ارسال: ۱۴۰۴/۱۲/۰۳
پذیرش: ۱۴۰۵/۰۱/۳۰
10.22034/nf.2026.577011.1518

طراحی یک پیکرهٔ گفتاری در مطالعات آواشناسی تجربی: تصمیم‌ها و چالش‌های روش‌شناختی

ندا موسوی * (دپارتمان علم گفتار و آواشناسی، دانشگاه مارتین لوتر هاله–ویتنبرگ و مؤسّسهٔ زیبایی‌شناسی تجربی ماکس پلانک، فرانکفورت)

چکیده: هدف این مقاله ارائهٔ مروری روش‌شناختی بر مراحل طراحی و ساخت یک پیکرهٔ گفتاری با تمرکز بر تصمیم‌ها و چالش‌های فرایند گردآوری داده‌های گفتاری استاندارد در مطالعات آواشناسی تجربی است. این مرور بر پایهٔ تجربه‌های حاصل از طراحی و پیاده‌سازی یک پیکرهٔ گفتاری در چارچوب رسالهٔ دکتری نویسنده انجام شده است. پیکرهٔ اصلی ماهیتی دوزبانه دارد، امّا تمرکز این مقاله صرفاً بر بخش فارسی آن است. در این نوشتار، مراحل مختلف ساخت پیکره، ازجمله انتخاب و طراحی وظایف برانگیزش گفتار، نمونه‌گیری از گویندگان، طراحی محیط و شرایط ضبط، سازمان‌دهی و نام‌گذاری داده‌ها و، درنهایت، پیش‌پردازش و تقطیع گفتار، به‌صورت نظام‌مند مورد بحث قرار می‌گیرند. همچنین، نشان داده می‌شود که طراحی پیکره فرایندی صرفاً اجرایی نیست، بلکه مجموعه‌ای از تصمیم‌های روش‌شناختی آگاهانه را در بر می‌گیرد که مستقیماً بر کیفیت داده‌ها، تفسیرپذیری نتایج و امکان بازتولید تحلیل‌ها اثر می‌گذارند. همچنین، مقاله به چالش‌های مرتبط با استفاده از روش‌های خودکار در تقطیع گفتار پیکره‌های فارسی می‌پردازد و محدودیت‌های ابزارهای بازشناسی خودکار گفتار را در بافت زبان‌های کم‌منبع برجسته می‌کند. در پایان، استدلال می‌شود که مستندسازیِ دقیق این تصمیم‌ها و مراحل خود بخشی اساسی از فرایند ساخت پیکره است و نقشی تعیین‌کننده در شفافیت روش، ارزیابی نتایج و امکان استفادهٔ علمی از داده‌ها ایفا می‌کند.

کلیدواژه‌ها: پیکرهٔ گفتاری، برانگیزش گفتار، تقطیع گفتار، بازشناسی خودکار گفتار، زبان‌های کم‌منبع.

---

* ndlmousavi@gmail.com

۱. مقدّمه

امروزه، با پیشرفت تکنولوژی‌های پردازش گفتار و دسترسی آسان‌تر به مدل‌های تحلیلی از طریق ابزارهای مبتنی بر کاربرد هوش مصنوعی، سرعت مطالعات در حوزه‌های مختلف پردازش گفتار به‌طور چشمگیری افزایش یافته است. فناوری‌های موجود امکان تهیه، ذخیره‌سازی و پردازش حجم بالایی از دادگان را فراهم می‌کنند و روش‌های پردازش خودکار، دسترسی به اطّلاعات جزئی در پیکره‌های گفتاری را آسان‌تر کرده‌اند. با این حال، آنچه همچنان نیازمند توجّه جدی است حساسیت در فرایند تهیهٔ داده‌های ورودی برای چنین مدل‌هایی است. بدیهی است که داده‌های غیراستاندارد، هرچند قابل‌تحلیل باشند، نتایجی به‌همراه خواهند داشت که از اعتبار کافی برخوردار نیستند.

با چنین مقدّمه‌ای، این نوشتار تلاش می‌کند به نکته‌ها و جزئیاتی بپردازد که یک آواشناس تجربی در فرایند گردآوری داده‌های پژوهشی باید به آنها توجّه داشته باشد. محتوای این مقاله بر مستندسازی یک پیکرهٔ گفتاری استوار است که نویسنده برای رسالهٔ دکتری خود با موضوع «ریتم فردی: رویکردی آکوستیکی-شناختی به تولید ریتم در وظایف گفتاری مختلف» در رشتهٔ «علم گفتار»[1] در دانشگاه مارتین لوتر[2] کشور آلمان تهیه کرده است (Mousavi 2025). پیکرهٔ اصلی با نام «پِرچ»[3] ماهیتی دوزبانه دارد و شامل داده‌هایی به زبان‌های آلمانی و فارسی است. در این مقاله، تنها بخش فارسی این پیکره مورد توجّه و بررسی قرار می‌گیرد.

با توجّه به دانش نویسنده، تعداد پیکره‌های گفتاری زبان فارسی که به‌صورت نظام‌مند طراحی و مستندسازی شده باشند همچنان محدود است. ازجمله منابع شناخته‌شده می‌توان به «پیکرهٔ گفتار محاوره‌ای زبان فارسی امروز» (بی‌جن‌خان ۱۳۹۵)، فارس‌دات (Bijankhan et al. 1994)، دادگان گفتار احساسی (Nezami et al. 2019)، یا دادگان گفتار لهجه‌دار سهند (Pilehvar and Sedaaghi 2008) اشاره کرد. وبگاه «پیکره‌گان»[4] مجموعه‌ای از پیکره‌های موجود و دردسترس را معرفی می‌کند. با این حال، تنوع پرسش‌های پژوهشی و طرح چالش‌های نو در مطالعات تجربی همواره نیاز به گردآوری داده‌های تازه را مطرح می‌کند. از این رو، طراحی و گردآوری پیکره‌های گفتاری با تمرکز بر

---

1) Speech Science
2) Martin Luther University Halle-Wittenberg, Germany
3) Pertsch: Persian-Deutsch Speech Corpus
4) https://peykaregan.ir/

نیازهای آواشناسیِ تجربی همچنان ضروری است و این فرایند نیازمند دقّت روش‌شناختی در تمامیِ مراحل گردآوری و آماده‌سازی داده‌ها است.

۲. گفتار طبیعی در برابر گفتار آزمایشگاهی

در پژوهش‌های گفتاری، دسترسی به داده‌هایی که بازتاب‌دهندهٔ کاربرد واقعی زبان باشند همواره با چالش‌های روش‌شناختی همراه بوده است. آنچه در بسیاری از مطالعات از آن با عنوان «گفتار طبیعی» یاد می‌شود به داده‌هایی اطلاق می‌شود که در بافت‌های ارتباطی باز و بدون دستورالعمل‌های صریح و ازپیش‌تعیین‌شده تولید می‌شوند و، از این‌رو، تنوع، ناپیوستگی‌ها و سازوکارهای تنظیم و تطبیق گفتار را به‌خوبی منعکس می‌کنند. اهمیّت چنین داده‌هایی نخستین‌بار در مطالعات طولی اکتساب زبان کودک برجسته شد، جایی که تحلیل گفتار طبیعی امکان شناسایی الگوهای بنیادین رشد زبانی را فراهم کرد (Brown 1973). در چنین رویکردی، گفتار طبیعی به‌عنوان منبعی ارزشمند در زبان‌شناسی کاربردی و آواشناسی تجربی مورد توجّه قرار گرفته است. با این حال، گردآوری داده‌های گفتاریِ طبیعی در عمل با محدودیت‌های متعددی همراه است. ضبط گفتار در شرایط واقعی اغلب به داده‌هایی با کیفیت صوتی ناهمگون، نویز محیطی بالا و فراوانی پدیده‌هایی چون مکث‌ها، آغازهای ناتمام و ناپیوستگی‌ها منجر می‌شود (Wagner et al. 2015). همچنین، در پژوهش‌هایی که بر پدیده‌های زبانی مشخص تمرکز دارند، داده‌های کاملاً طبیعی ممکن است تنها نمونه‌های محدودی از واحدهای هدف را در اختیار بگذارند و، درنتیجه، حجم قابل‌توجّهی از دادهٔ غیرمرتبط تولید شود (Baker and Hazan 2011; Faitaki and Murphy 2019).

درمقابل، ضبط گفتار در محیط‌های آزمایشگاهی امکان کنترل متغیرها، تکرارپذیری و دستیابی به داده‌هایی با کیفیت صوتی بالا را فراهم می‌کند، امّا اغلب به‌دلیل فاصله گرفتن از بافت‌های ارتباطی روزمره، به‌عنوان «دادهٔ غیرطبیعی» مورد انتقاد قرار گرفته است. با این حال، برخی پژوهشگران این دوگانه‌سازیِ میان گفتار طبیعی و گفتار آزمایشگاهی را ساده‌انگارانه می‌دانند. شو[1] (2010) استدلال می‌کند که محدودیت‌های گفتار آزمایشگاهی نه ذاتی، بلکه نتیجهٔ طراحی نامناسب وظایف گفتاری است. به‌طورمشابه، واگنر و دیگران (2015) تأکید می‌کنند که «طبیعی بودن» معیار دقیقی برای ارزیابی

1) Xu

کیفیت داده‌های گفتاری نیست، چرا که تمام داده‌های گفتاری، صرف‌نظر از بافت تولید، محصول کنش زبانی انسان‌اند. از این منظر، مسئلهٔ اصلی نه تقابل میان گفتار طبیعی و آزمایشگاهی، بلکه طراحی آن‌دسته از وظایف گفتاری است که ضمن حفظ امکان کنترل تحلیلی، بتوانند شرایط ارتباطی معنادار و انگیزه‌های واقعی تولید گفتار را بازسازی کنند.

به دنبال این مقدّمه، این مطالعه به یک دغدغهٔ روش‌شناختیِ اصلی می‌رسد: نقشِ «وظیفه‌های برانگیزش گفتار»[1] در گردآوری داده در پژوهش‌های آواشناسی تجربی. اصطلاح «برانگیزش»[2] ریشه در فعل لاتینی «elicere» دارد و به معنای «فراخواندن/ بیرون کشیدن/ برانگیختن» است (Senft 1995). در پژوهش‌های زبان‌شناسی و گفتار، وظیفه‌های برانگیزش گفتار به سناریوهای طراحی‌شده‌ای گفته می‌شود که شرکت‌کننده را به تولید نوعی دادهٔ زبانی (گفتاری یا نوشتاری) سوق می‌دهند تا خروجی آن برای تحلیل، ثبت و کدگذاری شود (Faitaki and Murphy 2019). مزیت اصلی این وظیفه‌ها در این است که می‌توانند میان دو نیازِ ظاهراً متضاد پل بزنند: از یک‌سو کنترل آزمایشی و مقایسه‌پذیری و، از سوی دیگر، حفظِ بخشی از ویژگی‌های بیانِ طبیعی در تولید زبان.

زنفت (1995) تأکید می‌کند که گردآوری داده اساساً تابع هدف و دامنهٔ پژوهش است؛ بنابراین، پژوهشگر لازم است دربارهٔ نوع گفتاری که می‌خواهد جمع‌آوری کند، ویژگی‌های نمونه (شرکت‌کنندگان)، محیط ضبط، و زمینهٔ زبانی-فرهنگی تصمیم‌های روشن بگیرد. همین هدف‌ها معیار انتخاب روش برانگیزش و شکل‌دادن به نمونه‌ای می‌شوند که تا حد امکان نمایندهٔ جامعهٔ هدف باشد. این انعطاف‌پذیری امکان می‌دهد وظیفه‌های برانگیزش، متناسب با هدف پژوهش، در امتدادی از شرایط کاملاً ساختارمند تا قالب‌های نسبتاً باز طراحی شوند. با این حال، زبان ذاتاً بافت‌مند است و همین بافت‌مندی کار را پیچیده می‌کند؛ به‌خصوص در محیط‌های کنترل‌شده‌ای مثل آزمایشگاه که بسیاری از جزئیاتش محصول طراحی پژوهشگر است. برای مثال، دستورالعمل‌هایی از نوع «متن را محاوره‌ای بخوانید» می‌توانند به‌صورت ذهنی و ناهمسان تفسیر شوند و، درنتیجه، خروجی‌هایی تولید کنند که نه کاملاً خوانده و متنی‌اند و نه کاملاً طبیعی، بلکه در حدفاصل این دو قرار می‌گیرند. این امر نشان

1) Speech Elicitation Tasks
2) elicitation

می‌دهد که تکرارپذیری و میزان نمایندگی گفتارِ واقعی به کیفیت و دقّت طراحی وظیفهٔ گفتاری وابسته است.

از منظر طبقه‌بندی روش‌ها، فایتاکی و مورفی (2019) نشان می‌دهند که وظیفه‌های برانگیزش گفتار را می‌توان بر روی یک پیوستار از نظر میزان ساختارمندی جای داد؛ پیوستاری که از شیوه‌های کم‌ساختار، مانند ضبط گفتار با حداقل هدایت یا کنترل، آغاز می‌شود و تا وظیفه‌های بسیار کنترل‌شده، مانند فعالیت‌های نام‌گذاری یا تصمیم‌گیری واژگانی با محدودیت زمانی، امتداد می‌یابد. در کنار این پیوستار، تمایز مهم دیگری نیز وجود دارد که به وضعیت ازپیش‌تدوین‌شده بودن وظیفه مربوط می‌شود. وظیفه‌هایی که برای جمعیتی مشخص طراحی شده و پیش‌تر از نظر پایایی و روایی آزموده شده‌اند، معمولاً دارای دستورالعمل‌ها، شرایط اجرا و شیوه‌های کدگذاری نسبتاً ثابت و ازپیش‌تعیین‌شده‌اند. درمقابل، بسیاری از وظیفه‌های مورداستفاده در پژوهش‌های تولید گفتار به‌طور خاص برای پاسخ به یک پرسش پژوهشی معین طراحی می‌شوند و میزان ساختارمندی و کنترل در آنها مستقیماً تابع هدف مطالعه است.

۳. طراحی وظایف برانگیزش گفتار برای یک پیکرهٔ گفتاری

یکی از چالش‌های اصلی در تهیهٔ یک پیکرهٔ گفتاری، طراحی وظایفی است که برای برانگیزش گفتار به کار می‌روند. مواجهه با این چالش در گرو پاسخ دادن به چند پرسش بنیادین دربارهٔ هدف آزمایش است: داده‌های حاصل از این پیکره قرار است به چه پرسش‌های پژوهشی پاسخ دهند؟ کدام جنبه‌های گفتار باید در داده‌ها بازنمایی شوند؟ چه میزان کنترل یا آزادی در تولید گفتار مورد نظر است؟ پاسخ به این پرسش‌ها است که مسیر طراحی را روشن می‌کند و امکان ارزیابی این را فراهم می‌آورد که یک وظیفه تا چه اندازه با اهداف پژوهش هم‌راستا است.

در پژوهش‌های گفتاری، طیف متنوعی از وظایف برانگیزش برای دستیابی به انواع مختلف داده به کار رفته‌اند. در یک سوی این طیف، وظایف خوانداری قرار دارند که شامل خواندن واژه‌ها، جمله‌ها یا متون پیوسته‌اند و بیشترین میزان کنترل را بر محتوای زبانی و شرایط تولید گفتار فراهم می‌کنند، مانند خواندن متون خبری یا دیالوگ‌های ازپیش‌طراحی‌شده (نک. Niebuhr & Michaud 2015). در سوی دیگر،

وظایف تعاملی و مسئله‌محور مانند «وظیفهٔ نقشه»[1] قرار می‌گیرند که در آن‌ها گفتار در جریان یک کنش ارتباطی هدفمند و میان دو مشارک تولید می‌شود (Anderson et. al. 1991). میان این دو قطب، انواع دیگری از وظایف قرار می‌گیرند، ازجمله روایت‌گری، توصیف تصویر، پرسش‌وپاسخ، تکمیل جمله و وظایف مبتنی بر حل مسئله، که هریک با درجه‌ای متفاوت از کنترل و طبیعی بودن، برای برانگیزش جنبه‌های خاصی از گفتار به کار می‌روند (Faitaki and Murphy 2019).

مرور این تنوع نشان می‌دهد که هیچ وظیفه‌ای به‌خودی‌خود «بهترین» گزینه برای گردآوری داده نیست، بلکه کارآمدی هر وظیفه در نسبت با هدف پژوهش و نوع تحلیلِ مدنظر تعیین می‌شود. در مطالعهٔ موردی حاضر، که به مستندسازی مراحل طراحی و ساخت یک پیکرهٔ گفتاری در چارچوب رسالهٔ دکتری نویسنده اختصاص دارد، تمرکز بر بررسی الگوهای ریتم گفتار در میان وظایف گفتاری متفاوت، با تأکید ویژه بر تفاوت‌های فردی در تولید ریتم بوده است. بر این اساس، طراحی پیکره بر این فرض استوار شد که برای مطالعهٔ ریتم فردی، اتکا به یک نوع واحد از وظایف گفتاری کفایت نمی‌کند و لازم است طیفی از بافت‌ها و سبک‌های تولید گفتار در داده‌ها بازنمایی شوند. همچنین، تمامی وظایف به‌صورت تک‌نفره طراحی شدند و شامل تعامل یا مکالمهٔ دوطرفه نبودند. درنتیجه، پیکره شامل هفت وظیفهٔ گفتاری است که به‌گونه‌ای انتخاب و طراحی شده‌اند تا دامنه‌ای از بافت‌های تولید گفتار را پوشش دهند: از گفتار خوانده و نسبتاً کنترل‌شده تا گفتار محاوره‌ای‌تر، و نیز پیام‌های صوتی با سطوح متفاوتِ رسمیت. این تنوع وظایف امکان مشاهده و مقایسهٔ الگوهای ریتم گفتار را در بافت‌های گفتاری گوناگون فراهم می‌کند.

دو وظیفهٔ نخست به گفتار خوانده اختصاص دارند و با هدف ایجاد شرایطی نسبتاً کنترل‌شده برای تولید گفتار طراحی شده‌اند. در این بخش، از شرکت‌کنندگان خواسته شد دو متن را با صدای بلند بخوانند: یک متن رسمی و یک متن غیررسمی. متن رسمی شامل یک خبر با ساختارِ استاندارد ژورنالیستی (تیتر، لید و بدنه) بود که با زبانی عینی و اطلاع‌رسان نوشته شده بود و به موضوعی مرتبط با محیط‌زیست می‌پرداخت. درمقابل، متن غیررسمی یک دستور پخت غذا بود که زبانی محاوره‌ای‌تر و کارکردی آموزشی داشت. انتخاب این دو متن با هدف ایجاد تمایز سبکی در گفتار خوانده و بررسی تأثیر رسمیت و نوع متن بر الگوهای گفتار انجام شد. هر دو متن از نظر طول قابل مقایسه بودند.

---

1) Map task

بخش گفتار محاوره‌ایِ پیکره شامل سه وظیفهٔ مجزا است که هریک با میزان متفاوتی از ساختارمندی و کنترل طراحی شده‌اند. در وظیفهٔ گفت‌وگو، از شرکت‌کنندگان خواسته شد دربارهٔ غذای موردعلاقهٔ خود صحبت کنند و توضیح دهند که آن را چگونه تهیه می‌کنند، چه زمانی آن را مصرف می‌کنند و چه ویژگی‌هایی برایشان اهمیّت دارد. این وظیفه با هدف برانگیزش گفتار نسبتاً آزاد و نزدیک به تعامل‌های روزمره طراحی شد و شرکت‌کنندگان تشویق شدند گفتار خود را به‌صورت توضیحی و پیوسته ارائه دهند.

در وظیفهٔ توصیف تصویر، یک تصویرِ انتخاب‌شده در اختیار شرکت‌کنندگان قرار گرفت و از آنها خواسته شد هرآنچه را در تصویر می‌بینند با جزئیات توصیف کنند. این تصویر برگرفته از یک پلتفرم شبکه‌های اجتماعی بود که مردی را در حال صرف صبحانه در اتاقی نسبتاً شلوغ نشان می‌داد و چهار گربه نیز در صحنه حضور داشتند. این تصویر به دلیل محتوای بصری و عاطفیِ نسبتاً غنی انتخاب شد.

در وظیفهٔ روایت‌گری، شرکت‌کنندگان خاطره‌ای شخصی را از دوران کودکی خود، به‌ویژه یک تعطیلات تابستانی، بازگو کردند. این وظیفه با هدف برانگیزش گفتار روایی طراحی شد.

دو وظیفهٔ پایانی پیکره به ضبط پیام‌های صوتی اختصاص داشت و با هدف بازنمایی تولید گفتار هدفمند در بافت‌های ارتباطی مشخص طراحی شد. در هر دو وظیفه، شرکت‌کنندگان پیامی صوتی خطاب به مخاطبی فرضی تولید کردند؛ با این تفاوت که سطح رسمیت در دو سناریو متفاوت بود. در وظیفهٔ پیام صوتی غیررسمی، از شرکت‌کنندگان خواسته شد دوستی را به یک مهمانی دعوت کنند و اطّلاعاتی مانند زمان، مکان و دلیل دعوت را در پیام خود بیان کنند. درمقابل، در وظیفهٔ پیام صوتی رسمی، سناریویی تعریف شد که در آن شرکت‌کننده از یک استاد درخواست نوشتن توصیه‌نامه می‌کرد. برای هر دو وظیفه، تنها چارچوب کلی سناریو مشخص شده بود و از شرکت‌کنندگان خواسته می‌شد پیام‌ها را بر اساس سبک فردی خود و با افزودن جزئیاتی که مرتبط می‌دانستند تکمیل کنند.

٤. انتخاب گویندگان و طراحی محیط ضبط صدا

پس از تصمیم‌گیری دربارهٔ نوع وظایف برانگیزش گفتار، مرحله‌ای است که توجّه به آن نقشی تعیین‌کننده در کیفیت داده‌های گردآوری‌شده ایفا می‌کند: انتخاب گویندگان و طراحی شرایط ضبط صدا. نیبور و میشو (2015) یادآوری می‌کنند که گردآوری داده‌های گفتاری چالشی است که اغلب دست‌کم گرفته می‌شود؛

زیرا اگر هدف ثبت گفتاری باشد که بازتاب‌دهندهٔ فرایندهای واقعی ارتباطی است، صِرف تعریف یک «وظیفه» کافی نخواهد بود و سایر وجوه فرایند برانگیزش نیز باید به‌دقّت مورد توجّه قرار گیرند، ازجمله خود گوینده و امکانات و شرایط ضبط. بر همین اساس، در طراحی این پیکره تلاش شد هم ویژگی‌های گویندگان به‌عنوان منبعی مهم از تفاوت‌های گفتاری به‌صورت شفاف تعریف شود و هم جزئیات تنظیمات فنّی و شرایط محیط ضبط با دقّت و حساسیت در نظر گرفته شده و مستند شوند.

تفاوت در صداهای گویندگان بخشی از واقعیت داده‌های گفتاری است. تفاوت‌های فیزیولوژیک بر مؤلّفه‌هایی چون بسامد پایه، سرعت گفتار و ویژگی‌های طیفی اثر می‌گذارند. به‌همین‌ترتیب، تفاوت‌های مرتبط با جنسیت هم ریشه‌های زیستی دارند و هم با رفتارهای اجتماعی و آموخته‌شده درهم تنیده‌اند، به‌گونه‌ای‌که تفکیک این دو منبع از یکدیگر همواره ساده نیست (Laver 1994). همچنین، در تعامل‌های گفتاری، پدیده‌هایی مانند «همگرایی یا همسازی آوایی»[1] می‌توانند موجب شوند دو گوینده در طول گفت‌وگو از نظر ویژگی‌های آوایی و نوایی به یکدیگر نزدیک‌تر شوند؛ به این معنا که حتی «هم‌نشینی دو نفر» نیز خود می‌تواند عاملی برای تغییر در داده‌های گفتاری باشد. نیبور و میشو (2015) همچنین تأکید می‌کنند که تجربهٔ زبانی گویندگان ازجمله دوزبانگی یا چندزبانگی، تماس زبانی و پیشینهٔ لهجه‌ای در بلندمدّت بر تولید گفتار اثر می‌گذارد و این اثرها به‌ویژه در سطوح نوایی و آهنگی نمود پیدا می‌کنند.

برای پیکره‌ای که هدف آن بررسی الگوهای فردی در گفتار است، توجّه به این ملاحظات اهمیّتی اساسی دارد؛ زیرا از همان ابتدا روشن می‌کند کدام متغیرها باید تا حد امکان کنترل شوند، کدام ویژگی‌ها نیازمند مستندسازی دقیق‌اند و کدام تفاوت‌ها دقیقاً قرار است موضوع تحلیل باشند. در نگاه نخست، ممکن است این تصور شکل بگیرد که هرچه تنوع داده‌ها بیشتر باشد، داده «بهتر» خواهد بود، امّا در عمل این پرسش مطرح می‌شود که چنین رویکردی تا چه اندازه امکان‌پذیر است و افزایش بیش‌ازحدِ تنوع چگونه می‌تواند تفسیر نتایج را دشوار یا حتی محل تردید کند. آنچه در این میان نقش تعیین‌کننده دارد، نسبت میان حجم پژوهش و تعداد متغیرهایی است که وارد طراحی پیکره می‌شوند. افزودن متغیرهای بیشتر، هرچند می‌تواند دامنهٔ داده‌ها را گسترش دهد، هم‌زمان کنترل شرایط پژوهش و ردیابی منشأ تغییراتِ مشاهده‌شده را پیچیده‌تر می‌کند.

---

1) vocal accommodation

از این رو، تصمیم‌گیری دربارهٔ میزان تنوع قابل‌قبول و انتخاب آگاهانهٔ متغیرها نه یک انتخاب اجرایی، بلکه بخشی جدایی‌ناپذیر از طراحی روش‌شناختی پیکره است که باید از همان آغاز مورد توجّه قرار گیرد.

در پیکرهٔ «پِرچ»، ۳۰ گویندهٔ فارسی‌زبان حضور دارند. در طراحی نمونهٔ گویندگان، سه معیار ورود در نظر گرفته شد: تک‌زبانه بودن، داشتن پیش‌زمینهٔ دانشگاهی و قرار گرفتن در بازهٔ سنی هجده تا چهل سال. نمونه از نظر جنسیت شامل ۱۵ زن و ۱۵ مرد می‌شد. همهٔ گویندگان مهاجران ساکن آلمان بودند؛ با این حال، بر پایهٔ ارزیابی شنیداری پژوهشگر، گفتار فارسی آنها نشانهٔ برجسته‌ای از تأثیر زبان دوم یا لهجه‌های منطقه‌ایِ فارسی نشان نمی‌داد.

نکتهٔ مهم دیگر، خودِ موقعیت ضبط است. نیبور و میشو (2015) تأکید می‌کنند که «اتاق ضبط» یک فضای خنثی نیست؛ شرایط آزمایشگاهی می‌تواند شیوهٔ سخن گفتن را تغییر دهد، به‌ویژه زمانی که مخاطب واقعی یا کنش ارتباطی مشخصی وجود نداشته باشد. آنها حتی نشان می‌دهند که حضور یا نبودِ یک شنونده در اتاق ضبط می‌تواند بر روانی و جریان روایت تاثیرگذار باشد. بر این اساس، «طبیعی بودن» گفتار معیار ساده‌ای نیست. گفتار همواره واکنشی طبیعی به یک موقعیت مشخص است، حتی اگر آن موقعیت از نظر پژوهشگر غیرعادی به نظر برسد. بنابراین، آنچه اهمیّت بیشتری دارد نه تمایز میان گفتار طبیعی و آزمایشگاهی، بلکه تعریف روشن موقعیت ارتباطی و قابل‌درک بودن آن برای گوینده است (Niebuhr & Michaud 2015; Wagener 1986). با توجّه به این ملاحظات، مرحلهٔ پیش‌ازضبط و شیوهٔ ارائهٔ دستورالعمل‌ها در این پیکره بخشی از طراحی روش‌شناختی محسوب شد، نه صرفاً یک مرحلهٔ اجرایی بدیهی.

در مرحلهٔ پیش‌ازضبط، ازآنجاکه کل فرایند در آزمایشگاه آکوستیک انجام می‌شد، تلاش شد فضا تا حد امکان آرام و بدون تنش باشد تا اثر اِسترس بر صدا به حداقل برسد. هر جلسه با یک گفت‌وگوی کوتاه آغاز می‌شد که صرفاً نمای کلی فرایند را توضیح می‌داد و از ورود به جزئیاتی که ممکن بود اجرای وظایف را جهت‌دهی کند پرهیز می‌شد. درصورت تمایلِ شرکت‌کننده، توضیحات تکمیلی به پس از جلسه موکول می‌شد. این گفت‌وگوی اوّلیه کارکرد عملی مهمی نیز داشت و به شرکت‌کنندگان کمک می‌کرد سطح صدای طبیعیِ خود را پیدا کنند؛ چراکه در فضای ساکتِ آزمایشگاه، بسیاری از افراد ناخودآگاه صدای خود را کاهش می‌دهند.

پس از توضیح کلیِ روند ضبط، فرم رضایت آگاهانه در اختیار مشارکان قرار می‌گرفت تا پس از مطالعه آن را تکمیل و امضا کنند. این فرم برای اعلام رضایت در استفاده از داده‌های صوتی صرفاً در چارچوب اهداف پژوهشی طراحی شده بود. در سال‌های اخیر، توجّه به ابعاد اخلاقی در حوزهٔ پردازش گفتار و مطالعات زبانی پررنگ‌تر شده است. باتلینر[1] و دیگران (2023) تأکید می‌کنند که آگاهی اخلاقی باید به‌صورت نظام‌مند در طراحی و به‌کارگیری سامانه‌های مبتنی بر داده‌های گفتاری لحاظ شود و اصولی مانند حریم خصوصی، مسئولیت‌پذیری، شفافیت و احترام به خودمختاری مشارکان بخشی از طراحی پژوهش به‌شمار آیند.

در این پژوهش نیز این ملاحظات از همان مرحلهٔ گردآوری داده مدنظر قرار گرفتند. برای هر مشارک یک شناسهٔ ناشناس بر اساس سال تولّد (میلادی)، شمارهٔ مشارک و جنسیت تعریف شد و به او اطّلاع داده شد که از این پس داده‌ها تنها با این شناسه ذخیره و پردازش می‌شوند و نام یا اطّلاعات هویتی در هیچ بخشی از پیکره ثبت نخواهد شد. داده‌های هویتی نیز جدا از داده‌های پژوهشی نگهداری شدند. بدین‌ترتیب، ارجاع و پیگیری داده‌ها برای تحلیل امکان‌پذیر بود، بدون آنکه هویت افراد در ساختار پیکره منعکس شود. مشارکت کاملاً داوطلبانه بود و امکان انصراف در هر مرحله وجود داشت. برای جبران زمان صرف‌شده، مبلغ ۱۰ یورو به مشارکان پرداخت شد؛ هرچند برخی از آنها ترجیح دادند بدون دریافت دستمزد در ضبط شرکت کنند.

پس از این مرحله، ضبط اصلی آغاز می‌شد. دستورالعمل همهٔ وظایف بر روی تبلت نمایش داده می‌شد و پیش از هر وظیفه، یک اسلاید کوتاه ارائه می‌شد که توضیح می‌داد از شرکت‌کننده چه انتظاری می‌رود. این اسلایدها فرصتی فراهم می‌کردند تا پیش از شروع هر بخش، پرسش‌ها یا ابهام‌های احتمالی برطرف شوند. برای پرهیز از ورود ناگهانی به تکالیف ساختارمند، یک مرحلهٔ گرم کردن کوتاه نیز در نظر گرفته شد که شامل خواندن سه جملهٔ ساده از روی صفحه بود. این مرحله به عادت‌کردن به تجهیزات و تنظیم سطح صدای راحت و طبیعی کمک می‌کرد. وظایف به‌صورت فایل‌های تصویری و با ترتیب ثابت برای همهٔ مشارکان ارائه می‌شدند و شرکت‌کنندگان بدون فشار زمانی، با سرعت خود میان اسلایدها پیش می‌رفتند. این روند عمداً انتخاب شد تا

1) Batliner

روانی گفتار تحت تأثیر شتاب یا اضطراب قرار نگیرد. مدّت هر جلسهٔ ضبط، بدون احتساب مرحلهٔ پیش‌ازضبط، حدود ۲۰ تا ۳۰ دقیقه بود.

ضبط‌ها در دو آزمایشگاه آکوستیک و با حضور خود پژوهشگر به‌عنوان ناظر و همچنین مخاطب انجام شد: در دانشگاه گوته فرانکفورت و در دانشگاه مارتین لوتر هاله-ویتنبرگ. هر دو محیط برای کاهش نویز و فراهم‌کردن کیفیت مناسب ضبط گفتار تجهیز شده بودند. در تمامی جلسات، از یک رکوردر (ضبط‌صوت) دیجیتال Olympus به‌همراه میکروفون هدست استفاده شد. میکروفون به‌صورت ثابت در سمت راست صورت و با فاصله‌ای کمتر از سی سانتی‌متر از دهان قرار می‌گرفت. انتخاب این ترکیب میکروفون-رکوردر پس از مقایسهٔ چند گزینهٔ مختلف انجام شد و معیارهای اصلی شامل وضوح صوتی، راحتی برای شرکت‌کننده و ثبات میان جلسات بودند. صدا به‌صورت تک‌کاناله (mono) با نرخ نمونه‌برداری ۴۴٫۱ کیلوهرتز و رزولوشن ۱۶ بیت ضبط شد و تمامی فایل‌ها با فرمت WAV و بدون فشرده‌سازی ذخیره شدند.

۵. سازمان‌دهی داده‌ها

پس از پایان مرحلهٔ ضبط، پژوهشگر با مجموعه‌ای از داده‌های خامِ آوایی روبه‌رو است که کارآمدی آنها تا حد زیادی به نحوهٔ مدیریت، ذخیره‌سازی و سازمان‌دهی آنها وابسته است. سازمان‌دهی مناسب داده‌ها نه‌تنها برای نگه‌داری و دسترسی مجدد به فایل‌ها ضروری است، بلکه نقش تعیین‌کننده‌ای در امکان استخراج و بازیابی اطّلاعات درون هر فایل، بر اساس نظام نام‌گذاری و فراداده‌های همراه، ایفا می‌کند. درمقابل، ذخیرهٔ داده‌های خام بدون ساختاری روشن و نظام‌مند، حتی اگر از کیفیت ضبط بالایی برخوردار باشند، می‌تواند در مراحل بعدی پردازش و تحلیل به سردرگمی، خطا در ارجاع و کاهش شفافیت منجر شود.

بر این اساس، تمامی فایل‌های صوتی مربوط به هر مشارک پس از ضبط به‌صورت دستی و در محیط نرم‌افزار «پرات» (Praat) (Boersma & Weenink 2024) بر اساس مرزهای هر وظیفهٔ گفتاری قطعه‌بندی شدند. این فرایند به استخراج ۲۱۰ واحد مجزای گفتاری از داده‌های فارسی منجر شد که هریک اجرای یک وظیفهٔ مشخص توسط یک گوینده را نمایندگی می‌کند. هر واحد صوتی بر اساس یک قرارداد استاندارد نام‌گذاری شد، به‌گونه‌ای‌که اطّلاعات اصلی فراداده‌ای مستقیماً در نام فایل

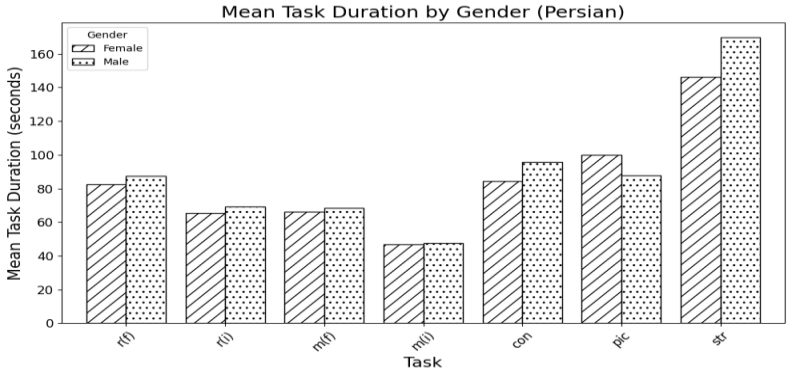
منعکس شوند. این اطّلاعات شامل زبان (G برای آلمانی و P برای فارسی)، سال تولّد به میلادی، جنسیت (F برای زنان و M برای مردان)، شناسهٔ گوینده (از عدد ۱ تا ۳۰ برای هر زبان) و نوع وظیفهٔ گفتاری بود. برای نمونه، نام فایل P_1996_F_19_m(f) به اجرای وظیفهٔ «پیام صوتی رسمی» توسط گویندهٔ شمارهٔ ۱۹، زنِ فارسی‌زبان و متولّد ۱۹۹۶ میلادی اشاره دارد. ضبط داده‌ها در سال ۲۰۲۴ میلادی انجام شد. مزیّت این شیوهٔ نام‌گذاری آن است که در مراحل بعدی تحلیل، هنگام فراخوانی فایل‌ها در محیط‌های پردازش داده، اطّلاعات جمعیّت‌شناختی گویندگان می‌تواند مستقیماً از نام فایل به مجموعه‌داده‌ها اضافه شود، بدون آنکه نیاز به رجوع جداگانه به فایل‌های فراداده باشد. برای حفظ انسجام در پردازش داده‌ها و شفافیت در ارائهٔ نتایج، در سراسر تحلیل از مجموعه‌ای ثابت از اختصارات برای اشاره به وظایف گفتاری استفاده شد (جدول).

جدول اختصارهای به‌کاررفته برای نام‌گذاری وظایف گفتاری

گفتار خوانداری	خواندن متن رسمی[1]	r(f)
	خواندن متن غیررسمی[2]	r(i)
گفتارمحاوره‌ای	گفت‌وگو[3]	con
	توصیف تصویر[4]	pic
	روایت‌گری[5]	str
پیام صوتی	پیام رسمی[6]	m(f)
	پیام غیررسمی[7]	m(i)

این نظامِ نام‌گذاری و سازمان‌دهی داده‌ها، دسترسی به فایل‌ها، ارجاع دقیق به آیتم‌ها در مراحل تحلیل و گزارش شفاف نتایج را تسهیل می‌کند و به‌عنوان بخشی اساسی در فرایند ساخت پیکره در نظر گرفته شده است. شکل ۱ میانگین مدّت‌زمانِ اجرای هریک از وظایف گفتاری را برای گویشوران زن و مرد در بخش فارسی پیکره نشان می‌دهد.

1) Formal read speech
2) Informal read speech
3) Conversation
4) Picture description
5) Story-telling
6) Formal voice message
7) Informal voice message



شکل ۱ میانگین مدّت‌زمان اجرای وظایف گفتاری (برحسب ثانیه) به تفکیک جنسیت در بخش فارسی پیکره

٦. پیش‌پردازش داده‌ها و تقطیع گفتار

در تبدیل داده‌های خامِ آوایی به واحدهایی که بتوان آنها را به‌طور نظام‌مند اندازه‌گیری و تحلیل کرد، مرحلهٔ تقطیع گفتار نقشی تعیین‌کننده پیدا می‌کند. محدودیت‌های تقطیعِ کاملاً دستیِ گفتار موضوعی است که از مدّت‌ها پیش در پژوهش‌های آواشناختی به آن اشاره شده است. مطالعات نشان می‌دهند که برچسب‌گذاریِ دستیِ گفتار می‌تواند تا ۴۰۰ برابرِ طول فایل صوتی زمان ببرد یا به‌طور متوسط برای هر واج بین ۱۱ تا ۳۰ ثانیه زمان نیاز داشته باشد (Godfrey et al. 1992). چنین هزینهٔ زمانی بالایی باعث می‌شود تقطیع دستی، به‌ویژه در مطالعاتی که با حجم قابل‌توجهی از داده سروکار دارند، به‌سختی قابل‌اتکا و عملاً مقیاس‌ناپذیر باشد. از سوی دیگر، پیشرفت‌های اخیر در فناوریِ گفتار امکان استفاده از روش‌ها و سامانه‌های خودکار برای تقطیع سیگنال صوتی را فراهم کرده‌اند و، درنتیجه، وابستگی به برچسب‌گذاریِ کاملاً دستی به‌طور محسوسی کاهش یافته است.

در پژوهش‌های آواشناختی، تقطیع گفتار معمولاً با تکیه بر سامانه‌های «هم‌ترازی»[1] متن و گفتار انجام می‌شود. در این چارچوب، ابزارها و پلت‌فرم‌هایی مانند «پرات»[2] (Boersma & Weenink 2024) و «وب‌ماوس»[3] (Kisler et al. 2017) سیگنال گفتار را به مرزهای واجی تقسیم می‌کنند. در کنار این رویکردهای واج‌محور، روش‌های جایگزینی نیز وجود دارند که تقطیع را بر پایهٔ ویژگی‌های

1) Forced alignment
2) Praat
3) WebMAUS

سیگنال مانند تغییرات پوش دامنه[1] انجام می‌دهند و بیشتر بر رویدادهای برجستهٔ ادراکی تکیه دارند (de Jong & Wempe 2009). در پژوهش‌های پیشینِ نویسنده نیز، با تمرکز بر شناسایی آغازه‌های واکه به‌عنوان نقاط مرجع زمانی، روش‌های هم‌ترازی خودکار و روش‌های مبتنی بر پوش دامنه مقایسه شده و نشان داده شده است که انتخاب روش تقطیع می‌تواند بر تعیین این نقاط و بر شاخص‌های زمانیِ مبتنی بر آنها اثر بگذارد (Mousavi & Grawunder 2024; Mousavi 2025). در این مطالعه و برای ساخت پیکرهٔ «پِرچ»، از سامانهٔ وب‌ماوس به‌عنوان ابزار تقطیع گفتار استفاده شد.

در سامانه‌های تقطیع مبتنی بر هم‌ترازی، علاوه بر سیگنال خام آوایی، وجود یک نسخهٔ متنی از گفتار شرط اساسی برای انجام تقطیع خودکار است. در پیکره‌های کوچک، این متن می‌تواند به‌صورت دستی تهیه شود؛ امّا در مواجهه با حجم بالای داده‌های گفتاری، استفاده از روش‌های خودکار مبتنی بر بازشناسی خودکار گفتار (ASR) به یک ضرورت عملی تبدیل می‌شود.

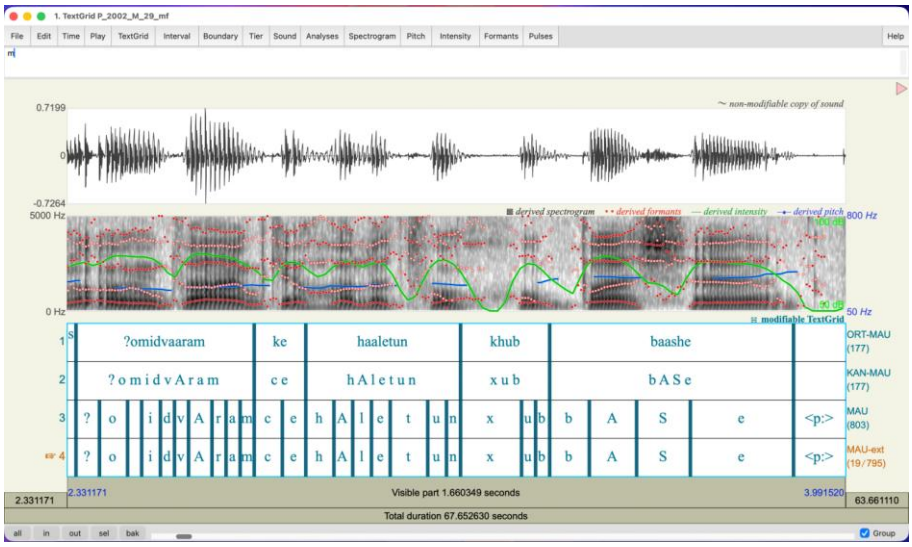
در این مطالعه، برای داده‌های فارسی ابتدا از مدل‌های بازشناسی گفتار مبتنی بر یادگیری عمیق ازجمله Wav2Vec2 (Grosman 2021) و Whisper (Radford et al. 2023) استفاده شد. امّا، نتایج نشان داد که خروجی این مدل‌ها برای فارسی، حتی در شرایط گفتار خوانده و نسبتاً کنترل‌شده، از دقّت لازم برای استفادهٔ مستقیم در فرایند تقطیع واجی برخوردار نیست. خطاهایی مانند حذف یا افزوده‌شدن واج‌ها، جایگزینی واژه‌ها و ناهماهنگی‌های نوشتاری، استفادهٔ بی‌واسطه از این ترانویسی‌ها را برای هم‌ترازی نامطمئن می‌کرد (Mousavi 2025).

همچنین، ازآنجاکه سامانهٔ وب‌ماوس برای تقطیع گفتار فارسی به ورودی متنیِ رومانیزه نیاز دارد، مرحلهٔ تبدیل متن به این قالب چالش دیگری ایجاد می‌کرد. در این پژوهش، ترانویسی گفتار فارسی مستقیماً به‌صورت رومانیزه انجام شد و با توجّه به نبود ابزارهای رومانیزه‌سازیِ خودکارِ قابل‌اتکا برای فارسی و به‌منظور جلوگیری از انباشت خطاهای زنجیره‌ای، این مرحله به‌صورت دستی توسط پژوهشگر انجام گرفت. این فرایند هرچند زمان‌بر بود، امّا برای حفظ دقّت و انسجام داده‌ها اجتناب‌ناپذیر به‌نظر می‌رسید.

پس از آماده‌سازی متن رومانیزه‌شده، داده‌های صوتی و متنی به سامانهٔ وب ماوس وارد شدند و تقطیع گفتار در مرحلهٔ اوّل در دو سطح واژه‌ای و واجی انجام گرفت. خروجی این فرایند،

1) amplitude envelope

به‌صورت فایل‌های TextGrid پرات، مبنای استخراج اطّلاعات آکوستیکی از واحدهای آوایی قرار گرفت.



شکل ۲ نمونه‌ای از خروجی تقطیع خودکار گفتار با استفاده از وب‌ماوس: سیگنال گفتار، نمایش طیفی، و مرزهای واژه‌ای و واجی در قالب فایل TextGrid در محیط پرات نشان داده شده‌اند.

پس از انجام تقطیع اوّلیه با استفاده از وب‌ماوس، فایل‌های TextGrid در محیط پرات بازبینی شدند. از آنجاکه این پیکره در ابتدا با هدف مطالعهٔ ریتم گفتار طراحی شده بود، لایه‌های تحلیلی دیگری بر پایهٔ برچسب‌های همخوان، واکه و آغازهٔ واکه به آنها افزوده شد. این فرایند به‌صورت خودکار و با استفاده از اسکریپت‌های نوشته‌شده در پرات انجام گرفت (Mousavi 2025). در این مرحله، قطعه‌های واجی به دو دستهٔ همخوان و واکه طبقه‌بندی شدند، بازه‌های زمانی متناظر با هریک تعیین شد و به‌ترتیب با برچسب‌های C و V نشانه‌گذاری شدند. افزون بر این، لایه‌ای مستقل برای نشانه‌گذاری آغازه‌های واکه تعریف شد تا استخراج منظم فواصل زمانی میان آغاز واکه‌ها امکان‌پذیر شود؛ در این لایه، تنها نقاط آغاز واکه‌ها با برچسب v مشخص شدند. نتیجهٔ این فرایند، مجموعه‌ای شامل ۲۱۰ فایل گفتاری به زبان فارسی به‌همراه TextGridهای متناظر است که افزون بر کاربرد در تحلیل ریتم گفتار، منبعی غنی و نظام‌مند برای تحلیل‌های آواشناختی است.

۷. نتیجه و بحث

هدف این مقاله ارائهٔ گزارشی تحلیلی از فرایند ساخت یک پیکرهٔ گفتاری است؛ گزارشی که نه بر اساس یک دستورالعملِ ازپیش‌تدوین‌شده، بلکه بر پایهٔ تجربه‌های عملی حاصل از طراحی، ضبط و آماده‌سازی داده‌ها شکل گرفته است. تمرکز اصلی بر این بوده است که نشان داده شود چگونه تصمیم‌هایی که در نگاه نخست جزئی به نظر می‌رسند، از انتخاب نوع وظایف گفتاری گرفته تا شیوهٔ نام‌گذاری فایل‌ها و آماده‌سازی داده‌ها برای تقطیع، می‌توانند پیامدهای مستقیمی برای کیفیت و تفسیرپذیری داده‌ها داشته باشند. از این منظر، فرایند ساخت پیکره نه صرفاً پیش‌نیاز تحلیل، بلکه خود بخشی از فعالیت پژوهشی تلقی شده است.

یکی از نتایج ملموس این رویکرد، شکل‌گیری مجموعه‌ای از داده‌های گفتاری با ساختاری روشن و قابل‌پیگیری بود. این شیوهٔ سازمان‌دهی نشان داد که مدیریت داده‌ها صرفاً یک مسئلهٔ فنّی یا اجرایی نیست، بلکه نقشی اساسی در امکان بازبینی، مقایسه و بازتحلیل داده‌ها ایفا می‌کند. توجّه به این سطح از جزئیات، به‌ویژه در پیکره‌هایی که برای تحلیل‌های چندمرحله‌ای طراحی می‌شوند، از بروز ابهام در مراحل بعدی جلوگیری می‌کند.

بررسی داده‌ها همچنین نشان داد که تفاوت‌های مشاهده‌شده میان وظایف گفتاری، بیش از آنکه تنها ناشی از تفاوت‌های فردی باشند، به ماهیت خود وظایف و نوع کنش ارتباطی آنها بازمی‌گردند. ثبت و مستندسازی این تفاوت‌ها اهمیّت ویژه‌ای دارد، زیرا چنین ویژگی‌هایی مستقیماً بر حجم دادهٔ قابل‌تحلیل و تعداد واحدهای استخراج‌شده اثر می‌گذارند و، درنتیجه، باید در تفسیر نتایج در نظر گرفته شوند. این موضوع نشان می‌دهد که توصیف دقیق مراحل تهیهٔ داده‌ها بخشی جدایی‌ناپذیر از تحلیل است.

در مرحلهٔ پیش‌پردازش و تقطیع، تجربهٔ کار با داده‌های فارسی اهمیّت توجّه به محدودیت‌های ابزارهای موجود را برجسته کرد. هرچند سامانه‌های تقطیع خودکار امکان کار با حجم بالای داده را فراهم می‌کنند، کیفیت خروجی آنها به‌شدّت به کیفیت و سازگاری داده‌های ورودی وابسته است. چالش‌هایی مانند دقّت محدود بازشناسی خودکار گفتار و نبود ابزارهای رومانیزه‌سازیِ قابل‌اتکا برای فارسی نشان داد که بخشی از فرایند استانداردسازی داده همچنان نیازمند مداخلهٔ انسانی است. تصمیم به انجام برخی مراحل به‌صورت دستی، هرچند زمان‌بر، درنهایت به شفافیت بیشتر داده‌ها و کاهش ابهام در مراحل بعدی انجامید.

درمجموع، این مطالعهٔ موردی نشان می‌دهد که ساخت یک پیکرهٔ گفتاری استاندارد فرایندی تدریجی و مبتنی بر تصمیم‌های روش‌شناختی است و کیفیت نهایی آن از مجموعه‌ای از انتخاب‌های به‌هم‌پیوسته شکل می‌گیرد. مستندسازی این انتخاب‌ها بخشی از خودِ پژوهش است، زیرا امکان ارزیابی انتقادی داده‌ها، بازخوانی نتایج و استفادهٔ مجدد از پیکره را فراهم می‌کند.

در کنار این موضوع، توجّه به محدودیت‌های عملی نیز ضروری است. دسترسی به جامعهٔ گویشوران، شرایط ضبط، منابع فنّی و زمانی و بافت اجتماعی مشارکان همگی بر شکل نهایی پیکره اثر می‌گذارند. این محدودیت‌ها لزوماً نشانهٔ ضعف طراحی نیستند، بلکه بخشی از واقعیت میدانی پژوهش‌اند و بهتر است که به‌صورت شفاف ثبت شوند. حتی در نبود شرایط کاملاً ایده‌آل، ثبت تصمیم‌ها و محدودیت‌ها به درک بهتر دامنهٔ اعتبار داده‌ها و تفسیر دقیق‌تر نتایج کمک می‌کند.

منابع

بی‌جن‌خان، محمود (۱۳۹۵)، «پیکرهٔ گفتار محاوره‌ای زبان فارسی امروز»، مجموعه‌مقالات دومین همایش ملّی زبان‌شناسی پیکره‌ای، تهران، نشر نویسه پارسی.

Anderson, A., A. Clark, & J. Mullin (1991), "Introducing Information in Dialogues: How young speakers refer and how young listeners respond", *Journal of Child Language*, 18, pp. 663-687.

Baker, R. and V. Hazan (2011), "DiapixUK: Task Materials for the Elicitation of Multiple Spontaneous Speech Dialogs", *Behavior Research Methods*, 43, pp. 761-770.

Batliner, A., Neumann, M., Burkhardt, F., Baird, A., Meyer, S., Vu, N. T., and Schuller, B. W. (2023), "Ethical Awareness in Paralinguistics: A Taxonomy of Applications", *International Journal of Human-Computer Interaction*, 39(9), pp. 1904-1921.

Bijankhan, M., Sheikhzadegan, J., Roohani, M. R., Samareh, Y., Lucas, C., and Tebyani, M. (1994), "FARSDAT-The Speech Database of Farsi Spoken Language", *The Proceedings of the Australian Conference on Speech Science and Technology*, 2, pp. 828-831.

Boersma, P. and Weenink, D. (2024), *Praat: Doing Phonetics by Computer* (Version 6.4.23) [Computer program], Retrieved October 27, 2024. http://www.praat.org/

Brown, R. (1973), "Development of the First Language in the Human Species", *American Psychologist*, 28(2), 97, American Psychological Association.

De Jong, N. H. and T. Wempe (2009), "Praat Script to Detect Syllable Nuclei and Measure Speech Rate Automatically", *Behavior Research Methods*, 41(2), pp. 385-390.

Faitaki, F. and V. A. Murphy (2019), "Oral Language Elicitation Tasks in Applied Linguistics Research", *The Routledge Handbook of Research Methods in Applied Linguistics*, pp. 360-369, Routledge.

Godfrey, J. J., E. C. Holliman & J. McDaniel (1992), "SWITCHBOARD: Telephone Speech Corpus for Research and Development", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 517-520, IEEE Computer Society.

Grosman, J. (2021), "Fine-tuned XLSR-53 Large Model for Speech Recognition in Persian", *Hugging Face*, pp. 410-411. https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53- Persian.

Kisler, T., U. Reichel & F. Schiel (2017), "Multilingual Processing of Speech via web Services", *Computer Speech & Language*, 45, pp. 326-347.

Laver, J. (1994), *Principles of Phonetics*, Cambridge, Cambridge University Press.

Mousavi N., S. Grawunder (2024), "The Influence of Signal Segmentation Methods on Rhythm-based Speaker Recognition", Baumann, T. (ed.), *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2024*, Dresden, *TUDpress*, pp. 225-232.

Mousavi, N. (2025, PhD Thesis), "Individual Speech Rhythm in Persian and German: An Acoustic–Cognitive Approach of Rhythm Production in Different Speaking Tasks", Martin-Luther Universität Halle-Wittenberg, Halle (Saale), Germany.

Nezami, M.O., P. Jamshid Lou & M. Karami (2019), "ShEMO: A Large-scale Validated Database for Persian Speech Emotion Detection", *Language Resources & Evaluation*, 53(1), pp. 1-16.

Niebuhr, O. and A. Michaud (2015), "Speech Data Acquisition: The Underestimated Challenge", *KALIPHO - Kieler Arbeiten zur Linguistik und Phonetik*, 3, pp. 1-42.

Pilehvar, A. & M. Sedaaghi (2008), *Documentation of the Sahand Accented Speech Database (SES)*, Department of Electrical Eng., Sahand University of Tech, Iran.

Radford, A., J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever (2023), "Robust Speech Recognition via Large-scale Weak Supervision", *International Conference on Machine Learning*, pp. 28492-28518, PMLR.

Senft, G. (1995), "Elicitation", *Handbook of Pragmatics: Manual*, pp. 577-581, John Benjamins.

Wagener, P. (1986), "Sind Spracherhebungen Paradox? Über die Möglichkeit, Natürliches Sprachverhalten Wissenschaftlich zu Erfassen", A. Schöne (ed.), Tübingen, Niemeyer, Akten des VII. IVG-Kongresses, Vol. 4, pp. 319-327.

Wagner, P., J. Trouvain & F. Zimmerer (2015), "In Defense of Stylistic Diversity in Speech Research", *Journal of Phonetics*, 48, pp. 1-12.

Xu, Y. (2010), "In Defense of Lab Speech", *Journal of Phonetics*, 38(3), pp. 329-336.

References
Sources are in Persian unless otherwise specified.

Bijankhan, Mahmoud (2016), "Persian Colloquial Speech Corpus", *Proceedings of the Second National Conference on Corpus Linguistics*, Tehran, Neveeseh Parsi Publishing.

❑