

استفاده از فناوری پردازش گفتار در ارزیابی خودکار روانی گفتار در سیستم‌های آموزش زبان خارجی

بهزاد بختیاری

محمد مهدی همایون پور

دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران

چکیده

در این مقاله روشی جدید برای ارزیابی خودکار روانی گفتار ارائه خواهد شد. از کاربردهای ارزیابی خودکار روانی گفتار می‌توان به آموزش زبان خارجی به کمک کامپیوتر^۱ اشاره نمود. بدین صورت که گفتار زبان آموز ارزیابی شده و سطح روان و سلیس بودن گفتار وی مشخص می‌گردد. با اختیار داشتن ارزیابی کننده خودکار روانی گفتار، زبان آموز می‌تواند از سطح روان بودن گفتار خود آگاه شده و با تکرار و تمرین سعی در بهتر نمودن سطح روانی گفتار خود نماید. برای این منظور معیارهای کمی مختلفی ارائه شده است که می‌توان به کمک آنها روانی گفتار را پیش‌بینی نمود. در اینجا با بررسی مناسب بودن این معیارها، روشی برای ارزیابی خودکار روانی گفتار ارائه می‌شود. برای بررسی صحت ارزیابی خودکار ارائه شده دادگانی از ترک زبانان تهیه شده و آزمایشات بر روی آن انجام گرفت. همبستگی بین نمرات داوران انسانی و ارزیابی کننده خودکار در نهایت نزدیک به ۰/۹۵ رسید.

کلمات کلیدی: آموزش زبان خارجی به کمک کامپیوتر، ارزیابی خودکار روانی گفتار، شناسایی خودکار گفتار.

۱- مقدمه

هزینه‌های گزافی دارند، دسترسی به سیستم‌های خودکار آموزش زبان سهل الحصول و کم هزینه می‌باشد.

عموماً آن مختصه‌های آوایی که پس از تقسیم یک زنجیره به کوچکترین واحد های تشکیل دهنده آن بدست می‌آیند را واحدهای زنجیری^۲ می‌گویند. واکه‌ها و همخوانها واحدهای زنجیری گفتار را تشکیل می‌دهند. آن دسته از مشخصات آوایی که در هیچ یک از واحدهای زنجیری نمی‌گنجند (زیرا این مشخصات اولاً همزمان با بیش از یک واحد زنجیری تولید می‌شوند، ثانیاً جایگاه آنها در زنجیره گفتار متغیر است) را واحدهای زیر زنجیری یا مشخصات نوایی^۳ گفتار^۴ گویند. اگر بخواهیم در یک تقسیم‌بندی کلی از مشخصه‌های نوایی^۴ گفتار نام ببریم، باید از تکیه^۵، زیر و بمی^۶، سلیس و روانی^۷ گفتار، درنگ و وزن طبیعی^۸ گفتار یا ریتم^۹ نام ببریم. در اینجا از بین مشخصه‌های نوایی به ارزیابی سلیس و روان بودن گفتار به کمک کامپیوتر خواهیم پرداخت. این مشخصه در ارتباطات شفاهی از اهمیت خاصی برخوردار است. سطح مربوط به آموزش این مشخصه در سطوح آموزش

امروزه با توجه به اهمیت بالای آموزش زبان، از کاربردهای مهم و مفید فناوری شناسایی خودکار گفتار می‌توان به 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 آموزش روانی گفتار مورد بررسی قرار خواهد گرفت. بطور کلی آموزش خودکار زبان دارای مزیت‌هایی نسبت به روشهای آموزش سنتی می‌باشد. به عنوان مثال در کلاس‌های سنتی، معلم دقت و انرژی خود را صرف همه زبان آموزان می‌کند و بر روی یک زبان آموز خاص تمرکز ندارد. در حالیکه در آموزش خودکار زبان، سیستم به سادگی بر روی یک زبان آموز تمرکز می‌کند. از مشکلات دیگری که در کلاسهای سنتی آموزش زبان وجود دارد، خجالت کشیدن زبان آموز در سؤال کردن‌های متوالی و یا کم حوصلگی و خستگی معلم می‌باشد. همچنین از آنجائیکه کلاسهای سنتی همیشه در دسترس نیستند و

در این مقاله نیز هدف از بررسی روانی گفتار در آموزش زبان، تأکید بر روی جنبه‌های زمانی روانی گفتار خواهد بود. در جهت تعریف جنبه‌های زمانی روانی گفتار اغلب فرض می‌کنیم که هدف آموزش زبان است که شامل تولید "گفتار گویندگان بومی با توجه به پارامترهای زمانی که حدالمقدور شامل وقفه‌های سکوت^{۱۴}، درنگ^{۱۵}، وقفه‌های پر شده^{۱۶}، شروع دوباره و اصلاح مجدد، ... نباشد [۵]. قابل ذکر است که مطالعات کمی صورت گرفته در مورد وقفه‌های واجی برای گویندگان بومی نتایج هموار و پیوسته‌ای را همیشه به دنبال ندارد. در واقع این مطلب نشان می‌دهد که درنگ‌های واجی برای تمایز بین گفتار بومی و غیربومی نمی‌تواند کافی باشد. آقای مول^{۱۷} برای حل این مشکل پیشنهاد کرد، توزیع تعداد وقفه‌ها در نظر گرفته شود [۱۳]. مطالعاتی که بر روی مقایسه گفتار دو زبان با گویندگان یکسان صورت گرفت نشان می‌دهد که می‌توان تفاوتی بین دو گفتار قائل شد [۱۲، ۱۳].

از ابزارهای ارائه شده برای این منظور می‌توان به کارهای ارائه شده در [۱، ۱۴] اشاره کرد. در [۱۴]، روانی در گفتار بصورت بیان همه کلمات در محدوده کشش^{۱۸} از پیش تعریف شده‌ای که از میانگین گفتار بومی بدست می‌آید، تعریف می‌شود. در [۱۱]، 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^{۱۹} وجود داشته و در نتیجه می‌توان به ارزیابی مبتنی بر کامپیوتر اعتماد نمود. در هر دو کار ذکر شده معیارهایی برای سلیس و روان بودن گفتار معرفی می‌شود ولی روشی برای ارزیابی خودکار میزان روان بودن گفتار زبان آموزان ارائه نمی‌شود.

همچنین در دانشگاه کالیفرنیا جنوبی تحقیقاتی در زمینه ارزیابی خودکار روانی گفتار صورت گرفته که به کمک آن کودکان مکزیکی به آموختن زبان انگلیسی پرداختند [۱۵]. در [۱۶] اشاره شده است که برای ارزیابی خودکار روانی گفتار دو دسته الگوریتم وجود دارد. در دسته اول هدف پیدا کردن مکان و زمان دقیق وقفه گفتار زبان آموز می‌باشد. از جمله این کارها می‌توان به [۱۷، ۱۸] اشاره نمود. در این کارها سیستم مبتنی بر شناسایی کننده‌های خودکار می‌باشد که فرض بر این است که نگارش آوایی گفتار نیز در دسترس می‌باشد. اما نوع دوم سعی می‌کند به میزان روانی گفتار زبان آموز امتیازی تخصیص دهند که از این جمله می‌توان به [۱۹، ۲۰] اشاره نمود. در این مقاله سعی گردید که هر دو هدف محقق شود به نحوی که کار صورت گرفته علاوه بر اینکه مکان‌های وقفه گفتار زبان آموز مشخص می‌شود در کل نمره‌ای نیز به میزان روانی گفتار وی داده می‌شود.

نمای کلی روش ارزیابی خودکار روانی گفتار را در شکل ۱ مشاهده می‌کنید. در ادامه، در بخش ۲ به دادگان غیربومی مورد استفاده خواهیم پرداخت. در بخش ۳، معیارهای کمی روانی گفتار را تعریف خواهیم نمود و به بررسی رابطه آنها با ارزیابی‌های انسانی خواهیم پرداخت. در بخش ۴، روشی برای ارزیابی خودکار روانی گفتار به کمک معیارهای کمی تعریف شده خواهیم پرداخت. بخش ۵ نیز به خلاصه و نتیجه‌گیری می‌پردازد.

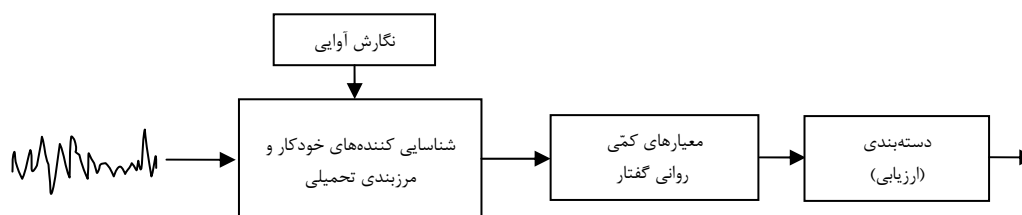
زبان، به زبان آموزانی با سطح بالای متوسط پیشنهاد می‌شود. بدیهی است که زبان آموزان مبتدی حتی در قسمت بیان یک کلمه و تلفظ صحیح آن نیز دچار اشکال می‌گردند و در نتیجه تمرین برای روانی گفتار یک جمله کاری دشوار برای زبان آموزان مبتدی خواهد بود. اما زبان آموزانی که سطح مبتدی را پشت سر گذاشته و به سطح متوسط می‌رسند می‌توانند با تمرین بیان جمله‌های مختلف در روان صحبت کردن مهارت لازم را کسب کنند.

به طور معمول اساتید و محققان، "روانی گفتار" را بصورت کارایی زبان بومی و غیربومی توصیف می‌کنند [۱]. در [۲، ۳، ۴، ۵، ۶] مروری بر روی واژه روانی گفتار و جنبه‌های مختلف آن شده است. در اینجا نشان خواهیم داد که بین "تعریف روانی گفتار با توجه به کارایی زبان بومی" و "روانی گفتار در محتوای آموزش زبان خارجی" تفاوت‌هایی وجود دارد. در حالت دوم روانی بصورت معیار مهمی از کارایی افراد غیربومی محسوب می‌شود که می‌توان آن را مورد ارزیابی و داوری قرار داد [۸]. این مطلب از این واقعیت ناشی می‌شود که روانی گفتار اغلب در ارزیابی‌ها و تست‌ها گنجانده می‌شود.

اگر کارایی گویندگان بومی در نظر گرفته شود، می‌توان از آن برای تعیین مشخصات کارایی گویندگان استفاده نمود. ولی در حقیقت نمی‌تواند نشان دهنده یک معیار ارزیابی باشد. همچنین از طرف دیگر واژه غیر روان بودن معمولاً به اتصال قسمتی از گفتار ناخواسته مانند، وقفه‌ها به گفتار اطلاق می‌شود. در نتیجه غیر روان بودن بصورت تعداد تکرارهای غیر عادی زیاد وقفه‌ها در جریان گفتار شناخته می‌شود [۹].

آقای فیلمور^۹ در [۳] چهار توانایی مختلفی را برای واژه روانی ذکر نموده است:

- ۱- توانایی برای صحبت کردن در طولی از زمان با تعداد وقفه‌های کم
 - ۲- توانایی صحبت کردن بطوری که بین جملات بیان شده ارتباط منطقی برقرار باشد
 - ۳- توانایی بکار گیری چیزهایی مناسب در صحبت کردن در یک زمینه وسیع از نظر محتوایی
 - ۴- توانایی در خلق و تصور چیزهایی در کاربردهای زبانی
- اما در مقوله آموزش زبان خارجی تعاریف دیگری نیز وجود دارد. از جمله، در آموزش زبان ارتباطی^{۱۰}، تأکید بر روی روانی نقطه مقابل دقت بشمار می‌آید. بر اساس تعریف آقای پرمفیت^{۱۱} روانی، تأثیر مضاعف عملیات سیستم آموزش زبان است که موجب اکتساب توسط زبان آموزان می‌گردد [۴]. با توجه به این تعریف، کارایی مشابه بودن به گویندگان بومی^{۱۲} هدفی نیست که بایستی به آن دست یابیم. همچنین کارایی شباهت به بومی بودن را می‌توان بصورت هدف نهایی تفسیر روانی در نظر گرفت. این تعریف می‌تواند در کاربردهای روزانه زبان توسعه پیدا کرده و نشان دهنده مهارت و کارایی زبان گردد [۵، ۷]. بالاخره، در حالت‌های خاص و محدود شده‌ای، واژه روانی برای مهارت در گفتار شفاهی و به طور خاص تر در جنبه‌های زمانی^{۱۳} بکار می‌رود [۵، ۶، ۸، ۱۰، ۱۱، ۱۲].



شکل ۱- دیگرام مراحل ارزیابی خودکار روانی گفتار

۲- دادگان غیربومی

۲-۲- ارزیابی روانی دادگان غیربومی توسط داوران انسانی و سازگاری آنها در ارزیابی

۲-۱- تهیه و طراحی دادگان غیربومی

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 مورد استفاده ۹ تن از دانشجویان کارشناسی ارشد و دکتری کامپیوتر بوده اند که زمینه فعالیت آنها پردازش گفتار می‌باشد. این دانشجویان آشنایی کامل به آواشناسی و مفاهیم پردازش گفتار دارند. برای داوری دادگان غیربومی نرم افزاری تهیه شد که فایل‌های گفتار را به ترتیب برای داور پخش می‌نمود. داوران به صورت انفرادی در محیط آرام توسط گوشی به گفتار گویندگان غیربومی گوش داده و سطح روانی گفتار مربوطه را به کمک نرم افزار ذخیره می‌کنند. این ارزیابی‌ها برای هر گوینده غیربومی، در دامنه ۱ تا ۵ صورت می‌گرفت که ۱ مربوط به روانترین و ۵ مربوط به بدترین گفتار از لحاظ سلیسی و روانی گفتار می‌باشد. مفهوم ۵ کلاس مذکور و امتیازات مربوطه در جدول ۲ آمده است.

از آنجائیکه با مقایسه نتایج ارزیابی داوران انسانی و نتایج سیستم ارزیابی خودکار، کارایی و صحت سیستم ارزیابی خودکار محاسبه خواهد شد و با توجه به اینکه داوری امری شخصی می‌باشد، بایستی ابتدا نحوه ارزیابی داوران انسانی بررسی شود. بدیهی است در حالتیکه نتایج یک داور نسبت به دیگر داوران به دلایلی مانند، عدم دقت کافی، سخت‌گیری، در نظر گرفتن جنبه‌های دیگری از روانی گفتار و غیره فاصله بسیار زیادی داشته باشد، بایستی از آن داور در ارزیابی‌ها استفاده نشود. در این قسمت نتایج ارزیابی داوران در ارزیابی یک مجموعه گفتار غیربومی مشترک مورد بررسی قرار می‌گیرد. برای اینکه سازگاری داوران در ارزیابی روانی گفتار مشخص گردد، امتیازات داده شده توسط هر داور نسبت به امتیازات دیگر داوران مورد بررسی قرار می‌گیرد.

جدول ۳ همبستگی بین یک داور و دیگر داوران را نشان می‌دهد. برای هر داور همبستگی آن با تک تک داوران محاسبه شده و در نهایت میانگین همبستگی‌ها در جدول منظور شده است. با دقت در این جدول مشاهده می‌شود که ارزیابی روانی گفتار توسط داوران انسانی دارای همبستگی بالایی می‌باشد که نشان دهنده سازگار بودن داوران در این ارزیابی می‌باشد. تمام مقدار همبستگی‌ها بالای ۰/۹۵ می‌باشند. این همبستگی بالا شاید به این دلیل باشد که اولاً مشخصه روانی گفتار، مشخصه پیچیده‌ای نیست و ارزیابی آن ساده و بدون از مشکلات خاص می‌باشد. ثانیاً می‌توان نتیجه گرفت داوران تقریباً به یک اندازه سخت‌گیری نموده‌اند و همچنین جنبه‌های یکسانی از روانی گفتار مدنظر همه داوران بوده است.

به منظور بررسی این مطلب که امتیازات مربوط به ارزیابی داوران در گفتار بومی و غیربومی متفاوت بوده و علاوه بر آن برای بررسی تفاوت روانی گویندگان بومی و گویندگان غیربومی، میانگین امتیازات دادگان غیربومی مذکور با میانگین یک مجموعه گفتار بومی مورد مقایسه قرار می‌گیرد.

جدول ۲- سطوح مختلف روانی گفتار و امتیازات مربوطه

| امتیاز | سطح روانی گفتار |
|--------|-----------------|
| ۱ | بسیار خوب |
| ۲ | خوب |
| ۳ | متوسط |
| ۴ | بد |
| ۵ | بسیار بد |

در این مقاله چون هدف آموزش زبان مقصد به زبان آموزشی است که زبان مادری‌شان با زبان مقصد متفاوت است، بایستی دادگان غیربومی 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 آموزش زبان فارسی تا کنون کمتر مورد بررسی قرار گرفتند و همچنین زبان ترکی به دلیل دسترس بودن ترک زبانان در کشور انتخاب شد.

با توجه به اینکه در این مقاله هدف آموزش زبان به کمک کامپیوتر در زمینه روانی گفتار می‌باشد، بایستی جملاتی انتخاب شود تا ویژگی‌های خاصی ارضاء گردد. از جمله ویژگی‌هایی که این جملات بایستی آن را دارا باشند می‌توان به پوشش دادن تمام واج‌های موجود در زبان فارسی اشاره نمود. همچنین مناسب بودن طول جملات نیز از ویژگی‌های مطلوب محسوب می‌شود. چرا که این جملات قرار است توسط افراد غیربومی بیان شود و بعنوان مثال، بلند بودن جمله ممکن است زبان آموز را در بیان آن دچار مشکل سازد. در اینصورت آموزش بدرستی انجام نمی‌گیرد و تعامل درستی بین سیستم و زبان آموز انجام نمی‌شود. علاوه بر این بهتر است جملات پیچیده، حاوی کلمات مهجور، کلمات ادبی سنگین و مشابه اینها نباشد تا زبان آموز به سادگی جملات را بیان نموده و از مشکلات سلیسی و روانی خود آگاه گردد. برای این منظور ۴۰ جمله از دادگان فارس دات ۱۰۰ انتخاب شدند. این دادگان شامل ۶۰۰ جمله می‌باشد که اکثر جملات به گونه‌ای طراحی شده‌اند که تمام واج‌های زبان فارسی را بپوشانند. لازم به ذکر است از این دادگان برای آموزش مدل‌ها نیز استفاده شده است که گفتار استفاده شده از این دادگان جملاتی به غیر از این ۴۰ 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 که جدول ۱ جنسیت، سن، میزان تحصیلات و تعداد جمله بیان شده را نمایش می‌دهد.

جدول ۱- اطلاعات مربوط به دادگان گویندگان غیربومی

| ردیف | شناسه گوینده | جنس | حدود سنی | میزان تحصیلات | تعداد جمله بیان شده |
|------|--------------|-----|----------|---------------|---------------------|
| ۱ | ۳ | مرد | ۲۵ | کارشناسی | ۱۰ |
| ۲ | ۴ | زن | ۵۵ | زیر دیپلم | ۶ |
| ۳ | ۵ | مرد | ۶۰ | دیپلم | ۱۱ |
| ۴ | ۹ | زن | ۴۰ | دیپلم | ۱۲ |
| ۵ | ۱۲ | مرد | ۲۵ | کارشناسی | ۱۸ |
| ۶ | ۱۷ | زن | ۳۵ | دیپلم | ۶ |
| ۷ | ۲ | مرد | ۲۵ | کارشناسی ارشد | ۱۰ |
| ۸ | ۶ | مرد | ۳۰ | کارشناسی ارشد | ۱۰ |
| ۹ | ۷ | مرد | ۳۰ | کارشناسی | ۱۰ |
| ۱۰ | ۱۳ | مرد | ۳۰ | کارشناسی | ۱۰ |
| ۱۱ | ۱۵ | مرد | ۲۵ | کارشناسی | ۱۰ |

جدول ۳- همبستگی داوران انسانی در ارزیابی روانی گفتار دادگان غیربومی

| داور | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۸ | ۹ |
|---------|------|------|------|------|------|------|------|------|------|
| همبستگی | ۰/۹۶ | ۰/۹۵ | ۰/۹۵ | ۰/۹۶ | ۰/۹۵ | ۰/۹۷ | ۰/۹۶ | ۰/۹۵ | ۰/۹۶ |

$$\#P = \text{تعداد وقفه‌های درون جمله‌ای بیشتر از } ۰/۲ \text{ ثانیه} \quad (۴)$$

کل زمان وقفه‌ها (TDP)^{۲۴}:

$$TDP = \text{کل زمان وقفه‌های درون جمله‌ای بیشتر از } ۰/۲ \text{ ثانیه} \quad (۵)$$

میانگین طول وقفه‌ها (MLP)^{۲۵}:

$$MLP = \text{میانگین کل زمان وقفه‌های درون جمله‌ای بیشتر از} \quad (۶)$$

۰/۲ ثانیه

میانگین طول تداوم (MLR)^{۲۶}:

$$MLR = \text{میانگین تعداد واج‌هایی که بین دو وقفه بیشتر از } ۰/۲ \quad (۷)$$

ثانیه رخ می‌دهد

هفت معیار مذکور، از معیارهای معروفی هستند که محققان زیادی بر روی آنها مطالعاتی انجام داده اند. در کارهای ابتدایی، این معیارها بصورت دستی توسط یک فرد خبره از گفتار استخراج می‌شد. اما در [۱] این معیارها بصورت خودکار توسط شناسایی کننده‌های خودکار از گفتار گویندگان استخراج می‌شود. در اینجا نیز با توجه به اینکه هدف ارزیابی خودکار روانی به کمک کامپیوتر می‌باشد از شناسایی کننده‌های خودکار برای این منظور استفاده خواهد شد. در [۱] این معیارها برای هلندی زبانان مورد بررسی قرار گرفت که در واقع گویندگان بومی هلندی زبان و گویندگان غیربومی دارای زبان‌های مختلفی بودند که هلندی صحبت می‌کنند. در اینجا معیارهای مذکور با دادگان غیربومی (آذری زبانان) و بومی (فارسی زبانان) مورد بررسی قرار می‌گیرد.

۳-۱- شناسایی کننده‌های خودکار و استخراج معیارهای کمی روانی گفتار

از آنجائیکه روش‌های ارزیابی خودکار ارائه شده مبتنی بر مدل مخفی مارکوف می‌باشند از شناسایی کننده‌های مبتنی بر مدل مخفی مارکوف استفاده خواهیم نمود. مدل مخفی مارکوف از ابزارهای قوی آماری برای شناسایی گفتار بشمار می‌رود که در مسائل شناسایی گفتار به کثرت مورد استفاده قرار می‌گیرد. برای شناسایی در سطح واج، هر واج با یک مدل مخفی مارکوف سه حالتی چپ به راست مدل می‌شود. از آنجائیکه بخش آزمایشات در این مقاله مربوط به آموزش زبان فارسی به ترک زبانان می‌باشد، بایستی شناسایی کننده‌هایی برای واج‌های زبان فارسی طراحی شود. برای این منظور ۲۹ واج زبان فارسی و یک برچسب برای سکوت در نظر گرفته شد.

مجموعه گفتار بومی از دادگان فارس دات انتخاب شد. برای این منظور از ۲۰ گوینده با لهجه تهرانی استفاده شد. جدول ۴ میانگین امتیازات دادگان غیربومی و امتیازات مجموعه گفتار بومی را نمایش می‌دهد. همانطور که در این جدول مشاهده می‌شود گویندگان بومی امتیازات بهتری از گویندگان غیربومی کسب نموده اند. پس می‌توان نتیجه گرفت که اولاً تفاوتی بین روانی گفتار گویندگان بومی و غیربومی وجود دارد. ثانیاً داوران نیز در ارزیابی روانی گویندگان بومی و غیربومی تفاوت قائل می‌شوند.

جدول ۴- میانگین امتیازات داوران انسانی در روانی گفتار دادگان بومی و دادگان غیربومی

| | |
|----------------|-----|
| دادگان بومی | ۲/۵ |
| دادگان غیربومی | ۴/۱ |

۳-۲- معیارهای کمی روانی گفتار

در این قسمت معیارهایی 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 پیش‌بینی (ارزیابی) نمود. محققان زیادی بر روی معیارهای مختلف کار نموده اند، از جمله می‌توان به [۵، ۸، ۱۰، ۱۱، ۱۲] اشاره نمود. معیارهای معرفی شده در این کارها، جنبه زمانی روانی گفتار را در نظر گرفته اند و معیارهایی کمی^{۲۰} می‌باشند. از بین معیارهای ارائه شده در این کارها، هفت معیار پایه و مهم زیر انتخاب شدند:

سرعت گفتار (ROS)^{۲۱}:

$$ROS = \frac{\text{تعداد واج‌های بیان شده}}{\text{کل زمان گفتار با در نظر گرفتن وقفه‌های درون جمله}} \quad (۱)$$

نسبت آوا به زمان (PTR)^{۲۲}:

$$PTR = \frac{\text{کل زمان گفتار بدون وقفه‌ها}}{\text{کل زمان گفتار با در نظر گرفتن وقفه‌های درون جمله}} \times ۱۰۰\% \quad (۲)$$

سرعت تعامل (ART)^{۲۳}:

$$ART = \frac{\text{تعداد واج‌های بیان شده}}{\text{کل زمان گفتار بدون وقفه‌ها}} \quad (۳)$$

تعداد وقفه‌های سکوت (#P):

با توجه به اینکه ضرایب کپسترال مبتنی بر معیار مل در محیط‌های عملیاتی و نویری 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 فوق ۳۹ می‌باشد. جهت استخراج ویژگی از فریم‌هایی با طول ۳۰ میلی‌ثانیه 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 فریم کمتر شود. پس از فریم‌بندی، 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.

برای آموزش مدل‌ها از دادگان فارس دات نسخه ۱۰۰ استفاده شده است. این دادگان شامل ۳۰۰ گوینده و ۶۰۰ جمله می‌باشد که هر گوینده ۱۰ جمله در محیطی کاملاً آرام و بدون نویز بیان نموده است. گویندگان دارای سن، جنس و تحصیلات مختلفی هستند. لهجه‌های مختلف زبان فارسی در این دادگان عبارتند از: تهرانی، اصفهانی، یزدی، شمالی، جنوبی، لوری، خراسانی، بلوچی، کردی، ترکی. در اینجا ۴۵ گوینده با لهجه تهرانی از این دادگان برای آموزش انتخاب شد. این گویندگان کاملاً از لحاظ جنسی و سنی مختلف انتخاب شدند. به کمک این شناسایی‌کننده‌ها و نگارش آوایی گفتار بیان شده، مرزبندی تحمیلی به کمک الگوریتم ویتربی انجام می‌گیرد و مرزهای واج‌ها و سکوت تعیین می‌شود. با داشتن مرزهای مربوط به واج‌ها و سکوت‌ها معیارهای کمی مذکور محاسبه می‌شوند. از آنجائیکه محل وقفه‌ها در گفتار زبان آموز مشخص نیست، در نگارش آوایی بعد از هر کلمه یک برچسب سکوت گذاشته می‌شود. بدیهی است در صورتیکه وقفه‌ای وجود نداشته باشد، طول تعیین شده توسط مرزبندی تحمیلی برای سکوت در نظر گرفته شده کوچک خواهد بود و در صورتیکه وقفه‌ای رخ دهد طول دقیق آن مشخص می‌گردد.

۲-۳- معیارهای کمی روانی گفتار و پیش‌بینی روانی گفتار

در این قسمت به بررسی ارتباط بین معیارهای کمی تعریف شده و روانی گفتار زبان آموزان خواهیم پرداخت. برای این منظور دادگان غیربومی که در بخش ۲ توضیح داده شد مورد استفاده قرار گرفته و معیارهای مذکور از این دادگان استخراج می‌شود، سپس مقادیر این معیارها با نمراتی که ۹ داور انسانی به روانی گفتار زبان آموزان داده‌اند مقایسه می‌شوند. جدول ۵ همبستگی بین نمرات داوران انسانی و معیارهای کمی روانی گفتار را نشان می‌دهد.

شکل ۲ نیز میانگین قدر مطلق همبستگی هر معیار را با ۹ داور انسانی نمایش می‌دهد. این شکل در واقع قدر مطلق ستون میانگین در جدول ۵ را نمایش می‌دهد. مقادیر منفی در این جدول نشان دهنده وجود ارتباط معکوس معیارها با روانی گفتار (نمرات داوران انسانی) می‌باشد. بعنوان مثال تعداد وقفه‌ها (#P) هر چقدر بیشتر باشد، گفتار بیان شده از روانی و سلیسی کمتری برخوردار است. با دقت در این جدول مشاهده می‌شود که میانگین همبستگی هر معیار با نمرات داوران، قابل قبول بوده و بین معیارها و ارزیابی‌های داوران انسانی همبستگی بالایی وجود دارد. بیشترین همبستگی مربوط به معیار سرعت گفتار (ROS) می‌باشد. شاید این معیار با ذهنیت ما در مورد روان صحبت کردن نیز مطابقت داشته باشد. چرا که معمولاً اگر فردی روان و سلیس صحبت کند، سرعت گفتار وی بایستی مناسب باشد و در صورتیکه بیان فرد روان نباشد گفتار او شامل مکث‌ها، وقفه‌ها، مین کردن‌ها و چیزهای مشابه اینها خواهد بود و در نتیجه سرعت گفتار وی کاهش پیدا خواهد کرد.

در نتیجه در این مثال معیار میانگین طول وقفه‌ها نمی‌تواند معیار مناسبی باشد. بطور کلی اگر چه معیار سرعت گفتار (ROS) معیار مناسب‌تری می‌باشد، اما نمی‌تواند به تنهایی پوشش دهنده تمام خصوصیات روانی گفتار باشد. شاید با در نظر گرفتن تعداد وقفه‌ها در کنار سرعت گفتار اطلاعات بیشتری بدست آید. سطر آخر جدول ۵ میانگین قدر مطلق همبستگی بین هر داور و معیارها را نشان می‌دهد. با دقت در این سطر مشاهده می‌شود که میانگین قدر مطلق همبستگی بین معیارهای کمی تعریف شده با هر داور مقدار قابل قبولی می‌باشد. بیشترین همبستگی مربوط به داوران ۲، ۵ و ۸ می‌باشد. میانگین سطر آخر برابر ۰/۹۰ می‌باشد این مقدار که در واقع میانگین میانگین‌های قدر مطلق همبستگی بین معیارها و هر داور می‌باشد، نشان‌دهنده این مطلب است که میانگین همبستگی که بین کل معیارها و هر داور انسانی وجود دارد برابر ۰/۹۰ می‌باشد که مقدار قابل قبولی می‌باشد. نتیجه مشابهی با میانگین‌گیری از ستون آخر جدول ۵ بدست می‌آید. میانگین این ستون نشان دهنده این واقعیت است که میانگین همبستگی بین هر معیار با داور انسانی برابر ۰/۹۰ می‌باشد.

میانگین قدر مطلق همبستگی هر معیار با امتیازات داوران در شکل ۲ نمایش داده شده است که بطور کلی می‌توان گفت تقریباً تمام همبستگی‌ها قابل قبول هستند و این معیارها با روانی گفتار به نوعی در ارتباط می‌باشند و در نتیجه می‌توان از این معیارها برای پیش‌بینی و ارزیابی خودکار روانی گفتار استفاده نمود.

در جدول ۶ همبستگی بین هر دو معیار آمده است. با توجه به جدول، مشاهده می‌شود که بین همه معیارها نسبت به هم تقریباً همبستگی بالایی وجود دارد. در این جدول، بالاترین همبستگی‌ها مربوط به معیار سرعت گفتار (ROS) و کمترین همبستگی‌ها مربوط به معیار کل زمان وقفه‌ها (TDP) می‌باشد. قابل ذکر است علی‌رغم وجود همبستگی‌های بالا بین معیارها، این معیارها متفاوت می‌باشند.

به منظور بررسی بیشتر، میانگین و انحراف معیار معیارهای کمی روانی گفتار و امتیازات داوران انسانی مربوط به دادگان غیربومی و مجموعه گفتار بومی بخش ۲-۱ محاسبه شد و نتایج آن در جدول ۷ آمده است. در این جدول NS مربوط به گویندگان بومی و NNS مربوط به گویندگان غیربومی می‌باشد. با دقت در این جدول مشاهده می‌شود که معیارهای کمی روانی گفتار برای گویندگان بومی و غیربومی مقادیر متفاوتی دارند. به عبارت دیگر، گویندگان بومی گفتاری سریع‌تری داشته و در گفتار خود وقفه‌های کمتری تولید می‌کنند.

۴- ارزیابی خودکار روانی گفتار

ستون آخر (میانگین) در جدول ۹ و شکل ۳ همبستگی کلی ارزیابی انسانی و ارزیابی خودکار را نمایش می‌دهد. با توجه به نتایج بدست آمده مشاهده می‌شود که معیار سرعت گفتار (ROS) بهترین ویژگی برای پیش‌بینی روانی گفتار می‌باشد. این مطلب با نتایج مقایسه همبستگی معیارهای کمتی روانی گفتار و نمرات داوران انسانی که در جدول ۵ آمده است، مطابقت دارد. کمترین همبستگی نیز مربوط به معیار، میانگین طول تداوم (MLR) می‌باشد. با توجه به جدول ۳، مشاهده می‌شود که ۹ داور انسانی در ارزیابی یک دادگان غیربومی همبستگی‌هایی نزدیک ۰/۹۵ داشتند. در واقع اگر یک ارزیابی کننده دیگر با داوران مذکور در ارزیابی دادگان مشابه نزدیک به این همبستگی را داشته باشد، می‌توان نتیجه گرفت که ارزیابی کننده مذکور مانند ارزیابی کننده‌های انسانی عمل نموده است. در جدول ۸ برای معیار سرعت گفتار (ROS) همبستگی ۰/۹۳ بدست آمد که نزدیک به ارزیابی انسانی می‌باشد.

می‌توان برای ارزیابی خودکار روانی گفتار از ترکیب معیارهای مختلف کمتی 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. بنابراین خواهیم داشت:

$$\min_i \sum_{j=1}^n |V_j - M_{i,j}| \quad 1 \leq i \leq 5 \quad (9)$$

که n تعداد معیارهای استفاده شده و $M_{i,j}$ میانگین معیار j ام مربوط به دسته i ام می‌باشد (جدول ۸). بردار V نیز مربوط به ویژگی‌های استخراج شده از گفتار می‌باشد. در اینصورت بردار V به دسته i ام تعلق خواهد داشت. در جدول ۱۰ ترکیب دوتایی معیارها و نتایج آنها آمده است. در این جدول معیارهای هر ردیف با معیاری از هر ستون ترکیب شده و نتیجه همبستگی ارزیابی انسانی و ارزیابی خودکار به کمک دو معیار مربوطه آمده است. در اینجا نیز ابتدا همبستگی‌های بین ارزیابی‌های خودکار و تک تک داوران انسانی محاسبه شده و سپس میانگین آنها برای هر مورد منظور می‌شود.

در بخش قبل معیارهای کمتی تعریف شده در [۱۱] 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 مذکور به ارزیابی روانی گفتار گویندگان غیربومی بپردازیم. با توجه به دیاگرام انجام مراحل ارزیابی خودکار روانی گفتار در شکل ۱، بعد از استخراج معیارهای کمتی روانی گفتار، توسط دسته‌بندی کننده ارزیابی روانی خودکار انجام می‌گیرد. خروجی ارزیابی خودکار روانی مشابه ارزیابی‌های داوران انسانی دارای ۵ کلاس خواهد بود که در جدول ۲ آمده است. برای تعیین سطوح آستانه و دسته‌بندی سطح روانی گفتار با توجه به معیارهای کمتی می‌توان از میانگین هر دسته مربوط به معیار کمتی 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 مذکور با تک تک میانگین هر دسته محاسبه شود، در نهایت به دسته‌ای تعلق خواهد گرفت که به آن نزدیک‌تر باشد. بنابراین خواهیم داشت:

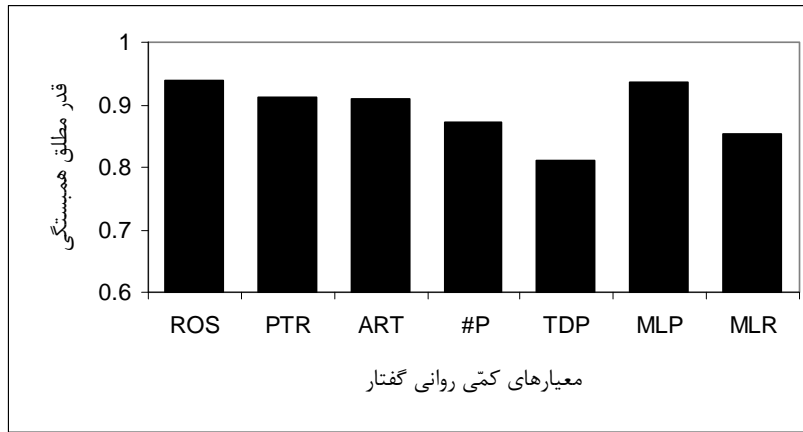
$$\min_i |V_x - M_{x,i}| \quad 1 \leq i \leq 5 \quad (8)$$

که V_x و $M_{x,i}$ به ترتیب برابر میانگین معیار x مربوط به دسته i ام و مقدار معیار x که از گفتار زبان آموز استخراج شده است، می‌باشد. بعنوان مثال، $\min_i |V_{ROS} - M_{ROS,i}|$ دسته‌بندی را برای معیار سرعت گفتار (ROS) انجام می‌دهد. مقدار i نتیجه دسته‌بندی را مشخص می‌نماید. جدول ۸ میانگین دسته‌های مختلف روانی گفتار را برای هر معیار نشان می‌دهد.

همانطور مشاهده می‌شود در اینجا برای تخمین و ارزیابی روانی گفتار فقط از یک معیار استفاده شده است. برای بدست آوردن مقادیر میانگین مربوط به هر معیار تقریباً نیمی از دادگان غیربومی استفاده شد. جدول ۹ نتیجه دسته‌بندی نیمی دیگر از دادگان و همبستگی آن را با امتیازات داوران انسانی نمایش می‌دهد.

جدول ۵- همبستگی بین نمرات داوران انسانی و معیارهای کمتی روانی گفتار

| | داور ۱ | داور ۲ | داور ۳ | داور ۴ | داور ۵ | داور ۶ | داور ۷ | داور ۸ | داور ۹ | میانگین |
|------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| ROS | ۰/۸۹۷۳ | ۰/۹۹۰۵ | ۰/۹۳۵۹ | ۰/۸۶۴۰ | ۰/۹۹۰۵ | ۰/۹۷۵۷ | ۰/۹۱۱۸ | ۰/۹۹۰۵ | ۰/۸۹۷۳ | ۰/۹۳۹۳ |
| PTR | ۰/۹۷۵۲ | ۰/۹۴۱۰ | ۰/۸۷۷۴ | ۰/۸۵۲۲ | ۰/۹۴۱۰ | ۰/۸۷۳۹ | ۰/۸۴۰۵ | ۰/۹۴۱۰ | ۰/۹۷۵۲ | ۰/۹۱۳۰ |
| ART | ۰/۸۰۰۵ | ۰/۹۶۷۵ | ۰/۹۱۵۲ | ۰/۸۵۴۸ | ۰/۹۶۷۵ | ۰/۹۷۹۴ | ۰/۹۲۴۹ | ۰/۹۶۷۵ | ۰/۸۰۰۵ | ۰/۹۰۸۶ |
| #P | -۰/۹۳۱۹ | -۰/۸۹۹۷ | -۰/۸۳۹۱ | -۰/۸۱۱۳ | -۰/۸۹۹۷ | -۰/۸۳۷۲ | -۰/۸۰۲۳ | -۰/۸۹۹۷ | -۰/۹۳۱۹ | -۰/۸۷۲۵ |
| TDP | -۰/۸۵۵۸ | -۰/۸۲۴۳ | -۰/۷۴۸۶ | -۰/۸۲۸۵ | -۰/۸۲۴۳ | -۰/۷۵۰۳ | -۰/۷۸۸۱ | -۰/۸۲۴۳ | -۰/۸۵۵۸ | -۰/۸۱۱۱ |
| MLP | -۰/۹۷۴۰ | -۰/۹۶۱۶ | -۰/۸۸۴۵ | -۰/۹۰۸۶ | -۰/۹۶۱۶ | -۰/۸۹۸۸ | -۰/۸۹۵۰ | -۰/۹۶۱۶ | -۰/۹۷۴۰ | -۰/۹۳۵۵ |
| MLR | ۰/۸۰۶۵ | ۰/۹۱۴۰ | ۰/۸۶۲۰ | ۰/۶۸۹۷ | ۰/۹۱۴۰ | ۰/۹۵۶۳ | ۰/۸۰۸۱ | ۰/۹۱۴۰ | ۰/۸۰۶۵ | ۰/۸۵۲۳ |
| میانگین قدر مطلق | ۰/۸۹۱۶ | ۰/۹۲۸۴ | ۰/۸۶۶۲ | ۰/۸۲۹۹ | ۰/۹۲۸۴ | ۰/۸۹۶۰ | ۰/۸۵۳۰ | ۰/۹۲۸۴ | ۰/۸۹۰۰ | |



شکل ۲- میانگین قدر مطلق همبستگی هر معیار با ۹ داور انسانی

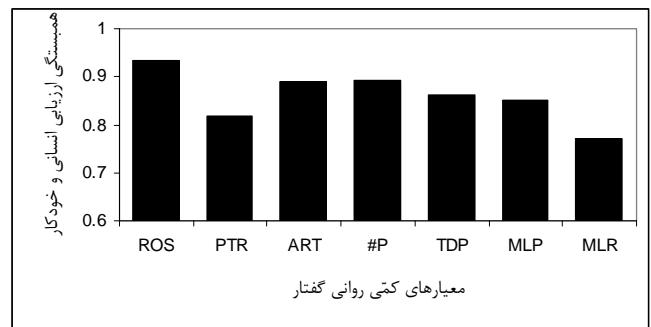
جدول ۶- همبستگی بین معیارهای مختلف روانی گفتار

| MLR | MLP | TDP | #P | ART | PTR | ROS | |
|---------|---------|---------|---------|--------|--------|-----|-----|
| ۰/۹۵۳۶ | -۰/۹۳۹۴ | -۰/۸۲۲۱ | -۰/۹۰۸۴ | ۰/۹۸۱۸ | ۰/۹۳۶۳ | ۱ | ROS |
| ۰/۸۲۷۳ | -۰/۹۷۸۳ | -۰/۹۳۷۰ | -۰/۹۸۵۲ | ۰/۸۵۸۶ | ۱ | | PTR |
| ۰/۹۵۴۱ | -۰/۸۷۳۸ | -۰/۷۴۷۱ | -۰/۸۳۱۶ | ۱ | | | ART |
| -۰/۷۹۸۴ | ۰/۹۳۲۵ | ۰/۹۷۰۰ | ۱ | | | | #P |
| -۰/۶۵۸۸ | ۰/۸۸۲۳ | ۱ | | | | | TDP |
| -۰/۸۲۱۴ | ۱ | | | | | | MLP |
| ۱ | | | | | | | MLR |

برای ترکیب سه معیار مشخص شده است. نتایج بدست آمده نشان می‌دهد که افزودن معیار کل زمان وقفه‌ها (TDP) 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 ۷- میانگین و انحراف معیار امتیازات داوران انسانی و معیارهای کمی روانی گفتار در دادگان غیربومی و مجموعه گفتار بومی

| sd NNS | \bar{x} NSS | sd NS | \bar{x} NS | |
|--------|---------------|-------|--------------|---------------|
| ۰/۹۸ | ۴/۱ | ۰/۳۵ | ۲/۵ | داوران انسانی |
| ۳/۸۷ | ۸/۱۷ | ۲/۲۳ | ۱۲/۳۳ | ROS |
| ۱۷/۸۵ | ۷۴/۱۲ | ۸/۷۷ | ۸۵/۴۵ | PTR |
| ۲/۸۲ | ۱۰/۶۱ | ۱/۱۹ | ۱۲/۰۱ | ART |
| ۱/۳۲ | ۱/۴۵ | ۰/۲۲ | ۰/۷۸ | #P |
| ۱/۸۷ | ۱/۳۹ | ۰/۱۲ | ۰/۲۱ | TDP |
| ۰/۵۵ | ۰/۵۸ | ۰/۲۵ | ۰/۳۲ | MLP |
| ۱۵/۰۳ | ۲۰/۵۶ | ۹/۹۶ | ۳۴/۲۶ | MLR |



شکل ۳- همبستگی ارزیابی انسانی و ارزیابی خودکار توسط معیارهای مختلف روانی گفتار

در جدول ۱۰ مشاهده می‌شود که ترکیب دو معیار سرعت گفتار (ROS) و سرعت تعامل (ART) بیشترین کارایی را نتیجه می‌دهد. این مقدار (۰/۹۴۱۳) از همبستگی‌های بدست آمده در ارزیابی خودکار با استفاده از یک معیار که در جدول ۹ آمده است، بیشتر می‌باشد.

از آنجائیکه در ترکیب دو تایی معیارها، بیشترین کارایی برای دو معیار سرعت گفتار (ROS) و سرعت تعامل (ART) بدست آمد، حال با افزودن معیارهای دیگر به این دو معیار نتایج را بررسی می‌کنیم. نتیجه بدست آمده از این ترکیب در جدول ۱۱ آمده است. در این جدول معیارهای دیگر به این دو اضافه شده و نتایج

نتایج حاصل شده از ارزیابی خودکار روانی نشان می‌دهد که ارزیابی کننده خودکار نیز می‌تواند مشابه یک داور انسانی عمل نموده و روانی گفتار را با دقتی مشابه ارزیابی نماید. چرا که همبستگی‌های بدست آمده برای ارزیابی خودکار نزدیک به همبستگی‌های بین داوران انسانی می‌باشد.

جدول ۱۰- 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 | + |
|--------|--------|--------|--------|--------|--------|-----|
| ۰/۸۸۰۵ | ۰/۹۳۷۳ | ۰/۹۳۷۳ | ۰/۹۳۷۲ | ۰/۹۴۱۳ | ۰/۷۲۴۶ | ROS |
| ۰/۸۵۹۷ | ۰/۸۱۸۶ | ۰/۸۱۸۶ | ۰/۸۱۸۶ | ۰/۸۶۲۳ | | PTR |
| ۰/۸۶۹۸ | ۰/۸۷۵۶ | ۰/۸۹۷۵ | ۰/۸۹۷۵ | | | ART |
| ۰/۸۳۹۷ | ۰/۹۰۷۷ | ۰/۹۰۳۵ | | | | #P |
| ۰/۸۳۹۷ | ۰/۸۹۷۳ | | | | | TDP |
| ۰/۸۳۹۵ | | | | | | MLP |

جدول ۱۱- همبستگی بین ارزیابی انسانی و ارزیابی خودکار روانی گفتار با استفاده از ترکیب سه تایی معیارهای کمی روانی گفتار

| MLR | MLP | TDP | #P | PTR | + |
|--------|--------|--------|--------|--------|---------|
| ۰/۸۳۲۷ | ۰/۹۴۱۳ | ۰/۹۴۳۵ | ۰/۹۳۸۸ | ۰/۷۴۵۴ | ART.ROS |

مراجع

[1] C. Cucchiari, H. Strik, and L. Boves, "Quantitative Assessment of Second Language Learners' Fluency by Means of Automatic Speech Recognition Technology," *Acoustical Society of America*, Vol. 107, pp. 989-999, 2000.

[2] R. Leeson, *Fluency and Language Teaching*, London, Longman, 1975.

جدول ۸- میانگین دسته‌های مختلف روانی گفتار برای هر معیار روانی گفتار

| دسته ۱ | دسته ۲ | دسته ۳ | دسته ۴ | دسته ۵ | |
|--------|--------|--------|--------|--------|-----|
| ۱۴/۸۶ | ۱۱/۹۴ | ۹/۰۲ | ۷/۴۰ | ۴/۱۴ | ROS |
| ۹۲/۸۳ | ۹۰/۲۸ | ۷۷/۹۰ | ۷۸/۶۰ | ۵۳/۸۳ | PTR |
| ۱۶/۰۱ | ۱۳/۲۰ | ۱۱/۵۹ | ۹/۴۴ | ۷/۸۶ | ART |
| ۰ | ۰/۳۳ | ۱/۱۴ | ۱/۲۵ | ۳ | #P |
| ۰ | ۰/۱۰ | ۰/۵۱ | ۱/۵۱ | ۳/۵۸ | TDP |
| ۰ | ۰/۰۸ | ۰/۴۲ | ۰/۴۱ | ۱/۲۲ | MLP |
| ۴۱/۴۱ | ۳۹/۲۹ | ۲۰/۸۵ | ۱۱/۳۷ | ۸/۶۲ | MLR |

۵- نتیجه گیری

در این مقاله به ارزیابی روانی خودکار گفتار با استفاده از معیارهای زمانی روانی گفتار پرداختیم. همانطور که در بخش ۱ توضیح داده شد، تعاریف مختلفی برای روانی گفتار وجود دارد. در اینجا جنبه‌های زمانی روانی گفتار در آموزش زبان مورد بررسی قرار گرفت. معیارهای کمی زمانی متنوعی برای روانی گفتار توسط بسیاری از محققین تعریف شده است که هفت مورد از معروفترین آنها برای زبان فارسی مورد بررسی قرار گرفت. نتایج ارزیابی داوران انسانی در مورد روانی گفتار دادگان بومی و غیربومی نشان داد که داوران در ارزیابی این دادگان با هم سازگاری مناسبی داشتند و گویندگان بومی نمرات بهتری از گویندگان غیربومی اخذ نمودند. این ارزیابی در ۵ سطح صورت گرفت. مناسب بودن معیارهای روانی تعریف شده برای زبان فارسی، با مقایسه امتیازات داوران انسانی مشاهده شد. به کمک این معیارها توانستیم روانی گفتار را پیش‌بینی نموده و ارزیابی روانی گفتار را بصورت خودکار انجام دهیم.

معیار سرعت گفتار (ROS) کارایی بالاتری را از دیگر معیارها از خود نشان داد. مشاهده شد که افزودن یک یا دو معیار دیگر باعث افزایش کارایی این معیار شد تا حدی که به نتایج همبستگی بین داوران انسانی بسیار نزدیک‌تر شد. همچنین مشاهده شد که افزودن معیارهای بیشتر موجب افزایش کارایی ارزیابی خودکار نگردد. که می‌توان نتیجه گرفت در بررسی جنبه‌های زمانی روانی گفتار، معیار سرعت گفتار ابعاد بسیاری از این مسأله را تحت پوشش خود قرار می‌دهد و معیار مناسبی می‌باشد.

جدول ۹- همبستگی بین نمرات داوران انسانی و ارزیابی خودکار روانی گفتار با استفاده از معیارهای مختلف کمی روانی گفتار

| میانگین | داور ۱ | داور ۲ | داور ۳ | داور ۴ | داور ۵ | داور ۶ | داور ۷ | داور ۸ | داور ۹ |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| ۰/۹۳۲۱ | ۰/۹۲۱۲ | ۰/۹۴۲۳ | ۰/۹۳۳۵ | ۰/۹۳۶۷ | ۰/۹۳۸۸ | ۰/۹۱۴۴ | ۰/۹۲۹۸ | ۰/۹۳۱۱ | ۰/۹۳۴۵ |
| ۰/۸۵۳۴ | ۰/۸۲۸۵ | ۰/۸۳۲۱ | ۰/۷۸۳۵ | ۰/۷۷۳۷ | ۰/۸۴۲۳ | ۰/۷۹۵۶ | ۰/۸۰۳۳ | ۰/۸۱۳۲ | ۰/۸۱۸۶ |
| ۰/۸۹۲۳ | ۰/۸۸۰۱ | ۰/۸۸۵۵ | ۰/۸۹۲۳ | ۰/۹۰۶۵ | ۰/۹۲۶۴ | ۰/۹۱۴۵ | ۰/۸۷۴۵ | ۰/۹۰۷۶ | ۰/۸۹۰۲ |
| ۰/۸۸۲۳ | ۰/۸۷۸۹ | ۰/۸۹۳۴ | ۰/۹۰۳۴ | ۰/۸۸۴۳ | ۰/۹۲۸۷ | ۰/۹۱۵۶ | ۰/۸۹۳۳ | ۰/۹۰۱۲ | ۰/۸۹۲۱ |
| ۰/۸۵۲۳ | ۰/۸۶۹۹ | ۰/۸۴۷۳ | ۰/۸۵۱۲ | ۰/۸۷۲۳ | ۰/۸۸۱۲ | ۰/۹۰۴۳ | ۰/۸۴۲۳ | ۰/۸۸۳۴ | ۰/۸۶۳۱ |
| ۰/۸۶۴۵ | ۰/۸۶۲۳ | ۰/۸۸۳۴ | ۰/۸۲۱۱ | ۰/۸۵۶۶ | ۰/۸۵۳۴ | ۰/۸۴۸۷ | ۰/۸۴۵۵ | ۰/۸۵۴۴ | ۰/۸۵۰۶ |
| ۰/۷۸۳۳ | ۰/۷۶۹۹ | ۰/۷۹۵۵ | ۰/۸۰۲۲ | ۰/۷۷۷۱ | ۰/۷۶۲۲ | ۰/۷۵۲۲ | ۰/۷۶۳۳ | ۰/۷۶۲۲ | ۰/۷۷۲۰ |

[18] K. Audhkhasi, K. Kandhway, O. D. Deshmukh, and A. Verma, "Formant-based technique for automatic filled-pause detection in spontaneous spoken English," *Proc, IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, pp. 4857-4860, 2009.

[19] O. D. Deshmukh, K. Kandhway, A. Verma, and K. Audhkhasi, "Automatic evaluation of spoken English fluency," *Proc, IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*, pp. 4829-4832, 2009.

[20] K. Zechner, and I. Bejar, "Towards automatic scoring of non-native spontaneous speech," *Proc, Human Language Technologies Conf.*, pp. 216-223, 2006.



بهزاد بختیاری در سال‌های ۱۳۷۸ تا ۱۳۸۲ به تحصیل در مقطع کارشناسی مهندسی کامپیوتر گرایش نرم‌افزار مشغول و در سال‌های ۱۳۸۲ تا ۱۳۸۴ در دانشگاه صنعتی امیرکبیر در مقطع کارشناسی‌ارشد مهندسی کامپیوتر گرایش هوش مصنوعی به تحصیل پرداخته است. نامبرده هم‌اکنون عضو هیات علمی گروه کامپیوتر موسسه آموزش عالی غیردولتی - غیرانتفاعی سجاد است. زمینه‌های مورد علاقه ایشان پردازش سیگنال گفتار، شبکه‌های عصبی و پردازش تکاملی می‌باشد.

آدرس پست‌الکترونیکی ایشان عبارت است از:

bakhtiari@sadjad.ac.ir



محمد مهدی همایون‌پور در سال ۱۳۳۹ در شهر شیراز متولد شد. تحصیلات تا مقطع دیپلم را در شهر شیراز سپری و دیپلم متوسطه خود را در سال ۱۳۵۸ دریافت کرد. وی تحصیلات خود در مقطع کارشناسی را در رشته مهندسی برق (الکترونیک) در دانشگاه صنعتی امیرکبیر (سال ۱۳۶۶)، کارشناسی‌ارشد را در رشته برق (مخابرات)، از دانشگاه خواجه‌نصیرالدین طوسی (سال ۱۳۶۹)، کارشناسی‌ارشد دوم خود را در زمینه فونیتیک (۱۳۷۴) در دانشگاه سوربون جدید در فرانسه و همزمان دوره دکترا خود را در دانشگاه پاریس ۱۱ در زمینه مهندسی برق (۱۳۷۴) بپایان رسانید. نامبرده از سال ۱۳۷۴ در سمت عضو هیات علمی دانشکده مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر به تدریس و تحقیق مشغول می‌باشد. زمینه‌های تخصصی مورد علاقه ایشان شامل پردازش سیگنال‌های دیجیتال، بازشناسی گفتار و گوینده، تبدیل متن به گفتار، کدینگ گفتار، پردازش زبان طبیعی، تشخیص نفوذ در سیستم‌ها و شبکه‌های کامپیوتری، اتوماسیون صنعتی، چند رسانه‌ای و طراحی ساخت‌افزار می‌باشد.

آدرس پست‌الکترونیکی ایشان عبارت است از:

homayoun@ce.aut.ac.ir

[3] C. J. Fillmore, "On Fluency in Individual Differences in Language Ability and Language Behavior," *Academic: New York*, pp. 85-101, 1979.

[4] C. Brumfit, *Communicative Methodology in Language Teaching: The Roles of Fluency and Accuracy*, Cambridge, Cambridge University Press, 1984.

[5] P. Lennon, "Investigating Fluency In EFL: A Quantitative Approach," *Language Learning* 3, pp. 387-417, 1990.

[6] R. Schmidt, "Psychological Mechanisms Underlying Second Language Fluency," *Stud. Second Language Acquisition*, Vol. 14, pp. 357-385, 1992.

[7] F. Chambers, "What do we mean by fluency," *System*, Vol. 25, No. 4, pp. 535-544, 1997.

[8] H. Riggenbach, "Toward an Understanding Of Fluency: A Microanalysis Of Non-Native Speaker Conversations," *Discourse Process*, Vol. 14, No. 4, pp. 423-441, 1991.

[9] T. J. Peters, and B. Guitar, *Stuttering. An Integrated Approach to its Nature and Treatment William and Wilkins*, Baltimore, 1991.

[10] P. Nation, "Improving Speaking Fluency," *System*, Vol. 17, No. 3, pp. 377-384, 1989.

[11] B. F. Freed, "What Makes Us Think That Students Who Study Abroad Become Fluent?," *Proc, Second Language Acquisition in a Study-Abroad Context*, pp. 123-148, 1995.

[12] R. Towell, R. Hawkins, and N. Bazergui, "The Development of Fluency in Advanced Learners of French," *Appl. Linguistics*, Vol. 17, No. 1, pp. 84-119, 1996.

[13] D. Mo'hle, "A Comparison Of The Second Language Speech Production Of Different Native Speakers," *Proc, Second Language Productions*, pp. 26-49, 1984.

[14] M. Eskenazi, "Using Automatic Speech Processing for Foreign Language Pronunciation Tutoring: Some and Prototype," *Language Learning and Technology*, Vol. 2, No. 2, pp. 62-76, 1999.

[15] M. Black, J. Tepperman, S. Lee, P. Price, and S. Narayanan, "Automatic detection and classification of disfluent reading miscues in young children's speech for the purpose of assessment," *Proc, Interspeech ICSLP*, pp. 206-209, 2007.

[16] K. Audhkhasi, "Automatic Evaluation of Fluency in Spoken Language," *IETE Technical Review*, Vol. 26, pp. 108-114, 2009.

[17] Y. Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper, "Enriching speech recognition with automatic detection of sentence boundaries and disfluencies," *IEEE Trans. Speech and Audio Processing*, Vol. 14, No. 5, pp. 1526-40, September 2006.

اطلاعات بررسی مقاله:

تاریخ ارسال: ۸۶/۰۶/۰۴

تاریخ اصلاح: ۸۹/۰۱/۲۴

تاریخ قبول شدن: ۸۹/۰۳/۲۳

نویسنده مرتبط: دکتر محمد مهدی همایون‌پور، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، ایران.

-
- ¹ CALL (Computer-Assisted Language Learning)
 - ² Segmental
 - ³ Supra Segmental
 - ⁴ Prosody
 - ⁵ Stress
 - ⁶ Intonation
 - ⁷ Fluency
 - ⁸ Rhythm
 - ⁹ Fillmore
 - ¹⁰ Communicative Language Teaching
 - ¹¹ Burmfit
 - ¹² Native-Speaker-Like Performance
 - ¹³ Temporal
 - ¹⁴ Silent Pauses
 - ¹⁵ Hesitations
 - ¹⁶ Filled Pauses
 - ¹⁷ Mohle
 - ¹⁸ Duration
 - ¹⁹ Correlation
 - ²⁰ Quantitative Measures
 - ²¹ Rate of Speech
 - ²² Phonation Time Ratio
 - ²³ Articulation Rate
 - ²⁴ Total Duration of Pauses
 - ²⁵ Mean Length of Pauses
 - ²⁶ Mean Length of Runs