

رتبه‌بندی واج‌های گفتار فارسی از نظر کارآیی در بازشناسی گوینده

جواد شیخ‌زادگان^۱

دانشیار پژوهشکده پردازش هوشمند علام

(از ص ۷۷ تا ص ۹۶)

تاریخ دریافت مقاله: ۹۴/۱۲/۹؛ تاریخ پذیرش مقاله: ۹۵/۳/۳۱

چکیده

در این مقاله، کارآیی واج‌های گفتار فارسی از نظر بازشناسی گوینده مورد مطالعه و پژوهش قرار گرفته و با توجه به میزان کارآیی‌ها، رتبه‌بندی واج‌ها صورت گرفته‌اند. جهت برآورد کارآیی واج‌ها، از یک معیاری که به صورت نسب «فاصله بین گوینده‌ای» واج‌ها به «فاصله در گوینده‌ای» تعریف شده است و ما آن را «نسبت تأثیرپذیری گوینده» نامیده‌ایم، استفاده شده است. آزمایش‌ها و محاسبات لازم برای کلیه واج‌های گفتار فارسی (باستثنای واج /ʔ/) با استفاده از دادگان گفتار فارسی «فارس‌دات» انجام شده و رتبه‌بندی‌ها براساس نتایج آزمایش‌ها و محاسبات هم در مورد دسته‌های کلی واجی و هم برای تک‌تک واج‌ها صورت گرفته‌اند. نتایج آزمایش‌ها و محاسبات نشان داده‌اند که در رتبه‌بندی دسته‌های کلی واجی، واکه‌ها و نیم‌واکه‌ها در رتبه اول، خیشومی‌ها، سایشی‌ها و روان‌ها در رتبه دوم و انسدادی‌ها و انفجاری‌ها در رتبه سوم از نظر کارآیی در بازشناسی گوینده قرار دارند. رتبه‌بندی تک‌تک واج‌ها نیز نشان می‌دهد که واج /ð/ در رتبه اول و واج /t/ در رتبه آخر از نظر کارآیی در بازشناسی گوینده قرار می‌گیرند. نتایج این تحقیق در مقایسه با نتایج پژوهش‌های انجام شده در مورد برخی از زبان‌های دیگر نظیر انگلیسی، آلمانی و دوچ از نظر رتبه‌بندی دسته‌های کلی واجی سازگاری بالایی دارد اما از نظر جزئیات رتبه‌بندی‌ها، تفاوت‌های قابل توجهی ملاحظه می‌شود.

واژه‌های کلیدی: رتبه‌بندی واج‌ها، گفتار فارسی، کارآیی واج‌ها و بازشناسی گوینده و نسبت

تأثیرپذیری

۱- مقدمه

واج‌ها واحدهای زبان‌شناختی هستند که باعث تمایز معنایی واژه‌های زبان از یکدیگر می‌شوند. به‌عنوان مثال، آنچه که باعث تمایز ذهنی دو واژه‌ی «پَر» و «پِر» یا دو واژه‌ی «بَر» و «پِر» می‌شود، در زوج واژه‌ی اولی تمایز در واحدهای زبانی /a/ و /o/ و در زوج واژه‌ی بعدی، تمایز در واحدهای زبانی /b/ و /p/ از یکدیگر است. درک گویشوران زبان از واج‌ها تابعی از یادگیری برخی از پارامترهای تولیدی یا صوتی در فرآیند اکتساب نظام آوایی زبان است. واج‌ها در سطح سیگنال به‌صورت یک تابع ایستا در واحد زمان ظاهر می‌شوند و به همین علت، در پردازش گفتار، مقادیر هر پارامتر بازنمایی در حوزه‌ی فرکانس در قاب‌های متوالی که متعلق به یک واج یا بخشی از واج است، اختلاف معنی‌داری از نظر آماری ندارند. به این دلیل، واج به‌عنوان یک واقعیت ساختاری در سیگنال گفتار اهمیت فوق‌العاده‌ای در پردازش گفتار پیدا کرده است (بی‌جن‌خان، ۱۳۷۶ الف: ۱). واج‌های زبان گفتاری از دیرباز به‌عنوان عناصر زبانی مورد توجه پژوهشگران و علمای علوم زبانشناسی بوده‌اند، اما در چند دهه اخیر، آنها همچنین به‌عنوان عناصر شناختی و عناصر پردازشی به شدت مورد توجه محققین و دانش‌پژوهان حوزه زبانشناسی رایانشی و پژوهشگران و مهندسين حوزه‌های مختلف مهندسی الکترونیک، مخابرات و رایانه که به‌نحوی با پردازش سیگنال‌های گفتاری و یا به‌طور کلی با پردازش زبان طبیعی سروکار دارند، قرار گرفته‌اند (بی‌جن‌خان، ۱۳۷۶ الف: ۲ - ۳). واج‌ها از دیدگاه آواشناسی و واج‌شناسی توسط زبانشناسان به دسته‌های کلی و جزئی زیادی طبقه‌بندی شده‌اند. به‌عنوان مثال، دو نمونه از عمده‌ترین دسته‌بندی‌های سنتی واج‌ها که براساس مشخصه‌های تولیدی آواها صورت گرفته‌اند عبارتند از: (۱) واکه‌ها^۱ و همخوان‌ها^۲ و اکدارها^۳ و بی-واک‌ها^۴ (ثمره، ۱۳۶۸) و یا در واج‌شناسی زایا از نظر نقش آواها در ساخت هجا، واج‌ها به دو دسته‌ی کلی واج‌های هجایی^۵ (یا رسا^۶) و واج‌های غیرهجایی (نارسا یا گرفته^۷) دسته‌بندی شده و به همین ترتیب دسته‌بندی‌های خیلی جزئی و ریز نیز صورت گرفته‌اند (ثمره، ۱۳۶۸) و (مشکوه‌الدینی، ۱۳۸۸: ۱۹). با توسعه علوم و فنون پردازش

1. Vowels
 2. Consonants
 3. Voiced
 4. Unvoiced
 5. Syllabic
 6. Sonorant
 7. Obstruent

سیگنال‌های گفتاری در دو سه دهه‌ی اخیر در حوزه‌های مهندسی الکترونیک، مخابرات و رایانه، اهمیت واج‌های زبان گفتاری به‌عنوان عناصر مفید و مؤثر در موضوعات کاربردی پردازش سیگنال‌های گفتاری به‌خصوص در بازشناسی گفتار^۱ و بازشناسی گوینده^۲ نمایان شد و از این‌رو مطالعات و تحقیقات قابل توجهی در راستای شناخت ابعاد، ویژگی‌ها و خصوصیات واج‌ها از منظر پردازش سیگنال‌های گفتاری در مورد برخی از زبان‌های خارجی از جمله زبان‌های انگلیسی، آلمانی، دوچ^۳ و... و نیز در مورد زبان فارسی صورت گرفتند که از آن جمله می‌توان به (ماتسویی و همکاران، ۱۹۹۲: ۶۰۳-۶۰۶)؛ (هیوول و همکاران، ۱۹۹۲: ۱۵۸۱-۱۵۸۴)؛ (ایتوک و همکاران، ۱۹۹۲: ۱۴۱۱-۱۴۱۴)؛ (ابه و همکاران، ۱۹۹۰: ۱۵۷-۱۶۰) در مورد زبان‌های خارجی و (بی‌جن‌خان، ۱۳۷۶ الف: ۱-۶)، (بیجن‌خان، ۱۳۷۶ ب: ۷-۱۲) (سیدصالحی، ۱۳۷۶: ۱۳-۱۸)، (شیخ‌زادگان، ۱۳۷۴ الف: ۲۷-۳۵) در مورد زبان فارسی اشاره کرد.

۲- مرور کارهای مرتبط

در بازشناسی خودکار گوینده توسط رایانه، ویژگی‌های شناسایی به‌نحوی از سیگنال صوتی گوینده استخراج می‌شوند. سیگنال‌های گفتاری تظاهر عینی عناصر ذهنی گفتار یعنی واج‌ها هستند. بنابراین ویژگی‌های هویتی گوینده به‌طور بنیادی به مجموعه‌ی واج‌های زبان وابستگی دارند و از این‌رو در دهه‌های اخیر به‌طور چشمگیری در طراحی و پیاده‌سازی سامانه‌های بازشناسی خودکار گوینده در هر دو روش «وابسته به متن^۴» و «مستقل از متن^۵»، به مجموعه واج‌های زبان معطوف گشته است (همان). به‌خصوص در روش مستقل از متن که بایستی داده‌های یادگیری و آزمون غیروابسته و متفاوت از هم باشند، اصولاً بایستی ویژگی‌های شناسایی گوینده از عناصری استخراج شوند که کمترین وابستگی را به متن داشته باشند. از آنجا که واحدهای بنیادی تشکیل‌دهنده گفتار، واج‌ها هستند، بنابراین واج‌ها به‌عنوان عناصر مستقل از متن در بازشناسی گوینده به روش مستقل از متن، مورد توجه ویژه قرار گرفتند (فوری، ۱۹۸۶؛ ۱۸۳-۱۹۷)؛ (ماتسویی، ۱۹۹۲: ۶۰۳-۶۰۶) و (شیخ‌زادگان، ۱۳۷۴ ب: ۲۷-۳۵).

1. Speech Recognition
2. Speaker Recognition
3. Dutch
4. Text dependent
5. Text Independent

می‌دانیم که هر زبان گفتاری مجموعه مشخصی از واج‌ها را دارا هستند و تحقیقات علوم زبانشناسی نشان می‌دهند که مشخصات کلی این مجموعه به‌طور کلی برای زبان‌های مختلف متفاوت هستند. به‌عنوان مثال، مجموعه واج‌های گفتار انگلیسی ۴۴ عنصر دارد در صورتی‌که مجموعه واج‌های گفتار فارسی دارای ۲۹ عنصر است (ثمره، ۱۳۶۸). قابل توجه است که مجموعه الفبای نوشتار انگلیسی دارای ۲۶ عنصر و مجموعه‌ی الفبای نوشتار فارسی ۳۲ عنصر دارد. در دهه‌های اخیر، تحقیقات قابل توجهی نیز در رابطه با ویژگی‌های واج‌های مختلف یک زبان از نقطه نظر بازشناسی خودکار گوینده صورت گرفته‌اند که بخش قابل توجهی از آنها در رابطه با رتبه‌بندی واج‌های یک زبان گفتاری از نظر مرتبه کارایی آنها در بازشناسی خودکار گوینده است که در ادامه‌ی این بخش، به برخی از آنها که ارتباط بیشتری با موضوع مقاله‌ی حاضر دارند، می‌پردازیم. در سال ۱۹۹۳ ایتوک^۱ و ماسون^۲ کارایی واج‌های زبان انگلیسی را از نظر بازشناسی گوینده مورد مطالعه و تحقیق قرار داده ترتیب درجه اهمیت طبقات کلی واج‌های زبان انگلیسی را مشخص کردند. نتیجه تحقیقات آنها نشان داد که از نظر بازشناسی گوینده، واکه‌های بلند در رتبه اول، خیشومی‌ها در رتبه دوم، واکه‌های کوتاه در رتبه سوم قرار دارند و انفجاری‌ها برای اهداف بازشناسی گوینده، از کمترین کارایی و اهمیت قرار داشته و در رتبه آخر قرار دارند (ایتوک و ماسون، ۱۹۹۲: ۱۴۱۱-۱۴۱۴). هیوول^۳ و ریتولد^۴ در همان سال تغییرپذیری مرتبط با گوینده را در قطعات گفتاری زبان هلندی^۵ مورد مطالعه و تحقیق قرار داده و نشان دادند که واج /I/ همانند واج‌های انفجاری زبان هلندی از نظر قدرت تفکیک و تمایز گوینده‌ها از کمترین اهمیت برخوردار بوده و در پایین‌ترین رتبه قرار دارد. در حالی‌که، براساس مطالعات و تحقیقات انجام شده توسط گلدستاین^۶ در سال ۱۹۷۶ و ایتوک و ماسون در سال ۱۹۹۲، واج /I/ در زبان‌های انگلیسی و آلمانی از نظر دارا بودن ویژگی‌های فردی گوینده غنای متوسطی دارد و از این‌رو در رتبه‌بندی واج‌ها از نظر بازشناسی گوینده، در مرتبه‌ی متوسط (بعد از واکه‌ها و خیشومی‌ها و قبل از انفجاری‌ها و روان‌ها) قرار دارد (گلدستاین، ۱۹۷۶: ۱۷۶-۱۸۲) و (ایتوک و ماسون، ۱۹۹۲: ۱۴۱۱-۱۴۱۴). از آنجا که واج‌ها به‌عنوان عناصر مستقل از متن جهت

1. J.P. Eatok

2. J.S.D. Mason

3. Heuvel

4. Rietveld

5. Dutch

6. U.G. Goldstein

شرکت در فرآیند بازشناسی گوینده به‌خصوص در روش مستقل از متن خیلی مناسب و مفید هستند، بنابراین لحاظ نمودن میزان کارآیی واج‌های مختلف زبان در فرآیند بازشناسی با یک طرح مناسب نیز در ارتقای کارآیی سامانه‌ی بازشناسی مؤثر و مفید خواهد بود. از این‌رو نکارنده‌ی مقاله حاضر، در (شیخ‌زادگان، ۱۳۷۴ الف: ۲۷-۳۵) درجه اهمیت واج‌های گفتار فارسی را از نظر قدرت تفکیک و تمایز گوینده‌ها مورد مطالعه و پژوهش قرار داده بود که این تحقیق در حقیقت نسخه توسعه‌یافته پژوهش مذکور است.

۳- معیار برآورد کارآیی

تحقق صوتی واج‌ها برآیند فرآیندهای واجی^۱، یعنی برآیند تأثیرگذاری‌های آوایی و تظاهرات ناشی از آنهاست (مشکوه‌الدینی، ۱۳۸۸: ۱۳۰). فرآیندهای واجی نمایانگر رابطه ویژگی‌های سطح صوتی زبان با سطح واج است که به‌طور کلی تحت تأثیر دو عامل اصلی هستند: بافت آوایی و شیوه تولید. ویژگی‌های سطح صوتی یک واج در بافت‌های مختلف برای یک گوینده مشخص متفاوت است و بدین ترتیب هر یک از واج‌های زبان گفتاری در سطح سیگنال، به‌صورت مختلف تظاهر پیدا می‌کنند که به آنها واج-گونه‌های^۲ آن واج گفته می‌شود. بنابراین در فضای ویژگی‌های سطح صوتی یک واج، ما تنها با یک بردار ویژگی^۳ مشخصی سروکار نداریم بلکه به تعداد واج‌گونه‌های واج مورد نظر بردار ویژگی خواهیم داشت. از طرف دیگر، شیوه تولید یک واج در یک بافت مشخص نیز می‌تواند ویژگی‌های متفاوتی در سطح صوت ایجاد کند و بدین ترتیب برای یک واج‌گونه مشخص (از نظر بافت)، سیگنال‌های صوتی متفاوتی به ازای شیوه‌های مختلف تولید، خواهیم داشت. شیوه تولید وابستگی زیادی به گوینده دارد. بنابراین برای یک واج در یک بافت مشخص (یعنی یک واج‌گونه مشخص) نیز به تعداد گوینده‌های جمعیت مورد نظر، سیگنال‌های متفاوتی خواهیم داشت. تفاوت‌های اخیر در ویژگی‌های سطح صوتی واج‌ها (متأثر از گوینده‌ها)، در حقیقت منشاء تفکیک و تمایز گوینده‌ها در سامانه‌های بازشناسی گوینده است که به ویژه در بازشناسی گوینده به روش مستقل از متن، می‌تواند خیلی مفید واقع شود. در مقابل تفاوت‌ها در ویژگی‌های سطح صوتی واج ناشی از بافت، ابهام در تفکیک و تمایز گوینده‌ها ایجاد می‌کنند. پس معیار کارآیی واج‌ها در تفکیک و تمایز گوینده‌ها و به عبارت دیگر رتبه‌ی کارآیی واج‌ها در بازشناسی

^۱ Phonological processes

^۲ Allophones

^۳ Feature vector

گوینده، رابطه مستقیم با تفاوت‌های ویژگی‌های سطح صوتی واج‌ها که متأثر از گوینده-ها هستند، دارد و با تفاوت‌های ویژگی‌های سطح صوتی واج‌ها که متأثر از بافت هستند، این رابطه معکوس است. پس ما برای فرموله کردن معیار کارآیی، ابتدا بایستی تفاوت-های مذکور را درخصوص ویژگی‌های سطح صوتی واج‌ها (ناشی از بافت و متأثر از گوینده‌ها)، مدل کنیم.

۳-۱- فاصله درگوینده‌ای^۱ (IraSD)

در مدل‌های ریاضی، تفاوت بین دو بردار را با معیار «فاصله» بیان می‌کنند. بنابراین، تفاوت‌های ویژگی‌های سطح صوتی واج‌ها را که ناشی از بافت هستند، می‌توان با یک معیار فاصله بین ویژگی‌های سطح صوتی مربوط به یک واج در بافت‌های مختلف بیان کرد که آن را «فاصله درگوینده‌ای» می‌نامیم. ویژگی‌های صوتی مختلفی در بازشناسی گوینده مورد بهره‌برداری قرار می‌گیرند که در این خصوص در بخش‌های بعدی به‌طور مجمل صحبت خواهیم کرد، اما در اینجا برای مدل کردن، فاصله درگوینده‌ای، فرض می‌کنیم که تعداد ویژگی‌های سطح صوتی برای هر نمونه یا واحد پردازشی (قاب زمانی^۲) n بوده و ویژگی‌ها را با C_l^i ($l = 1, 2, \dots, n$) نمایش دهیم. به ازای یک گوینده مشخص، فاصله دو نمونه‌ی مختلف در دو بافت متمایز از یک واج مشخص (نمونه‌های i و j) را می‌توان توسط رابطه زیر تعیین کرد که در آن C_l^i و C_l^j به ترتیب ویژگی‌های l ام از نمونه‌های i و j هستند.

$$\sum_{l=1}^n (c_l^j - c_l^i)^2$$

با انتخاب تعداد زیادی نمونه از واج مورد نظر از یک دادگان گفتاری غنی از واج‌گونه‌ها، می‌توان در سطح قابل قبولی مطمئن شد که تفاوت‌های ناشی از بافت در ویژگی‌های سطح صوتی واج‌ها، لحاظ خواهند شد. از این‌رو، هرگاه ما K نمونه (که نمونه‌ها واحدهای پردازش هستند و K عدد بزرگی است) از کلیه واج‌گونه‌های موجود در دادگان مورد استفاده برای یک گوینده مشخص در نظر بگیریم و فاصله‌ی دوبردوی آنها را محاسبه کنیم، در این‌صورت میانگین فاصله برای K نمونه می‌تواند طبق رابطه زیر محاسبه شود که این در حقیقت فاصله‌ی درگوینده‌ای مورد نظر برای واج معین (P_l) و به ازای یک گوینده‌ی مشخص (مثلاً گوینده‌ی a از جمعیت M نفری گوینده‌ها) است:

^۱ Intra speaker Distance (IraSD)

^۲ Frame

$$\text{IraSD}_{P_t}^a = \frac{1}{\frac{K}{2}(K-1)} \sum_{j=1}^K \sum_{\substack{i=1 \\ i>j}}^K \sum_{l=1}^n (c_l^j - c_l^i)^2$$

اگر رابطه فوق یعنی $\text{IraSD}_{P_t}^a$ را به ازای گوینده‌ها مختلف جمعیت مورد نظر محاسبه کرده و میانگین بگیریم، در آن صورت به همان «فاصله در گوینده‌ای» مورد نظر خواهیم رسید. یعنی داریم:

$$\text{IraSD}_{P_t} = \frac{1}{M} \sum_{a=1}^M \text{IraSD}_{P_t}^a$$

۲-۳- فاصله بین گوینده‌ای^۱ (IerSD)

مشابه بحثی که در مورد فاصله در گوینده‌ای (IraSD) داشتیم، می‌توان تفاوت‌های ویژگی‌های سطح صوتی واج‌ها را که ناشی از مشخصه‌ای انفرادی گوینده‌هاست^۲، با یک معیار فاصله بین ویژگی‌های سطح صوتی مربوط به یک واج از گوینده‌های مختلف، بیان کرد که ما آن را «فاصله بین گوینده‌ای» نامیده‌ایم. نظیر مفروضات بخش ۳-۱، اگر تعداد ویژگی‌های سطح صوتی را در هر واحد پردازشی n و ویژگی‌ها را با C_l نشان دهیم، فاصله دو نمونه از یک واج مشخص که متعلق به دو گوینده مختلف هستند (نمونه‌های i و j) را می‌توان توسط رابطه:

$$\sum_{l=1}^n (c_l^j - c_l^i)^2$$

بیان کرد. قابل ذکر است که در اینجا نیز نمونه‌ها در حقیقت فریم‌های پردازشی هستند. حال اگر تعداد K_a نمونه از واج‌گونه‌ها مختلف یک واج مورد نظر از یک گوینده مشخص (گوینده a) و همچنین تعداد $K_{\bar{a}}$ نمونه از واج‌گونه‌های مختلف همان واج مورد نظر از گوینده‌های دیگر را از دادگان مورد استفاده در نظر گرفته و فاصله دوبروی آنها را حساب کنیم، در این صورت میانگین فاصله‌ها برای واج مورد نظر و به ازای یک گوینده مشخص (همان گوینده a) از دیگر گوینده‌های جمعیت مورد نظر توسط رابطه زیر قابل بیان است:

$$\text{IerSD}_{P_t}^a = \frac{1}{K_{\bar{a}} \cdot K_a} \sum_{j=1}^{K_{\bar{a}}} \sum_{i=1}^{K_a} \sum_{l=1}^n (c_l^j - c_l^i)^2$$

در حقیقت فاصله بین گوینده‌ای واج است P_t که در آن میانگین فاصله واج‌گونه‌های گوینده a نسبت به واج‌گونه‌های بقیه گوینده‌های جمعیت مورد نظر

^۱ Inter Speaker Distance (IerSD)

^۲ Speaker Individual Characteristics

محاسبه شده است. اگر تعداد گوینده‌های جمعیت مورد نظر را M در نظر بگیریم و میانگین فاصله واج‌گونه‌های هر یک از گوینده‌ها را نسبت به بقیه گوینده‌ها محاسبه نماییم (که آن را با $IerSD_{P_t}^S$ نشان می‌دهیم)، آنگاه میانگین $IerSD_{P_t}^S$ بین M گوینده در حقیقت همان «فاصله گوینده‌ای» مدنظر خواهد بود. یعنی داریم:

$$IerSD_{P_t} = \frac{1}{M} \sum_{S=1}^M IerSD_{P_t}^S = \frac{1}{M} \sum_{S=1}^M \left[\sum_{j=1}^{K_S} \sum_{i=1}^{K_S} \sum_{l=1}^n (C_i^j - C_i^l)^2 \right]$$

۳-۳- معیار کارآیی واج‌ها در بازشناسی

حال با توجه به بحث‌هایی که در ابتدای بخش ۳ و زیربخش‌های آن یعنی ۳-۱ و ۳-۲ به عمل آمد، می‌توانیم معیاری را جهت برآورد کارآیی واج‌ها از نظر بازشناسی گوینده به صورت نسبت «فاصله بین‌گوینده‌ای» به «فاصله درگوینده‌ای» فرموله کنیم. ما این نسبت را که در حقیقت یک نسبت «تمایزگر^۱» برای واج‌ها از جهت قدرت تفکیک و تمایز گوینده‌هاست و به عبارت دیگر میزان تأثیرپذیری گوینده را به ازای یک واج مشخص بیان می‌کند، نسبت «تأثیرپذیری گوینده^۲» (تأثیرپذیری گوینده از واج) می‌نامیم:

$$SAR_{P_t} = \frac{IerSD_{P_t}}{IraSD_{P_t}}$$

SAR_{P_t} همان معیاری است که برآوردی از کارآیی واج P_t را از نظر بازشناسی گوینده، بدست می‌دهد.

۳-۴- ویژگی‌های سطح صوتی در بازشناسی گوینده

تا به حال انواع زیادی از بازنمایی‌ها^۳ جهت استخراج ویژگی‌ها از سیگنال گفتار به طور کلی در کاربردهای مختلف پردازش گفتار و به طور خاص در بازشناسی گوینده مورد توجه و بهره‌برداری پژوهشگران و مهندسين قرار گرفته‌اند که تعدادی از متداول‌ترین و مناسب‌ترین بازنمایی‌ها که به وفور مورد استفاده قرار گرفته‌اند عبارتند از:

– بازنمایی بانک فیلتر^۴ (پروزانسکی و ماتیوز، ۱۹۶۴: ۲۰۴۱-۲۰۴۷)؛ (لی و هوگس، ۱۹۷۴: ۸۳۳-۸۳۸).

^۱ Discriminant

^۲ Speaker Affectability Ratio (SAR)

^۳ Representation

^۴ Filter bank representation

- بازنمایی سازه‌ها^۱ (دودینگتون، ۱۹۷۰)؛ (گلدشتاین، ۱۹۷۶: ۱۷۶-۱۸۲) و (لمیس، ۱۹۷۸: ۸۰-۸۹).
- بازنمایی ضرایب پیشگویی خطی^۲ و انرژی باقیمانده در پیش‌گویی خطی^۳ (سامبور، ۱۹۷۶: ۲۸۳-۲۸۹) و (ورنچ، ۱۹۸۳: ۵۵۳-۵۵۸).
- بازنمایی ضرایب انعکاس^۴ و نسبت‌های لگاریتم سطح^۵ (مارکل و همکاران، ۱۹۷۷: ۳۳۰-۳۳۷) و (شوارتز و همکاران، ۱۹۸۲: ۱۶۴۹-۱۶۵۲).
- بازنمایی کپسترال^۶ (اتل، ۱۹۷۴: ۱۳۰۴-۱۳۱۲) و (شریده‌هار و همکاران، ۱۹۸۱: همکاران، ۱۹۸۱: ۱۹۷-۲۰۴) بازنمایی زوج خطوط طیفی^۷ (LSP) (لیا و همکاران، ۱۹۹۰: ۲۷۷-۲۸۰) و (پالیوال، ۱۹۸۸: ۴۸۵-۴۸۸).
- بازنمایی‌های مرتبط با منبع تحریک چاکنایی^۸ نظیر منحنی تغییرات فرکانس گام^۹، منحنی تغییرات انرژی^{۱۰} و مشخصات موج چاکنایی^{۱۱} (اتل، ۱۹۷۲: ۱۶۸۷-۱۹۷۲) (۱۹۷۲: ۱۹۷۲)؛ (یگنارایانا، ۱۹۹۴: ۱۸۶۷-۱۸۷۰) و (ماتسویی و فورویی، ۱۹۹۰: ۱۳۷-۱۴۰).

۳-۵- استخراج ویژگی‌های سطح صوتی از بازنمایی کپسترال

از آنجا که از بین انواع بازنمایی‌های مذکور، ضرایب کپسترال به‌عنوان یکی از بهترین بازنمایی‌ها جهت استخراج ویژگی‌های سطح صوتی در بازشناسی گوینده شناخته شده است (رز و رینولدز، ۱۹۹۰) از این‌رو ما در این پژوهش از ۱۲ ضریب کپستروم مستخرج از ضرایب پیش‌گویی خطی (LPCC) با استفاده از روابط زیر بهره می‌گیریم (اتل، ۱۹۷۴: ۱۳۰۴-۱۳۱۲):

$$C_0 = \ln \bar{E}$$

$$C_1 = a_1 \quad , C_j = \sum_{k=1}^{j-1} \left(1 - \frac{k}{n}\right) a_j C_{j-k} + a_j \quad : \text{ برای } 1 < j < p$$

1. Formants representation
2. Linear prediction coefficients
3. Linear prediction residual energy
4. Reflection Coefficients
5. Log Area Ratio
6. Cepstral Coefficients
7. Line Spectral pairs
8. Glottal excitation
9. Pitch frequency contour
10. Energy contour
11. Glottal wave characteristics

$$C_j = \sum_{k=1}^{j-1} \left(1 - \frac{k}{n}\right) a_j C_{j-k} \quad : \text{ برای } j > p$$

$$\hat{x}(n) = \sum_{k=1}^p a_k x(n-k) \quad , \quad E = \sum_{n=0}^{N-1} \left[x(n) - \sum_{k=1}^p a_k x(n-k) \right]^2$$

در روابط فوق $x(n)$ سیگنال گفتار، a_k ضرایب پیشگویی خطی، p مرتبه پیشگویی، E انرژی باقیمانده در پیشگویی، N تعداد نمونه‌های سیگنال گفتار در یک قاب زمانی و C_j ضرایب کپستروم مستخرج از ضرایب پیشگویی خطی یعنی LPC ها هستند. ضرایب مذکور برای قاب‌های زمانی ۳۲ میلی ثانیه‌ای با پنجره‌ی همینگ^۱ و با قدم‌های ۸ میلی ثانیه‌ای از روی نمونه‌های دیجیتالی گفتار استخراج می‌شوند که در آن سرعت نمونه‌برداری سیگنال آنالوگ ۱۴/۷ کیلوهرتز و نمونه‌ها ۱۶ بیتی هستند. قابل ذکر است که جهت فراهم کردن شرایط پادهم‌پوشی^۲، سیگنال گفتار قبل از نمونه‌برداری، توسط یک فیلتر پایین‌گذر با فرکانس قطع حدود ۷ کیلو هرتز، فیلتر می‌شود.

۴- آزمایش‌ها و نتایج

برای انجام آزمایش‌ها، قبل از هر چیز بایستی از دادگان مناسبی که دارای حجم قابل توجهی از واج‌گونه‌های مختلف انواع واج‌های زبان فارسی است و توسط تعداد قابل قبولی از گوینده‌های متنوع تولید شده‌اند، نمونه‌های لازم از واج‌گونه‌های هر واج را از گوینده‌های مختلف انتخاب کرد و آنگاه با استفاده از روابطی که در بخش ۳ و زیربخش-های آن ارائه شدند، ابتدا ویژگی‌های سطح صوتی از بازنمایی کپسترال سیگنال گفتاری واج‌گونه‌ها با استفاده از روابط زیربخش ۳-۵ استخراج شده و سپس با استفاده از روابط زیربخش‌های ۳-۱ الی ۳-۳ محاسبات لازم را با استفاده از «نسبت تأثیرپذیری گوینده» (SAR) انجام و کارآیی هر واج را از نظر بازشناسی گوینده تعیین می‌کنیم. بنابراین لازم است که در ادامه یک معرفی کوتاهی در خصوص دادگان مورد استفاده در این تحقیق داشته باشیم.

۴-۱- دادگان گفتار

دادگان گفتاری مورد استفاده در این تحقیق دادگان «فارس‌دات^۳» (بیجن‌خان و همکاران، ۱۹۹۴: ۳۹۷-۴۰۳) است که از آن حدود ۳۰۰۰ واج‌گونه مختلف برای ۲۸ واج

^۱ Hamming window

^۲ Antialiasing

^۳ FARSDAT-Farsi Spoken Language Database

گفتار فارسی از ۱۲ گوینده‌ی بومی فارسی زبان انتخاب شده‌اند و ترکیب جمعیت گوینده‌ها به صورت ۸ نفر مذکر و ۴ نفر مؤنث می‌باشد. قابل ذکر است که طول زمانی واج /ʔ/ خیلی کوتاه است و معمولاً شبیه واکه‌ی مجاورش بوده و لذا در این تحقیق به‌طور مستقل و جداگانه مورد توجه قرار نگرفته است. بنابراین به‌طور تقریبی از هر گوینده برای هر واج حدود ۱۰ واج‌گونه در نظر گرفته شده است. از آنجا که در این تحقیق ما از ۱۲ ضریب کپسترال استخراج شده از ضرایب پیش‌گویی خطی^۱ (LPCC) استفاده می‌کنیم و این ضرایب برای فریم‌های ۳۲ میلی ثانیه‌ای با قدم‌های ۸ میلی ثانیه‌ای از روی نمونه‌های دیجیتال گفتار استخراج می‌شوند لذا برای محاسبه فاصله‌های درگوینده‌ای و بین‌گوینده‌ای حدود ۱۰۰ نمونه‌ی فریمی از واج‌گونه‌های هر واج را برای هر گوینده از گوینده‌های جمعیت مورد نظر خواهیم داشت.

۲-۴-۲- انجام محاسبات

همان‌طوری که در بخش ۳-۳ گفته شد، ما از یک معیاری بنام « نسبت تأثیرپذیری گوینده » (SAR) جهت برآورد کارآیی واج‌ها از نظر بازشناسی گوینده استفاده می‌کنیم که برای تعیین مقدار این نسبت برای هر واج، ابتدا بایستی « فاصله درگوینده‌ای » و « فاصله بین‌گوینده‌ای » را برای واج مورد نظر با استفاده از روابطی که در زیربخش‌های ۱-۳ و ۲-۳ ارائه شدند، محاسبه کنیم. برای انجام محاسبات فاصله درگوینده‌ای برای واج $(IraSD_{p_t}) p_t$ از زیربخش ۱-۳ داریم:

$$IraSD_{p_t} = \frac{1}{M} \sum_{a=1}^M IraSD_{p_t}^a, \quad IraSD_{p_t}^a = \frac{1}{\frac{k}{2}(k-1)} \sum_{j=1}^K \sum_{i=1}^K \sum_{l=1}^n (c_i^j - c_l^i)^2, \quad i > j$$

براساس معلومات ارائه شده در بخش ۱-۴، در انجام محاسبات توسط روابط فوق داریم:
 $K=100$, $M=12$, $n=12$
 G_t ها همان LPCCهایی هستند که توسط روابط ارائه شده در بخش ۳-۵ برای هر فریم ۳۲ میلی ثانیه‌ای از سیگنال واج‌گونه‌ها محاسبه می‌شوند. حال روابط فوق را برای ۲۸ واج زبان فارسی (به جز واج /ʔ/) و به ازای معلومات اخیرالذکر محاسبه می‌کنیم. نتایج حاصل از این محاسبات، در جدول ۱ به صورت مرتب‌شده براساس مقادیر $IraSD_{p_t}$ ارائه شده‌اند.

^۱ Linear Prediction Cepstral Coefficient

جدول ۱: « فاصله درگوبندهای واج‌های گفتار فارسی

فاصله درگوبندهای	واج‌ها	فاصله درگوبندهای	واج‌ها	فاصله درگوبندهای
۰/۳۴۷	س	s	ک	c
۰/۳۳۶	آ	/	ج	ج
۰/۳۳۱	ش	.	چ	'
۰/۳۳۰	ای	i	ر	r
۰/۳۲۶	د	d	ت	t
۰/۳۲۴	ا	a	ح	h
۰/۳۲۲	ژ	[خ	x
۰/۳۱۶	او	u	گ	g
۰/۲۹۷	ی	j	ز	z
۰/۲۸۰	ن	n	ل	l
۰/۲۷۵	م	m	غ و ق	q
۰/۲۳۰	و	v	آ	o
۰/۲۲۹	ب	b	پ	p
۰/۲۲۲	ف	f	!	e

جهت انجام محاسبات « فاصله بین گوبندهای $(IerSD_{p_i})$ نیز از زیربخش ۳-۲ داریم:

$$IerSD_{p_i} = \frac{1}{M} \sum_{s=1}^M [\sum_{j=1}^{K_a} \sum_{i=1}^{K_a} \sum_{l=1}^n (c_l^j - c_l^i)^2]$$

C_1 ها و مقادیر M و n دقیقاً همانهایی هستند که در مورد « فاصله در گوبندهای » داشتیم، اما برای مقادیر K_a و $k_{\bar{a}}$ داریم:

$$K_a=100, \quad k_{\bar{a}}=11 \times 100=1100$$

حال رابطه اخیر را نیز برای ۲۸ واج مورد نظر و به‌ازای معلومات فوق‌الذکر محاسبه می‌کنیم و نتایج حاصل را به‌صورت مرتب‌شده بر اساس مقادیر $IerSD_{p_i}$ در جدول (۲) می‌آوریم.

(جدول ۲): « فاصله بین گوبندهای واج‌های گفتار فارسی

فاصله بین گوبندهای	واج‌ها	فاصله بین گوبندهای	واج‌ها
۱/۳۴۲	او	u	ج
۱/۳۳۷	خ	x	آ
۱/۳۱۸	س	s	ک
۱/۱۷۹	غ و ق	g	ر
۱/۰۴۲	ت	t	ل
۱/۰۲۳	ش	.	ی
۱/۰۰۹	م	m	آ
-/۹۹۷	و	v	چ
-/۹۸۳	ن	n	را
-/۹۵۶	ژ	[ح
-/۸۶۲	ف	f	ز
-/۸۳۵	د	d	آ
-/۷۸۵	پ	p	گ
-/۵۴۶	ب	b	ای

همانطوری که در بخش ۳-۳ بحث شد، ما کارایی واج‌ها را از نظر بازشناسی گوبنده توسط معیاری به‌نام « نسبت تأثیرپذیری گوبنده (SAR_{p_i}) برآورد می‌کنیم که توسط رابطه زیر محاسبه می‌شود:

$$SAR_{p_t} = \frac{IraSDp_t}{IraSDp_t}$$

ما SAR_{p_t} را به ازای ۲۸ واج گفتار فارسی محاسبه کرده و نتایج را در جدول (۳) می‌آوریم. از آنجا که در برخی از پژوهش‌های مرتبط با موضوع این تحقیق (نظیر بازشناسی گفتار و بازشناسی گوینده)، نشان داده شده است که بکارگیری فیلتر پیش‌تأکید^۱ در مرحله پیش‌پردازش که منتهی به استخراج ویژگی‌های سطح صوتی از سیگنال گفتار می‌شود (در اینجا استخراج ۱۲ ضریب LPCC)، تأثیر مثبتی در افزایش دقت بازشناسی دارد، از اینرو ما محاسبه «نسبت تأثیرپذیری گوینده» را برای واج‌های گفتار فارسی یعنی SAR_{p_t} ها را یکبار بدون فیلتر پیش‌تأکید و یکبار نیز با فیلتر پیش‌تأکید محاسبه کرده و نتایج را به صورت مرتب شده بر اساس مقادیر SAR_{p_t} در جدول (۳) می‌آوریم. قابل ذکر است که تابع انتقال فیلتر پیش‌تأکید در اینجا $H(z)=1-0.95z^{-1}$ می‌باشد.

(جدول-۳): «نسبت تأثیرپذیری گوینده» برای واج‌های گفتار فارسی

(کارایی واج‌های گفتار فارسی از نظر بازشناسی گوینده)

واج‌ها	SARها		واج‌ها	SARها	
	بدون فیلتر پیش‌تأکید	با فیلتر پیش‌تأکید		بدون فیلتر پیش‌تأکید	با فیلتر پیش‌تأکید
/	۵/۸۵۸	۵/۶۱۳	ک	۳/۰۴۸	۳/۰۸۳
y	۵/۱۳۴	۵/۰۰۲	گ و ق	۳/۰۱۳	۳/۰۰۹
v	۴/۳۳۴	۴/۳۴۳	ز	۲/۹۷۲	۳/۰۴۸
a	۴/۳۱۷	۴/۱۱۸	خ	۲/۹۴۹	۲/۹۹۴
u	۴/۲۵۳	۴/۱۸۰	ح	۲/۸۸۰	۲/۷۹۸
i	۴/۱۲۳	۳/۹۵۲	ر	۲/۶۷۵	۲/۶۶۵
e	۴/۱۱۶	۴/۱۰۴	ک	۲/۶۴۱	۲/۶۳۷
o	۴/۰۶۵	۴/۱۲۰	د	۲/۴۶۰	۲/۳۹۷
l	۴/۰۴۹	۳/۶۷۹	ب	۲/۳۸۶	۲/۴۳۶
f	۳/۸۸۰	۳/۸۱۲	چ	۲/۳۵۶	۲/۳۸۲
m	۳/۶۶۲	۳/۵۵۸	پ	۲/۲۳۲	۲/۲۷۶
s	۳/۵۱۰	۳/۵۶۴	ت	۲/۰۴۱	۱/۸۶۰
n	۳/۵۰۶	۳/۳۵۹	ث	۳/۲۰۱	۳/۲۰۵
.	۳/۱۰۰	۳/۰۶۵	ش	۳/۰۹۱	۲/۹۵۵

۵- تحلیل داده‌های حاصل از محاسبات

مقادیر مندرج در جدول (۱) «فاصله در گوینده‌ای» واج‌های گفتار فارسی را نشان می‌دهد. همانطوریکه قبلاً نیز اشاره شد، «فاصله در گوینده‌ای» یک واج در حقیقت بیانگر تفاوت‌ها در شیوه تولید آن واج در بافت‌های مختلف توسط یک گوینده است و از آنجا که در بازشناسی گوینده تفاوت‌ها در شیوه تولید واج‌ها توسط گوینده‌های مختلف ملاک تفکیک و تمایز گوینده‌ها از یکدیگر قرار می‌گیرند، از اینرو «فاصله در گوینده‌ای

« ابهام در تفکیک و تمایز گوینده‌ها ایجاد می‌کند. بنابراین هرچه مقدار « فاصله درگوینده‌ای » برای یک واج کمتر باشد، موجب ابهام کمتری در تفکیک و تمایز گوینده‌ها می‌شود و در نتیجه از نظر بازشناسی گوینده می‌تواند کارا تر باشد. حال با توجه به جدول (۱) ملاحظه می‌شود که میانگین « فاصله درگوینده‌ای » برای واکه‌ها حدود ۰/۳۳۵ و برای همخوان‌ها حدود ۰/۴۰۷ است. بنابراین بطور کلی می‌توان گفت که واکه‌ها در بازشناسی گوینده ابهام کمتری ایجاد کرده و در نتیجه کارایی بالاتری خواهند داشت. به همین صورت از مقادیر مندرج در جدول (۱) می‌توان استنتاج کرد که متوسط « فاصله درگوینده‌ای » برای واج‌های واکدار حدود ۰/۳۶۵ و برای بی‌واک‌ها حدود ۰/۴۴۷ است. پس در حالت کلی می‌توان نتیجه‌گیری کرد که واج‌های واکدار در مقایسه با بی‌واک‌ها موجب ابهام کمتری در بازشناسی گوینده می‌شوند و از اینرو می‌توان نتیجه گرفت که در حالت کلی واج‌های واکدار در بازشناسی گوینده نسبت به بی‌واک‌ها کارا تر هستند. حال با توجه به اینکه میانگین « فاصله درگوینده‌ای » برای واکه‌ها حدود ۰/۳۳۵ و برای کل واج‌های واکدار حدود ۰/۳۶۵ است، بنابراین از نظر دسته‌بندی‌های کلی واج‌ها، می‌توان گفت که واکه‌ها از نظر بازشناسی گوینده کارا ترین دسته بوده و در رتبه اول قرار دارند. جالب توجه است که با دقت در جدول (۱) ملاحظه می‌شود که کمترین « فاصله درگوینده‌ای » به ترتیب از آن واج‌های /f/، /b/، /v/، /m/، /n/ و /y/ است که هیچکدام واکه نیستند و بیشترین « فاصله درگوینده‌ای » به ترتیب از آن واج‌های /c/، /، /، /، /، /v/ و /h/ است که همگی جزو دسته همخوان‌ها هستند و این حقیقت تا حدودی با نتیجه قبلی که « واکه‌ها از نظر بازشناسی گوینده کارا ترین دسته هستند » مغایرت دارد. اما همانطوریکه در بخش ۳ بحث شد، کارایی واج‌ها را از نظر بازشناسی گوینده نمی‌توان صرفاً بر اساس مقادیر « فاصله درگوینده‌ای » نتیجه گرفت و بایستی به « فاصله بین گوینده‌ای » آنها نیز توجه داشت. جدول (۴) خلاصه بحث‌های فوق را نشان می‌دهد.

(جدول-۴): رتبه‌بندی دسته‌های کلی واج‌ها از نظر کارایی در بازشناسی گوینده با توجه به «

فاصله‌های درگوینده‌ای « واج‌ها

رتبه	میانگین « فاصله درگوینده‌ای »	دسته‌های کلی واج‌ها
۱	۰/۳۳۵	واکه‌ها
۲	۰/۳۶۵	واکدارها
۳	۰/۳۷۹	همخوان‌های واکدار
۴	۰/۴۰۷	کل همخوان‌ها (واکدار و بی‌واک)
۵	۰/۴۴۷	همخوان‌های بی‌واک

حال مقادیر « فاصله بین گوینده‌ای » واج‌ها را در جدول (۲) آورده‌ایم، مورد تجزیه و تحلیل قرار می‌دهیم. چنانکه در بخش ۳-۲ نیز گفته شد، « فاصله بین گوینده‌ای » واج‌ها تفاوت‌های ویژگی‌های صوتی واج‌ها را که ناشی از مشخصه‌های انفرادی گوینده‌هاست، نشان می‌دهد و لذا رابطه مستقیم با میزان تفکیک و تمایز گوینده‌ها دارد. بنابراین هرچه « فاصله بین گوینده‌ای » واجی بیشتر باشد، انتظار می‌رود که کارایی آن واج در بازشناسی گوینده بالاتر باشد. مقادیر مندرج در جدول (۲) نشان می‌دهند که میانگین « فاصله بین گوینده‌ای » برای واکه‌ها حدود ۱/۴۲۹ و برای همخوان‌ها حدود ۱/۲۳۶ است. پس بطور کلی می‌توان گفت که واکه‌ها در تفکیک و تمایز گوینده‌ها مؤثرتر از همخوان‌ها هستند. با استفاده از مقادیر مندرج در جدول (۲) میانگین « فاصله بین گوینده‌ای » برای واج‌های واکدار حدود ۱/۳۵۶ و برای بی‌واک‌ها حدود ۱/۲۲۲ برآورد می‌شود. بنابراین در حالت کلی می‌توان گفت که واج‌های واکدار از نظر بازشناسی گوینده در مقایسه با بی‌واک‌ها مؤثرترند. با توجه به اینکه میانگین « فاصله بین گوینده‌ای » برای واکه‌ها ۱/۴۲۹ و برای کل واج‌های واکدار حدود ۱/۳۵۶ است. بنابراین از نظر « فاصله بین گوینده‌ای » نیز واکه‌ها برای بازشناسی گوینده کارا تر از دیگر دسته‌های کلی واج‌ها هستند. با دقت در جداول ۱ و ۲ ملاحظه می‌شود که هر دو « فاصله در گوینده‌ای » و « فاصله بین گوینده‌ای » برای سه همخوان /C/، /l/ و /r/ بالاست و در مقابل هر دو فاصله مذکور برای دو همخوان /f/ و /b/ پایین است و از اینرو نبایستی انتظار داشت که این واج‌ها در بازشناسی گوینده کارایی قابل توجهی داشته باشند. نتایج تحلیل‌های اخیر در جدول (۵) خلاصه شده است.

(جدول-۵): رتبه‌بندی دسته‌های کلی واج‌ها از نظر کارایی در بازشناسی گوینده با توجه به «

فاصله‌های بین گوینده‌ای « واج‌ها

رتبه	میانگین « فاصله بین گوینده‌ای »	دسته‌های کلی واج‌ها
۱	۱/۴۲۹	واک‌ها
۲	۱/۳۵۶	واک‌دارها
۳	۱/۲۴۵	همخوان‌های واکدار
۴	۱/۲۳۶	کل همخوان‌ها (واکدار و بی‌واک)
۵	۱/۲۲۲	همخوان‌های بی‌واک

جالب توجه است که نتایج حاصل از تحلیل « فاصله در گوینده‌ای » واج‌ها و « فاصله بین گوینده‌ای » واج‌ها با هم سازگار بوده و مؤید همدیگرند.

چنانکه در بخش ۳-۳ نیز گفته شد، هیچکدام از فواصل در گوینده‌ای و بین گوینده‌ای واج‌ها به تنهایی نمی‌توانند به‌عنوان معیار کارایی آنها در بازشناسی گوینده تلقی شوند و

بایستی با توجه به رابطه معکوس « فاصله در گوینده‌ای » واج‌ها با توانایی تفکیک و تمایز گوینده‌ها و بالعکس رابطه مستقیم « فاصله بین گوینده‌ای » واج‌ها در تفکیک و تمایز گوینده‌ها، معیار کارایی واج‌ها را از نظر بازشناسی گوینده، بصورت نسبت آن دو در نظر بگیریم که ما در بخش ۳-۳ آن را تحت عنوان « نسبت تأثیرپذیری گوینده » از واج مطرح کرده و به صورت زیر فرموله کردیم:

$$SAR_{p_t} = \frac{IerSDp_t}{IraSDp_t}$$

و مقادیر SAR_{p_t} ‌ها را به ازای ۲۸ واج مورد نظر یکبار بدون فیلتر پیش‌تأکید و یکبار نیز با فیلتر پیش‌تأکید محاسبه و در جدول (۳) می‌آوریم. اولین نکته‌ای که با توجه به جدول (۳) جلب توجه می‌کند اینست که در نظر گرفتن فیلتر پیش‌تأکید در مرحله پیش‌پردازش تأثیر ناچیزی در نتایج حاصل داشته و نتایج کلی را تغییر نمی‌دهد و تنها در جزئیات نتایج تأثیرگذار است. به عنوان مثال از جدول (۳) ملاحظه می‌شود که هم با فیلتر پیش‌تأکید و هم بدون فیلتر پیش‌تأکید، واکه‌ها در رتبه‌های اولی و انفجاری‌ها در رتبه‌های آخری قرار دارند و از این نظر می‌توان گفت که فیلتر پیش‌تأکید بی‌تأثیر بوده است. اما وقتی که در جزئیات دقت کنیم ملاحظه می‌کنیم که مثلاً رتبه واج /a/ با فیلتر پیش‌تأکید ۵ و بدون فیلتر پیش‌تأکید ۴ است و یا اینکه رتبه‌های دو واج انفجاری /b/ و /d/ با فیلتر پیش‌تأکید و بدون فیلتر پیش‌تأکید جایگزین یکدیگر می‌شوند. نکته جالب توجه دیگر در جدول (۳) اینست که دو واج /y/ و /v/ در ردیف واج‌های رتبه اول یعنی واکه‌ها قرار گرفته‌اند. می‌دانیم که واج‌های /y/ و /w/ از جمله واج‌های « غلت^۱ » هستند که زبان به‌هنگام تولید این واج‌ها به سرعت به سوی وضعیت واکه مجاور حرکت می‌کند و سیگنال صوتی تولیدی مشابه سیگنال واکه می‌شود و از اینرو به این واج‌ها، نیم‌واکه^۲ نیز گفته می‌شود (مشکوه‌الدینی، ۱۳۸۸: ۳۴). در زبان فارسی واج /w/ وجود ندارد اما فارسی زبانان واج /v/ را در موارد قابل توجهی به صورت تقریبی یا « قریب به صحت^۳ » تولید می‌کنند و سیگنال صوتی واج /v/ مشابه سیگنال صوتی واج /w/ تولید می‌شود. از اینرو، واج‌های /y/ و /v/ از نظر رتبه کارایی در بازشناسی گوینده، در ردیف واکه‌ها آمده‌اند. حال با توجه به مقادیر مندرج در جدول (۳) و نیز با توجه به بخش‌های فوق نتایج نهایی را توسط جداول (۶) و (۷) ارائه می‌کنیم.

1 Glide

2 Semi- vowel

3. Approximation

(جدول-۶): رتبه‌بندی دسته‌های کلی واج‌ها از نظر کارایی در بازشناسی گوینده

با توجه به «نسبت تأثیرپذیری گوینده (SAR_{Pt})»

رتبه	دسته‌های کلی واج‌ها
۱	واکه‌ها و نیم‌واکه‌ها
۲	خیشومی‌ها، سایشی‌ها و روان‌ها
۳	انسدادی‌ها ^۱ و انفجاری‌ها ^۲

(جدول-۷): رتبه‌بندی واج‌های گفتار فارسی از نظر کارایی در بازشناسی گوینده

با توجه به «نسبت تأثیرپذیری گوینده» (بدون فیلتر پیش‌تأکید)

رتبه	واج	رتبه	واج	رتبه	واج	رتبه	واج	رتبه	واج
۱	آ /	۸	ا	۱۵	ج	۲۲	ر	r	
۲	ی	۹	ل	۱۶	ش	۲۳	ک	c	
۳	و	۱۰	ف	۱۷	گ	۲۴	د	d	
۴	آ	۱۱	م	۱۸	غ و ق	۲۵	ب	b	
۵	او	۱۲	س	۱۹	ژ	۲۶	چ	'	
۶	ای	۱۳	ن	۲۰	خ	۲۷	پ	p	
۷	!	۱۴	ز	۲۱	ح	۲۸	ت	t	

۶- نتیجه‌گیری و کارهای آتی

همانطوریکه در بخش ۲ گفته شد، واج‌ها به‌عنوان عناصر مستقل از متن در برخی از طرح‌های بازشناسی گوینده بخصوص در بازشناسی گوینده به روش مستقل از متن مورد توجه مهندسين و پژوهشگران قرار گرفته‌اند و به‌همین جهت تحقیقات قابل توجهی نیز در خصوص کارایی واج‌های زبان‌های گفتاری مختلف از نظر بازشناسی گوینده در جهان صورت گرفته است و در همین راستا ما نیز در این مقاله کارایی واج‌های گفتار فارسی مورد مطالعه و پژوهش قرار داده و بر اساس میزان کارایی واج‌ها آنها را رتبه‌بندی کردیم. در این تحقیق ما با انجام آزمایش‌ها و محاسبات نشان دادیم که از نظر دسته‌های کلی واجی، واکه‌ها و نیم‌واکه‌ها در رتبه اول، خیشومی‌ها، سایشی‌ها و روان‌ها در رتبه دوم و انسدادی‌ها و انفجاری‌ها در رتبه سوم به‌لحاظ کارایی در بازشناسی گوینده قرار می‌گیرند و همچنین رتبه‌بندی کلیه واج‌های گفتار فارسی (به‌استثنای واج /ʔ/) را توسط جدول (۷) ارائه کردیم. توجه به رتبه کارایی واج‌ها در طراحی سامانه‌های بازشناسی گوینده مبتنی بر واج‌ها، می‌تواند در ارتقای کیفیت سامانه مؤثر و مفید واقع شود. به‌عنوان مثال، هرگاه در یک طرح بازشناسی گوینده تنها از واکه‌ها و نیم‌واکه‌ها (دسته واج‌های رتبه اول از نظر کارایی در بازشناسی گوینده) استفاده شود. زمان پردازش در مرحله آزمون می‌تواند تا ۷۰٪ کاهش یابد و در عین حال دقت بازشناسی کاهش نیافته و حتی افزایش داشته باشد. همچنین با توجه به نتایج این تحقیق می‌توان در سامانه‌های بازشناسی گوینده به روش وابسته به متن، جهت ارتقای کیفیت سامانه،

متونی را انتخاب کرد که از نظر واج‌های با رتبه بهتر، غنی‌تر باشند. به‌علاوه در سامانه‌های بازشناسی گوینده که با گویه‌های کوتاه^۱ سروکار دارند، توجه به رتبه کارایی واج‌ها می‌تواند هم در دقت و هم در سرعت عمل سامانه مفید باشد. با توجه به مطالب بخش ۱-۴ ملاحظه می‌شود که در انجام آزمایش‌ها و محاسبات در این تحقیق، برای محاسبه معیار کارایی واج‌ها در بازشناسی گوینده یعنی SAR_{Pt}ها، به‌ازای هر گوینده و هر واج، ۱۰ نمونه واج‌گونه‌ای (و یا ۱۰۰ نمونه فریمی) در نظر گرفته شده است و با توجه به اینکه تعداد کمی از واج‌های گفتار فارسی دارای بیش از ۱۰ واج‌گونه هستند (بیجن‌خان، ۱۳۷۱ج)، از اینرو برای اینگونه واج‌ها، ۱۰ نمونه احتمالاً کفایت نمی‌کند و این نقیصه ممکن است نتایج را کمی تغییر داده باشد. از اینرو، یکی از کارهای آتی در این زمینه، می‌تواند افزایش تعداد نمونه‌ها برای هر گوینده و به‌ازای هر واج باشد. همچنین با توجه به مطالب بخش ۱-۴ ملاحظه می‌شود که ما در این تحقیق تعداد گوینده‌های جمعیت مورد نظر تحقیق را ۱۲ نفر (۸ مذکر و ۴ مؤنث) در نظر گرفته‌ایم و می‌توان در پژوهش‌های آتی تعداد گوینده‌ها را افزایش داد که مسلماً با تعداد گوینده‌های بیشتر، اعتبار نتایج تحقیق بالاتر خواهد بود. از جمله کارهای دیگر در این زمینه اینست که با لحاظ کردن رتبه‌های مندرج در جداول (۶) و (۷) در پیاده‌سازی سامانه‌های بازشناسی گوینده نتایج علمی و کاربردی این تحقیق را مورد ارزیابی قرار داد.

منابع

- بی‌جن‌خان، محمود و سید صالحی، سیدعلی (۱۳۷۶ الف). واج به‌عنوان یک عنصر زبانی، شناختی و پردازشی، اولین مجموعه مقالات پژوهشکده پردازش هوشمند علائم ۱-۶.
- بیجن‌خان، محمود و غفوریان، محمدعلی (۱۳۷۶ ب). آموزش و بازشناسی خودکار طبقات واجی در گفتار پیوسته فارسی با استفاده از منطق فارسی، اولین مجموعه مقالات پژوهشکده پردازش هوشمند علائم، ۷-۱۲.
- بیجن‌خان، محمود و سید صالحی، سیدعلی (۱۳۷۶ ج). بررسی واج‌گونه‌های زبان فارسی و استخراج فرکانس سازه‌ها، گزارش پژوهشی، مرکز تحقیقات پردازش هوشمند علائم.
- ثمره، یدالله (۱۳۶۸). آواشناسی زبان فارسی، مرکز نشر دانشگاهی، چاپ دوم.
- سید صالحی، سیدعلی و همکاران (۱۳۷۶). بازشناخت مستقل از گوینده واج‌های گفتار پیوسته فارسی با استفاده از ویژگی‌های تولیدی، اولین مجموعه مقالات پژوهشکده پردازش هوشمند علائم، ۱۳-۱۸.

شیخ‌زادگان، جواد (۱۳۷۴ الف). بررسی درجه اهمیت واج‌های زبان فارسی گفتاری از نقطه نظر بازشناسی گوینده، مجموعه مقالات دهمین کنفرانس بین‌المللی مهندسی برق ایران، ۱۸۰-۱۸۷.

شیخ‌زادگان، جواد (۱۳۷۴ ب). تعیین هویت گوینده بصورت مستقل از متن، رساله دکتری، دانشگاه تربیت مدرس، ۲۷-۳۵.

مدرسی قوامی، گلناز (۱۳۹۲). آواشناسی: بررسی علمی گفتار، انتشارات سمت، چاپ دوم. مشکوه‌الدینی، مهدی (۱۳۸۸). ساخت آوایی زبان، انتشارات دانشگاه فردوسی مشهد، چاپ ششم.

ABE, M. & Sagayam, S. 1990. Statistical Study on voice Individual Conversion Across Different Languages, ICSLP.

Atal, B.S. 1972. Automatic speaker recognition based on pitch contours, Acoust, Soc, Amer, 52:1972-1687.

Atal, B.S. 1974. Effectiveness of linear predication characteristics of the speech wave for Automatic speaker Identification and verification, JASA, 55, 6: 1304- 1312.

Bijankhan, M. Sheikhzadegan, J. Roohani, M.R. Samareh, Y. Lucas, K.. & Tebyani, M. 1994. FARSDAT – The speech Database of Farsi spoken Language, Proceeding SST – 94, vol. 11, Des-.

Doddington, G.R. 1970. A computer Method of speaker verification, Ph.D. dissertation, department of Electrical Engineering, University of Wisconsin Madison.

Eatok, J.P. & Mason, J.S.D. 1992. Phoneme performance in speaker Recognition, ICSLP.

Furui, S. 1986. Research on individuality features in speech waves and automatic speaker recognition techniques, Speech communication, 5, 2: 183- 197.

Goldstein, U.G. 1976. Speaker identification feature based on formant tracks, JASA, vol. 59, no. 1: 176-182, January.

Heuvel, H.V.D. & Rietveld, T. 1992. Speaker Related Variability in cepstral Representation of Dutch Speech Segments, ICSLP.

Li, K.P. & Hughes, G.W. 1974. Talker Differences as they Appear in correlation Matrices of continuous speech spectra, JASA, vol.55, No. 4: 833- 837.

Li, K.P. & Wrench, Jr.E.H. 1983. An Approach To Text- Independent Speaker Recognition with short utterances, proc. IEEE, Int. Conf. Acoust. Speech signal processing, Boston, MA, 1209: 555-558.

Lin, C.S. etal. 1990. Study of line spectrum pair frequencies for speaker Recognition, proc. ICASSP 90, vol.1: 277- 280.

Lumms, R.C. 1975. speaker verification by computer using speech Intensity for Temporal Registration, IEEE Trans. Audio Electroacoust vol.63, pp. 561- 580.

Markel, J.D. etal. 1977. Long Term Feature Averaging for speaker Recognition, IEEE Trans. ASSP, vol. PSSP- 25, No. 4: 330- 337.

Mastui, t. & Furui, S. 1992. Speaker Recognition Using Concatenated phoneme Models, ICSLP.

- Matsui, T. & Furui, S. 1990. Text Independent speaker Recognition using Vocal Tract and pitch Information, proc. ICSLP 90, vol. 1: 137- 140.
- Nolan, F. 1983. The phonetic basis of speaker recognition, Cambridge University press.
- Paliwal, K.K.. 1988. A study of line spectrum pair frequencies for speech Recognition, proc. ICASSP 88, vol. 1: 485- 488.
- Paul, J. & Rabinowit, A. 1979. Development of analytical methods for a semi- automatic speaker Identification system, Automatic speech and Speaker Recognition, IEEE Press: 390.
- Pruzcmsky, S. & Mathews, M.V. 1964. Talker Recognition Based on Analysis of variance, JASA, vol. 36, No. 11: 2041- 2047.
- Rose, R.C. & Reynolds, D.A. 1990. Text – indepent speaker Identification using Automatic Acoustic segmentation, ICASSP.
- Rose, R.C. & Reynolds, D.A. 1990. Text – Independent speaker Identification using Automatic Acoustic segmentation, proc. ICASSP 90, 551.
- Sambur, M.R. 1976. Speaker Recognition using orthogonal linear predication, IEEE Trances. ASSP, vol. ASSP 24, No. 4: 283- 289.
- Sambur, M.R. 1972. Selection of acoustic feature for speaker identification", IEEE Trans. ASSP – 23.
- Schwartz, R. etal. 1982. The Application of Probability Density Estimation to Text – Independent speaker Identification, proc. ICASSP 82, vol. 2: 1649- 1652.
- Shridhar, M. etal. 1981. Text- Independent speaker Recognition using orthogonal linear prediction, ICASSP – 81: 197- 204.
- SU, I.S. & etel. 1974. Identification of speaker by use of nasal coarticulation JASA, vol. 56, no. 6: 1876- 1882, December.
- Tou, J.T. & Gonzalez, R.C. 1974. : Pattern Recognition Principles, Addison Wesley Pulishing Company.
- Wolf, J.J. 1972. Efficient acoustic parameters for spesker recognition, JASA, vol. 51, no, 6, pp. 2044-2056, June.
- Yegnanarayana, B. etal. 1994. A speaker verification system using prosodic feature, ICSLP 94, vol. 4, pp. 1867-1870.