

ارائه روشهایی برای آموزش تلفظ و نوای گفتار زبان خارجی به کمک فناوری پردازش گفتار

بهزاد بختیاری^۱؛ محمد مهدی همایونپور^۲

چکیده

در این مقاله تحقیقات انجام گرفته در راستای طراحی و ساخت سیستمی است که در آن از فناوری پردازش گفتار برای آموزش تلفظ و آموزش نوای گفتار یک زبان به زبان آموزان استفاده شده است. روش‌های ارائه شده در این تحقیق قابل گسترش برای آموزش تلفظ و نوای گفتار یک زبان به نو آموزانی است که زبان مادری آنها هر زبان دیگر باشد. در این سیستم ارزیابی تلفظ در سطح واج انجام می‌گیرد. برای این منظور با استفاده از گویندگان بومی زبان مقصد، مدل‌های مرجعی ساخته می‌شود سپس گفتار زبان‌آموزان با این مدل مرجع مقایسه شده و امتیازاتی به واحدهای مختلف گفتار داده می‌شود و مکان‌هایی که در آن خطا رخ داده شده مشخص می‌گردند. در زمینه آموزش نوای گفتار، ارزیابی خودکار روانی گفتار و ارزیابی خودکار تغییرات فرکانس پایه گفتار مورد بررسی قرار می‌گیرد. برای ارزیابی خودکار روانی گفتار از ویژگی‌های زمانی مانند، سرعت گفتار، نسبت آوا به زمان، تعداد وقفه‌ها استفاده می‌شود و در ارزیابی خودکار تغییرات فرکانس پایه نیز از منحنی تغییرات گام استفاده شد.

کلمات کلیدی

آموزش زبان خارجی به کمک کامپیوتر، ارزیابی خودکار تلفظ و نوای گفتار، میزان بومی بودن گفتار.

Presentation of Methods for a Foreign Language Pronunciation and Prosody Teaching Using Speech Processing Technology

Behzad Bakhtiyari, Mohammad Mehdi Homayounpour

ABSTRACT

This paper concerns Computer-Assisted Language Learning. Speech processing technology was used to develop techniques to teach correct phoneme and intonation pronunciation to those who want to learn a new language. Evaluation of correct pronunciation is done in phoneme level. For this purpose, phoneme models are trained using speech data from speakers of the destination language. Speech data from a student is compared to the phoneme models and the goodness of pronunciation of each phoneme is scored. The assessment is made at phoneme levels of the learners' speech, it will be possible to advise them of the exact errors. Fundamental frequency contour and fluency parameters are used to evaluate the goodness of intonation. Fluency is evaluated using speech temporal features, speech rate, rate of uttered phonemes and number of inserted pauses. Phoneme pronunciation and intonation scores are also used to evaluate the amount of nativeness of speakers and to evaluate how good a speaker utters a new language. The proposed techniques can be applied to any other languages.

^۱ کارشناس ارشد هوش مصنوعی و رباتیک، دانشگاه صنعتی امیرکبیر، دانشکده مهندسی کامپیوتر، Email: bakhtiyari@ce.aut.ac.ir

^۲ عضو هیات علمی دانشگاه صنعتی امیرکبیر، دانشکده مهندسی کامپیوتر، آزمایشگاه سیستم‌های هوشمند صوتی گفتاری،

Email: homayoun@aut.ac.ir

KEYWORDS

Computer-Assisted Language Learning, Pronunciation, Intonation, Nativeness evaluation, Speech Processing.

تلفظ مقایسه می‌گردد.

۱- مقدمه

در زمینه ارزیابی و آموزش نوای گفتار به کمک کامپیوتر، کار زیادی انجام نشده است. از بین ویژگی‌های نوای گفتار می‌توان به سلیس و روان بودن گفتار اشاره نمود. این مشخصه در ارتباطات شفاهی از اهمیت خاصی برخوردار است. از کارهای ارائه شده می‌توان به [۷]، [۸]، [۹] اشاره کرد. در [۷]، روانی در گفتار بصورت بیان همه کلمات در محدوده کشش از پیش تعریف شده‌ای که از میانگین گفتار بومی بدست می‌آید، تعریف می‌شود. در [۸]، [۹] ۹ ویژگی مختلف شامل کشش بخش‌ها با یکدیگر ترکیب می‌شوند تا معیاری را برای روانی گفتار تشکیل دهند. در این کار نشان داده شده که بین ارزیابی خودکار روانی گفتار و ارزیابی داوران انسانی در روانی گفتار همبستگی وجود داشته و در نتیجه می‌توان به ارزیابی مبتنی بر کامپیوتر اعتماد نمود.

در زمینه آموزش تغییرات زیر و بمی گفتار به کمک کامپیوتر، اکثر کارهای صورت گرفته، بدین صورت است که با نمایش تصاویری مانند منحنی تغییرات گام مربوط به زبان آموز و منحنی تغییرات گام درست سعی شده است تا به زبان آموز در بیان صحیح نوای گفتار کمک شود. اما هیچ‌گونه ارزیابی یا امتیازدهی نسبت به نوای گفتار زبان آموز صورت نمی‌گیرد. از کارهای انجام شده در این زمینه می‌توان به [۱۰]، [۱۱] اشاره نمود. در این کارها تنها منحنی تغییرات گام زبان آموز و منحنی تغییرات گام مرجع نمایش داده می‌شود و زبان آموز بایستی خود سعی کند جمله را به نحوی بیان نماید تا منحنی تغییرات گام وی شبیه منحنی تغییرات گام مرجع گردد.

در [۱۱] روشی برای آموزش نوای گفتار در دو سطح هجا و جمله ارائه شده است که با پردازش گفتار زبان آموز و گفتار مرجع، محل‌هایی از گفتار مانند تکیه نابجا را که زبان آموز در آنها دچار اشتباه شده است، مشخص می‌نماید.

۲- دادگان غیر بومی

از آنجا که هدف پژوهش حاضر آموزش زبان مقصد به زبان‌آموزانی است که زبان مادری‌شان با زبان مقصد متفاوت است، بایستی دادگان غیر بومی از گویندگان زبان مبدأ تهیه شود تا بتوان کارایی الگوریتم‌های ارائه شده را بررسی نمود. برای این منظور زبان مبدأ، زبان ترکی و زبان مقصد، زبان فارسی انتخاب شد. دلیل انتخاب زبان فارسی این بوده است که الگوریتم‌های ارائه شده در آموزش زبان به کمک کامپیوتر در

امروزه با توجه به اهمیت بالای آموزش زبان خارجی، می‌توان به کاربردهای مهم و مفید فناوری پردازش گفتار و استفاده از این فناوری در آموزش زبان خارجی اشاره نمود. از این فناوری می‌توان در دو زمینه آموزش تلفظ و آموزش نوای گفتار استفاده نمود. در این مقاله روش‌های موجود در زمینه آموزش تلفظ و آموزش نوای گفتار مورد بررسی و روش‌های جدیدی نیز ارائه شده است. در اینجا هدف اینست که سیستم ارزیابی خودکاری ارائه شود تا گفتار زبان‌آموزان را ارزیابی نموده و به درستی گفتار آنها از نظر تلفظ، روانی گفتار و تغییرات فرکانس پایه گفتار، امتیازی داده شود.

تحقیقاتی در مورد چگونگی و صحت استفاده از شناسایی خودکار گفتار برای آموزش زبان خارجی انجام شده است [۱]. همچنین سیستمی تحت عنوان "Tell Me More French" ارائه شده است که امتیازی به تلفظ‌ها داده می‌شود و علاوه بر آن شکل موج گفتار زبان‌آموزان نیز نمایش داده می‌شود [۲]. محققان در به این نتیجه رسیدند که گاهی اوقات گویندگان بومی نمره کمتری نسبت به گویندگان غیر بومی از سیستم دریافت می‌کنند، زیرا حروف انتهای گفتار از بین می‌روند که در نتیجه برای سیستم شناسایی خودکار گفتار، قابل قبول نخواهد بود [۳].

در حالت ایده‌آل، سیستم بایستی بتواند جزئیات مشکل تلفظ را اطلاع داده و جهت رفع آن پیشنهادهایی را مانند یک معلم به زبان‌آموز بدهد. در این مورد آقای هینکس توضیح داد که اکثر سیستم‌های تجاری به دنبال عیب‌یابی‌های خاصی هستند و بازخوردهای ارزیابی شده‌ای را به جای بازخوردهای اصلاح کننده به کاربر می‌دهند [۴]. همچنین سعی شده است که بازخوردهای اصلاح کننده‌ای به زبان‌آموز داده شود که این امر مفید است [۵].

در [۶] که از کارهای معروف انجام شده در زمینه ارزیابی تلفظ در سطح واج می‌باشد، برای ارزیابی از الگوریتم میزان خوبی تلفظ استفاده شده است. ارزیابی‌های خودکار تلفظ که در سطح واج انجام می‌گیرد نسبت به روش‌هایی که در سطوح بزرگتری مانند جمله، مفیدتر می‌باشند، زیرا زبان‌آموزان از محل دقیق خطای خود آگاه می‌شوند و سعی در رفع آن می‌نمایند. در اینجا روش جدیدی برای ارزیابی تلفظ در سطح واج ارائه می‌شود که کارایی آن با کارایی الگوریتم میزان خوبی

آموزش زبان فارسی تا کنون کمتر مورد بررسی قرار گرفته‌اند و همچنین زبان ترکی به دلیل دسترس بودن ترک زبانان در کشور انتخاب شد.

برای این منظور، ۳۴ جمله از دادگان فارس دات ۱۰۰ انتخاب شدند. این دادگان تقریباً شامل ۴۰۰ جمله می‌باشد که بعضی از جملات به گونه‌ای طراحی شده‌اند که تعداد زیادی از واجهای زبان فارسی را بپوشانند. لازم به ذکر است از این دادگان برای آموزش مدلها نیز استفاده شده است. البته گفتار استفاده شده از این دادگان جملاتی به غیر از این ۳۴ جمله می‌باشد. برای تهیه دادگان مناسب سعی شد از ترک زبانان افرادی با تحصیلات، جنسیت و سنین مختلف انتخاب شود. در کل گفتار ۱۱ گوینده ترک زبان ضبط شد.

بعد از تهیه دادگان مذکور بایستی جملات بیان شده توسط داوران انسانی مورد ارزیابی قرار گیرند تا نتایج ارزیابی خودکار و ارزیابی انسانی با هم مقایسه گردند. داوران ۹ تن از دانشجویان کارشناسی ارشد و دکتری رشته کامپیوتر با زمینه فعالیت پردازش گفتار بوده‌اند. این دانشجویان آشنایی کامل به آواشناسی و مفاهیم پردازش گفتار داشته‌اند.

این ارزیابی‌ها برای هر گوینده غیر بومی، در چهار زمینه تلفظ، روانی، تغییرات فرکانس پایه و میزان بومی بودن گفتار صورت می‌گیرد. در زمینه تلفظ، ارزیابی‌ها در سطح واج بوده و داوران درستی یا نادرستی هر واج را تعیین می‌کنند. در زمینه روانی گفتار، ارزیابی‌ها در دامنه ۱ تا ۵ صورت می‌گیرد که ۱ مربوط به روان‌ترین و ۵ مربوط به بدترین گفتار از لحاظ روانی گفتار می‌باشد. در زمینه ارزیابی تغییرات فرکانس پایه، ارزیابی‌ها در دامنه ۱ تا ۳ صورت گرفت که ۱ مربوط به صحیح‌ترین و ۳ مربوط به نادرست‌ترین گفتار از لحاظ تغییرات فرکانس پایه گفتار می‌باشد.

۳- ارزیابی خودکار تلفظ

در این قسمت، روش‌های ارزیابی خودکار تلفظ بررسی می‌شوند. از جمله روش‌هایی که به منظور ارزیابی خودکار تلفظ در سطح واج استفاده می‌شود، روش میزان خوبی تلفظ است [۶]. علاوه بر این در اینجا روش جدیدی به نام N-Best ارائه می‌شود و سپس نتایج آن با الگوریتم میزان خوبی تلفظ مقایسه می‌شود. از آنجا که برای ارزیابی خودکار نیاز به شناسایی‌کننده‌های خودکار می‌باشد، شناسایی‌کننده‌هایی مبتنی بر مدل مخفی مارکوف طراحی می‌شود که نحوه ساخت آنها در ادامه آمده است.

۳-۱- شناسایی‌کننده‌های مبتنی بر مدل مخفی مارکوف

برای شناسایی در سطح واج، هر واج با یک مدل مخفی مارکوف مدل می‌شود. از آنجا که انجام آزمایش‌ها در این مقاله مربوط به آموزش زبان فارسی به ترک زبانان می‌باشد، بایستی شناسایی‌کننده‌هایی برای واج‌های زبان فارسی طراحی شود. در اینجا ۲۹ واج زبان فارسی و یک برچسب برای سکوت در نظر گرفته شده است. هر واج با یک مدل مخفی مارکوف سه حالتی چپ به راست مدل می‌شود.

با توجه به این که ضرایب کپسترال مبتنی بر معیار مل در محیط‌های عملیاتی و نویزی از سایر ضرایب بهتر عمل می‌کنند، این ضرایب به همراه لگاریتم انرژی و مشتقات اول و دوم آن جهت بردارهای ویژگی انتخاب شده‌اند. بُعد بردارهای یاد شده ۳۹ می‌باشد. جهت استخراج ویژگی از قاب‌هایی با طول ۲۰ میلی ثانیه و همپوشانی ۲۰ میلی‌ثانیه استفاده شده است که بر روی آنها پنجره در طرفین کاهنده همینگ اعمال شده است تا اثر ناپیوستگی در ابتدا و انتهای قاب کمتر شود. پس از قاب‌بندی، طیف سیگنال گفتار به کمک یک فیلتر پیش‌تأکید که دارای ضریب ۰/۹۷۵ می‌باشد، هموار شده است. تعداد فیلترهای مورد استفاده در بانک فیلتر مل برابر ۱۸ بوده است.

از چالش‌های مهم در این مطالعه می‌توان به شناسایی خودکار گفتار غیر بومی اشاره نمود. چون مدل‌های شناسایی خودکار گفتار مربوط به زبان مبدأ می‌باشند و با گفتار بومی آموزش می‌بینند، بیشتر این شناسایی‌کننده‌ها کارایی کمتری برای شناسایی گفتار غیر بومی دارند. مشابه [۱۲] در اینجا نیز برای افزایش کارایی شناسایی‌کننده‌های گفتار زبان فارسی، مدل‌ها با گفتار زبان ترکی تطبیق داده می‌شوند. برای این منظور از روش MLLR استفاده می‌شود. از هر گوینده موجود در دادگان غیر بومی نزدیک ۳ جمله انتخاب شد که در مجموع ۲۸ جمله غیر تکراری از ۳۴ جمله مورد استفاده قرار گرفت.

۳-۲- الگوریتم میزان خوبی تلفظ

در این روش شناسایی‌کننده‌ها مبتنی بر مدل مخفی مارکوف هستند. برای این منظور مجموعه‌ای از مدل‌های مخفی مارکوف برای واج‌های موجود در زبان مقصد با استفاده از گفتار بومی آموزش می‌بینند. با اختیار داشتن این مجموعه مدل‌ها می‌توان $P(O^{(q)} | q)$ را برای هر واج q محاسبه نمود، که $O^{(q)}$ بخش آکوستیکی مربوط به واج q می‌باشد. با توجه به بخش آکوستیکی $O^{(q)}$ ، میزان خوبی تلفظ برای واج p بصورت رابطه (۱) تعریف می‌شود:

$$GOP(p) = \left| \log \left(\frac{P(O^{(p)}|p)}{\sum_{q \in Q} P(O^{(p)}|q)} \right) \right| / NF(p) \quad (1)$$

که Q مجموعه تمام واحه‌های موجود و $NF(p)$ تعداد قاب‌های بخش آکوستیکی $O^{(q)}$ می‌باشد.

بعد از محاسبه معیار میزان خوبی تلفظ توسط یک سطح آستانه درستی یا نادرستی تلفظ واج مورد نظر تعیین می‌شود. پیشنهاد ساده و کارآمد برای تعیین سطح آستانه استفاده از ابزارهای آماری می‌باشد. بعنوان مثال، می‌توان از میانگین و انحراف معیار مربوط به مقادیر میزان خوبی تلفظ آوای p در تلفظ‌های صحیح استفاده نمود. برای این منظور دادگانی را که در آن کلمات به درستی تلفظ شده‌اند، مورد ارزیابی قرار داده و از میانگین و انحراف معیار امتیازات هر واج برای تعیین سطح آستانه واج مورد نظر استفاده می‌شود.

۳-۳- ارزیابی N-Best

به طور کلی اگر نوعی تلفظ، تلفظ صحیح در نظر گرفته شود، هدف ارزیابی تلفظ، بررسی میزان شباهت گفتار تلفظ شده با تلفظ صحیح فرض شده می‌باشد. با الهام از این واقعیت، الگوریتم N-Best طراحی می‌شود. در ارزیابی خودکار تلفظ، تلفظ‌های صحیح، همان مدل‌هایی از واج هستند که با گفتار زبان بومی آموزش داده شده‌اند. یعنی مدل‌های واحه‌های زبان بومی، مدل‌های مرجع هستند که نشان دهنده تلفظ صحیح می‌باشند. در نتیجه هدف، بررسی میزان شباهت گفتار بیان شده با این مدل‌های مرجع می‌باشد. این میزان شباهت را می‌توان با شناسایی‌کننده N-Best انجام داد. شناسایی‌کننده یاد شده میزان شباهت ورودی به مدل‌های مرجع را به ترتیب از بیشترین شباهت تا کمترین شباهت نشان می‌دهد.

در این روش هر واج توسط یک مدل مخفی مارکوف مدل می‌شود. این مدل‌ها با گفتار زبان بومی آموزش می‌بینند. ابتدا گفتار زبان آموزان به رشته‌ای از ضرائب کپسترال مبتنی بر معیار مل^۲ تبدیل می‌شود. سپس با استفاده از نگارش آوایی گفتار و الگوریتم مرزبندی احتمالی مرز واحه تعیین می‌گردد. با اختیار داشتن مرز واحه، ویژگی‌های مربوط به هر واج جدا می‌شود. در مرحله بعد بر روی هر واج جدا شده یک شناسایی N-Best صورت می‌گیرد. این شناسایی‌کننده، شناسایی‌کننده واحه‌های زبان بومی می‌باشد که شناسایی را بصورت N-Best انجام می‌دهد. در این حالت واج ورودی توسط شناسایی‌کننده N-Best به N واج شناسایی می‌شود که عبارت است از اولین (بهترین) واج شناخته شده تا N امین واج. در بهترین حالت، اگر واجی درست تلفظ شود این واج ورودی اولین واج شناخته شده

در خروجی شناسایی‌کننده خواهد بود و به نسبت هر چقدر از درستی تلفظ آن کاسته شود، واج مذکور در خروجی‌های بعدی شناسایی‌کننده قرار خواهد گرفت. در نتیجه امتیازی که در این روش به تلفظ واحه داده می‌شود برابر رتبه آن در شناسایی‌کننده N-Best می‌باشد.

البته در حالتی که واج درست هم بیان شود ممکن است در خروجی به عنوان اولین واج شناخته نشود که این امر ناشی از عوامل مختلفی چون شباهت برخی واحه مانند $n/$ و $m/$ یا عدم دقت شناسایی‌کننده می‌باشد. به همین دلیل نباید انتظار داشت که همیشه واحه‌های درست تلفظ شده در اولین خروجی ظاهر شوند. در نتیجه با اعمال یک سطح آستانه می‌توان در نهایت درستی یا نادرستی تلفظ هر واج را تعیین نمود. از آنجا که مدل مخفی مارکوف یک واج با واحه‌های دیگر متفاوت است، استفاده از یک سطح آستانه عمومی موجب کاهش کارایی می‌شود. در نتیجه برای هر واج یک سطح آستانه تعریف می‌شود. یک روش ساده و کارآمد برای تعیین سطح آستانه برای هر واج، در نظر گرفتن سطح آستانه‌ای است که با آن سطح آستانه درصد دلخواهی از دادگان ارزیابی شده مورد پذیرش قرار گیرد. برای این منظور دادگانی که در آن کلمات به طور صحیح تلفظ شده‌اند، مورد ارزیابی قرار گرفته و برای هر واج تعیین می‌شود که در قسمت شناسایی N-Best چندمین واج شناسایی شده است. حال سطح آستانه برای آن واج طوری تعیین می‌شود که درصد دلخواهی از تعداد واحه‌های موجود در این دادگان پذیرفته شوند.

در واقع تفاوت روش میزان خوبی تلفظ با روش یاد شده در اینست که در این روش اعدادی که به عنوان احتمال رخداد واحه توسط مدل‌های مخفی مارکوف (امتیاز درست‌نمایی) بدست می‌آیند بطور مستقیم و کمی مورد استفاده قرار نمی‌گیرد بلکه تنها از ترتیب کمی این احتمال‌ها برای محاسبه امتیاز استفاده می‌شود و از آنجا که این احتمال‌ها از مدل‌های مخفی مارکوف بدست می‌آیند و بسیار به نویز و گوینده وابسته هستند این امر موجب می‌شود که در روش میزان خوبی تلفظ که از مقادیر بطور مستقیم استفاده می‌کند نتوان سطح آستانه مناسب با قابلیت اطمینان بالا پیدا نمود و کارایی آفت پیدا می‌کند.

علاوه بر این در [۱۳] استدلالی شده است که، مخرج کسر رابطه (۱) که حاصل جمع تمام امتیازات واحه‌های موجود می‌باشد، موجب از بین رفتن تأثیر آکوستیک زبان مقصد می‌شود و در واقع امتیازات به کمک آکوستیک زبان مبدأ نرمالیزه می‌شود. حال ایرادی که به روش میزان خوبی تلفظ وارد است این است که این الگوریتم از امتیازات یاد شده

استفاده می‌کند و برای تشخیص درستی یا نادرستی هر واج از سطح آستانه‌ای استفاده می‌شود که این سطوح از امتیازات شناسایی گفتار بومی حاصل شده‌اند. از آنجا که در شناسایی گفتار غیر بومی تغییری در این امتیازات رخ می‌دهد و این تغییر برای هر واج مختلف می‌باشد در نتیجه دلیلی ندارد که همه واج‌ها به یک اندازه تغییر یابند و در اینصورت سطوح آستانه از قبل تعریف شده قابل اطمینان نخواهد بود و عمل نرمالیزه کردن نیز فایده‌ای نخواهد داشت.

۳-۴- آزمایشات

جدول ۱، کارایی دو روش ارزیابی خودکار را نشان می‌دهد. تعداد کل واجها در این آزمایش برابر ۴۶۷۳ می‌باشد. با توجه به این جدول مشاهده می‌شود که روش ارائه شده در این مقاله کارایی بالاتری نسبت به الگوریتم میزان خوبی تلفظ دارد. برای محاسبه کارایی ارزیابی کننده‌های خودکار بدین صورت عمل می‌شود که هر بار یک داور انسانی مرجع قرار می‌گیرد و نتیجه حاصل از ارزیابی کننده خودکار با ارزیابی داور انسانی مورد مقایسه قرار گرفته و درصد درستی (کارایی) ارزیابی کننده خودکار محاسبه می‌شود. برای این منظور نسبت تعداد پذیرش و عدم پذیرش واجها که در ارزیابی خودکار و ارزیابی انسانی یکسان هستند به تعداد کل واجها محاسبه می‌شود. سپس مقادیر بدست آمده از ۹ داور میانگین گیری می‌شود و نتیجه حاصل به عنوان کارایی (دقت ارزیابی) ارزیابی کننده خودکار در نظر گرفته می‌شود.

با دقت در جدول ۱ مشاهده می‌شود که معیار پذیرش نادرست برای دو روش تفاوت چندانی ندارد (حدود ۱ درصد) اما معیار عدم پذیرش نادرست حدود ۸ درصد اختلاف دارد که موجب افزایش کارایی روش N-Best شده است.

جدول (۱): نتایج کارایی دو روش ارزیابی خودکار تلفظ

	N-Best +MLLR	N-Best	GOP +MLLR	GOP	
دقت ارزیابی (SA)	۹۱/۹۸	۹۱/۲۵	۸۵/۰۰	۸۲/۸۸	
پذیرش نادرست (FA)	۴/۹۴	۵/۴۶	۴/۶۰	۴/۲۴	
عدم پذیرش نادرست (FR)	۲/۰۸	۲/۲۹	۱۰/۴۰	۱۱/۷۸	

همچنین مشاهده می‌شود عمل تطبیق موجب افزایش کارایی بیشتری در روش میزان خوبی تلفظ نسبت به روش N-Best شده است. می‌توان نتیجه گرفت که روش میزان خوبی تلفظ حساسیت بیشتری نسبت به آکوستیک زبان مقصد و مدل‌های مرجع دارد و علاوه بر این نیز چون حساسیت بیشتری به مدل‌های مرجع وجود دارد پس شاید بتوان گفت حساسیت بیشتری نسبت به نویز هم خواهد داشت. همانطور که در بخش ۳-۳

اشاره شد دلیل این امر اینست که روش میزان خوبی تلفظ از مقادیر احتمال رخداد یک واج که از مدل مخفی مارکوف مربوط به خود بدست می‌آید، بطور مستقیم استفاده می‌نماید و از آنجا که این اعداد با تغییرات آکوستیکی و نویز بسیار متغیر می‌باشند و از طرف دیگر سطوح آستانه از قبل تعیین شده درستی یا نادرستی تلفظ واج را تعیین خواهند نمود با تغییر این اعداد، سطوح آستانه تعریف شده مناسب نخواهند بود و کارایی کاهش می‌یابد.

۴-۱- ارزیابی روانی گفتار

هدف از بررسی روانی گفتار در آموزش زبان، تأکید بر روی جنبه‌های زمانی روانی گفتار است. در جهت تعریف جنبه‌های زمانی روانی گفتار اغلب فرض می‌شود که هدف آموزش زبان است که شامل تولید گفتار گویندگان بومی با توجه به پارامترهای زمانی که جای ممکن شامل وقفه‌های سکوت، درنگ، وقفه‌های پر شده، شروع دوباره و اصلاح مجدد، و مانند آن نباشد.

۴-۱-۱- معیارهای کمی روانی گفتار

در این قسمت معیارهایی برای روانی گفتار معرفی خواهند شد. این معیارها بایستی به نحوی باشند که بتوان آنها را بصورت خودکار استخراج نمود. هدف از این معیارها اینست که بتوان به کمک آنها روانی گفتار را ارزیابی نمود. معیارهای معرفی شده در [۸]، [۹] جنبه زمانی روانی گفتار را در نظر گرفته‌اند و معیارهایی کمی می‌باشند. از بین معیارهای ارائه شده در این کارها، هفت معیار مهم انتخاب شده که عبارتند از:

۱- سرعت گفتار (ROS):

$$ROS = \frac{\text{تعداد واجهای بیان شده}}{\text{کل زمان گفتار با در نظر گرفتن وقفه‌های درون جمله}} \quad (2)$$

۲- نسبت آوا به زمان (PTR):

$$PTR = \frac{\text{کل زمان گفتار بدون وقفه‌ها}}{\text{کل زمان گفتار با در نظر گرفتن وقفه‌های درون جمله}} \times 100\% \quad (3)$$

۳- سرعت تعامل (ART):

$$ART = \frac{\text{تعداد واجهای بیان شده}}{\text{کل زمان گفتار بدون وقفه‌ها}} \quad (4)$$

۴- تعداد وقفه‌های سکوت (#P):

$$\#P = \text{تعداد وقفه‌های درون جمله‌ای بیشتر از } 0.2 \text{ ثانیه} \quad (5)$$

۵- کل زمان وقفه‌ها (TDP):

(۶) کل زمان وقفه‌های درون جمله‌ای بیشتر از $TDP = 0.2$ ثانیه

۶- میانگین طول وقفه‌ها (MLP):

(۷) میانگین کل زمان وقفه‌های درون جمله‌ای $MLP =$ بیشتر از 0.2 ثانیه

۷- میانگین طول تداوم (MLR):

(۸) میانگین تعداد واجهایی که بین دو وقفه $MLR =$ بیشتر از 0.2 ثانیه رخ می‌دهد.

هفت معیار یاد شده، از معیارهای معروفی هستند که محققان زیادی بر روی آنها مطالعاتی انجام داده‌اند. در اینجا با توجه به اینکه هدف ارزیابی خودکار روانی به کمک کامپیوتر می‌باشد از شناسایی کننده‌های خودکار برای این منظور استفاده خواهد شد.

۴-۲- ارزیابی کننده خودکار روانی گفتار

بعد از اینکه معیارهای کمی تعریف شده در بخش ۴-۱ مورد بررسی قرار گرفت و مناسب بودن آنها برای تخمین و ارزیابی روانی گفتار برای زبان فارسی مشاهده شد. در این بخش سعی خواهد شد با استفاده از معیارهای گفته شده به ارزیابی روانی گفتار گویندگان غیر بومی پرداخته شود. ابتدا معیارهای کمی روانی گفتار از گفتار استخراج شده و سپس توسط دسته‌بندی کننده ارزیابی روانی خودکار انجام می‌گیرد. خروجی ارزیابی خودکار روانی مشابه ارزیابی‌های داوران انسانی دارای ۵ کلاس خواهد بود.

به کمک شناسایی‌کننده‌ها (بخش ۳-۱) و نگارش آوایی گفتار بیان شده، مرزبندی تحمیلی به کمک الگوریتم ویتربی^۲ انجام می‌گیرد و مرز واج‌ها و سکوت تعیین می‌شود. با داشتن مرزهای مربوط به واج‌ها و سکوت‌ها معیارهای کمی یاد شده محاسبه می‌شوند. از آنجا که محل وقفه‌ها در گفتار زبان‌آموز مشخص نیست، در نگارش آوایی بعد از هر کلمه یک برچسب سکوت گذاشته می‌شود. بدیهی است در صورتی که وقفه‌ای وجود نداشته باشد، طول تعیین شده توسط مرزبندی تحمیلی برای سکوت در نظر گرفته شده کوچک خواهد بود و در صورتی که وقفه‌ای رخ دهد طول دقیق آن مشخص می‌گردد.

برای تعیین سطوح آستانه و دسته‌بندی سطح روانی گفتار با توجه به معیارهای کمی می‌توان از میانگین هر دسته مربوط به معیار کمی استخراج شده استفاده نمود. بدین صورت که دادگانی که توسط داوران انسانی مورد ارزیابی قرار گرفته

است، در نظر گرفته می‌شود و سپس میانگین هر دسته محاسبه می‌شود. در نتیجه برای هر دسته از سطح روانی گفتار مقدار میانگین معیار مربوط به آن دسته در اختیار خواهد بود. حال برای دسته‌بندی مقدار یک معیار به سطح روانی مربوطه کفایت فاصله مقدار معیار با تک تک میانگین هر دسته محاسبه شود، در نهایت به دسته‌ای تعلق خواهد گرفت که به آن نزدیک‌تر باشد. برای این منظور از ۴۰ درصد گفتار هر گوینده موجود در دادگان استفاده شد و از باقیمانده گفتارها برای آزمایش استفاده شد.

جدول ۲، همبستگی کلی ارزیابی انسانی و ارزیابی خودکار را نمایش می‌دهد. برای حصول این نتایج هر بار یک داور به عنوان مرجع در نظر گرفته شده و با مقایسه همبستگی بین ارزیابی او با نتایج ارزیابی خودکار کارایی محاسبه شده و در نهایت کارایی‌های بدست آمده توسط ۹ داور میانگین‌گیری شده تا کارایی نهایی بدست آید. با توجه به نتایج بدست آمده مشاهده می‌شود که معیار سرعت گفتار (ROS) بهترین ویژگی برای پیش‌بینی روانی گفتار می‌باشد. کمترین همبستگی نیز مربوط به معیار، میانگین طول تداوم (MLR) می‌باشد.

می‌توان برای ارزیابی خودکار روانی گفتار از ترکیب معیارهای مختلف کمی تعریف شده نیز استفاده نمود. در واقع در این حالت برای تخمین روانی گفتار از ویژگی‌های بیشتری استفاده می‌شود و بعد ورودی افزایش می‌یابد. برای دسته‌بندی، میانگین معیارهای استفاده شده بصورت برداری در نظر گرفته می‌شود و بردار ناشناخته به دسته‌ای تعلق می‌گیرد که فاصله کمتری با آن دسته دارد. با ترکیب دو تایی معیارها، بیشترین کارایی برای دو معیار سرعت گفتار (ROS) و تعداد وقفه‌ها (#P) بدست آمد که نتایج آن در جدول ۳ آمده است. حال با افزودن معیارهای دیگر به این دو معیار نتایج را بررسی می‌کنیم. نتیجه بدست آمده از این ترکیب در جدول ۴ آمده است. در این جدول معیارهای دیگر به ایز دو اضافه شده و نتایج برای ترکیب سه معیار مشخص شده است. نتایج بدست آمده نشان می‌دهد که افزودن معیار میانگین کل وقفه‌ها (MLP) به دو معیار اخیر فقط اندکی کارایی را افزایش می‌دهد. افزودن معیارهای دیگر نیز یا موجب کاهش کارایی می‌گردند و یا کارایی را تغییر نمی‌دهند.

جدول (۲): همبستگی بین ارزیابی انسانی و ارزیابی خودکار روانی گفتار با استفاده از معیارهای مختلف کمی روانی گفتار

#P	ROS	PTR	ART	TDP	MLP	MLR
۰/۸۹۲۱	۰/۹۳۴۵	۰/۸۱۸۶	۰/۸۹۰۲	۰/۸۶۳۱	۰/۸۵۰۶	۰/۷۷۲۰

جدول (۳): همبستگی بین ارزیابی انسانی و ارزیابی خودکار روانی گفتار با استفاده از ترکیب دو به دو معیارهای کمی روانی گفتار

	MLR	MLP	TDP	#P	ART	PTR	+
ROS	۰/۸۴۷۲	۰/۹۳۱۹	۰/۹۲۶۶	۰/۹۴۲۰	۰/۹۰۴۵	۰/۹۱۹۸	
PTR	۰/۸۷۲۰	۰/۸۱۸۶	۰/۸۱۸۶	۰/۸۱۸۶	۰/۹۲۷۷		
ART	۰/۸۲۶۹	۰/۸۹۷۵	۰/۸۹۷۵	۰/۹۰۱۱			
#P	۰/۷۹۹۷	۰/۸۹۹۴	۰/۸۹۵۶				
TDP	۰/۸۱۱۳	۰/۸۷۹۴					
MLP	۰/۷۹۰۷						

جدول (۴): همبستگی بین ارزیابی انسانی و ارزیابی خودکار روانی گفتار با استفاده از ترکیب سه تایی معیارهای کمی روانی گفتار

	MLR	MLP	TDP	ART	PTR	+
ROS, #P	۰/۸۴۷۲	۰/۹۴۵۸	۰/۹۲۷۱	۰/۹۰۱۹	۰/۹۱۹۸	

پخش می‌شود که به عنوان مرجع در نظر گرفته می‌شود و گوینده بایستی جمله را از لحاظ تغییرات فرکانس پایه به آن صورت بیان نماید و ارزیابی تغییرات فرکانس پایه وی با تغییرات فرکانس پایه جمله مرجع مقایسه می‌گردد. در نتیجه ارزیابی‌های انجام شده در اینجا دارای محدودیت‌هایی هستند که از جمله آنها می‌توان به محدود شدن نحوه بیان جمله از لحاظ نوایی اشاره نمود.

به طور کلی شاید بتوان ارزیابی تغییرات فرکانس پایه گفتار را به دو روش سراسری و محلی تقسیم‌بندی نمود. در ارزیابی سراسری تغییرات فرکانس پایه، تغییرات فرکانس پایه کل گفتار، فرازها و نشیب‌ها در طول کل گفتار مدنظر است در حالی که در ارزیابی تغییرات فرکانس پایه بصورت محلی، تغییرات فرکانس پایه در سطوح کوچک‌تری مانند، کلمه و هجا انجام می‌شود.

۱-۵- روش‌های سراسری

در این بخش با استفاده از معیار همبستگی بین منحنی تغییرات گام گفتار مرجع و منحنی تغییرات گام گفتار زبان‌آموز ارزیابی تغییرات فرکانس پایه صورت می‌گیرد. در این روش حتی برای تکرار یک جمله توسط یک گوینده نیز نباید انتظار داشت که همبستگی بین منحنی‌های تغییرات گام آنها ۱۰۰ درصد یا خیلی نزدیک به ۱۰۰ درصد شود. به عبارت دیگر اگر یک گوینده یک جمله را چندین بار به طور مشابه تکرار نماید، همواره همبستگی ۱۰۰ درصد بدست نخواهد آمد.

از مشکلات این روش می‌توان به عدم تساوی طولی دو منحنی تغییرات گام در محاسبه همبستگی آن دو اشاره نمود. در این حالت بایستی دو منحنی تغییرات گام مورد نظر از لحاظ طولی یکسان شوند تا بتوان همبستگی آن دو را محاسبه نمود. برای این منظور بعد از محاسبه منحنی تغییرات گام مربوط به زبان‌آموز، طول منحنی تغییرات بدست آمده با طول منحنی تغییرات گام گفتار مرجع مقایسه می‌شود و در صورت نامساوی بودن طول‌ها، منحنی تغییرات گام زبان‌آموز به حوزه فرکانس منتقل شده و عمل نمونه‌برداری با فرکانس نمونه-برداری (m/n) انجام می‌شود که n طول منحنی تغییرات گام مرجع و m طول منحنی تغییرات زبان‌آموز است. سپس یک فیلتر پائین‌گذر اعمال شده و نتیجه به حوزه زمان بازگردانده می‌شود. در واقع برای نمونه‌های موجود در حوزه زمان مربوط به منحنی تغییرات گام، عمل نمونه‌برداری دوباره انجام می‌گیرد. در اینجا هدف این است که منحنی تغییرات گامی که نسبت به منحنی تغییرات گام مرجع کوچک‌تر است را بکشیم تا بزرگ‌تر گردد و در حقیقت در اینجا شکل دو منحنی تغییرات با یکدیگر مقایسه می‌گردد.

با توجه به نتایج ارزیابی‌ها معلوم گردید که معیار سرعت گفتار (ROS) کارایی بالاتری نسبت به دیگر معیارها دارد. افزودن یک یا دو معیار دیگر باعث افزایش کارایی این معیار شد تا حدی که به نتایج همبستگی بین داوران انسانی بسیار نزدیک‌تر گردید. همچنین مشاهده شد که افزودن معیارهای بیشتر موجب افزایش کارایی ارزیابی خودکار نمی‌گردد.

۵- ارزیابی تغییرات فرکانس پایه گفتار به کمک

منحنی تغییرات گام

در این بخش روش‌هایی برای ارزیابی تغییرات فرکانس پایه توسط منحنی تغییرات گام ارائه خواهد شد. همانطور که در بخش‌های قبلی ذکر شد منحنی تغییرات گام حاوی اطلاعات مهمی از تغییرات فرکانس پایه (زیر و بمی) گفتار است. البته، هر چند که منحنی تغییرات گام حاوی اطلاعات مهم و ارزشمندی از تغییرات فرکانس پایه گفتار می‌باشد، با این همه از اشکالات اساسی آن اینست که تکرارهای یک جمله توسط یک گوینده همیشه منحنی تغییرات گام یکسانی را نتیجه نمی‌دهد. حال در ارزیابی تغییرات فرکانس پایه افراد مختلف کار بسیار دشوارتر از حالت یک گوینده خواهد شد. در اینجا برای ارزیابی تغییرات فرکانس پایه توسط منحنی تغییرات گام، برای هر جمله یک مرجع در نظر گرفته می‌شود و منحنی تغییرات گام مربوط به آن با منحنی تغییرات گام گفتار زبان‌آموزان مقایسه می‌شود. برای این منظور با توجه به این که دادگان غیر بومی حاوی ۳۴ جمله است، توسط یک گوینده فارسی زبان ۳۴ جمله با نوای درست بیان شد. از منحنی تغییرات گام این ۳۴ جمله در ارزیابی نوای گفتار زبان‌آموزان استفاده شد.

توجه شود که در اینجا برای گویندگان غیر بومی گفتاری

لازم به ذکر است که در این قسمت هم ردیف کردن زمانی طول دو گفتار صورت نمی‌گیرد، چون همانطور که اشاره شد ارزیابی‌ها در این بخش محدود به جملاتی خاص با نوای مشخصی هستند که در گفتار مرجع وجود دارد و علاوه بر این جملات بیشتر کوتاه هستند و هدف در این قسمت تنها مقایسه منحنی تغییرات گام دو گفتار می‌باشد. به همین دلیل فقط سعی شده است که شکل دو منحنی تغییرات با یکدیگر مقایسه شود و در صورت یکسان نبودن طول منحنی تغییرات گام زبان‌آموز، طول آن با طول منحنی تغییرات گام مرجع یکسان‌سازی می‌شود.

۲-۵- روش‌های محلی

هدف از ارائه روش‌های محلی در این بخش مقایسه بین منحنی تغییرات گام زبان‌آموز و منحنی تغییرات گام مرجع در سطوح کوچک‌تری مانند هجا (بجای کل جمله) است و ارزیابی نهایی نوای گفتار جمله مربوطه به کمک مقایسه انجام شده صورت می‌گیرد. در واقع در این روش‌ها ارزیابی در سطوح کوچک‌تر انجام نمی‌شود بلکه به کمک مقایسه‌های محلی ارزیابی کلی از نوای گفتار جمله مربوطه انجام می‌شود. در این بخش دو روش ارائه خواهد شد.

روش اول بطور مستقیم از منحنی تغییرات گام استفاده نموده و شیب آن را در هر هجا تقریب می‌زند. سپس شیب‌های بدست آمده از منحنی تغییرات گام زبان‌آموز با شیب‌های منحنی تغییرات گام مرجع مقایسه می‌شود و در نهایت توسط یک سطح آستانه ارزیابی نهایی انجام می‌شود. این روش "ارزیابی تقریب شیب" نامیده می‌شود. اما در روش دوم، منحنی‌های تغییرات گام (مرجع و زبان‌آموز) بصورت پارامتری مدلسازی می‌شوند و ارزیابی توسط پارامترهای مدل مربوطه انجام می‌شود.

۲-۵-۱ ارزیابی تقریب شیب

در این روش برای هر هجا در جمله، منحنی تغییرات گام هجای مربوطه جدا می‌شود و سپس منحنی تغییرات گام مربوط به هر هجا با یک خط تقریب زده می‌شود. این کار هم برای منحنی تغییرات گام مرجع و هم برای منحنی تغییرات گام زبان‌آموز انجام می‌شود. در نهایت با مقایسه شیب خط‌های بدست آمده ارزیابی صورت می‌گیرد. ایده و هدف اصلی این روش اینست که برخی از جزئیات منحنی تغییرات گام در نظر گرفته نشوند زیرا ممکن است برخی از این جزئیات به گوینده یا پارامترهای ناشناخته‌ای وابسته باشند و به همین دلیل نباید آنها را در ارزیابی درستی یا نادرستی تغییرات فرکانس پایه گفتار دخالت داد. هر چند وابستگی پارامترهایی که در اینجا نیز

استفاده می‌شوند به گوینده یا پارامترهای دیگر نامشخص است. به عبارت دیگر باید دید استفاده یا عدم استفاده از برخی جزئیات چه تأثیری بر ارزیابی خواهد داشت.

بعد از محاسبه شیب هجاهای موجود در گفتار زبان‌آموز و گفتار مرجع شیب‌های بدست آمده برای هر مورد به صورت یک بردار در نظر گرفته می‌شود. برای نمونه اگر جمله‌ای دارای N هجا باشد، دو بردار با N مؤلفه وجود دارد. حال برای ارزیابی و امتیازدهی به کمک این روش بدین صورت عمل می‌شود که میانگین مربع خطای دو بردار مربوط به شیب‌ها نرمالیزه شده و به عنوان امتیاز در نظر گرفته می‌شود.

۲-۲-۵ ارزیابی به کمک مدلسازی منحنی تغییرات گام

در این قسمت سعی خواهد شد تا با مدلسازی منحنی تغییرات گام مرجع و زبان‌آموز به ارزیابی تغییرات فرکانس پایه پرداخته شود. برای این منظور از مدل تیلت استفاده خواهد شد. در ادامه به بررسی مدلسازی تیلت و نحوه ارزیابی به کمک آن اشاره می‌گردد.

الف- مدلسازی منحنی تغییرات گام به کمک مدل تیلت

مدل تیلت جزء مدل‌های پارامتری پیوسته می‌باشد. در این مدل‌ها آهنگ به شکل رویدادهایی توصیف می‌شود و مشابه مدل‌های دیگر جزئیات بین رویدادها در نظر گرفته نمی‌شود. واحد پایه در مدل تیلت رویداد آهنگی یا رویداد زیر و بمی است. رویدادهای اصلی عبارتند از تکیه زیر و بمی و نواخت-های مرزی.

مدل تیلت از مجموعه‌ای از پارامترهای پیوسته، به نام پارامترهای تیلت برای تعیین شکل محلی منحنی تغییرات گام مربوط به یک رویداد، استفاده می‌کند. مدل تیلت بر مبنای یک مدل ساده‌تر به نام RFC ساخته شده است.

در مدل RFC هر رویداد با یک خیز که یک آفت به دنبال آن است، مدل می‌شود. هر قسمت دارای یک دامنه و یک کشش است و زمان و مقدار گام هر رویداد نیز به عنوان دو پارامتر دیگر در نظر گرفته می‌شود. مدل تیلت چهار پارامتر مربوط به دامنه و کشش آفت و خیزها را به سه پارامتر تبدیل می‌کند:

- ۱- دامنه (برحسب هرتز): مجموع دامنه‌های آفت و خیز.
- ۲- کشش (برحسب ثانیه): مجموع کشش آفت و خیز.
- ۳- عدد تیلت: یک عدد حقیقی بدون بُعد در فاصله ۱- تا ۱ که شکل کلی رویداد را مستقل از دامنه و کشش آفت و خیز، مشخص می‌کند.

ب- ارزیابی به کمک پارامترهای تیلت

مدلسازی منحنی تغییرات گام توسط پارامترهای تیلت در

بخش قبلی مشاهده شد. حال برای ارزیابی تغییرات فرکانس پایه گفتار به کمک پارامترهای تیلت طوری عمل می‌شود که منحنی تغییرات گام مرجع و منحنی تغییرات گام مربوط به گفتار زبان‌آموز به روش تیلت مدل و در هر هجا رُخداد یک رویداد در نظر گرفته شود. در این صورت برای جمله‌ای با N هجا، N بردار وجود خواهد داشت که هر بردار متناظر با هر هجا مقادیر پارامترهای تیلت مربوط به آن هجا را دارا می‌باشد. در ساده‌ترین حالت می‌توان فقط از یکی از چهار پارامتر استفاده نمود که در این صورت بردارها یک بُعدی خواهند بود یا در حالتی دیگر می‌توان از چهار پارامتر دامنه، عدد تیلت، فرکانس پایه و کشش استفاده نمود.

مشابه روش تقریب شیب، پارامترهای مربوط به هر هجا در منحنی تغییرات گام مرجع و زبان‌آموز با هم مقایسه شده و در نهایت به کمک نتایج این مقایسات امتیازی به تغییرات فرکانس پایه گفتار داده می‌شود. برای این منظور در حالتیکه فقط از یک پارامتر استفاده می‌شود میانگین مربع خطای دو بردار مربوط به هجاهای موجود در جمله، نرمالیزه شده و به عنوان امتیاز در نظر گرفته می‌شود و در حالتیکه تمام پارامترها در نظر گرفته می‌شوند بعد از محاسبه امتیاز حاصل از هر پارامتر به صورت جداگانه، میانگین امتیاز آنها در نظر گرفته می‌شود.

۳-۵ نتایج آزمایشات

در بخش‌های قبل سه روش برای ارزیابی تغییرات فرکانس پایه گفتار ارائه شد که اولی به صورت سراسری و دوتای دیگر بصورت محلی بودند. این روش‌ها به تغییرات فرکانس پایه گفتار زبان‌آموزان امتیازاتی می‌دهند. حال به کمک این امتیازات بایستی ارزیابی تغییرات فرکانس پایه برای گفتار انجام شود. در واقع با داشتن امتیازات مربوط به هر جمله عمل دسته‌بندی صورت می‌گیرد و تعیین می‌شود که تغییرات فرکانس پایه گفتار بیان شده در کدام کلاس قرار دارد.

برای تعیین سطوح آستانه و دسته‌بندی سطح درستی تغییرات فرکانس پایه گفتار زبان‌آموزان با توجه به امتیازات خودکار بدست آمده، بدین صورت عمل می‌شود که دادگانی که توسط داوران انسانی مورد ارزیابی قرار گرفته است، در نظر گرفته می‌شود و سپس با روش‌های قید شده به هر جمله امتیازی به میزان درستی تغییرات فرکانس پایه گفتار داده می‌شود. سپس از آنجا که کلاس درستی تغییرات فرکانس پایه گفتار هر جمله با دآوری انسانی مشخص است، میانگین امتیازات هر دسته بطور جداگانه محاسبه می‌شود. در نتیجه برای هر دسته از سطح درستی تغییرات فرکانس پایه گفتار مقدار میانگین امتیاز مربوط به آن دسته با توجه به روش

استخراج امتیاز آن در اختیار خواهد بود.

حال با داشتن این سطوح آستانه برای هر روش در زمان دسته‌بندی امتیاز یک جمله، کفایت فاصله مقدار امتیاز بدست آمده با تک تک میانگین هر دسته محاسبه شود و در نهایت کلاس درستی تغییرات فرکانس پایه گفتار جمله یاد شده به دسته‌ای تعلق خواهد گرفت که به آن نزدیک‌تر باشد.

کارایی روش اول برابر ۶۸٪ بدست آمد. کارایی به صورت تعداد دسته‌بندی‌های درست (از نظر داوران انسانی) نسبت به تعداد کل، محاسبه می‌شود. برای این منظور هر بار یک داور به عنوان مرجع در نظر گرفته می‌شود و کارایی با توجه به ارزیابی‌های آن داور محاسبه می‌شود. سپس میانگین کارایی‌های بدست آمده توسط داوران مختلف به عنوان کارایی روش منظور می‌شود. دسته‌بندی دو روش دیگر نیز مشابه می‌باشد. برای روش تقریب شیب نیز بیشترین کارایی برابر ۰/۶۸ بدست آمد.

جدول ۵، کارایی را برای حالتی که چهار امتیاز یاد شده با یکدیگر ترکیب شده‌اند را نشان می‌دهد. این امتیازات به ترتیب از راست به چپ مربوط به امتیازات حاصل از پارامترهای عدد تیلت، کشش، دامنه و فرکانس پایه می‌باشند. کارایی حاصل شده خیلی به کارایی داوران انسانی نزدیک نیست که این امر ناشی از پیچیدگی ارزیابی نوای گفتار می‌باشد.

جدول (۵): نتایج کارایی روش مدلسازی منحنی تغییرات گام با

ترکیب امتیازات	
امتیازات	$Score_{fre} + Score_{amp} + Score_{dur} + Score_{til}$
کارایی	۰/۷۰۲۱

۶- خلاصه و نتیجه‌گیری

در زمینه ارزیابی تلفظ به کمک کامپیوتر، دو روش میزان خوبی تلفظ و ارزیابی N-Best پیاده‌سازی شد. روش اول جزء معروف‌ترین روش‌های ارزیابی تلفظ در سطح واج می‌باشد. روش دوم نیز روشی جدید است که در اینجا ارائه شد و کارایی آن با کارایی روش میزان خوبی تلفظ مقایسه گردید. نتایج نشان می‌دهند که روش ارائه شده کارایی بالاتری دارد. این روش به نوعی مشکلات روش میزان خوبی تلفظ را بر طرف می‌نماید. همچنین مشاهده شد که استفاده از عمل تطبیق گوینده نیز موجب افزایش کارایی می‌گردد. به طور کلی نیز می‌توان گفت کارایی بدست آمده در زمینه ارزیابی خودکار تلفظ، کارایی قابل قبول و مناسبی است و بسیار نزدیک به ارزیابی‌کننده انسانی می‌باشد.

در زمینه ارزیابی خودکار روانی گفتار از برخی معیارهای کمی روانی گفتار که بیشتر جنبه‌های زمانی روانی گفتار را در

گام به دو روش سراسری و محلی تقسیم شد و روش‌های مختلفی مورد بررسی قرار گرفت. کارایی‌های بدست آمده در این بخش قدری قابل قبول بودند اما کارایی بدست آمده مانند کارایی‌های ارزیابی‌کننده‌های خودکار تلفظ و روانی گفتار خیلی به کارایی‌های انسانی نزدیک نبود. توجهی که می‌توان داشت این است که ارزیابی‌های نوای گفتار دارای پیچیدگی بیشتری بوده و مانند تلفظ یا روانی گفتار تعریف خاصی ندارد. چه بسا که حتی کارایی ارزیابی‌های انسانی در ارزیابی‌های نوای گفتار نیز مانند کارایی‌های ارزیابی‌های تلفظ و روانی گفتار بالا نیست. در پایان می‌توان گفت با توجه به کارایی‌های بدست آمده، بویژه در قسمت تلفظ و روانی، استفاده از ارزیابی‌کننده‌های خودکار برای آموزش زبان خارجی مناسب و قابل قبول می‌باشد. البته گفتنی است از آنجا که آموزش زبان روالی اجتماعی محسوب می‌شود و حضور معلم انسانی و تأثیر محیط آموزش سنتی را هرگز نمی‌توان نادیده گرفت، استفاده از کامپیوتر و بطور مشخص ارزیابی‌کننده‌های خودکار ارائه شده تنها در کنار آموزش سنتی پیشنهاد می‌شود.

- [۹] Cucchiari. C.; Strik. H.; Boves. L.; "Quantitative Assessment of Second Language Learners' Fluency: Comparisons Between Read And Spontaneous Speech", Journal of the Acoustical Society of America, Vol. 111 (6), p.p. 2862-2873, 2002.
- [۱۰] Neri. A.; Cucchiari. C.; Strik. H.; "Effective Feedback on L2 Pronunciation in ASR-Based CALL", Proc. Computer Assisted Language Learning, p.p. 40-48, 2001.
- [۱۱] Dorothy M., "Signal Analysis Software For Teaching Discourse Intonation", Language Learning & Technology, vol. 2, no. 1, p.p. 74-93, 1998.
- [۱۲] Ehsani. F.; "NTT-data Japanese-English ATC ASR System Description", Technical report, Entropic Inc, 1996.
- [۱۳] Neumeyer. L.; Franco. H.; Digalakis. V.; Weintraub. M.; "Automatic Scoring of Pronunciation Quality", Speech Communication, vol. 30, p.p. 83-93, 2000.
- [۱۴] Taylor. P.; "Analysis and Synthesis of Intonation Using the Tilt Model", Center for Speech Technology Research, University of Edinburgh, 2000.
- [۱۵] Franco. H.; Neumeyer. L.; Digalakis. V.; Ronen. O.; "Combination of Machine Scores for Automatic Grading of Pronunciation Quality", Speech Communication, vol. 30, p.p. 121-130,

بر می‌گیرند، استفاده شد. کارایی حاصل از ارزیابی‌کننده خودکار روانی گفتار ارائه شده مناسب بوده و بسیار نزدیک به ارزیابی‌کننده انسانی می‌باشد. لازم به ذکر است که ارزیابی خودکار روانی گفتار ارائه شده نیز مانند ارزیابی خودکار تلفظ، مستقل از متن بوده و با توجه به روش ارائه شده کفایت برای جمله مورد نظر فقط نگارش آوایی آن در دسترس باشد. در کارهای قبلی آموزش نوای گفتار، گفتار زبان‌آموز مورد ارزیابی خاصی قرار نمی‌گیرد و فقط با نمایش منحنی‌هایی مانند منحنی تغییرات گام به زبان‌آموز سعی می‌شود تا به زبان‌آموز کمک شود که گفتار خود را طوری بیان نماید که منحنی تغییرات گام وی مشابه منحنی تغییرات مرجع گردد. نشان داده شده است که این روش کارایی پایینی دارد. به همین دلیل در اینجا سعی شده است که روشی برای ارزیابی نوای گفتار زبان‌آموز ارائه شود. برای ارزیابی خودکار نوای گفتار از منحنی تغییرات گام مربوط به گفتار زبان‌آموز و منحنی تغییرات گام مرجع جمله بیان شده استفاده شد. برای نوای گفتار ۳ سطح در نظر گرفته شد. مقایسات دو منحنی تغییرات

۷- مراجع

- [۱] Neri. A.; Cucchiari. C.; Strik. H.; "Automatic Speech Recognition for Second Language Learning: How and Why It Actually Works", Proc. of 15th ICPhS, p.p. 1157-1160, 2003.
- [۲] Reesner. T.; "Tell Me More French," Software review, CALICO Journal, vol. 19, p.p. 419-428, 2002.
- [۳] Wildner. S.; "Learn German Now! Version 8" , Software review, CALICO Journal, vol. 20, p.p. 161-174, 2002.
- [۴] Hincks. R.; "Speech Recognition For Language Teaching And Evaluating: A Study Of Existing Commercial Products", Proc. ICSLP, p.p. 733-736, 2002.
- [۵] Neri. A.; Cucchiari. C.; Strik. H.; "ASR Corrective Feedback on Pronunciation: Does It Really Work?", Proc. of ICSLP, p.p. 1982-1985, 2006.
- [۶] Witt. S. M.; Young. S. J.; "Phone-level Pronunciation Scoring and Assessment for Interactive Language Learning", Speech Communication, vol. 30, p.p. 95-108, 2000.
- [۷] Eskenazi. M.; "Using Automatic Speech Processing for Foreign Language Pronunciation Tutoring: Some and Prototype", Language Learning and Technology, vol. 2, no. 2, p.p. 62-76, 1999.
- [۸] Cucchiari. C.; Strik. H.; Boves. L.; "Quantitative Assessment of Second Language Learners' Fluency by Means of Automatic Speech Recognition Technology", Acoustical Society of America, vol. 107, p.p. 989-999, 2000.

۸- زیر نویس‌ها

- ¹ Haming Window
² Mel Frequently Cepstral Coefficient
³ Viterbi Algorithm
⁴ Tilt