بازشناسی خودکار واج‌های فارسی با استفاده از مدل‌سازی واج‌گونه‌ها

طاهره احمدی[۱]، حسین کارشناس[۲]، باقر باباعلی[۳] و بتول علی‌نژاد*[۴]
[۱و۴] دانشکده زبان‌های خارجی، دانشگاه اصفهان، اصفهان، ایران
[۲] دانشکده کامپیوتر، دانشگاه اصفهان، اصفهان، ایران
[۳] دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تهران، تهران، ایران

چکیده

یکی از مراحل زیربنایی در بازشناسی خودکار گفتار، بازشناسی واج‌ها و از موانع جدی برای بازشناسی واج‌ها، هم‌تولیدی است. یک روش برای جبران تأثیر هم‌تولیدی، استفاده از مدل‌های وابسته به بافت در بازشناسی واج‌هاست. در این پژوهش، از یک روش زبان‌شناختی برای مدل‌سازی واج‌گونه‌ها استفاده شده است. بدین‌منظور ابتدا قواعد وقوع واج‌گونه‌ها در زبان فارسی استخراج و مشخص شده است که هر واج چه واج‌گونه‌هایی دارد. برای مدل‌سازی و شناسایی واج‌گونه‌ها، یک پیکرهٔ واج‌گونه‌ای لازم است که به‌منظور تولید آن، از پیکرهٔ فارس‌دات کوچک استفاده و برچسب‌گذاری واج‌گونه‌ای آن انجام و از این پیکره، برای مدل‌سازی و سپس شناسایی واج‌گونه‌های مختلف گفتار ورودی استفاده شده است. درنهایت، با قرارگرفتن هر یک از واج‌گونه‌های شناسایی‌شده در دستهٔ واجی مربوط به خود، بازشناسی واج‌ها از مسیر واج‌گونه‌ها انجام شده است. با این روش، دقت بازشناسی واج‌ها در زبان فارسی در مقایسه با بهترین نتایج گزارش‌شده تاکنون، بهبود قابل‌ملاحظه‌ای نشان داده است.

واژگان کلیدی: بازشناسی خودکار گفتار، بازشناسی خودکار واج، مدل‌های وابسته به بافت، واج، واج‌گونه، هم‌تولیدی.

Allophone-based acoustic modeling for Persian phoneme recognition

Tahere Ahmadi[1], Hossein Karshenas[2], Bagher Babaali[3], Batool Alinezhad*[4]
[1, 4] linguistics Department, foreign languages Faculty, University of Isfahan
[2] Artificial intelligence Department, computer Faculty, University of Isfahan
[3] Faculty of Mathematics, Statistics and Computer Science, University of Tehran

Abstract

Phoneme recognition is one of the fundamental phases of automatic speech recognition. Coarticulation which refers to the integration of sounds, is one of the important obstacles in phoneme recognition. In other words, each phone is influenced and changed by the characteristics of its neighbor phones, and coarticulation is responsible for most of these changes. The idea of modeling the effects of speech context, and using the context-dependent models in phoneme recognition is a method which used to compensate the negative effects of coarticulation. According to this method, if two similar phonemes in speech have different contexts, each of them constitute a separate model. In this research, a linguistic method called allophonic modeling has been used to model context effects in Persian phoneme recognition. For this purpose, in the first phase, the rules required for occurrence of various allophones of each phoneme, are extracted from Persian linguistic resources. So each phoneme is considered as a class, consisting of its various context-dependent forms named allophones. The necessary prerequisites for modeling and identifying allophones, is an allophonic corpus. Since there was no such corpus in Persian language, SMALL FARSDAT corpus has been used. This corpus is segmented and labelled manually for each sentence, word and phoneme. So the phonological and lingual context required for the realization of allophones, is implemented in this corpus.

* Corresponding author
*نویسندهٔ عهده‌دار مکاتبات

• تاریخ ارسال مقاله: ۱۳۹۷/۰۷/۰۶ • تاریخ پذیرش: ۱۳۹۸/۰۳/۰۱ • تاریخ انتشار: ۱۳۹۹/۰۹/۱۵ • نوع مطالعه: کاربردی

[DOI: 10.29252/jsdp.17.3.37]

For example, the syllabification has been performed on the corpus and then, for each phoneme, its position (first, middle and end) in the word and syllable is specified using different numeric tags. In the next step, allophonic labeling has been performed by searching for each of the allophonic contexts in the corpus. These allophonic corpus is used to model and recognize the allophones of input speech. Finally, each allophone is assigned to a proper phonemic class so phoneme recognition has been done using allophones. The experimental results show a high accuracy of the proposed method in phenome recognition, indicating a significant improvement comparing with other state-of-the-art methods.

Keywords: automatic speech recognition, automatic phoneme recognition, context-dependent models, phoneme, allophone, coarticulation.

۱- مقدمه

با توجه به نقش مهم بازشناسی گفتار در تسهیل ارتباط انسان‌ها با یکدیگر و نیز ارتباط انسان‌ها با ماشین، از حدود پنج دههٔ پیش، پژوهش در این حوزه به موضوعی قابل‌توجه در جهان تبدیل شده است. در سال‌های اخیر، این فناوری منجر به تغییرات مثبتی در زندگی انسان‌ها شده و امکان ارتباط با بسیاری از ابزارها را فراهم ساخته است [26,27]. در راستای این تحولات مهم، مطالعات بسیاری در زمینهٔ بازشناسی گفتار در زبان‌های مختلف صورت گرفته است [28-36]. در زبان فارسی نیز پژوهش‌های مختلفی از حدود دو دههٔ پیش در این حوزه آغاز شده و با وجود پیشرفت‌هایی چشم‌گیر، تلاش در جهت بهبود آن، همواره موردتوجه بوده و ایده‌های مختلفی در این رابطه مطرح شده است [37-47].

یکی از مهم‌ترین گام‌ها در بازشناسی خودکار گفتار، بازنمایی واجی گفتار است؛ علاوه‌براین، برخی مسائل دیگر در علوم رایانه اعم از مدل‌سازی گوینده و تبدیل متن به گفتار هم نیازمند این نوع بازنمایی است. به‌منظور آشنایی بیشتر با بازنمایی واجی گفتار، معرفی مختصر واج‌ها و واج‌گونه‌ها ضروری به نظر می‌رسد. طبق تعریف رُوچ[1] (۲۰۱۰) واج‌ها به‌عنوان کوچکترین واحدهای ممیز معنا، مجموعه‌ای از واحدهای انتزاعی هستند که مبنای گفتار را تشکیل می‌دهند. گرچه این واحدها نیز مانند حروف الفبا انتزاعی هستند، اما همان‌گونه که هر یک از حروف الفبا را می‌توان به صورت‌های نوشتاری مختلفی نمایش داد، صورت‌های گفتاری مختلفی نیز برای نمایش هر یک از واج‌ها وجود دارد. در حالت‌های سادهٔ بازشناسی، هر واج به‌طور مستقل مدل می‌شود و واج‌های اطراف آن، موردتوجه قرار نمی‌گیرند؛ این در حالی است که در شرایط واقعی، خصوصیات طیفی واج‌ها در حین گفتار دچار تغییرات گوناگونی می‌شود و یک واج خاص، در دو موقعیت مختلف، حتی در گفتار یک فرد واحد، یکسان تلفظ نمی‌شود.

دسته‌ای از این تغییرات، ناشی از بافت وقوع واج‌هاست. در زبان‌شناسی، واحدهای واجی وابسته به بافت، واج‌گونه نام دارند و هر واج، واحدی است که ممکن است، دربرگیرندهٔ واج‌گونه‌های مختلفی باشد. از دیدگاه زبان‌شناختی، بسیاری از تغییرات طیفی واج در بافت‌های مختلف و در مجاورت واج‌های مختلف، ناشی از هم‌تولیدی[2] هستند. هم‌تولیدی به یک‌پارچه‌شدن صدا اشاره و به عوامل بافتی بسیاری از جمله موقعیت واج هدف در واژه و هجا، و نیز نوع واج‌های مجاور و گاهی غیرمجاور واج هدف، بستگی دارد. از روش‌های جبران تأثیر منفی هم‌تولیدی بر بازشناسی واج‌ها، مدل‌سازی اثر بافت[3] گفتار و استفاده از ایدهٔ مدل‌های وابسته به بافت در بازشناسی واج‌هاست. بر اساس این ایده، دو واج مشابه در گفتار، اگر دارای همسایه‌های متفاوتی باشند، هر کدام یک مدل جداگانه را تشکیل می‌دهند؛ بنابراین به‌منظور مدل‌سازی درست گفتار برای بازشناسی، لازم است، بافت صحبت را نیز در این کار دخالت داد [23, 48, 49]. بر اساس تجربه، به‌کارگیری اطلاعات وابسته به بافت، خطای بازشناسی را به‌میزان قابل‌توجهی کاهش می‌دهد.

برای مدل‌سازی اثرات بافت گفتار، روش‌های مختلفی به کار رفته‌اند و مدل‌سازی گفتار بر اساس واحدهای واجی وابسته به بافت متنوعی انجام و اثرات آن بر افزایش دقت بازشناسی سنجیده شده است. یکی از واحدهای واجی وابسته به بافت که می‌توان با استفاده از قواعد زبان‌شناختی آن‌ها را استخراج کرد، واج‌گونه‌ها[4] هستند [23, 49]. در پژوهش حاضر، از واج‌گونه‌ها برای بهبود دقت یک سامانهٔ بازشناسی گفتار پیوستهٔ فارسی، استفاده شده است. در گام‌های بعدی، به‌منظور افزایش دقت بازشناسی، سایر واحدهای واجی نیز به کار گرفته شده‌اند. با به‌کارگیری این روش، دقت بازشناسی واج‌های فارسی در مقایسه با بهترین نتایج به‌دست‌آمده تا کنون، افزایش یافته است.

[1] Roach, P.

[2] Co-articulation
[3] Context
[4] Allophones

[DOI: 10.29252/jsdp.17.3.37]

در ادامه در بخش ۲ مقاله، ایدهٔ واحدهای آوایی وابسته به بافت، توصیف و در بخش ۳، برخی پژوهش‌های پیشین در این زمینه مطرح می‌شود. در بخش ۴، جزئیات روش پیشنهادی در این مطالعه مورد بررسی قرار می‌گیرد. بخش ۵ به گزارش نتایج به‌کارگیری این روش در سامانهٔ بازشناسی گفتار کلدی و بخش ۶ به جمع‌بندی اختصاص خواهد یافت.

۲- واحدهای آوایی وابسته به بافت

ایدهٔ واحدهای آوایی وابسته به بافت این است که تلفظ واژه‌ها یا واحدهای کوچک‌تر از واژه مانند هجا، واج و مواردی از این قبیل، به‌شدّت به بافت وقوع آن‌ها وابسته است؛ بنابراین تلفظ و خصوصیات طیفی یک واج (مستقل از بافت)، در متون مختلف، تغییر می‌کند. این تغییرات به‌صورت وسیع توسط آواشناسان و زبان‌شناسان در زبان‌های مختلف مانند فارسی [1-21] مورد مطالعه قرار گرفته‌اند و قواعد واج‌شناختی مختلفی نیز برای توصیف این تغییرات استخراج شده است. بر اساس این تغییرات، برای هر واج به جای یک حالت، می‌توان چند حالت کلی درنظرگرفت. به عبارت دیگر، یک واحد آوایی، به‌عنوان مثال /p/ بسته به بافتی که در آن وقوع می‌یابد، می‌تواند حالات گوناگونی داشته باشد؛ پس بهتر است به جای مدل‌سازی یک واحد آوایی مستقل از بافت، انواع گوناگون وابسته به بافت آن را در مدل‌سازی و بازشناسی به کار برد [49]. مهم‌ترین فرایند در ایجاد واحدهای آوایی وابسته به بافت، هم‌تولیدی است که روند وقوع آن در فرایندهای واجی در ادامه توصیف می‌شود.

۲-۱- ویژگی‌های تولید نخستین و دومین

هر رشتهٔ گفتاری از واحدهایی به نام واحدهای زنجیری و واحدهای زبرزنجیری تشکیل شده است. هر واحد زنجیری، دسته‌ای از مختصات آوایی است که در توالی گفتار، باهم و به‌صورت ترکیب‌شده ظهور می‌کنند و یک آوای مستقل را تشکیل می‌دهند. برای مثال [b]، [ɒ]، [z]، [ɟ]، [a]، [ʃ] و [t] واحدهای زنجیری واژهٔ «بازگشت» هستند. واحدهای زنجیری به دو دستهٔ واکه‌ها و همخوان‌ها تقسیم می‌شوند. هر یک از دو طبقهٔ واکه‌ها و همخوان‌ها، با ویژگی‌های تولید نخستین و دومین خاصی شناسایی می‌شوند. تولید نخستین همخوان‌ها شامل ترکیب سه مختصهٔ جایگاه تولید، شیوهٔ تولید و واکداری یا بی‌واکی است. تولید نخستین واکه‌ها حاصل ترکیب سه مختصهٔ ارتفاع زبان، شکل لب‌ها و جایگاه فعال زبان است. یک واکه یا همخوان، ممکن است، تنها حاصل تولید نخستین باشد یا علاوه‌بر تولید نخستین، حاصل مختصات آوایی دیگری ناشی از تولید دومین نیز باشد. از مهمترین تولیدهای دومین ویژه همخوان‌ها می‌توان به دمش، واک‌رفتگی، هجایی‌شدگی، لبی‌شدگی، کامی‌شدگی، نرم‌کامی‌شدگی، حلقومی‌شدگی و چاکنایی‌شدگی اشاره کرد؛ و تنها مختصهٔ ناشی از تولید دومین برای واکه‌ها خیشومی‌شدگی است. همچنین تولیدهای دومین مشترک بین همخوان و واکه، سختی، نرمی و کشش هستند [4].

۲-۲- هم‌تولیدی

از ویژگی‌های بنیادی گفتار انسانی که به‌صورت ناآگاهانه اتفاق می‌افتد، این است که در حین گفتار، واج‌ها به‌عنوان واحدهای مجزا تلفظ نمی‌شوند و اندام‌های گویایی، زمانی برای تولید یک واج آماده می‌شوند که هنوز در مراحل تولیدی واج قبلی قرار دارند؛ یعنی بین تلفظ یک توالی واجی به‌صورت مجزا و تلفظ آن‌ها در زنجیرهٔ گفتار، تناظر یک به یک وجود ندارد و نمی‌توان به‌طور دقیق، نقطهٔ پایان تولید یک واج و نقطهٔ آغاز تولید واج بعدی را مشخص کرد. علت این رویداد، آن است که واج‌ها در محیط‌های آوایی متفاوت، یکسان تلفظ نمی‌شوند و ویژگی‌های یک واج، ویژگی‌های واج‌های دیگر را تحت تأثیر قرار می‌دهد. واج‌ها در زنجیرهٔ گفتار، از نظر زمان‌بندی و ویژگی‌های تولیدی، با واج‌های مجاور یا گاهی غیرمجاور خود هم‌پوشی پیدا می‌کنند و بر یکدیگر تأثیر می‌گذارند. این تأثیرگذاری که ممکن است بر واج‌های قبلی یا بعدی باشد، هم‌تولیدی نام دارد. هم‌تولیدی باعث می‌شود نتوان برای یک واج خاص، در تمام بافت‌ها ویژگی‌های تولیدی و آکوستیکی یکسانی را متصور شد [50-52]. به‌عنوان مثال، /c/ (ک) در زبان فارسی، قبل از واکه[1]های پیشین یعنی /i, e, a/ (آ ، اِ ، ای)، در جایگاه کامی تلفظ می‌شود (مانند "ک" در کلمهٔ *کتاب* [cetɒb]) و قبل از واکه‌های پسین یعنی /o, u, ɒ/ (آ ، او ، اُ)، در جایگاه نرم‌کامی تولید می‌شود (مانند "ک" در کلمه‌ی *کار* [kɒr]).

مجرای گفتار برای تولید تمام واج‌ها، مشترک است و توالی واج‌ها، با تغییر شکل مجرای صوتی به‌صورت منظم و پیوسته تولید می‌شود؛ بنابراین شکل مجرای گفتاری در هر لحظه، تحت تأثیر واج‌های مختلفی تغییر می‌کند. با استفاده از طیف‌نگاشت می‌توان تأثیرات آکوستیکی ناشی از هم‌تولیدی را مشاهده و بررسی کرد و صورت‌ها و درجات مختلف تأثیر آواهای مجاور را بر یکدیگر نشان داد [52-50].

[1] Vowel

[DOI: 10.29252/jsdp.17.3.37]

پردازش موازی اطلاعات چندین واج به‌صورت هم‌زمان به‌دنبال هم‌تولیدی، باعث می‌شود که درک گفتار، سریع‌تر از حد انتظار صورت گیرد. هم‌تولیدی ناشی از ویژگی‌های حرکتی سازوکار گفتار است و می‌تواند یکی از جهانی‌های زبان محسوب شود. این پدیده در تمام زبان‌هایی که تاکنون مورد تحلیل و بررسی قرار گرفته‌اند، مشاهده شده است؛ گرچه ممکن است تظاهر آن در زبان‌های مختلف متفاوت باشد؛ بنابراین نمی‌توان آن را در چارچوب دستور هر زبان، توصیف کرد [52-50]. به هریک از صورت‌هایی که واج‌ها بر اثر هم‌تولیدی، در بافت‌های آوایی مختلف به خود می‌گیرند، واج‌گونه گفته می‌شود. چگونگی ایجاد واج‌گونه‌ها را فرایندهای آوایی/ واجی توصیف می‌کنند [56, 53, 50].

مدل‌سازی هم‌تولیدی، یکی از مهم‌ترین بخش‌ها در سامانه‌های بازشناسی گفتار و تبدیل متن به گفتار است. در این سامانه‌ها مقادیر پارامترها در هر قاب، توسط یک بخش کنترل‌کننده محاسبه می‌شود؛ بدین‌منظور، دو رویکرد وجود دارد: ۱. رویکرد مبتنی بر قاعده، که از دانش موجود استفاده می‌کنند؛ ۲. رویکرد مبتنی بر داده، که مجموعه‌ای از واحدهای آوایی ترکیبی با طول‌های مختلف را به‌کار می‌گیرد. هر یک از این دو رویکرد، مزایایی خاصی دارند. رویکرد مبتنی بر قاعده، راهی برای دست‌یابی پژوهش‌گر به اصول زیربنایی است؛ درحالی‌که رویکردهای مبتنی بر پیکره، نتایج طبیعی‌تری تولید می‌کنند. تحلیل پیکرهٔ گفتاری، می‌تواند منبع مناسبی برای جستجوی قواعد هم‌تولیدی و تغییرات وابسته به بافت باشد [52-50]

۳-۲- فرایندهای آوایی/ واجی

همان‌گونه که اشاره شد، فرایندهای واجی، هم‌تولیدی‌ها را توصیف می‌کنند و انواع مختلفی دارند که مهم‌ترین آنها عبارتند از: همگونی[1]، ناهمگونی[2]، حذف[3]، درج[4] و قلب[5][55].

- **همگونی**

همگونی، نتیجهٔ هم‌تولیدی است و به‌نوعی سازگاری زبان با محدودیت‌های گفتاری تلقی می‌شود. این فرایند، به تغییرات وابسته به بافت اصوات گفتاری اشاره دارد و زمانی اتفاق می‌افتد که در یک محیط آوایی، یک واج تحت تأثیر یک واج دیگر (واج دوم)، به واج دیگری (متفاوت با دو واج نخست) تبدیل شود. در نتیجهٔ این فرایند، ویژگی‌های آوایی واج‌های مجاور یا نزدیک به هم، تا حدودی به هم شبیه یا یکسان می‌شود؛ و هر قدر دو واج در زنجیرهٔ گفتار، به یکدیگر نزدیک‌تر باشند، احتمال شباهت آن‌ها افزایش می‌یابد. به‌عنوان یک مثال می‌توان به تلفظ واژهٔ /نَبر (/ʔɒnbor/) در زبان فارسی اشاره کرد که در آن /n/، تحت تأثیر /b/ به /m/ تبدیل می‌شود. یک راه‌کار برای تشخیص فرایند همگونی، مقایسهٔ تلفظ یک آوای مجزا با تلفظ آن آوا در مجاورت آواهای دیگر است. فلسفهٔ وقوع همگونی را می‌توان چنین توصیف کرد که حرکت اندام‌های گفتاری در هنگام تولید گفتار، نیاز به تلاش و صرف انرژی دارد؛ افزایش مشابهت واج‌های مجاور یا نزدیک به هم، میزان حرکت اندام‌های گفتار و در نتیجه میزان تلاش و انرژی لازم برای تولید آن‌ها را کاهش می‌دهد. همگونی، یک ویژگی زبانی است و فرایندهای همگونی در حوزهٔ دستور زبان قرار دارند و بنابراین، زبان‌ویژه هستند [57, 56, 54, 51, 50, 11].

- **ناهمگونی**

ناهمگونی به معنای اعمال برخی فرایندها بر واج‌های مجاور، با هدف افزایش تفاوت میان آن‌ها و تسهیل تمایز آن‌ها از یکدیگر است. این فرایند بر مبنای عدم‌مطابقت، و عکس فرایند همگونی است. علت همگونی، گرایش به سمت طبیعی‌ترشدن تولید و علت ناهمگونی، مربوط به « تأثیر پیچش زبان[6]» است. یکی از دلایل دشواربودن تلفظ در پیچش زبان این است که تولید اصوات تکراری یا اصواتی با درجه شباهت زیاد، برای اندام‌های تولید گفتار خسته‌کننده است [58, 57]. این تأثیر، گاهی در گفتار مشاهده می‌شود؛ به‌عنوان مثال، واژهٔ فارسی «اَنداخت» /ʔandɒχt/ گاهی به‌صورت «اِنداخت» [ʔendɒχt] تلفظ می‌شود که در آن، فرایند ناهمگونی در ارتفاع واکه اتفاق می‌افتد.

- **حذف**

حذف بدین معناست که یک واج تحت شرایط خاصی در زنجیرهٔ گفتار تلفظ نشود. حذف را می‌توان به نوعی «جانشینی یک واج با صفر[7]» در نظر گرفت که در آن، واج با قرارگرفتن در بافت خاصی، غیرقابل شناسایی می‌شود و بدین‌ترتیب می‌توان گفت که با صفر، جایگزین می‌شود. این پدیده به‌طورمعمول در گفتار پیوسته که واژه‌ها در کنار هم قرار می‌گیرند، رخ

[1] Assimilation
[2] Dissimilation
[3] Deletion
[4] Insertion
[5] Metathesis

[6] tongue-twister effect
[7] alternations with zero

[DOI: 10.29252/jsdp.17.3.37]

می‌دهد و به‌طورمعمول زمانی اتفاق می‌افتد که هم‌نشینی واج‌ها با هم، ترکیبی ایجاد کند که تلفظ آن دشوار باشد. برای رفع این اشکال، یک واج از زنجیرهٔ گفتار حذف می‌شود [56, 59]؛ به‌عنوان مثال، در زبان فارسی، واژهٔ «بِنِشین» به «بِنشین» تبدیل می‌شود (/beneʃin/ → [benʃin]) و /e/ حذف می‌شود [4].

- **درج**

در فرایند درج، عکس فرایند حذف اتفاق می‌افتد و به‌منظور تسهیل تولید، یک واج به زنجیرهٔ گفتار اضافه می‌شود [60]. در فارسی، تبدیل واژهٔ «یادگار» /jɒdgɒr/ به «یادِگار» [jɒdegɒr] که در آن /e/ به زنجیرهٔ گفتار اضافه می‌شود، نمونه‌ای از فرایند درج است [4].

- **قلب**

فرایند قلب مربوط به زمانی است که بر اثر هم‌نشینی واج‌ها، جای دو همخوان[1] با یکدیگر عوض شود یا زمانی که یک واج از مکان اصلی خود به موقعیتی در سمت چپ یا راست، تغییر مکان دهد [37]. تبدیل واژهٔ «کبریت» /cebrit/ به «کربیت» [cerbit] در زبان فارسی، نمونه‌ای از فرایند قلب است [4].

۳- پژوهش‌های پیشین

اهمیت و کاربرد سامانه‌های بازشناسی گفتار در حوزه‌های مختلف، امری انکارناپذیر است. به‌عنوان نمونه می‌توان به کاربرد این سامانه‌ها، به‌منظور دست‌یابی به اهداف بالینی و آموزشی مختلف در سال‌های اخیر اشاره کرد. در این زمینه، هارف و همکاران در [61] برای نخستین‌بار نشان داده‌اند که با استفاده ثبت الکتریکی امواج مغزی داخل جمجمه[2]، می‌توان گفتار پیوسته را به واژه‌های تشکیل‌دهندهٔ آن، رمزگشایی کرد و لَسلو و همکاران، چنان که در [29] گزارش شده است، از سامانه‌های بازشناسی گفتار برای تحلیل گفتار خودانگیخته به‌منظور تشخیص اختلالات شناختی خفیف، استفاده کرده‌اند. به‌منظور تأمین اهداف آموزشی، میرزایی و همکاران در [62] از خطاهای سامانهٔ بازشناسی گفتار، برای پیش‌بینی دشواری اجزای گفتار استفاده کرده و این نکته را در آموزش مهارت شنیدنِ زبان دوم[3]، برای زبان‌آموزان به کار گرفته‌اند. بسیاری از سامانه‌های بازشناسی گفتار امروزی که دارای واژگان بزرگ هستند، از شبکه‌های عصبی و مدل‌های مخفی مارکوف استفاده کرده‌اند [63]. شبکه‌های عصبی مورداستفاده در مدل‌سازی آکوستیکی را می‌توان در دو گروه کلی، شامل پیش‌خور و بازگشتی مورد بررسی قرار داد. محمد[4] و همکاران در [64] معتقدند که بهترین نتایج اخیر در حوزهٔ بازشناسی گفتار، حاصل به‌کارگیری مدل‌های آکوستیکی مبتنی بر حافظهٔ کوتاه‌مدت ماندگار[5] است که این مدل‌ها نیز درواقع نوعی شبکهٔ عصبی بازگشتی[6] هستند و قادرند پدیده‌های طولانی‌مدت[7] مربوط به طیف ورودی را برای واحدهای زبانی، مدل‌سازی کنند. باهداناو و همکاران در [63] نیز روشی مبتنی بر یک سازوکار توجه را برای مدل‌سازی یک زنجیره، پایه‌ریزی کرده‌اند که در این روش، از یک شبکهٔ عصبی بازگشتی برای موازی‌سازی زنجیره‌هایی از قاب‌های ورودی و برچسب‌های خروجی استفاده شده است. مون[8] و همکاران در [65] با دیدگاهی به‌نسبه مشابه معتقدند در حوزهٔ برنامه‌ها و ابزارهایی مانند بازشناسی گفتار، بازشناسی دست‌خط و ترجمهٔ ماشینی، شبکه‌های عصبی بازگشتی بهترین عملکرد را دارند. هیماوان[9] و همکاران در [66] و سان[10] و همکاران در [28] از شبکه‌های عصبی عمیق، برای مدل‌سازی آکوستیکی در بازشناسی گفتار استفاده کرده‌اند. رِووری[11] در [67] مدل‌های آکوستیکی مبتنی بر شبکه‌های عصبی عمیق‌ماشین بردار پشتیبان با ساختار نهان[12] را به‌عنوان جایگزینی برای مدل‌های آکوستیکی ترکیبی به‌دست‌آمده با شبکه‌های عصبی عمیق مدل مخفی مارکوف پیشنهاد داده و معتقد است این مدل‌های جایگزین، مبانی نظری قوی‌تری دارند. یکی دیگر از روش‌های بازشناسی خودکار گفتار، انجام این کار با استفاده از مدل‌سازی واجی است. سجان و ویجایا[13] در [34] به استفاده از این روش برای بازشناسی خودکار گفتار در زبان کانادایی پرداخته‌اند. شاوکت[14] و همکاران در [35] رویکردی برای تهیهٔ یک سامانه به‌منظور شناسایی واژه‌های مجزای زبان اردو معرفی کرده‌اند. آن‌ها با استخراج ضرایب مل کپستروم و مبناقراردادن یک دادگان گفتاری واژگانی ۲۵۰ واژه‌ای زبان اردو، رویکرد خود را با استفاده از مدل مخفی مارکوف، توسعه داده‌اند [35]. در [68] سامانه‌ای توسط چان

[1] Consonant
[2] Intracranial electrocorticographic recordings
[3] Second language listening skill

[4] Mohamed, A. R.
[5] Long Short-Term Memory (LSTM)
[6] Recurrent Neural Network (RNN)
[7] Long- span phenomena
[8] Moon, T.
[9] Himawan, I.
[10] Sun, S.
[11] Ravuri, S.
[12] Deep Neural Network- Latent Structured Support Vector Machine
[13] Sajjan, S. C., & Vijaya, C.
[14] Shaukat, A.

[DOI: 10.29252/jsdp.17.3.37]

و همکاران معرفی شده است که بدون استفاده از بخش‌های مرسوم سامانه‌های بازشناسی گفتار، مانند مدل‌های تلفظی و مدل مخفی مارکوف و ... گفتار را به‌طورمستقیم آوانگاری می‌کند. از جمله رویکردهای مؤثر در بهبود عملکرد سامانه‌های بازشناسی گفتار، استفاده از فیلترهای خطی چندکاناله است. ونگ[1] و همکاران در [69] به انجام یک مطالعهٔ آزمایشی در زمینهٔ استفاده از فیلترهای خطی در عملکرد خاصی از پردازش گفتار، پرداخته‌اند. چِن و همکاران در [33]، مدلی برای بازشناسی گفتار هم‌پوشانی‌شده تک‌کاناله بی‌نظارت ارائه داده‌اند. لایلیکات[2] و همکاران در [32] به بازشناسی گفتار محاوره‌ای تلفنی در زبان لیتوانیایی پرداخته‌اند. زبان لیتوانیایی دارای ۵۶ واج است. در پژوهش آن‌ها برای جبران کمبود نمونه‌های لازم جهت مدل‌سازی این تعداد واج (زیاد)، از فهرست‌های واجی مختلفی برای مدل‌سازی واج‌های مرکب، دیفتانگ‌ها و رساها، استفاده و نیز تأثیر استفاده از داده‌های وب در مدل‌سازی زبانی و داده‌های صوتی آوانگاری‌نشده برای آموزش نیمه‌نظارتی، بررسی شده است. از آن‌جا که بازشناسی گفتار تحت تأثیر نوفه با دشواری‌هایی همراه است و لازمهٔ کارایی سامانه‌های بازشناسی گفتار در محیط‌های معمولی، دارابودن حدی از قدرت، برای مقابله با درجاتی از نوفه و بازتاب صوت است، بخشی از پژوهش‌های حوزهٔ بازشناسی گفتار، در جهت کاهش نوفه بوده است. در همین راستا، هووای‌یو و اِمِای[3] در [71] روشی را برای کاهش نسبت سیگنال به نوفه پیشنهاد و دقت بازشناسی گفتار را بهبود داده‌اند. بارفوس[4] و همکاران در [70] نیز از روش‌های تقویت طیفی، برای بازشناسی گفتار در محیط‌های ناسازگار با دنیای واقعی استفاده کرده‌اند و با به‌کارگیری یک فیلتر ثانویه، دقت بازشناسی سامانه‌های مبتنی بر یادگیری عمیق را افزایش داده‌اند. یکی دیگر از رویکردها، استفاده از چند میکروفون برای تقویت گفتار است. مُور و همکاران در [10] با استفاده از این روش، تأثیر درجات مختلفی از نوفه و بازتاب را بر قدرت سامانه‌های بازشناسی گفتار، بررسی کرده‌اند. به‌منظور بهبود عملکرد و افزایش دقت سامانه‌های بازشناسی گفتار، استفاده از روش‌های وابسته به بافت، همواره موردتوجه پژوهش‌گران زبان‌های مختلف بوده است. در پژوهش [72] که یکی از پژوهش‌های انجام‌شده در زبان انگلیسی است، برای غلبه بر تغییرات تلفظی ناشی از هم‌تولیدی، از آوانگاری املایی و مجموعه‌ای از مدل‌های آکوستیکی از پیش ساخته‌شده کمک گرفته شده و گزارش شده است که با به‌کارگیری قواعد هم‌تولیدی، دقت سامانه آوانگاری تفصیلی خودکار، نسبت به روش بازشناسی تقویت‌شده افزایش می‌یابد. در زبان‌های پیوندی مانند ترکی، ژاپنی، کره‌ای و ... ، بسیاری از واژه‌ها از ترکیب تکواژها با یکدیگر حاصل و این باعث تلفظ متفاوت واج‌ها در بافت‌های مختلف می‌شود؛ بنابراین واج‌ها ملاک مناسبی برای نمایش تغییرات تلفظی موجود در گفتار پیوسته محسوب نمی‌شوند و نمی‌توان در مدل‌سازی آوایی آن‌ها، از تبدیل سادهٔ نویسه به واج استفاده کرد. جی، جِلین و یان[5] در [36] به‌منظور بازشناسی خودکار گفتار در این زبان‌ها، رویکردی را به نام *اشتقاق خودکار واج‌گونه‌ها*[6]، معرفی کرده‌اند که بدون نیاز به دانش زبان‌شناسی قبلی، با استفاده از یک الگوریتم بی‌نظارت عمل می‌کند. شناسایی واج‌گونه‌ها همچنین نقش به‌سزایی در بهبود عملکرد سنتزگرهای گفتار دارد. ایمِدجی‌داوِبن و هاویسِن[7] در [73] قواعد لازم را برای تبدیلات واجی- واج‌گونه‌ای در زبان عربی، شناسایی و معرفی کرده و سپس به پیاده‌سازی سامانه‌ای برای تولید واج‌گونه‌ها از واج‌ها پرداخته‌اند.

در همین چارچوب، در زبان فارسی، نخستین سامانهٔ بازشناسی گفتار پیوسته با نام *شنوا*، توسط الماس‌گنج و همکاران معرفی شده است. در این سامانه، مدل استفاده‌شده برای بازشناسی، مدلی ترکیبی از شبکه‌های عصبی، قواعد آوایی زبان فارسی و خصوصیات نوایی گفتار فارسی در سطح واج‌هاست و از مدل زبانی آماری از نوع «بایگرمِ دسته‌بندی‌شده[8]» استفاده شده است [80]. هم‌زمان، سامانهٔ بازشناسی گفتار پیوسته فارسی با واژگان بزرگ نیز توسط صامتی و همکاران ارائه شده است. در این طرح، از مدل مخفی مارکوف به‌عنوان مدل اصلی گفتار، استفاده و سامانه بازشناسی گفتار *نویسا ۱* تهیه شده است [38, 39]. برای بهبود دقت بازشناسی گفتار پیوسته، بحرانی و صامتی استخراج و مدل‌سازی واحدهای آوایی وابسته به بافت را پیشنهاد داده‌اند. بر اساس این ایده هر واج به چند نوع گوناگون دسته‌بندی و هر دسته جداگانه مدل‌سازی می‌شود. دسته‌بندی واج‌ها به‌صورت بی‌نظارت و با استفاده از الگوریتم Kمیانگین[9] انجام شده است [23].

[1] Wang, Z.
[2] Lileikyte, R.
[3] Huai YOU, CH. & MA, B.
[4] Barfuss, H.

[5] Ji, X., Jielin, P. & Yan, J.
[6] Automatic allophone deriving (AAD)
[7] Imedjdouben, F. & Houacine, A.
[8] Classified bigram
[9] K-means

[DOI: 10.29252/jsdp.17.3.37]

از حدود دو دهۀ قبل، در عرصۀ بازشناسی گفتار، سامانه‌های متن‌باز متعددی عرضه شده است که در بسیاری از آن‌ها مانند JULIUS [74]، SONIC [75]، SPHINX [76, 77] و KALDI [78]، تغییرات تلفظی ناشی از هم‌تولیدی در نظر گرفته شده و از مدل‌سازی وابسته به بافت واجی، در بازشناسی استفاده شده است. در زمینۀ استفاده از روش‌های وابسته به بافت برای بهبود دقت بازشناسی گفتار در زبان فارسی، سامانۀ متن‌باز کلدی[1] برای استفاده در زبان فارسی، توسط باباعلی مناسب‌سازی و در [25] گزارش آن ارائه شده است. انگیزۀ انتخاب کلدی، نیاز جدی به یک سامانۀ بازشناسی گفتار متن‌باز و به‌روز برای انجام پژوهش‌ها در حوزۀ بازشناسی گفتار فارسی و البته امتیازات این سامانه نسبت به سایر سامانه‌های متن‌باز بوده است.

۴- جزئیات روش پیشنهادی

تاکنون روش‌های مختلفی برای افزایش دقت سامانه‌های بازشناسی گفتار، پیشنهاد شده است. با توجه به این‌که بازشناسی واج‌ها، یکی از مهم‌ترین مراحل در بازشناسی خودکار گفتار است، یکی از روش‌های مرسوم برای افزایش دقت این سامانه‌ها، افزایش دقت بازشناسی واج‌هاست. همان‌گونه که پیش از این اشاره شد، تلفظ هر واج در محیط‌های آوایی مختلف تغییر می‌کند؛ بنابراین لازم است به‌منظور بهبود دقت بازشناسی واج‌ها، به این نکته توجه کرد و بافت گفتار را به‌نحوی در مدل‌سازی دخالت داد. به عبارت دیگر، لازم است به جای مدل‌سازی مستقیم واج‌ها، از مدل‌سازی واحدهای واجی وابسته به بافت استفاده کرد. تجربه نشان داده است که این اطلاعات وابسته به بافت، نقش مهمی در بازشناسی گفتار ایفا می‌کنند و با به‌کارگیری آن‌ها، خطای بازشناسی، به میزان قابل توجهی کاهش می‌یابد. برای مدل‌سازی اثرات بافت گفتار، روش‌های وابسته به بافت متنوعی (مانند دو-آوایی[2]، سه-آوایی[3]، هجا، نیم‌هجا، واحدهای آوایی چندگانه[4] و ...) انجام شده و اثرات آن بر افزایش دقت بازشناسی سنجیده شده است [23].

یکی از روش‌های وابسته به بافت، مدل‌سازی واحدهای آوایی چندگانه است. در این روش، از صورت‌های مختلفی که هر واج در بافت‌های مختلف به خود می‌گیرد، برای مدل‌سازی استفاده می‌شود. رویکردهای مختلفی برای دست‌یابی به این صورت‌های واجی و اعمال آن‌ها در سامانه‌های بازشناسی خودکار گفتار وجود دارد که می‌توان آن‌ها را در دو دستۀ بانظارت[5] و بی‌نظارت[6] قرار داد. در روش بی‌نظارت، از الگوریتم خوشه‌بندی k-means یا سایر الگوریتم‌های خوشه‌بندی و ... استفاده می‌شود. یکی از چالش‌های موجود در روش‌های بی‌نظارت، تعیین تعداد صورت‌های مختلف برای هر واج است [23]. در روش بانظارت، استفاده از الگوریتم‌های دسته‌بندی[7]، کارایی بالایی دارد.

در این پژوهش برای دسته‌بندی حالت‌های گوناگون یک واج، از روش‌های ساختاری مبتنی بر قواعد استخراج‌شده از دانش آواشناسی کمک گرفته شده و انتخاب واج‌گونه‌های مناسب برای هر واج، بر اساس این قواعد انجام شده است.

با استفاده از این قواعد، می‌توان تأثیرات بافت را بر یک واج، بررسی کرد و بر اساس آن، واج را در دستۀ خاصی قرار داد. در غیاب اطلاعات آوایی، راهی برای تعیین تعداد دقیق صورت‌های واجی وجود ندارد؛ بنابراین بهترین راه برای دسته‌بندی این صورت‌ها، روش زبان‌شناختی، و به بیان دقیق‌تر، استفاده از مدل زبانی واج‌گونه‌ای است. استفاده از این قواعد، اجازۀ تعیین نوع و میزان تأثیر هر واج از واج‌های مجاور یا نزدیکش را فراهم می‌کند و به این ترتیب می‌توان برای هر واج، صورت‌های مختلفی در نظر گرفت. به هر یک از صورت‌های استخراج‌شده به این روش، یک واج‌گونه گفته می‌شود. درواقع حضور واج‌ها در بافت‌های مختلف، منجر به پدیدآمدن حالت‌های واجی مختلف می‌شود که هر یک از این حالت‌ها، یک واج‌گونه هستند. به‌کارگیری قواعد آوایی به‌منظور تعیین طبقات واجی مختلف و واج‌گونه‌های متعلق به هر طبقۀ واجی، درواقع یک روش بانظارت است [49]. در پژوهش حاضر، ابتدا از واج‌گونه‌ها به‌عنوان یک ساختار وابسته به بافت مناسب، برای بهبود دقت بازشناسی واج‌های گفتار پیوستۀ فارسی استفاده شده است. در گام‌های بعدی، به‌منظور افزایش هرچه‌بیشتر دقت بازشناسی، سایر روش‌های وابسته به بافت، مانند سه‌آوایی‌ها نیز به کار گرفته شده‌اند.

به‌منظور مدل‌سازی واج‌گونه‌ها جهت استفاده از آن‌ها در بازشناسی گفتار، پیکره‌ای که علاوه‌بر تقطیع در سطح واج (یعنی پیکرۀ واجی)، دارای برچسب‌های واج‌گونه‌ای نیز باشد (یعنی پیکرۀ واج‌گونه‌ای) موردنیاز است. در صورت عدم

[1] Kaldi
[2] Biphone
[3] triphone
[4] Multiple Phone Units
[5] Supervised
[6] Unsupervised
[7] Classification

[DOI: 10.29252/jsdp.17.3.37]

وجود پیکرهٔ واج‌گونه‌ای در یک زبان، می‌توان از یک پیکرهٔ واجی کمک گرفت و آن‌گاه با انجام برخی پردازش‌ها، برچسب‌گذاری واج‌گونه‌ای را بدان افزود. در پژوهش حاضر نیز از پیکرهٔ واجی فارس‌دات کوچک، استفاده و برچسب‌گذاری واج‌گونه‌ای به آن اضافه شده است.

۱-۴- آماده‌سازی پیکرهٔ واج‌گونه‌ای

برای افزودن برچسب‌های واج‌گونه‌ای به پیکرهٔ واجی فارس‌دات کوچک، ضمن درنظرداشتن امکانات فعلی این پیکره، نیاز به مجموعه‌ای از قواعد زبان‌شناختی برای تبدیل واج‌ها به واج‌گونه‌ها در زبان فارسی است. این قواعد تعیین می‌کنند که هر واج با قرارگرفتن در بافت‌های مختلف، چه صورتی را به خود می‌پذیرد و به عبارت دیگر، به چه واج‌گونه‌ای تبدیل می‌شود.

در این مطالعه معیار انتخاب واج‌گونه‌ها، ویژگی‌های تولید نخستین واکه‌ها و هم‌خوان‌ها، به‌علاوهٔ ویژگی‌های تولید دومین واکه‌ها و همخوان‌ها در زبان فارسی و مطابقت‌دادن این ویژگی‌ها با جداول مربوطه در نظام الفبای آوانگاری بین‌المللی[1] (IPA) است.

بر این اساس، با مطالعهٔ منابع موجود در حوزهٔ آواشناسی و واج‌شناسی فارسی و به‌طور خاص، مطالعات واج‌گونه‌ای، به استخراج واج‌گونه‌های فارسی و بافت وقوع هر یک از آن‌ها پرداخته شده است؛ سپس داده‌های حاصل از منابع مختلف با یکدیگر مقایسه و موارد مشترک آن‌ها با هم مطابقت داده شده است. خروجی این مرحله، با جداول IPA مقایسه شده است. در برخی منابع، برای واج‌گونه‌ها نامی متفاوت با نام استاندارد معرفی‌شده در جداول IPA در نظر گرفته شده است. به‌عنوان‌مثال برای برخی واج‌ها، واج‌گونه‌ای به‌نام "بدون‌انجام" یا "تولید ناقص" معرفی شده [3, 10] که با توجه به بافت پیش‌فرض وقوع این واج‌گونه، در فرایند تطبیق‌دهی، معادل با "No audible release" در جداول IPA در نظر گرفته شده و نام "بدون‌رهش یا نارهیده" به آن اختصاص یافته است. به‌دلیل مشابه، واج‌گونهٔ "گرد" در ثمره [3] معادل "Labialized" جداول IPA دانسته شده و نام "لبی‌شده" برای آن به کار رفته و همچنین برای واج‌گونه‌های "خیشومی بدون‌انجام" و "کناری بدون‌انجام" ثمره، بر اساس جداول IPA به‌ترتیب نام‌های "رهش خیشومی[2]" و "رهش کناری[3]" در نظر گرفته شده است. در ادامه و در تعیین واج‌گونه‌های نهایی مختص هر واج، واج‌گونه‌ای به‌نام "سایر" در نظرگرفته شده است و هر یک از واج‌ها در بافتی به‌جز بافت‌های واج‌گونه‌ای مشخص‌شده برای آن‌ها، در این دسته قرار گرفته‌اند و علامت پایهٔ هر واج به آن‌ها اختصاص یافته است. از معرفی واج‌گونه‌هایی که تعریف شرایط وقوع آن‌ها نیازمند امکاناتی فراتر از امکانات فعلی پژوهش بوده، صرف‌نظر شده است تا این واج‌گونه‌ها نیز در دستهٔ سایر قرار گیرند؛ به‌عنوان‌مثال، با توجه به امکانات فعلی دادگان مورداستفاده (دادگان گفتاری زبان فارسی یا همان فارس‌دات کوچک)، تشخیص هجاهای تکیه‌دار ممکن نبوده است؛ بنابراین بافت‌هایی که بر اساس هجاهای تکیه‌دار تعریف می‌شده‌اند، نادیده گرفته شده و واج‌گونه‌های مربوط به آن‌ها، در دستهٔ سایر قرار گرفته‌اند. واج‌گونه‌هایی که برای آن‌ها معادلی در جداول IPA یافت نشده و نیز بافت‌های مربوط به آن‌ها از فهرست اولیهٔ واج‌گونه‌ها حذف شده‌اند و مصادیق آن‌ها در صورت عدم تطابق با دسته‌های واج‌گونه‌ای موجود، در دستهٔ سایر قرار گرفته‌اند. روند استخراج واج‌گونه‌های فارسی، در شکل (۱) نمایش داده شده است.

(شکل-۱): مراحل استخراج واج‌گونه‌های زبان فارسی
(Figure-1): Steps for extraction of Persian allophones

به این ترتیب، طی یک بررسی مقایسه‌ای، فهرستی از واج‌گونه‌های فارسی که با جداول IPA نیز مطابقت داده شده، تهیه شده است. این فهرست شامل ۱۱۳ واج‌گونهٔ معرفی‌شده در ستون سوم جدول (۱) است.

[1] International Phonetic Alphabet

[2] Nasal release
[3] Lateral release

[DOI: 10.29252/jsdp.17.3.37]

(جدول-۱): نمایش دسته‌های واجی و واج‌گونه‌های متعلق به هر دسته

(Table-1): Display phonemic categories and allophones belonging to each category

دسته‌های واجی	واج‌گونه‌های هر دسته	دسته‌های واجی	واج‌گونه‌های هر دسته
p	p , p^h, p^w, p°, p^n, p^l	h	h , h^w
b	b , $b̪$, b°, b^w, b^n, b^l	$t\int$	$t\int^\circ$ $t\int̪$, $t\int^w$,
t	t , $t̪^h$, $t̪^w$, $t̪^\circ$, $t̪^n$, $t̪^l$, t°	dʒ	dʒ, $d̪ʒ̪$, $dʒ^w$, $dʒ^\circ$
d	d, $d̪$, d°, d^w, d^n, d^l, d°	r	r, $r̪$, ɾ, r^w
c	c, k, c^h, c°, k°, c^n, c^l, k^h	l	l, $l̪$, l^w
ɟ	ɟ, $ɟ̪$, $ɟ^\circ$, g°, $ɟ^n$, $ɟ^l$, $g̪$, g	m	m, $m̪$, m^w, ɱ
ɢ	ɢ, $ɢ̪$, $ɢ^w$, $ɢ^\circ$, $ɢ^n$, $ɢ^l$, ʁ	n	n, $n̪$, n^w, $n̪$, $n̪$, $n̪$, ɴ
ʔ	ʔ , $ʔ^\circ$, $ʔ^n$, $ʔ^l$	j	j , $j̪$, j^w
s	s, s^w	i	i , ĩ
z	z , $z̪$, z^w	a	a , ã
ʃ	ʃ, $ʃ^w$	e	e , ẽ
ʒ	ʒ, $ʒ̪$, $ʒ^w$	o	o , õ
f	f , f^w	u	u , ũ
v	v, $v̪$, v^w	ɒ	ɒ , ɒ̃
χ	χ, $χ^w$	ø	ø

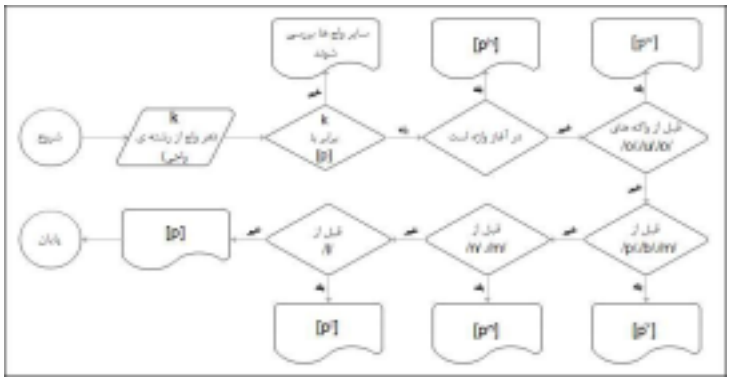
با دست‌یابی به فهرست واج‌گونه‌ها و قواعد واجی تولید آن‌ها در زبان فارسی، تعداد دسته‌های هر یک از واج‌ها به‌صورت بانظارت تعیین و صورت‌های واج‌گونه‌ای ممکن برای هر واج، مشخص شده است. حال با توجه به بافت واجی و زبانی لازم برای تحقق هر یک از واج‌گونه‌های زبان فارسی، پیش‌نیازهای لازم برای شناسایی هر یک از بافت‌های واج‌گونه‌ای، بر روی دادگان فارس‌دات واجی، اعمال شده است. به‌عنوان مثال، برای اعمال قواعد زبانی تولید واج‌گونه‌ها از واج‌ها و برچسب‌گذاری واج‌گونه‌ای دادگان واجی فارس‌دات، مشخص‌بودن مواردی مانند موقعیت واج در هجا، موقعیت واج در واژه، واج‌های مجاور واج هدف و ... الزامی است؛ بنابراین قبل از اعمال این قواعد، برخی قابلیت‌ها مانند تقطیع هجایی، به دادگان واجی فارس‌دات افزوده شده [24] و موقعیت هر واج در هجا و همچنین در واژه، تعیین شده است؛ سپس با پیاده‌سازی قواعد آوایی استخراج‌شده برای انجام تبدیلات واجی-واج‌گونه‌ای، برنامهٔ تولیدکنندهٔ واج‌گونه‌ها آماده شده است. این برنامه، واج‌ها را به‌عنوان ورودی دریافت و با تکیه بر قواعد یادشده، آن‌ها را به واج‌گونه تبدیل می‌کند. به‌منظور آشنایی بیشتر با روند تولید واج‌گونه‌های واج‌های مختلف و چگونگی توصیف وقوع واج‌گونه‌ها با استفاده از قواعد واجی، در شکل (۲) واج /p/ به‌عنوان نمونه بررسی شده است. برای /p/ شش واج‌گونه شامل [p^h]، [p^w]، [p°]، [p^n]، [p^l] و [p]، در نظر گرفته شده است. واج‌گونهٔ [p]، مربوط به حالتی است که هیچ‌یک از پنج موقعیت دیگر، برای /p/ برقرار نباشد. برای استخراج واج‌گونه‌های سایر واج‌ها نیز چنین الگوریتمی متناسب با هر واج، پیاده‌سازی شده است. با اجرای این برنامه بر روی پیکرهٔ فارس‌دات واجی، تمامی واج‌های موجود در این پیکره، در سطح واج‌گونه‌ای برچسب‌گذاری شده‌اند و پیکرهٔ فارس‌دات واج‌گونه‌ای تولید شده است. از این پیکره، برای آموزش مدل‌های آکوستیکی و بازشناسی در سطح واج‌گونه‌ها در یک سامانهٔ متن‌باز بازشناسی گفتار با نام کلدی[1] (که در بخش ۲-۴ به آن خواهیم پرداخت،) استفاده شده است. مدل‌سازی آکوستیکی هم بر مبنای تک‌آوایی‌ها[2] و هم بر مبنای سه‌آوایی‌ها بر روی پیکرهٔ واج‌گونه‌ای (متشکل از ۱۱۳ واج‌گونه) صورت گرفته و بازشناسی واج‌گونه‌ها بر اساس هر دو مدل تک‌آوایی و سه‌آوایی و با به‌کارگیری الگوریتم‌های مختلف، انجام شده است. در مدل‌سازی سه‌آوایی، برای هر نوع از سه‌آوایی‌های پیکره واج‌گونه‌ای، یک مدل جداگانه استخراج و از این مدل‌ها در فرایند بازشناسی رشتهٔ واج‌گونه‌ای گفتار پیوسته استفاده شده است. به این ترتیب و با استفاده از مدل‌سازی سه‌آوایی‌های واج‌گونه‌ای، بازشناسی رشتهٔ واج‌گونه‌ای گفتار پیوسته، با دقت بالاتری نسبت به مدل‌سازی تک‌آوایی‌های واج‌گونه‌ای انجام می‌شود.

پس از بازشناسی رشتهٔ واج‌گونه‌ای (به‌تفکیک مدل‌سازی تک‌آوایی یا سه‌آوایی)، برای بررسی تأثیر بازشناسی واج‌گونه‌ها بر بازشناسی واج‌ها، هر یک از واج‌گونه‌های شناسایی‌شده در سامانهٔ کلدی، به دستهٔ واجی خاصی مطابق با جدول (۱)، نسبت داده شده و برچسب

[1] Kaldi
[2] monophones

[DOI: 10.29252/jsdp.17.3.37]

واجی گرفته‌اند. به عبارت دیگر، تمام ۱۱۳ واج‌گونهٔ شناسایی‌شده توسط سامانهٔ بازشناسی گفتار، درنهایت به سی دسته (شامل ۲۹ دستهٔ واجی و یک دسته مربوط به مکث‌ها و سکوت‌ها) نگاشت شده‌اند. بنابراین می‌توان گفت، شناسایی واج‌ها از مسیر شناسایی واج‌گونه‌ها، منجر به افزایش دقت بازشناسی شده است.



(شکل-۲): الگوریتم وقوع واج‌گونه‌های مختلف /p/

(figure-2): The algorithm of occurrence of different allophones for /p/

به‌طور معمول، هر واج‌گونه تنها به یک دستهٔ واجی تعلق دارد؛ اما گاهی در مواردی که واج‌گونه‌ای در چندین دستهٔ واجی قرار داشته باشد، به‌منظور جلوگیری از بروز مشکل در روند بازشناسی، لازم است که هر واج‌گونه تنها به یک گروه واجی نسبت داده شود؛ بنابراین در این مرحله هر واج‌گونه تنها به دستهٔ واجی‌ای که شباهت بیشتری با آن دارد، و از نظر طبیعت آوایی به آن نزدیک‌تر است، نسبت داده شده و نتیجه در قالب جدول (۱) تنظیم شده است. به‌عنوان مثال، واج‌گونهٔ [t4] هم به دستهٔ واج /t4/ و هم به دستهٔ واج /d4/ تعلق دارد؛ یعنی هم واج‌گونه‌ای از /t4/ و هم واج‌گونه‌ای از /d4/ محسوب می‌شود؛ اما به‌طورطبیعی این واج‌گونه در درجهٔ نخست، به واج /t4/ نسبت داده می‌شود.

به این ترتیب، شناسایی واج‌ها از مسیر واج‌گونه‌ها صورت گرفته است. با استفاده از این روش، دقت بازشناسی واج‌ها در زبان فارسی، بهبود قابل‌توجهی داشته است.

۲-۴ سامانهٔ بازشناسی گفتار کلدی

کلدی یک سامانهٔ متن‌باز برای بازشناسی گفتار است که به زبان C++ توسعه داده شده و در سال ۲۰۰۹ توسط دانشگاه جان هاپکینز[1] معرفی شده است. کتابخانهٔ کلدی، مدل‌سازی ابعاد مختلف آوایی، مدل‌سازی آکوستیکی با مدل‌های مختلفی مانند مدل مخلوط گوسی زیرفضا[2] و مدل‌های مخلوط گوسی[3] را پشتیبانی می‌کند. اجزای بازشناسی گفتارِ این سامانه عبارتند از: استخراج ویژگی، مدل‌سازی آکوستیکی، درخت‌های تصمیم آوایی، مدل‌سازی زبانی و رمزگشاها. در سامانهٔ کلدی، به‌منظور استخراج ویژگی، به‌صورت پیش‌فرض از ویژگی‌های MFCC و PLPاستفاده می‌شود. برای محاسبهٔ MFCC، ابتدا انجام برخی پیش‌پردازش‌ها لازم است؛ بنابراین قبل از هر چیز سیگنال گفتار، قاب‌بندی می‌شود (به‌طورمعمول طول قاب‌ها ۲۵ میلی‌ثانیه و هم‌پوشانی میان آن‌ها ۶۰٪ در نظر گرفته می‌شود)؛ سپس از هر قاب، داده‌های موردنیاز استخراج و عامل پیش‌تأکید بر آن اعمال و آن‌گاه در یکی از توابع پنجره‌گذاری مانند همینگ[4] ضرب می‌شود. پس از آن با انجام برخی پردازش‌های دیگر برای هر قاب (مانند: محاسبهٔ انرژی، تبدیل فوریه و محاسبهٔ طیف قدرت، محاسبهٔ انرژی در هر مِل، محاسبهٔ لگاریتم انرژی‌ها و به‌دست‌آوردن تبدیل کسینوسی) ضرایب کپسترال فرکانسی به‌دست می‌آید. درقبل اشاره شد که مدل‌سازی آکوستیکی در کلدی با به‌کارگیری مدل‌های مخلوط گوسی، مدل‌های مخلوط گوسی زیرفضا و شبکه‌های عصبی عمیق انجام می‌شود. تطبیق گوینده[5] و سایر تبدیل‌های خطی مانند تبدیل خطی بیشینه شباهت[6] نیز با استفاده از برخی رده‌ها اعمال می‌شوند. این تطبیق، بر اساس مدل، با استفاده از رگرسیون خطی بیشینه شباهت[7] و بر اساس ویژگی، با استفاده از رگرسیون خطی بیشینه شباهت مبتنی بر ویژگی[8]، پشتیبانی

[1] Johns Hopkins University
[2] Gaussian Mixture Models (GMMs)
[3] Subspace Gaussian Mixture Models (SGMMs)

[4] Hamming
[5] Speaker adaptation
[6] maximum likelihood linear transform (MLLT)
[7] maximum likelihood linear regression (MLLR)
[8] feature-space MLLR (fMLLR)

[DOI: 10.29252/jsdp.17.3.37]

می‌شود. این سامانه همچنین با هنجارسازی طول تارهای صوتی[1] و هنجارسازی جنسیت[2]، به هنجارسازی گوینده می‌پردازد. برای آموزش تطبیقی گوینده[3]، امکان استفاده از هر دو روش VTLN و fMLLR وجود دارد که استفاده از fMLLR متداول‌تر است. در کلدی کد ساخت درخت تصمیم آوایی، به این صورت است که به‌ازای هر حالت HMM مربوط به هر تک‌آوایی، یک درخت تصمیم وجود دارد که بر مبنای سؤالاتی در مورد آواهای سمت چپ و راست آوای هدف، تصمیم‌گیری می‌کند. سؤالات می‌توانند بر اساس دانش زبان‌شناسی طراحی شوند یا این‌که به‌صورت خودکار، بر مبنای خوشه‌بندی درختی آواها[4] تولید شوند. این سؤالات اغلب در زمینهٔ مواردی مانند تکیه، جایگاه آوا در واژه و ... هستند [25].

۵- آزمایش‌ها، تحلیل‌ها و نتایج

فعالیت‌های صورت‌گرفته در این پژوهش و نتایج حاصل از بررسی‌ها، در سه زیربخش، خلاصه شده‌اند. در ۵-۱، به معرفی اجمالی پیکرهٔ فارس‌دات واجی پرداخته شده است. در ۵-۲، چگونگی به‌کارگیری پیکرهٔ فارس‌دات واج‌گونه‌ای در سامانهٔ کلدی و در ۵-۳، نتایج حاصل از بازشناسی واج‌گونه‌ها توسط این سامانه گزارش شده است.

۵-۱- پیکرهٔ فارس‌دات واجی

در پژوهش حاضر، برای مدل‌سازی آکوستیکی، از پیکرهٔ گفتاری فارس‌دات کوچک استفاده شده است. پیکرهٔ فارس‌دات کوچک، نخستین پیکرهٔ گفتاری استاندارد در زبان فارسی است که با هدف مطالعه مبانی و مدل‌سازی آکوستیکی زبان فارسی به‌منظور استفاده در پروژه‌های پردازش گفتار، توسط پژوهشگاه توسعه فناوری‌های پیشرفته خواجه نصیر طوسی تولید شده است. این پیکره متشکل از ۶۰۸۰ جمله است که توسط ۳۰۴ گویندهٔ فارسی‌زبان، با یکی از ده لهجهٔ رایج تهرانی، ترکی، اصفهانی، شمالی، جنوبی، لری، کردی، یزدی، خراسانی و بلوچی، با سن، جنسیت و میزان تحصیلات مختلف است [79]. استفاده از دادگان دارای گویش‌های رایج یک زبان، باعث می‌شود که برنامهٔ تهیه‌شده، قابلیت استفاده را برای گویش‌های مختلف آن زبان داشته باشد [81]. نسبت گویندگان مرد به زن، دو به یک است. هر گوینده، بیست جمله را در دو جلسهٔ مجزا در یک اتاق آکوستیک به سبک خوانداری رسمی بیان کرده است. تمام جملات این پیکره، در سطح واژه و واج، تقطیع و برچسب‌گذاری شده‌اند و به همین علت، در پژوهش حاضر، از آن به‌عنوان پیکرهٔ واجی نیز یاد شده است. سرعت متوسط گفتار گویندگان، ۴/۲ هجا در ثانیه است. جملات به‌صورت شانزده بیتی با فرکانس نمونه‌برداری ۲۲۰۵۰ هرتز و با نسبت سیگنال به نوفه[5] متوسط ۳۲ دسی‌بل ضبط شده‌اند. تعداد کل جملات فارس‌دات ۴۰۵ جمله است که دو جملهٔ آن، شامل کل واج‌های زبان فارسی (به‌جز /f/) است و این دو جمله توسط همهٔ گویندگان بیان شده است. درواقع جملات به‌گونه‌ای انتخاب شده‌اند که شامل تمامی دوآوایی‌های رایج زبان فارسی باشند [79].

یکی از محدودیت‌های پژوهش حاضر این است که منابع به‌کاررفته برای استخراج واج‌گونه‌ها، بر فارسی معیار متمرکز بوده‌اند؛ درحالی‌که داده‌های تنها پیکره گفتاری قابل استفاده در این پژوهش یعنی فارس‌دات کوچک، برگرفته از گویش‌های مختلف و داده‌های مربوط به فارسی معیار در این پیکره، حدود یک دهم کل پیکره است؛ که این میزان داده، به‌تنهایی کفایت لازم را برای مدل‌سازی ندارد. از طرفی از آن‌جا که سبک گفتاری مورد استفاده در پیکرهٔ فارس‌دات کوچک، سبک خوانداری رسمی است و درواقع، جملات بیان‌شده، روخوانی جملات ازپیش‌نوشته‌شده است، تصور می‌شود که تفاوت‌های گویشی و لهجه‌ای در این پیکره به کمینه برسد. بنابراین مدل‌سازی بر مبنای کل پیکره انجام شده است.

۵-۲- استفاده از پیکره فارس‌دات واج‌گونه‌ای در سامانهٔ کلدی

در ارزیابی‌های متداول در حوزهٔ بازشناسی گفتار، سه مجموعهٔ آموزش[6]، توسعه[7] و ارزیابی یا آزمون[8] برای دادگان گفتاری در نظر گرفته می‌شود. بر مبنای مجموعهٔ آموزش، مدل‌های آکوستیکی واحدهای آوایی آموزش داده می‌شوند. مجموعهٔ توسعه برای تنظیم و بهینه‌سازی پارامترها ازجمله وزن مدل زبانی استفاده می‌شود؛ درنهایت، بر اساس مدل‌های آموزش‌داده‌شده و پارامترهای تنظیم‌شده، دقت، بر روی مجموعهٔ آزمون، با فرض مشخص و ثابت‌شدن پارامترها

[1] Vocal Tract Length Normalization
[2] gender normalization
[3] speaker adaptive training (SAT)
[4] tree-clustering of the phones

[5] Signal to noise ratio
[6] Training Set
[7] Development Set
[8] Evaluation or Test Set

[DOI: 10.29252/jsdp.17.3.37]

بر مبنای مجموعهٔ توسعه گزارش می‌شود. سه مجموعهٔ یادشده، در [25] برای دادگان فارس‌دات تعریف شده است. در پژوهش حاضر برای آموزش مدل‌های آکوستیکی، تعداد ۳۹۹۴ جمله گفته‌شده توسط ۲۲۴ گوینده، به‌عنوان مجموعهٔ آموزش، تعداد ۴۷۵ جمله از پنجاه گوینده دیگر، به‌عنوان مجموعهٔ توسعه و تعداد ۲۸۷ جمله مربوط به سی گوینده متفاوت با گویندگان دو مجموعه قبل، برای مجموعهٔ آزمون، در نظر گرفته شده است.

در کلیه آزمایش‌ها، مدل‌سازی آکوستیکی واحدهای آوایی مبتنی بر HMM-GMM و HMM-SGMM بوده است. برای مدل‌سازی هر واحد آوایی (واج یا سه‌آوایی) از یک مدل مخفی مارکوف چپ به راست[1] با سه حالت استفاده شده است که تعداد گوسی‌های هر حالت طبق روند خودکار شکست-ادغام[2] حین آموزش مشخص می‌شود. در مدل‌سازی سه‌آوای از روش گره‌زدن[3] حالت‌ها به‌کمک درخت تصمیم[4] آوایی بهره گرفته شده است.

۳-۵- نتایج

در این پژوهش، فرایند بازشناسی بر روی مجموعهٔ آزمون، اعمال شده و نتایج مختلفی به‌ازای روش‌های مختلف موجود برای استخراج ویژگی و مدل‌سازی آکوستیکی در سامانهٔ کلدی، به‌دست آمده است. پیش از این نیز، در مطالعه‌ای مشابه در [25]، باباعلی به محاسبهٔ دقت بازشناسی واج به‌صورت مستقیم و با مدل‌سازی انجام‌شده بر اساس پیکرهٔ فارس‌دات واجی (بدون استفاده از پیکرهٔ فارس‌دات واج‌گونه‌ای) با استفاده از سامانهٔ کلدی پرداخته است. نتایج حاصل از پژوهش یادشده، در ستون دوم جدول (۲) گزارش شده است. در نخستین مرحله از آزمایش‌های انجام‌شده بر روی مجموعهٔ آزمون، به منظور استخراج ویژگی، برای هر قاب، سیزده ضریب مل-کپستروم و مشتق زمانی نخست و دوم این ضرایب، به‌دست آمده است. حاصل این کار، تولید یک بردار ویژگی ۳۹ بعدی به‌ازای هر قاب است. در این مرحله همچنین مدل‌سازی آکوستیکی با استفاده از مدل مخفی مارکوف و بر مبنای تک‌آوایی‌ها صورت گرفته است؛ بدین‌معنا که هر یک از واج‌ها به‌صورت مستقل مدل‌سازی شده و مرزهای بین واج‌ها در مدل‌سازی دخالت داده نشده‌اند. بر این اساس، نرخ خطای بازشناسی محاسبه‌شده که قبل از اعمال رویکرد واج‌گونه‌ها ۳۰/۵ درصد بوده، بعد از اعمال این رویکرد، به ۲۹/۵ درصد رسیده است. در دومین آزمایش، فقط مدل‌سازی از تک‌آوایی به سه‌آوایی تغییر کرده است. طبق انتظار، بهبود قابل‌توجهی در کاهش خطا، حاصل شده، به‌نحوی‌که میزان خطای بازشناسی واج‌ها به ۲۳/۳ درصد، کاهش یافته است. میزان خطای گزارش‌شده برای این مرحله، قبل از به‌کارگیری رویکرد واج‌گونه‌ها، ۲۵/۱ درصد بوده است. در آزمایش سوم، در مرحلهٔ استخراج ویژگی، بر روی هر بردار ویژگی مل-کپستروم سیزده مؤلفه‌ای به همراه سه بردار ویژگی قبل و بعد که تشکیل یک بردار ویژگی ۹۱ مؤلفه‌ای می‌دهد، دو تبدیل خطی LDA[5] و MLLT[6] به‌ترتیب اعمال و یک بردار چهل‌مؤلفه‌ای حاصل شده و مدل‌سازی سه‌آوایی بر مبنای آن صورت گرفته است. با این روش، خطای بازشناسی واج، که تا قبل از به‌کارگیری رویکرد واج‌گونه‌ها ۲۳/۳درصد بوده، به ۲۱/۸ درصد رسیده است. با مقایسهٔ نتایج این مرحله و مرحلهٔ قبل، این نتیجه حاصل شده که ویژگی MFCC به‌علاوهٔ تبدیلات خطی LDA و MLLT، نسبت به ویژگی MFCC به‌علاوهٔ مشتقات زمانی نخست و دوم آن، در کاهش خطای بازشناسی واج، موفق‌تر بوده است؛ بنابراین در مراحل بعدی، به جای مل-کپستروم و مشتقات نخست و دوم آن، از این ویژگی استفاده شده است. در آزمایش چهارم، در دو مرحله از روش fMLLR[7] استفاده شده است. یکی در مرحلهٔ آموزش، که با استفاده از این روش، آموزش تطبیقی به گوینده (SAT[8]) صورت گرفته و دیگری در مرحلهٔ بازشناسی که با روش fMLLR، ویژگی‌های هر گوینده مطابقت داده شده است. به‌کارگیری این رویکرد، باعث افزایش ۲/۶ درصدی دقت شده و خطای بازشناسی واج را از ۲۱/۸درصد قبل از اعمال رویکرد واج‌گونه‌ها به ۱۹/۲درصد بعد از اعمال این رویکرد، رسانده است. به‌کارگیری مدل‌سازی SGMM[9] در آزمایش پنجم، باز هم موجب بهبود ۱/۲درصدی دقت شده است. خطای بازشناسی واج، قبل از رویکرد واج‌گونه‌ها، در این آزمایش، ۱۹/۹درصد بوده که بعد از آن به هجده‌درصد رسیده است. اعمال آموزش تمایزی MMI[10] در ششمین آزمایش، کاهش دقت را به‌همراه داشته که نتایج مربوط به قبل و بعد از اعمال رویکرد واج‌گونه‌ها برای این مرحله نیز،

[1] Left-to-right
[2] Split and Merge
[3] Tying
[4] Decision Tree

[5] linear discriminant analysis
[6] maximum likelihood linear transform
[7] Feature-space Maximum Likelihood Linear Regression
[8] Speaker Adaptive Training
[9] Sub-space Gaussian Mixture Model
[10] Maximum mutual information

[DOI: 10.29252/jsdp.17.3.37]

در جدول (۲) نمایش داده شده است. همان‌گونه که ملاحظه می‌شود، بالاترین میزان دقت گزارش‌شده، قبل از به‌کارگیری رویکرد واج‌گونه‌ها، برابر با ۸۰/۲درصد (مربوط به خطای ۱۹/۸درصد) و بعد از اعمال رویکرد واج‌گونه‌ها، برابر با ۸۲درصد (مربوط به خطای ۱۸درصد) است. مقایسهٔ داده‌های ستون دوم و چهارم جدول (۲)، نشان‌دهندهٔ بهبود ۱/۸درصدی دقت بازشناسی واج، با استفاده از رویکرد واج‌گونه‌هاست. از جمله دلایل این بهبود دقت را می‌توان جزئی‌ترشدن واحدهای مورد بازشناسی (یعنی استفاده از واج‌گونه‌ها به جای واج‌ها) و در نظر گرفته‌شدن تغییرات واج‌ها در بافت‌های مختلف دانست که خود نتیجهٔ مستقیم مدل‌سازی واج‌گونه‌هاست. میزان دقت بازشناسی گفتار فارسی، تا قبل از به‌کارگیری این رویکرد نیز بسیار بالا بوده است و طبیعتاً بهبوددادن این دقت بالا، دستاوردی مهم و قابل‌توجه در حوزهٔ بازشناسی گفتار تلقی می‌شود. در این بخش، ضمن گزارش نتایج مربوط به هر یک از روش‌ها، به مقایسهٔ نتایج قبل و بعد از به‌کارگیری رویکرد واج‌گونه‌ها در هر روش پرداخته می‌شود.

(جدول-۲): درصد خطای بازشناسی واج، قبل و بعد از به‌کارگیری رویکرد واج‌گونه‌ها

(Table-2): Percent of phoneme error rate (PER) before and after using allophones approach

روش مدل‌سازی و استخراج ویژگی	قبل از به‌کارگیری رویکرد واج‌گونه‌ها		بعد از به‌کارگیری رویکرد واج‌گونه‌ها	
	آزمون	توسعه	آزمون	توسعه
+ HMM Mono-phone + MFCC − delta + delta-delta	30.5	29.8	29.5	29.3
Tri-phone + MFCC + delta +delta-delta −HMM	25.1	25.8	23.3	23.5
Tri-phone + MFCC + LDA+MLLT+HMM	23.3	24.8	21.8	21.7
Tri-phone + MFCC + LDA+MLLT+SAT+HMM	21.8	21.2	19.2	19.4
Tri-phone + MFCC + LDA+MLLT+SAT+SGMM	19.9	20.2	18.0	17.9
Tri-phone + MFCC + LDA+MLLT+SAT+SGMM+MMI	19.8	20.3	18.4	18.4

۶- نتیجه‌گیری

در این مقاله به بازشناسی واج‌های فارسی با استفاده از یک روش ساختاری مبتنی بر قواعد استخراج‌شده از دانش زبان‌شناسی پرداخته شده و درواقع از مدل زبانی واج‌گونه‌ای برای بهبود دقت بازشناسی واج‌ها استفاده شده است. بدین‌منظور ابتدا قواعد تولید واحدهای آوایی وابسته به بافت (واج‌گونه‌ها) در زبان فارسی استخراج و سپس بافت وقوع هر یک از این واج‌گونه‌ها بر روی پیکرهٔ واجی فارس‌دات، تعریف و پیاده‌سازی شده است. حاصل این فرایند، تولید پیکرهٔ واج‌گونه‌ای فارس‌دات است. از تقسیمات مربوط به مجموعه‌های آموزش، توسعه و ارزیابی در [25] برای پیکرهٔ واج‌گونه‌ای فارس‌دات نیز استفاده شده و با به‌کارگیری سامانهٔ بازشناسی گفتار کلدی برای این مجموعه‌ها، مدل‌سازی و بازشناسی واج‌گونه‌های فارسی انجام شده است. خروجی این مرحله، واج‌گونه‌های بازشناسی‌شده است. در گام بعدی، واج‌های زبان فارسی به‌عنوان دسته‌های واجی در نظر گرفته شده و هر یک از واج‌گونه‌های یادشده، به تنها یکی دسته‌های واجی اختصاص داده شده است؛ بدین‌ترتیب تعداد واج‌گونه‌های اختصاص‌یافته به دسته‌های واجی، به‌الزام یکسان نیست. با استفاده از این روش، بهترین میزان خطای بازشناسی واج در زبان فارسی، در مجموعهٔ توسعه، ۱۷/۹درصد و در مجموعهٔ آزمون، هجده‌درصد بوده است و با توجه به این که بهترین میزان خطای بازشناسی گزارش‌شده برای واج تا قبل از این پژوهش، ۲۰/۳درصد برای مجموعهٔ توسعه و ۱۹/۸درصد برای مجموعهٔ آزمون بوده، با به‌کارگیری این روش، دقت بازشناسی واج‌ها، با ۱/۸ درصد بهبود، به رقم ۸۲ درصد رسیده است.

۷ - منابع

[۱] بی‌جن‌خان، محمود، واج‌شناسی نظریهٔ بهینگی، تهران، انتشارات سمت، ۱۳۸۴.

[1] M. Bijankhan, "The phonology of optimality theory", Tehran: Samt, 2005.

[۲] بی‌جن‌خان، محمود، نظام آوایی زبان فارسی، تهران، انتشارات سمت، ۱۳۹۲.

[2] M. Bijankhan, "Phonetic system of Persian language", Tehran: Samt, 2013.

[۳] ثمره، یدالله، آواشناسی زبان فارسی (ویرایش دوم)، تهران، مرکز نشر دانشگاهی، صفحات ۳۷-۷۹، ۱۳۷۸.

[3] Y. Samareh, "Phonetics of Persian language", Tehran: Academic publishing center, 1999.

[۴] حق‌شناس، علی‌محمد، آواشناسی(فونتیک)، تهران، انتشارات آگه، صفحات ۶۹-۱۳۱، ۱۵۶، ۱۵۹، ۱۳۹۲.

[DOI: 10.29252/jsdp.17.3.37]

[4] A. M. Haghshenas, "Phonetic", Tehran: Agah, 2013.

[۵] دیهیم، گیتی، درآمدی بر آواشناسی عمومی، تهران، انتشارات دانشگاه ملی ایران «۱۵۹»، ۱۳۵۸.

[5] G. Deihaim, "An introduction to General Phonetics", Tehran: National university of Iran, 159, 1979.

[۶] سپنتا، ساسان، آواشناسی فیزیکی زبان فارسی، اصفهان، انتشارات گل‌ها، ۱۳۷۷.

[6] S. Sepanta, "Acoustic phonetics of Persian language", Isfahan: Golha, 1998.

[۷] علی‌نژاد، بتول و حسینی بالام، فهیمه، مبانی آواشناسی آکوستیکی، اصفهان، انتشارات دانشگاه اصفهان، ۱۳۹۲.

[7] B. Alinejad, F. Hosseini Balam, Fundamentals of acoustic phonetics", Isfahan: university of Isfahan, 2013.

[۸] علی‌نژاد، بتول، مبانی واج‌شناسی، اصفهان، انتشارات دانشگاه اصفهان، ۱۳۹۵.

[8] B. Alincjad, Fundamentals of phonology, Isfshsn: university of Isfahan, 2016.

[۹] کرد زعفرانلو کامبوزیا، عالیه، واج‌شناسی رویکردهای قاعده بنیاد، تهران، انتشارات سمت، ۱۳۹۲.

[9] A. Kodr Zafaranloo Kambozia, "phonology rule-based approach", Tehran: Samt, 2013.

[۱۰] مدرسی قوامی، گلناز، آواشناسی: بررسی علمی گفتار، تهران، انتشارات سمت، صفحه ۷۲، ۱۳۹۰.

[10] G. Modarresi Ghavami, "Phonetics: The scientific study of speech", Tehran: Samt, 2011.

[۱۱] مشکوةالدینی، مهدی، ساخت آوایی زبان (ویرایش سوم)، مشهد، انتشارات دانشگاه فردوسی مشهد، صفحه ۱۳۱، ۱۳۸۸.

[11] M. Meshkato Dini, "The sound pattern of language (third edition)", Mashhad: Ferdowsi University of Mashhad, p. 131, 2009.

[۱۲] یارمحمدی، لطف‌الله، درآمدی به آواشناسی، تهران، مرکز نشر دانشگاهی، ۱۳۶۴.

[12] L. Yarmohammadi, "An iIntroduction to phonetics", Tehran: university publication center, 1985.

[۱۳] میرسعیدی، عاطفه‌سادات، "بررسی صوت‌شناختی فرایندهای واجی همگونی و ناهمگونی در فارسی محاوره"، پایان‌نامه‌ی دکتری، دانشگاه اصفهان، اصفهان، ۱۳۹۰.

[13] A. S. Mirsaeidi, "phonetic study of phonological process assimilation and dissimilation in Persian", PH.D dissertation, fgn, Isf., Isfahan., 2011.

[۱۴] بی‌جن‌خان، محمود، "نظام واج‌گونه‌های زبان فارسی در چارچوب نظریه‌ی واج‌شناسی تولیدی"، مجله‌ی دانشکده‌ی ادبیات و علوم انسانی دانشگاه تهران، زمستان، صفحات ۹۵- ۱۱۷، ۱۳۷۹.

[14] M. Bijankhan, "Persian allophones system in the framework of articulatory phonemics theory", *Journal of the faculty of literature and humanities*, winter, pp. 95-117, 2001.

[۱۵] زاهدی، کیوان و فخاریان، فیضیه، "همگونی همخوان‌ها در زبان فارسی نوین: رویکرد واج‌شناسی هندسه‌ی مشخصه‌ها"، مجله‌ی پژوهش‌های زبان‌شناسی دانشگاه اصفهان، پاییز و زمستان، شماره‌ی ۲، صفحات ۴۷- ۶۴، ۱۳۹۰.

[15] K. Zahedi, F. Fakharian, " Consonantal Assimilation in Modern Persian: A Feature Geometry Approach", *Journal of researches in linguistics, autumn-winter*, Issue 2, pp. 47-64, 2010.

[۱۶] صادقی، وحید، "تأثیر دمش بر تقابل واک‌داری- بی‌واکی انسدادی‌های فارسی"، مجله‌ی زبان و زبان‌شناسی. صفحات ۶۵- ۸۴، ۱۳۸۶.

[16] V. Sadeghi, "The effect of aspiration on Persian stop voicing contrast", *Journal of language and linguistics*, pp. 65-84, 2007.

[۱۷] صادقی، وحید، "آواشناسی و واج‌شناسی همخوان‌های چاکنایی"، مجله‌ی پژوهش‌های زبان‌شناسی دانشگاه اصفهان، بهار و تابستان، شماره‌ی ۲، صفحات ۴۹- ۶۲، ۱۳۸۹.

[17] V.Sadeghi, "The phonetics and phonology of Persian glottal consonants", *Journal of researches in linguistics, spring and summer, issue 2*, pp. 49-62, 2010.

[۱۸] علی‌نژاد، بتول، "واک‌داری و دمش در زبان فارسی بر اساس نظریه‌ی واج‌شناسی حنجره‌ای"، فصل‌نامه‌ی علمی-پژوهشی پژوهش‌های زبان‌شناسی دانشگاه اصفهان، بهار و تابستان، شماره‌ی ۲، صفحات ۶۳- ۸۰، ۱۳۸۹.

[DOI: 10.29252/jsdp.17.3.37]

[18] B. Alinejad, "Persian aspiration and voicing in laryngeal phonology", *Journal of researches in linguistics, spring and summer*, issuc 2, pp. 63-80, 2010.

[۱۹] مدرسی قوامی، گلناز، "خنثی‌شدگی تقابل همخوان‌های انسدادی واک‌دار و بی‌واک در زبان فارسی"، مجموعه مقالات دانشگاه علامه طباطبایی، شماره‌ی ۲۱۹، صفحات ۴۴۱- ۴۵۴، ۱۳۸۶.

[19] G. Modarresi Ghavami, "Neutralization of contradiction between voiced and unvoiced stop consonants in Persian language", *Journal of proceeding of Allameh Tabatabaee university,* issuc 219, pp. 441-454, 2007.

[۲۰] نوربخش، ماندانا، "همخوان ملازی در فارسی معیار"، فصل‌نامه‌ی علمی- پژوهشی زبان‌پژوهی دانشگاه الزهرا، تابستان، شماره‌ی ۱۵، صفحات ۱۵۱- ۱۷۰، ۱۳۹۴.

[20] M. Norbakhsh, "uvular consonants in standard Persian", *Journal of language research Zabanpazhuhi*, issuc 15, pp. 151-170, 2015.

[۲۱] شریفی آتشگاه، مسعود و صادقی، وحید، "طراحی الگوریتم بازشناسی واج‌ها با به‌کارگیری همبسته‌های آکوستیکی مشخصه‌های واجی"، فصل‌نامه‌ی علمی- پژوهشی پردازش علائم و داده‌ها، شماره‌ی ۱۶، صفحات ۱۳-۲۸، ۱۳۹۰.

[21] M. Sharifi, V. Sadeghi, "phoneme recognition algorithm design using the acoustic correlates of the phonological features", *Journsl of signal and data processing*, Vol. 2 (SERIAL 16), pp. 13-28, 2011.

[۲۲] الماس‌گنج، فرشاد، سیدصالحی، سید علی و بیجن‌خان، محمود، "نرم‌افزار بازشناسی گفتار پیوسته فارسی: شنوا۲"، اولین کارگاه پژوهشی زبان فارسی و رایانه، صفحات ۷۷-۸۲، تهران، ۱۳۸۳.

[22] F. Almasganj, SA. Seyyed Salehi, M. Bijankhan, "Shenava 2: a Persian continuous speech recognition softwarc", *in the first workshop on Persian language and computer*, pp. 77-82, Tehran, 2004.

[۲۳] صامتی، حسین و بحرانی، محمد، "استخراج و مدل‌سازی واحدهای آوایی وابسته به بافت برای بهبود دقت بازشناسی گفتار پیوسته با روش دسته‌بندی واج‌ها"، نشریه مهندسی برق و مهندسی کامپیوتر ایران، سال ۳، شماره ۱، تهران، ۱۳۸۴.

[23] H. Sameti, M. Bahrani, "Extraction and modeling context dependent phone units for improvement of continuous speech recognition accuracy by phonemes clustering", *Journal of electrical engineering and computer engineering of Iran*, spring-summer, year 3, No. 1, pp. 45-51, 2005.

[۲۴] احمدی، طاهره، کارشناس، حسین، علی‌نژاد، بتول و نقوی راوندی، مصطفی، " تقطیع هجایی خودکار واژه‌های زبان فارسی بر اساس اصول هجابندی پولگرام"، مقاله‌ی ارائه‌شده در پنجمین کنفرانس بین‌المللی مطالعات زبان، ایران، دانشگاه علامه طباطبایی، ۱۳۹۶.

[24] T. Ahmadi, H. Karshenas, B. Alinejad, M. Naghavi Ravandi, "Automatic syllabification of Persian words based on Pulgram principles", *In the fifth international conference of language studies*, Iran, Allameh Tabatabaee university, 2017.

[۲۵] باباعلی، باقر، "پایه‌گذاری بستری نو و کارآمد در حوزه بازشناسی گفتار فارسی"، فصل‌نامه‌ی علمی-پژوهشی پردازش علائم و داده‌ها، ۱۳(۳)، صفحات ۵۱-۶۲، ۱۳۹۵.

[25] B. Babaali, "A state-of-the-art and efficient framework for Persian speech recognition", *Research center of intelligent signal processing,* Vol. 13(3), pp. 51-62, 2016.

[26] D. Yu, L. Deng, "Automatic speech recognition, a deep learning approach". Springer, pp. 1-2, London, 2016.

[27] S. Karpagavalli, E. Chandra, "A Review on Automatic Speech Recognition Architecture and Approaches", *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 9(4), pp. 393-404, 2016.

[28] S. Sun, B. Zhang, L. Xie, Y. Zhang, "An unsupervised deep domain adaptation approach for robust speech recognition", *Neurocomputing*, pp. 79-87, Sep 27, 2017.

[29] L. Toth, I. Hoffmann, G. Gosztolya, V. Vincze, G. Szatloczki, Z. Banreti, M. Pákáski, J. Kalman, "A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech", *Current Alzheimer Research*, vol. 15(2), pp. 130-138, Feb 1, 2018.

[30] S. Sinha, SS. Agrawal, A. Jain, "Continuous density hidden markov model for hindi speech recognition", *GSTF Journal on Computing (JoC),* vol. 3(2), Jan 19, 2018.

[31] CH. You, MA. Bin, "Spectral-domain speech enhancement for speech recognition", *Speech Communication*, pp. 30-41, Nov 1, 2017.

[DOI: 10.29252/jsdp.17.3.37]

systems via nonlinear dynamical features evaluated from the recurrence plot of speech signals", *Computers & Electrical Engineering*, pp. 215-226, 2017.

[44] MM. Goodarzi, F. Almasganj, "Model-based clustered sparse imputation for noise robust speech recognition", *Speech Communication*, pp. 218-229, Feb 1, 2016.

[45] Y. Shekofteh, F. Almasganj, A. Daliri, "MLP-based isolated phoneme classification using likelihood features extracted from reconstructed phase space", *Engineering Applications of Artificial Intelligence*, pp. 1-9, Sep 1, 2015.

[46] M. Bahrani, H. Sameti, "Building statistical language models for persian continuous speech recognition systems using the peykare corpus", *International Journal of Computer Processing of Languages*, vol.23(01), pp. 1-20, Mar, 2011.

[47] M. Sharifi Atashgah, V. Sadeghi, "A phoneme recognition algorithm design using the acoustic correlates of the phonological features", *Journsl of signal and data processing*, Vol.2, (SERIAL 16), pp. 13-28, 2011.

[48] P. Roach, "English Phonetics and Phonology Fourth Edition: A Practical Course", Ernst Klett Sprachen, pp. 42-43, 2010.

[49] C. Gussenhoven, H. Jacobs, "Understanding phonology", Routledge, 2017.

[50] WJ. Hardcastle, J. Laver, FE. Gibbon, "The handbook of phonetic sciences", John Wiley & Sons, Feb 22, pp.316-356, 500, 783-784, 793), 2010.

[51] D. Recasens, "Coarticulation and sound change in Romance", John Benjamins Publishing Company, Apr 15, p. ix, 3, 2014.

[52] B. Kühnert, F. Nolan, "The origin of coarticulation. Coarticulation: Theory, data and techniques", pp. 7-30, 1999.

[53] R. Kennedy, "Phonology: A Coursebook", Cambridge University Press, 2017.

[54] P. Ladefoged, K. Johnson, "A course in phonetics", Nelson Education, Jan 3, p. 71, 111, 277, 2014.

[55] B. Heselwood, "Phonetic transcription in theory and practice", Edinburgh University Press, Oct 31, p. 151, 2013.

[56] BS. Collins, IM. Mees, "Practical phonetics and phonology: A resource book for students", Routledge, Feb 11, p. 123, 2013.

[57] RM. Millar, L. Trask, "Trask's historical linguistics", Routledge, Feb 20, pp. 49-51, 2015.

[58] WG. Bennett, "Assimilation, dissimilation, and surface correspondence in Sundanese", *Natural Language & Linguistic Theory*, vol. 33(2), pp. 371-415, May 1, 2015.

[32] R. Lileikytė, L. Lamel, JL. Gauvain, A. Gorin, "Conversational telephone speech recognition for Lithuanian", *Computer Speech & Language*, pp. 71-82, May 31, 2018.

[33] Z. Chen, J. Droppo, J. Li, W. Xiong, Z. Chen, J. Droppo, J. Li, W. Xiong, "Progressive joint modeling in unsupervised single-channel overlapped speech recognition", *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol.26(1), pp. 184-196, Jan 1, 2018.

[34] SC. Sajjan, C. Vijaya, "Continuous Speech Recognition of Kannada language using triphone modeling", *In International Conference: Wireless Communications, Signal Processing and Networking (WiSPNET),* Mar 23, 2016, pp. 451-455.

[35] A. Shaukat, H. Ali, U. Akram, "Automatic Urdu Speech Recognition using Hidden Markov Model", *In International Conference: Image, Vision and Computing (ICIVC),* Aug 3, 2016, pp. 135-139.

[36] J. Xu, J. Pan, Y. Yan, "Agglutinative language speech recognition using automatic allophone deriving", *Chinese Journal of Electronics,* vol.25(2), pp. 328-333, Mar 1, 2016.

[37] B. Baba Ali, H. Sameti. "The sharif speaker-independent large vocabulary speech recognition system", *in The 2nd Workshop on information Technology & Its Disciplines (WITID 2004),* Feb 24. 2004, pp. 24-26.

[38] H. Sameti, H. Veisi, M. Bahrani, B. Babaali, K, Hosseinzadeh, "Nevisa: a Persian continuous speech recognition system", *In Advances in Computer Science and Engineering*, Springer, pp. 485-492, Berlin, Heidelberg, 2008.

[39] H. Sameti, H. Veisi, M. Bahrani, B. Babaali, K. Hosseinzadeh, "A large vocabulary continuous speech recognition system for Persian language", *EURASIP Journal on Audio, Speech, and Music Processing*, Dec 1, 2011.

[40] KE. Kafoori, SM. Ahadi. "Bounded cepstral marginalization of missing data for robust speech recognition", *Computer Speech & Language*, pp. 1-23, Mar 1, 2016.

[41] HR. Seresht, SM. Ahadi, S. Seyedin. "Spectro-temporal power spectrum features for noise robust ASR", *Circuits, Systems, and Signal Processing,* vol. 36(8), pp. 3222-3242, Aug 1, 2017.

[42] KE. Kafoori, SM. Ahadi, "Robust Recognition of Noisy Speech Through Partial Imputation of Missing Data", *Circuits, Systems, and Signal Processing*, vol.37(4), pp. 1625-1648, Apr 1, 2018.

[43] SG. Firooz, F. Almasganj, Y. Shekofteh, "Improvement of automatic speech recognition

[DOI: 10.29252/jsdp.17.3.37]

real- world environments", *Computer Speech & Language*, pp. 388-400, Nov 1, 2017.

[71] AH. Moore, PP. Parada, PA. Naylor, "Speech enhancement for robust automatic speech recognition: Evaluation using a baseline system and instrumental measures", *Computer Speech & Language*, pp. 574-84, Nov 1, 2017.

[72] A. Veiga, S. Candeias, L. Sá, F. Perdigão, "Using coarticulationrules in automatic phonetic transcription", *In Proceedings of PROPOR*, April, 2010.

[73] F. Imedjdouben, A. Houacine, "Generation of allophones for speech synthesis dedicated to the Arabic language", *In First International Conference on New Technologies of Information and Communication (NTIC),* 2015, pp. 1-4, Nov 8.

[74] A. Lee, T. Kawahara, K. Shikano, "Julius---an open source real-time large vocabulary recognition engine", 2001.

[75] "in the CU SONIC ASR system for noisy speech: The SPINE task", *In IEEE International Conference: Acoustics, Speech, and Signal Processing (ICASSP'03),* Vol. 1, 2003, pp. I-I.

[76] KF. Lee, HW. Hon, R. Reddy, "An overview of the SPHINX speech recognition system", *In Readings in speech Recognition*, pp. 600-610, 1990.

[77] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, J. Woelfel, "Sphinx-4: A flexible open source framework for speech recognition", 2004.

[78] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, "The Kaldi speech recognition toolkit", *In IEEE workshop on automatic speech recognition and understanding (No. EPFL-CONF-192584), IEEE Signal Processing Society*, 2011.

[79] M. Bijankhan, MJ. Sheikhzadegan, MR. Roohani, "SMALL FARSDAT-The speech database of Farsi spoken language", *In Proceedings of the 5th Australian International Conference on speech science and technology*, Perth, Australia, December, 1994, pp. 826-829.

[80] F. Almasganj, SA. Seyedsalehi, M. Bijankhan, H. Sameti, J. Sheikhzadegan, "SHENAVA-1: Persian spontaneous continuous speech recognizer", *In Proceedings of the International Conference on Electrical Engineering*, 2001, pp. 101-106.

[81] M. Caballero, A. Moreno, A. Nogueiras, "Multidialectal Spanish acoustic modeling for speech recognition", *Speech Communication,* vol. 51(3), pp. 217-229, 2009.

[59] RA. Knight, "Phonetics: A coursebook", Cambridge University Press, Jan 26, pp. 90, 103, 192-193, 2012.

[60] D. Jacques, "Generative and non-linear phonology", Routledge, Sep 25, pp. 298, 2014.

[61] C. Herff, D. Heger, A. De Pesters, D. Telaar, P. Brunner, G. Schalk, T. Schultz, "Brain-to-text: decoding spoken phrases from phone representations in the brain", *Frontiers in neuroscience*, pp. 217, Jun 12, 2015.

[62] MS. Mirzaei, K. Meshgi, T. Kawahara, "Exploiting automatic speech recognition errors to enhance partial and synchronized caption for facilitating second language listening", *Computer Speech & Language*, pp. 17-36, May 1, 2018.

[63] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, Y. Bengio. "End-to-end attention-based large vocabulary speech recognition", *In IEEE International Conference: Acoustics, Speech and Signal Processing (ICASSP),* Mar 20, 2016, pp. 4945-4949.

[64] AR. Mohamed, F. Seide, D. Yu, J. Droppo, A. Stoicke, G. Zweig, G. Penn, "Deep bi-directional recurrent networks over spectral windows", *In IEEE Workshop: Automatic Speech Recognition and Understanding (ASRU)*, pp. 78-83, Dec 13, 2015.

[65] T. Moon, H. Choi, H. Lee, I. Song, "Rnndrop: A novel dropout for rnns in asr", *In IEEE Workshop: Automatic Speech Recognition and Understanding (ASRU),* pp. 65-70, Dec 13, 2015.

[66] I. Himawan, P. Motlicek, D. Imseng, S. Sridharan, "Feature mapping using far-field microphones for distant speech recognition", *Speech Communication*, pp. 1-9, Oct 1, 2016.

[67] S. Ravuri, "Hybrid dnn-latent structured SVM acoustic models for continuous speech recognition", *In IEEE Workshop: Automatic Speech Recognition and Understanding (ASRU),* pp. 37-44, Dec 13, 2015.

[68] W. Chan, N. Jaitly, Q. Le, O. Vinyals, "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition", *In IEEE International Conference: Acoustics, Speech and Signal Processing (ICASSP)*, Mar 20, 2016, pp. 4960-4964.

[69] Z. Wang, E. Vincent, R. Serizel, Y. Yan, "Rank-1 constrained multichannel Wiener filter for speech recognition in noisy environments", *Computer Speech & Language*, pp. 37-51, May 1, 2018.

[70] H. Barfuss, C. Huemmer, A. Schwarz, W. Kellermann, "Robust coherence-based spectral enhancement for speech recognition in adverse

[DOI: 10.29252/jsdp.17.3.37]

طاهره احمدی در سال ۱۳۹۶ در مقطع کارشناسی ارشد رشته زبان‌شناسی گرایش رایانشی فارغ‌التحصیل شد و هم‌اکنون دانشجوی مقطع دکترای زبان‌شناسی در دانشگاه اصفهان است. زمینه‌های پژوهشی مورد علاقه وی آواشناسی و واج‌شناسی است.

نشانی رایانامه ایشان عبارت است از:

pazhvak.ta@gmail.com

حسین کارشناس در سال ۱۳۹۲ در مقطع دکترای هوش مصنوعی از دانشگاه پلی‌تکنیک مادرید اسپانیا فارغ‌التحصیل شدند و هم‌اکنون عضو هیئت علمی گروه هوش مصنوعی دانشکده رایانهٔ دانشگاه اصفهان هستند. زمینه‌های پژوهشی مورد علاقه ایشان هوش محاسباتی، بهینه‌سازی، یادگیری ماشین و داده‌کاوی است.

نشانی رایانامه ایشان عبارت است از:

h.karshenas@eng.ui.ac.ir

باقر باباعلی فارغ‌التحصیل مقطع دکترای هوش مصنوعی از دانشگاه صنعتی شریف در سال ۱۳۸۹ و هم‌اکنون عضو هیئت علمی دانشکده ریاضی، آمار و علوم رایانهٔ دانشگاه تهران هستند. زمینه‌های پژوهشی مورد علاقه ایشان، یادگیری ماشین و بازشناسی الگوی آماری، بازشناسی گفتار، بازشناسی الگوهای دنباله‌ای، یادگیری ژرف و کاربردهای آن است.

نشانی رایانامه ایشان عبارت است از:

babaali@ut.ac.ir

بتول علی‌نژاد فارغ‌التحصیل مقطع دکترای زبان‌شناسی از دانشگاه اصفهان در سال ۱۳۸۲ و هم‌اکنون عضو هیئت علمی گروه زبان‌شناسی دانشکده زبان‌های خارجی دانشگاه اصفهان هستند. زمینه‌های پژوهشی مورد علاقه ایشان، آواشناسی و واج‌شناسی است.

نشانی رایانامه ایشان عبارت است از:

b.alinezhad@fgn.ui.ac.ir

[DOI: 10.29252/jsdp.17.3.37]