, K. (2015). *A course in phonetics* (7th ed.). Cengage Learning.

Laver, J. (1968). Voice quality and indexical information. *International Journal of Language & Communication Disorders*, *3*(1), 43–54. https://doi.org/10.3109/13682826809011440

Lee, Y., & Kreiman, J. (2022a). Acoustic voice variation in spontaneous speech. *The Journal of the Acoustical Society of America*, *151*(5), 3462. https://doi.org/10.1121/10.0011471

Lee, Y., & Kreiman, J. (2022b). Linguistic versus biological factors governing acoustic voice variation. *Interspeech Proceedings*. http://doi.org/10.21437/Interspeech.2022-10847

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. https://doi.org/10.1121/1.5125134

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

Murton, O., Hillman, R., & Mehta, D. (2020). Cepstral peak prominence values for clinical voice evaluation. *American Journal of Speech-Language Pathology*, *29*(3),1596–1607. https://doi.org/10.1044/2020_AJSLP-20-00001

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

Park, S. J., Sigouin, C., Kreiman, J., Keating, P. A., Guo, J., Yeung, G., & Alwan, A. (2016). Speaker identity and voice quality: Modeling human responses and automatic speaker recognition. In *Interspeech* (pp. 1044-1048). http://doi.org/10.21437/Interspeech.2016-523

Park, S. J., Yeung, G., Kreiman, J., Keating, P. A., & Alwan, A. (2017). Using voice quality features to improve short-utterance, text-independent speaker verification systems. In *Interspeech* (pp. 1522-1526). http:/doi.org/10.21437/Interspeech.2017-157

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Rose, P. (2002). Forensic speaker identification. cRc Press.

Shama, K., Krishna, A., & Cholayya, N. U. (2006). Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. *EURASIP Journal on Advances in Signal Processing*, 1-9. https://doi.org/10.1155/2007/85286

Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis–jitter, shimmer and hnr parameters. *Procedia Technology*, *9*, 1112-1122. https://doi.org/10.1016/j.protcy.2013.12.124

Wagner, A., & Braun, A. (2003). Is voice quality language-dependent? Acoustic analyses based on speakers of three different languages. *In Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)* (pp. 651-654). Barcelona, Spain.

Woubie, A., Koivisto, L., & Bäckström, T. (2021). Voice-quality features for deep neural network based speaker verification systems. In 29th *European Signal Processing Conference (EUSIPCO)* (pp.176-180). IEEE. https://doi.org/10.23919/EUSIPCO54536.2021.9616242