

طراحی و جمع‌آوری دادگان تلفنی فارسی برای تصدیق هویت گوینده وابسته به متن

حسین زینلی^۱، حسین صامتی^۲ و هادی ویسی^۳

^۱ دانشگاه صنعتی شریف، zeinali@ce.sharif.edu

^۲ دانشگاه صنعتی شریف، sameti@sharif.edu

^۳ دانشگاه تهران، دانشکده علوم و فنون نوین، hadi_veisi@ut.ac.ir

چکیده - در این مقاله یک دادگان گفتاری تلفنی جدید معرفی می‌شود که برای استفاده در سیستم‌های تصدیق هویت گوینده وابسته به متن طراحی شده است. این دادگان که آن را *AGPVerif* می‌نامیم، از تعداد ۶۲ گوینده به زبان فارسی با پوشش انواع لهجه‌ها، سن و جنسیت جمع‌آوری شده است. صدای هر کدام از گوینده‌ها از کانال‌های مختلف مخابراتی شامل حداقل دو مکالمه تلفن ثابت و دو مکالمه تلفن همراه، با تنوع فاصله‌ای، ضبط شده است. هر گوینده ۱۴ جمله را بیان کرده است که در نتیجه این دادگان را برای کاربردهای گذرواژه ثابت و تصادفی قابل استفاده کرده است. برای تست این دادگان، با استفاده از گذرواژه تصادفی از روشی بر پایه مدل مخفی مارکوف استفاده شده که برای گذرواژه‌های رقمی نرخ خطای برابر (EER)، ۴.۲ درصد حاصل شده است. کلید واژه- دادگان گفتاری تلفنی فارسی، تصدیق هویت گوینده وابسته به متن، مدل مخفی مارکوف

۱- مقدمه

گفته است. به طور کلی دقت روش‌های وابسته به متن بیشتر از روش‌های مستقل از متن است و از همین‌رو در سیستم‌هایی که امنیت بالایی نیاز دارند، از این دسته روش‌ها استفاده می‌شود. روش‌های مستقل از متن بیشتر در سیستم‌های شناسایی هویت گوینده^۵ استفاده می‌شوند.

در روش وابسته به متن، سناریوهای متفاوتی استفاده می‌شود. این روش‌ها نسبت به یکدیگر برتری‌هایی دارند که در کاربردهای مختلف اهمیت پیدا می‌کنند. از سناریوهای که معمولاً استفاده می‌شوند، می‌توان به موارد زیر اشاره کرد [۱]:

استفاده از گذرواژه یکسان برای گویندگان: در این روش تمام گویندگان گذرواژه یکسانی دارند. این حالت ساده‌ترین روش این مجموعه است که مشکلات متعددی دارد. از این مشکلات می‌توان به قابلیت تقلید صدا و یا پخش صدای ضبط شده اشاره کرد.

استفاده از گذرواژه اختصاصی برای هر گوینده: در این حالت متن گذرواژه گویندگان با هم متفاوت است و هر گوینده باید گذرواژه مخصوص خود را بیان کند. در این حالت نیز به دلیل ثابت بودن گذرواژه، امکان ضبط صدا وجود دارد.

در سال‌های اخیر با پیشرفت تکنولوژی و افزایش نیاز به امنیت، استفاده از سیستم‌های زیست‌سنجی^۱ فراگیر شده است. از روش‌های مختلف زیست‌سنجی می‌توان به استفاده از اثر انگشت، چشم، صورت و گفتار اشاره کرد. مزیت اصلی گفتار نسبت به سایر روش‌های زیست‌سنجی قابلیت استفاده از آن در کاربردهای راه دور، مثل پشت خط تلفن است.

تصدیق هویت گوینده^۲، روشی است که به سؤال "آیا این گفتار توسط گوینده مشخصی بیان شده است؟" پاسخ می‌دهد. برای این کار سیستم با مقایسه گفتار ورودی با مدلی که در مرحله آموزش برای گوینده ساخته شده است، به این سؤال پاسخ می‌دهد. تصدیق هویت به دو دسته وابسته به متن^۳ و مستقل از متن^۴ تقسیم می‌شود. در دسته اول متنی که توسط گوینده بیان می‌شود اهمیت زیادی دارد و گوینده باید متن مشخصی را بیان کند. در حالی که در دسته دوم مهم نیست که گوینده چه چیزی

¹ Biometric

² Speaker verification

³ Text dependent

⁴ Text independent

⁵ Speaker identification

| نام | تعداد گوینده | محتوی |
|------------|--------------|--|
| [2] BANKA | ۲۰۸ | اطلاعات شخصی یک دنباله ثابت از ارقام |
| [3] BIOMET | ۹۱ | اطلاعات شخصی |
| [4] MIT | ۸۸ | عبارات ثابت ۲ یا ۳ کلمه‌ای |
| [5] Valid | ۱۰۶ | یک عبارت ثابت یک دنباله ثابت از ارقام |

همان‌طور که از جدول ۱ مشخص است، معمولاً دادگان‌هایی که جمع‌آوری شده‌اند فقط برای یکی از حالات وابسته به متن قابل استفاده هستند. به همین دلیل سعی شده تا دادگانی که در اینجا جمع‌آوری می‌شوند قابل استفاده در چند حالت مختلف باشند. زبان تمامی دادگان معرفی شده انگلیسی است. برای زبان فارسی دادگان‌های متعددی وجود دارد که از آنها می‌توان برای کاربرد تصدیق هویت گوینده استفاده کرد [۶ و ۷]. تمامی این دادگان مستقل از متن هستند و به همین خاطر نیاز به دادگان وابسته به متن فارسی احساس می‌شود.

۳- مشخصات کلی دادگان

این دادگان به منظور استفاده در سیستم‌های تصدیق هویت گوینده فارسی وابسته به متن جمع‌آوری شده است. کار مطالعه و ضبط آن حدود ۳ ماه طول کشید. تعداد کل گویندگانی که در ضبط این دادگان شرکت کرده‌اند ۶۲ نفر است. از این تعداد ۳۴ نفر مرد و ۲۸ نفر زن هستند. گویندگان این مجموعه به گونه‌ای انتخاب شده‌اند که تا حد ممکن تعدادی از لهجه‌های موجود در کشور را پوشش دهند. این لهجه‌ها شامل فارسی، ترکی، کردی، لری و جنوبی است که فراوانی تعداد گویندگان لهجه فارسی از بقیه بیشتر است. علاوه بر لهجه، گویندگان از شهرهای مختلفی انتخاب شده‌اند، به گونه‌ای که فاصله متفاوتی تا مکان ضبط صدا داشته باشند. دلیل این امر تأثیر فاصله بر کیفیت صدا در شبکه تلفن ثابت است. سن گویندگان انتخابی بین ۱۸ تا ۵۹ سال است که فراوانی محدوده‌های سنی مختلف در جدول ۲ آورده شده است.

تعداد ۵۰ نفر از کل گویندگان حداقل چهار مرتبه (دو بار با تلفن ثابت و دو بار با تلفن همراه) و بقیه بین یک تا سه بار فرآیند ضبط صدا را انجام داده‌اند. فاصله زمانی بین هر فرآیند ضبط برای هر گوینده بین یک هفته تا دو ماه بوده است. این

استفاده از متن تصادفی: در اینجا از گوینده خواسته

می‌شود که متنی که توسط سیستم پخش می‌شود را بیان کند. معمولاً متنی که در این حالت ساخته می‌شود ترکیبی از کلماتی محدود و مشخص است. یکی از مزیت‌های این روش این است که نیازی نیست گوینده چیزی را حفظ کند و همچنین امکان پخش مجدد صدای ضبط شده را نیز کاهش می‌دهد.

یکی از مشکلاتی که در سیستم‌های تصدیق هویت وجود دارد تأثیر نویزهای مختلف بر دقت این سیستم‌ها است. از این نویزها می‌توان به انواع نویزهای جمع‌شونده^۶ مثل سروصدای محیط و همچنین نویزهای پیچشی^۷ اشاره کرد. نویزهای پیچشی به دلایل مختلفی به وجود می‌آیند که از آنها می‌توان به تفاوت در میکروفون‌های مختلف، تفاوت در کانال‌های انتقال و همچنین تفاوت بودن سیستم‌های ضبط صدا اشاره کرد. نویزهای پیچشی به دلیل اینکه معمولاً تأثیرات غیرخطی بر گفتار دارند، دقت سیستم‌های تصدیق هویت را بیشتر کاهش می‌دهند.

هدف اصلی در این مقاله طراحی و ضبط دادگانی است که بتوان از آن برای آزمایش حالت‌های مختلف استفاده کرد. دادگانی که در اینجا تهیه شده است، برای هر سه روش تصدیق هویت وابسته به متن مذکور قابل استفاده است. علاوه بر این، برای اینکه بتوان تأثیر کانال‌های مختلف را بر دقت سیستم‌های تصدیق هویت بررسی کرد در ضبط این دادگان از دو کانال تلفن ثابت و تلفن همراه استفاده شده است.

این مقاله در ادامه شامل بخش‌های زیر است. در بخش ۲، به صورت خلاصه، تعدادی از دادگان موجود برای کاربرد وابسته به متن معرفی می‌شوند. سپس در بخش ۳ مشخصات دادگان جمع‌آوری شده بیان خواهد شد. در بخش ۴، روش تصدیق هویت استفاده شده برای آزمایش دادگان توضیح داده شده و نتایج این روش در بخش ۵ آورده شده است. در بخش انتهایی نیز نتیجه‌گیری و کارهای آینده بیان شده است.

۲- معرفی دادگان‌های موجود

در سال‌های اخیر دادگان‌های مختلفی برای تصدیق هویت گوینده وابسته به متن، معرفی شده‌اند. تعدادی از این دادگان‌ها به صورت تجاری بوده و در دسترس دیگران نیست و مابقی برای کاربردهای علمی جمع‌آوری شده‌اند. در جدول ۱ مشخصات تعدادی از این دادگان‌ها به صورت خلاصه آورده شده است.

^۶ Additive noise

^۷ Convolutional noise

| ردیف | جمله |
|------|------------------------------------|
| ۱ | صدای من نشان دهنده هویت من است. |
| ۲ | بنی آدم اعضای یکدیگرند. |
| ۳ | چو ایران نباشد تن من مباد. |
| ۴ | صدای من رمز عبور من است. |
| ۵ | نیوشا یکی از مصادیق تولید ملی است. |
| ۶ | سعیدیا مرد نکونام نمیرد هرگز. |
| ۷ | تابستان گرمترین فصل سال است. |
| ۸ | صدای هر کس منحصر به فرد است. |

۳-۲- شرایط ضبط

این دادگان در محیط‌های واقعی ضبط شده است، بدین معنی که گویندگانی که صدای آنها جمع‌آوری شده، در زمان تماس، در محیط‌های مختلفی مثل منزل، اداره، خیابان و غیره حضور داشته‌اند و به این ترتیب، نویزهای مختلفی در بعضی از فایل‌های صوتی وجود دارد. برای اینکه بتوان تأثیرات کانال‌های مختلف و نویزهای پیش‌پیشی را بر کارایی سیستم‌های تصدیق هویت بررسی کرد، برای جمع‌آوری این دادگان از تلفن‌های ثابت شهری و شبکه تلفن همراه استفاده شده است. گویندگان مختلف از گوشی‌های تلفن ثابت و تلفن همراه شخصی برای ضبط صدا استفاده کرده‌اند و تعدادی از گوشی‌های تلفن ثابت استفاده شده بی‌سیم بوده‌اند. از این‌رو، دادگان شامل تنوع گوشی‌های تلفن نیز می‌باشد.

گویندگان مختلف در شرایط طبیعی و به صورتی که احساس راحتی کرده‌اند، گوشی تلفن و موبایل را به کار برده‌اند؛ بدین معنی که برخی از آنها گوشی خود را روی بلندگو گذاشته و در برخی دیگر فاصله دهنی گوشی تا دهان گوینده زیاد بوده است. بنابراین، کیفیت گفتار ضبط شده در مواردی تغییرات زیادی دارد. برای ضبط صداها، در تمام موارد از کارت‌های تلفنی آنالوگ Dialogic مدل Diva استفاده شده است.

۳-۳- بررسی صحت دادگان و پالایش آن

برای اطمینان از اینکه گویندگان جملات را درست بیان کرده باشند، تمامی فایل‌های ضبط شده دو بار گوش داده شده‌اند و در صورتی که اشتباهی رخ داده باشد، عملیات ضبط صدا دوباره تکرار شده است. سکوت‌های موجود در ابتدا و انتهای فایل‌های ضبط شده به صورت دستی حذف شده‌اند. ارقام و

بازه‌ها به صورتی انتخاب شده‌اند که تا حد ممکن حالات گوینده در هر ضبط، نسبت به ضبط دیگر متفاوت باشد.

جدول ۲- توزیع گویندگان دادگان ضبط شده با توجه به محدوده سنی آنها

| محدوده سنی | مرد | زن | تعداد کل | درصد |
|------------|-----|----|----------|------|
| ۱۸-۲۴ | ۵ | ۷ | ۱۲ | ۱۹٪ |
| ۲۵-۲۹ | ۱۶ | ۱۷ | ۳۳ | ۵۳٪ |
| ۳۰-۴۰ | ۱۱ | ۳ | ۱۴ | ۲۳٪ |
| ۴۰-۵۹ | ۲ | ۱ | ۳ | ۵٪ |
| مجموع | ۳۴ | ۲۸ | ۶۲ | ۱۰۰٪ |

۳-۱- جملات ضبط شده

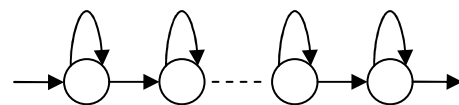
در جمع‌آوری این دادگان، هر گوینده در هر مرتبه از ضبط صدا، ۱۴ جمله را بیان کرده است. این جملات به صورتی طراحی شده‌اند که این دادگان قابل استفاده در هر سه نوع تصدیق هویت گوینده وابسته به متن باشد. برای این منظور از کل این جملات، ۸ جمله برای کاربرد گذرواژه یکسان بین گویندگان و یا گذرواژه اختصاصی برای هر گوینده طراحی شده است. در این حالت هر گوینده می‌تواند از بین این ۸ جمله یکی را به عنوان گذرواژه انتخاب کند. بقیه جملات برای کاربرد متن تصادفی طراحی شده‌اند. این جملات شامل ارقام فارسی از صفر تا نه و همچنین اسامی ماه‌های هجری-شمسی هستند. در هر مرتبه ضبط، گوینده ارقام صفر تا نه را به صورت شمرده شمرده و پشت سر هم بیان می‌کند. این دنباله دوبار ضبط شده است که بین هر مرتبه، گوینده چند جمله دیگر را بیان کرده است. اسامی ماه‌های سال به دو دسته تقسیم شده‌اند. ماه‌های فروردین تا شهریور به ترتیب پشت سر هم یکی از جملات را تشکیل می‌دهند و ماه‌های مهر تا اسفند نیز جمله دیگری از این دادگان هستند. جملات مربوط به ماه‌های سال نیز در هر مرتبه دوبار ضبط شده‌اند.

جملاتی که به منظور گذرواژه ثابت طراحی شده‌اند دارای طول‌های متفاوتی هستند. از آنجایی که پوشش واکه‌ای و بسامد رخداد آنها تأثیر زیادی در دقت سیستم‌های تصدیق هویت گوینده دارد، در این جملات سعی شده است که تعداد واکه‌ها در هر جمله متفاوت باشد تا بتوان تأثیر افزایش تعداد واکه‌ها بر دقت سیستم‌های تصدیق هویت را بررسی کرد. در جدول ۳ جملات بکار رفته در جمع‌آوری این دادگان، آورده شده است.

اسامی ماه‌های سال نیز بررسی گردیده که به صورت جداگانه و به ترتیب بیان شده باشند. مرز این کلمات مشخص نشده است و در کاربردهایی که مرز کلمات مورد نیاز باشد، مرز کلمات باید به صورت خودکار تشخیص داده شوند. در ادامه روشی برای این کار توضیح داده خواهد شد.

۴- سیستم تصدیق هویت گوینده

برای استفاده از این دادگان، در این مقاله یک سیستم تصدیق هویت گوینده مبتنی بر مدل مخفی مارکوف (HMM)^۸ پیاده‌سازی شده است. در این دست از روش‌ها از مدلی استفاده می‌شود که به آن مدل پس‌زمینه جهانی (UBM)^۹ گفته می‌شود. از مدل جهانی به عنوان مدل پایه برای آموزش مدل گویندگان استفاده می‌شود. در سیستم پیاده‌سازی شده از روش گذرواژه تصادفی با ارقام فارسی بهره گرفته شده که در ابتدا، برای هر کدام از ۱۰ رقم فارسی و سکوت (به عنوان مدل یازدهم) یک مدل مخفی مارکوف به عنوان مدل جهانی آموزش داده شده است. دلیل آموزش مدل سکوت این است که معمولاً بین هر دو رقم بیان شده سکوت کوتاهی وجود دارد که باید به صورت جداگانه مدل شود. برای آموزش این مدل‌های جهانی، از داده‌های دیگری که برای بازشناسی ارقام منفصل تلفنی توسط همین گروه جمع‌آوری شده بود، استفاده شد. مدل‌های به‌کار برده شده از نوع چگالی پیوسته و با تلفیق‌های گاوسی هستند. در این مدل‌ها از توپولوژی چپ به راست استفاده شده که در آن از هر حالت فقط می‌توان به همان حالت و یا حالت سمت راست رفت. در شکل ۱ توپولوژی این مدل‌ها نشان داده شده است.



شکل ۱- توپولوژی مدل‌های استفاده شده

کاربرد دیگر مدل جهانی ساخته شده در این مرحله استفاده برای نرمال‌سازی امتیاز بدست آمده در مرحله تست است. بعد از اینکه تمامی مدل‌ها ساخته شدند، با استفاده از داده‌های آموزشی هر گوینده و الگوریتم بیشینه‌سازی احتمال پسین (MAP)^{۱۰}، مدل‌های جهانی بروز می‌شوند. با این کار برای هر گوینده یازده

مدل مخفی مارکوف ساخته خواهد شد.

بعد از اینکه مدل‌های هر گوینده آموزش داده شد، نوبت به تست روش می‌رسد. برای این کار نیاز است که در هر فایل مرز بین ارقام مشخص باشد تا بتوان ارقام را از هم جدا کرد. از آنجایی که در این دادگان مرز بین ارقام مشخص نیست، در این مقاله این مرزها به صورت خودکار بدست آمده‌اند. برای این منظور، ابتدا با استفاده از پشت سر هم قرار دادن مدل ارقام و همچنین مدل سکوت مربوط به مدل جهانی، یک مدل مخفی بزرگ برای هر فایل ساخته می‌شود. سپس الگوریتم ویتربی^{۱۱} روی این مدل و با استفاده از دادگان هر فایل اجرا می‌شود. به دلیل اینکه توپولوژی مدل‌های استفاده شده چپ به راست است، به راحتی می‌توان مرز بین ارقام را پیدا کرد. برای اینکار زمانی که از حالت آخر یک مدل به حالت اول مدل دیگر منتقل شدیم، مرز دو رقم متوالی پیدا شده است.

بعد از پیدا شدن مرز بین ارقام، برای هر گوینده تعدادی عدد با تعداد ارقام دلخواه به عنوان داده تست تولید می‌شود. گفتار مربوط به هر یک از این اعداد، با پشت هم قرار دادن فایل صوتی ارقام آن (با ترتیب مشخص) تولید می‌شود. در نهایت گفتار مربوط به اعداد بدست آمده برای هر گوینده، برای تست سیستم استفاده می‌گردد.

۵- آزمایش‌های انجام شده و نتایج آنها

همان‌طور که بیان شد، در این مقاله از مدل‌های مخفی مارکوف با چگالی پیوسته استفاده شده است. برای مدل‌های ارقام و مدل سکوت از ۸ حالت استفاده شده که هر حالت دارای ۴ تلفیق گاوسی است. برای هر مدل احتمال شروع از حالت اول برابر یک و این احتمال برای بقیه حالات صفر است. ویژگی‌هایی که از گفتار استخراج شده‌اند، ضرایب کپسترال فرکانسی در مقیاس مل (MFCC)^{۱۲} هستند. هر بردار ویژگی شامل ۱۲ ضریب است که ضریب صفرم این ویژگی استفاده نشده است. علاوه بر این ۱۲ ضریب، از انرژی نیز استفاده شده که جمعاً ۱۳ ضریب را تشکیل می‌دهند. برای بهبود دقت از مشتقات اول و دوم این ۱۳ ضریب نیز استفاده شده که بدین ترتیب بردارهای ویژگی استفاده شده ۳۹ بعدی هستند. این بردارها از پنجره‌هایی به طول ۲۵ میلی‌ثانیه با هم‌پوشانی ۱۵ میلی‌ثانیه استخراج شده‌اند. برای استخراج بردارهای ویژگی و همچنین آموزش مدل‌های

^{۱۱} Viterbi Algorithm

^{۱۲} Mel Frequency Cepstral Coefficients

^۸ Hidden Markov Model

^۹ Universal Background Model

^{۱۰} Maximum A Posteriori

۶- خلاصه و نتیجه‌گیری

در این مقاله دادگان جدیدی برای تصدیق هویت گوینده با گفتار تلفنی فارسی معرفی شد. همچنین نتیجه یک آزمایش بر روی این دادگان گزارش شد. این دادگان به نحوی طراحی و جمع‌آوری شده که بتواند شرایط واقعی را شبیه‌سازی کند و بتوان با آن حالات مختلف تصدیق هویت را تست گرفت. از این دادگان می‌توان برای تصدیق هویت وابسته به متن و همچنین برای تصدیق هویت با استفاده از متن تصادفی استفاده کرد. به دلیل اینکه در جمع‌آوری این دادگان از دو کانال انتقال تلفنی موجود در کشور استفاده شده، از این دادگان می‌توان برای آزمایش روش‌های مقاوم‌سازی نسبت به تغییرات کانال نیز استفاده کرد.

برای ادامه کار قصد داریم که دادگانی جمع‌آوری کنیم تا بتوان با استفاده از آن گذرواژه‌های متغیر برای هر گوینده را آزمایش کرد. در این روش در هر مرتبه از گوینده خواسته می‌شود تا جمله‌ای را تکرار کند. این جمله در هر مرتبه می‌تواند با جملات قبلی متفاوت باشد.

سپاسگزاری

از شرکت عصر گویش پرداز به خاطر در اختیار قرار دادن امکانات و شرایط ضبط صدا و همچنین از خانم کبیری که در تهیه این دادگان ما را یاری کردند، کمال تشکر را داریم.

مراجع

- [1] H. Beigi, "Fundamentals of Speaker Recognition," Springer 2011.
- [2] E. Bailly-Bailliere, S. Bengio, et al., "The BANCA database and evaluation protocol," *Lecture Notes in Computer Science (LNCS)*, vol. 2688, pp. 625-638, 2003.
- [3] S. Garcia-Salicetti, C. Beumier, et al., "BIOMET: A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities," *Lecture Notes in Computer Science*, vol. 2688/2003, pp. 845-853, 2003.
- [4] R. H. Woo, A. Park, and T. J. Hazen, "The MIT Mobile Device Speaker Verification Corpus: Data Collection and Preliminary Experiments," in *Speaker and Language Recognition Workshop (IEEE Odyssey)*, 2006.
- [5] N. A. Fox, B. A. O'Mullane, and R. B. Reilly, "The Realistic Multi-modal VALID database and Visual Speaker Identification Comparison Experiments," in *International Conference of Audio and Video-Based Person Authentication, AVBPA*, New York (US), July 2005.
- [6] M. Bijankhan, M. J. Sheikhzadegan, et al., "FARSDAT - The Farsi Spoken Language Database," In *Proceedings of International Conference on Speech Sciences and Technology*, vol. 2, pp. 826-829, 1994.
- [7] M. Bijankhan, et al., "TFARSDAT - The Telephone Farsi Speech Database," *proc. of the EUROSPEECH'2003*, vol.2, pp. 1526-1528.
- [8] S. Young, G. Evermann, et al., "The HTK Book (for HTK Version 3.4)", Cambridge University Engineering Department, 2006.

مخفی مارکوف از بسته نرم افزاری HTK استفاده شده است [۸]. برای مقاوم‌سازی نسبت به تغییرات کانال، تفریق میانگین کپسترال (CMS)^{۱۳} نیز روی ویژگی‌ها اعمال شده است.

برای آموزش مدل‌های هر گوینده از سه فایل شامل ارقام صفر تا نه استفاده شده است. این فایل‌ها از گفتارهای ضبط شده از کانال تلفن ثابت انتخاب شده‌اند. سایر فایل‌هایی که از همین کانال ضبط شده‌اند، گفتارهای تست هر گوینده را تشکیل می‌دهند.

بعد از پیدا کردن مرز بین ارقام برای هر فایل تست، تعداد ۸ عدد ۴ رقمی به صورت تصادفی از این ارقام تولید شده است. به دلیل اینکه در فایل‌های ضبط شده ارقام از کوچک به بزرگ بیان شده‌اند، اعداد تصادفی نیز به گونه‌ای تولید شدند که ارقام آنها از کوچک به بزرگ باشند. با استفاده از ارقام عدد تولید شده و همچنین گفتارهای مشخص شده برای هر رقم، گفتار هر عدد تولید گشته است. این گفتار در مدل گویندگان مختلف، از جمله مدل گوینده هدف، قرار داده شده و احتمال شباهت گفتار ورودی به هر مدل محاسبه شده است. گفتار هر گوینده در مدل خودش به عنوان گوینده هدف در نظر گرفته شده که باید به درستی تصدیق شود. این گفتار برای مدل سایر گویندگان به عنوان داده گوینده نادرست در نظر گرفته می‌شود که باید رد تصدیق شود. در نهایت خطاهای تصدیق نادرست (FA)^{۱۴} و رد نادرست (FR)^{۱۵} محاسبه شده‌اند. برای تصمیم‌گیری در مورد اینکه آیا یک گفتار متعلق به یک گوینده هست یا نه، امتیاز بدست آمده بعد از نرمال‌سازی با یک آستانه مقایسه می‌شود و در صورتی که بزرگ‌تر از آستانه باشد، گوینده تصدیق می‌شود. معمولاً این آستانه را جایی در نظر می‌گیرند که دو خطای بالا، با هم برابر شوند و نرخ خطای برابر (EER)^{۱۶} بدست آید. در این مقاله هم همین رویکرد استفاده شده است.

نتایج آزمایش تصدیق هویت با استفاده از دادگان جمع‌آوری شده، نرخ خطای برابر معادل با ۴.۲ درصد (میانگین روی ۵۰ گوینده) را نتیجه داده است. در این آزمایشات فقط از ۵۰ نفری که حداقل چهار مرتبه در فرآیند ضبط شرکت کرده‌اند، استفاده شده است.

¹³ Cepstral Mean Subtraction

¹⁴ False Accept

¹⁵ False Reject

¹⁶ Equal Error Rate