بسم الله الرحمن الرحيم

١٣٨٧

٩٦٢٤٩

دانشکدهٔ مهندسی برق و کامپیوتر
گروه مهندسی مخابرات

پایان‌نامه
برای دریافت درجهٔ کارشناسی ارشد در رشتهٔ مهندسی برق- مخابرات (سیستم)

عنوان

تصدیق هویت گوینده با استفاده از مدل مخلوط گوسی

استاد راهنما
دکتر مسعود گراوانچی‌زاده

استاد مشاور
دکتر بهزاد مظفری تازه‌کند

پژوهش‌گر
امیر کریم‌پور

۱۳۸۷ /۶/ ۲۵

خرداد۸۶

۹۶۲۴۹

دانشگاه: **تبریز**

دانشکده: **مهندسی برق و کامپیوتر**

عنوان :

تصدیق هویت گوینده با استفاده از مدل مخلوط گوسی

Speaker Verification using Gaussian Mixture Models

استاد راهنما: **دکتر مسعود گراوانچی‌زاده**

دانشجو: **امیر کریم‌پور**

این پروژه تحت قرارداد پژوهشی شماره ۵۰۰/۱۰۷۸۷/ت مورخ ۸۶/۷/۲۳ از پشتیبانی مادی و معنوی مرکز تحقیقات مخابرات ایران بهره‌مند شده است.

تقدیم به

پدر و مادر عزیز و مهربانم

که در سختی‌ها و دشواری‌های زندگی

همواره یاوری دلسوز و فداکار

و پشتیبانی محکم و مطمئن برایم بوده‌اند.

تقدیر و تشکّر

حمد و سپاس خداوند را که به نام او کار را آغاز کردیم و به یاری حضرتش به پایان رساندیم. در اینجا بر خود وظیفه می‌دانم تا از زحمات و راهنمایی‌های افرادی که در طول انجام پایان‌نامه هر یک به نوعی مرا یاری کردند، سپاسگزاری نمایم.

از استاد راهنمای گرامی جناب آقای دکتر مسعود گراوانچی‌زاده که در تمام مراحل پژوهش، یاری‌گر و پشتیبان بنده بوده‌اند، کمال تشکر و قدردانی را دارم.

از استاد مشاور گرانقدر جناب آقای دکتر بهزاد مظفری تازه‌کند به خاطر راهنمایی‌های ارزنده‌شان تشکر و قدردانی می‌نمایم.

از پدر، مادر و خواهر عزیزم که در این مدت با مهربانی در کنارم بوده‌اند و هر یک به نحوی در انجام این پایان‌نامه یاریم کرده‌اند، قدردانی می‌نمایم.

از دوستان عزیزم آقایان مهندس سعید مشگینی و مهندس مرتضی فرهید که در مراحل سخت کار مشوق من بوده‌اند، تشکر و قدردانی می‌کنم و از خدای متعال برایشان سعادت و بهروزی طلب می‌نمایم.

نام خانوادگی: **کریم‌پور** نام: **امیر**

عنوان پایان‌نامه: **تصدیق هویت گوینده با استفاده از مدل مخلوط گوسی**

استاد راهنما: **دکتر مسعود گراوانچی‌زاده**

استاد مشاور: **دکتر بهزاد مظفری تازه‌کند**

مقطع تحصیلی: **کارشناسی ارشد** رشته: **مهندسی مخابرات** گرایش: **سیستم**

دانشگاه: **دانشگاه تبریز** دانشکده: **مهندسی برق و کامپیوتر**

تاریخ فارغ‌التحصیلی: **خرداد ۸۷** تعداد صفحه: **۱۱۰**

کلید واژه‌ها: **تصدیق خودکار هویت گوینده (ASV)، مدل مخلوط گوسی (GMM)، مدل پیش‌زمینه کلی (UBM)، بیشینه‌سازی متوسط (EM)، ضرائب کپسترال فرکانس مِل (MFCC)، ضرائب کپسترال فاکتور انسانی (HFCC)، نرخ احتمال لگاریتمی (LLR)، نرخ خطای برابر (EER)، منحنی توازن خطای آشکارسازی (DET)**

چکیده:

تصدیق هویت گوینده در واقع فرآیند قبول یا رد کردن هویت ادعا شده توسط گوینده، بر اساس مشخصات صوتی او می‌باشد. یک سیستم تصدیق هویت گوینده می‌تواند در بسیاری از سیستم‌های حفاظتی از قبیل دسترسی به تلفن بانک، دسترسی به مناطق امنیتی در ساختمآن‌ها، دسترسی به پست‌های صوتی و همچنین در زمینه های جُرم‌شناسی، شناسایی گفتار، محیطهای چند گوینده (برای شناسایی یا ردیابی یک گوینده مشخص در بین چند گوینده) و ... به کار برده شود. در مقایسه با سایر روش‌های زیست‌سنجی مثل شناسایی اثر انگشت و یا شناسایی چهره، سیستم تصدیق هویت گوینده به تجهیزات خاص گران‌قیمت نیازی ندارد و از این رو به لحاظ اقتصادی یک روش مقرون به صرفه می باشد. همچنین برای تأیید هویت، لازم نیست که شخص حضور فیزیکی در محل داشته باشد، بلکه از راه دور نیز می توان عمل تصدیق را انجام داد.

یک سیستم تصدیق هویت گوینده در حالت کلی سه بخش استخراج ویژگی، مدل‌سازی و تصمیم‌گیری را شامل می‌شود، جایی که مدل‌سازی در آن بدنه اصلی سیستم را تشکیل می‌دهد. در یک سیستم تصدیق هویت گوینده مستقل از متن (جایی‌که که اطلاعات قبلی در مورد کلمات بیان شده توسط گوینده وجود ندارد)، مدل مخلوط گوسی (GMM) کارایی بالایی را نسبت به سایر مدل‌ها از خود نشان می‌دهد، به نحوی که حجم محاسباتی نیز پایین می‌باشد.

در این پایان‌نامه ابتدا سیستم پایه تصدیق هویت گوینده مبتنی بر مدل مخلوط گوسی پیاده‌سازی می‌شود و چگونگی به دست آوردن پارامترهای بهینه مدل GMM برای دستیابی به کارایی بیشینه مورد بررسی و آزمایش قرار گرفته است. سپس کارایی این سیستم در مقابل داده‌های نویزی ارزیابی می‌شود. در شرایط واقعی یک سیستم تصدیق هویت گوینده در مرحله تست معمولاً با داده‌های آلوده به نویز مواجه می‌شود، بنابراین میزان کارایی سیستم در چنین شرایطی یک فاکتور مهم برای ارزیابی سیستم می‌باشد. ویژگی جدیدی به نام ضرائب کپسترال فاکتور انسانی (HFCC) که اخیراً در سیستم‌های شناسایی گفتار پیشنهاد شده است، کارایی بالایی را نسبت به ضرائب رایج MFCC، در شرایط نویزی از خود نشان می‌دهد. در این پایان نامه برای اولین بار از این روش در سیستم‌های تصدیق هویت گوینده استفاده می‌شود. با توجه به نتایج شبیه‌سازی متوسط نرخ خطای برابر (EER) برای روش پیشنهادی، در مقدار نرخ سیگنال به نویز (SNR) برابر ۱۲ دسی‌بل، حدود ۳۰ درصد کاهش یافته که مقدار چشم‌گیری است. در قسمت آخر یک ارزیابی مقایسه‌ای بین روش تصمیم‌گیری کلاسیک نرخ احتمال لگاریتمی و روش جدید ماشین بردارهای پشتیبان (SVM)، در سیستم‌های ASV مبتنی بر GMM انجام شده است. با توجه به نتایج به دست آمده، روش SVM با تابع هسته خطی، مقدار خطای EER را به اندازه دو درصد کاهش داده است. با وجود بار محاسباتی اضافی که استفاده از روش‌های تصمیم‌گیری مبتنی بر SVM بر سیستم تحمیل می‌کنند، نتایج به دست آمده بهبودی کارایی چندانی را نشان نمی‌دهد.

فهرست مطالب

بخش اول: بررسی منابع و پژوهش‌های پیشین

فصل اول: پیش‌گفتار..... ۱

۱-۱ مقدمه..... ۲

۱-۲ ساختار پایان‌نامه ۴

فصل دوم: مروری بر سیستم تصدیق هویت گوینده..... ۷

۲-۱ مقدمه..... ۸

۲-۲ اصطلاحات و مفاهیم اولیه..... ۸

۲-۳ اجزاء تشکیل دهنده سیستم تصدیق هویت گوینده..... ۹

۲-۳-۱ استخراج ویژگی..... ۱۲

۲-۳-۲ مدل‌سازی و تطابق الگو..... ۱۴

۲-۳-۳ تصمیم‌گیری..... ۱۵

۲-۴ کاربردهای سیستم تصدیق هویت گوینده..... ۱۶

۲-۵ مزایا و معایب..... ۱۹

۲-۶ بررسی منابع..... ۲۰

بخش دوم: مواد و روش‌ها

فصل سوم: استخراج ویژگی‌ها..... ۲۵

۳-۱ مقدمه..... ۲۶

۳-۲ به دست آوردن سیگنال گفتار دیجیتال..... ۲۶

۳-۳ ویژگی‌های گفتار وابسته به گوینده..... ۲۷

۳-۴ ویژگی‌های ضرائب MFCC..... ۳۰

۳-۵ نحوه استخراج ضرائب MFCC ۳۱
۳-۵-۱ فریم‌بندی سیگنال گفتار ۳۱
۳-۵-۲ پنجره کردن ۳۲
۳-۵-۳ تبدیل فوریه گسسته (DFT) ۳۳
۳-۵-۴ انتقال به مقیاس فرکانسی Mel ۳۴
۳-۵-۵ کپستروم ۳۶
۳-۵-۶ مشتقات ضرائب MFCC ۳۷
۳-۶ ضرائب کپسترال فاکتور انسانی (HFCC) ۳۸
۳-۶-۱ محدودیت MFCC و اصلاح آن در HFCC ۳۹
۳-۶-۲ فیلتر بانک در HFCC ۴۰
فصل چهارم: مدل‌سازی و تطبیق الگو ۴۳
۴-۱ مقدمه ۴۴
۴-۲ مدل کردن و تطابق الگو در سیستم ASV ۴۴
۴-۳ مروری بر دسته‌بندی‌کننده‌های مختلف در سیستم ASV ۴۶
۴-۴ تصدیق هویت گوینده با استفاده از آشکارسازی نرخ احتمال ۴۹
۴-۵ مدل مخلوط گوسی ۵۲
۴-۵-۱ توصیف مدل ۵۲
۴-۵-۲ تخمین پارامترهای GMM ۵۴
۴-۵-۳ تفسیر مدل ۵۶
۴-۵-۴ مدل کردن سایر گویندگان (وانمودکنندگان) ۶۰
۴-۵-۴-۱ مدل پیش‌زمینه کلی (UBM) ۶۲
۴-۶ نتیجه‌گیری ۶۵

فصل پنجم: تصمیم‌گیری..... ۶۶

۵-۱ مقدمه..... ۶۷

۵-۲ تصمیم‌گیری با استفاده از نرخ احتمال لگاریتمی..... ۶۷

۵-۳ تعمیم تابع تصمیم گیری به توابع تفکیکی..... ۶۹

۵-۳-۱ تابع تصمیم‌گیری بازگشت خطی..... ۷۰

۵-۳-۲ تابع تصمیم‌گیری ماشین بردارهای پشتیبان..... ۷۱

۵-۴ نتیجه‌گیری..... ۷۲

بخش سوم: نتیجه‌گیری و بحث

فصل ششم: نتایج شبیه‌سازی و بحث..... ۷۳

۶-۱ مقدمه..... ۷۴

۶-۲ معیار ارزیابی..... ۷۴

۶-۳ پایگاه دادگان استفاده شده برای ارزیابی سیستم..... ۸۰

۶-۴ پیش‌تنظیمات..... ۸۱

۶-۵ انجام آزمایشات بر روی پایگاه دادگان و نتایج شبیه‌سازی..... ۸۲

۶-۵-۱ ارزیابی سیستم ASV پایه..... ۸۳

۶-۵-۲ ارزیابی سیستم ASV پایه در مقابل داده‌های نویزی..... ۸۷

۶-۵-۳ مقایسه دو روش تصمیم‌گیری..... ۹۰

فصل هفتم: نتیجه‌گیری و پیشنهادها..... ۹۳

۷-۱ نتیجه‌گیری..... ۹۴

۷-۲ پیشنهادهایی برای کارهای آینده..... ۹۶

فهرست منابع و مراجع..... ۹۸

فهرست شکل‌ها

شکل ۲-۱ نمودار بلوکی فاز یادگیری..... ۱۰
شکل ۲-۲ نمودار بلوکی فاز تست..... ۱۱
شکل ۲-۳ سیستم صوتی انسان..... ۱۳
شکل ۳-۱ نمودار بلوکی سیستم به دست آوردن سیگنال گفتار دیجیتال..... ۲۷
شکل ۳-۲ نحوه انتقال ساختار فیزیکی ارگان‌های تولید صدا در طیف صحبت..... ۲۹
شکل ۳-۳ نمایش تفاوت طیف صحبت افراد مختلف برای بیان یک صدای مشابه..... ۲۹
شکل ۳-۴ نمودار بلوکی استخراج کننده ضرائب MFCC..... ۳۱
شکل ۳-۵ چگونگی فریم‌بندی و پنجره کردن سیگنال صحبت..... ۳۲
شکل ۳-۶ فیلتر بانک طراحی شده اولیه توسط Davis و Mermelstain..... ۳۵
شکل ۳-۷ مراحل استخراج ضرائب MFCC روی قسمتی از سیگنال گفتار..... ۳۶
شکل ۳-۸ ساختار فیلتر بانک HFCC..... ۴۲
شکل ۴-۱ ساختار یک سیستم ASV پایه..... ۴۵
شکل ۴-۲ مقایسه دسته‌بندی‌کننده‌های تولیدی با دسته‌بندی‌کننده‌های تفکیکی..... ۴۸
شکل ۴-۳ ساختار سیستم ASV مبتنی بر نرخ احتمال در فاز تست..... ۵۰
شکل ۴-۴ نمایش چگالی مخلوط گوسی با M مؤلفه..... ۵۳
شکل ۴-۵ نمایش چگونگی انعکاس ساختارهای ناحیه صوتی توسط مجموعه کلاسهای صوتی..... ۵۶
شکل ۴-۶ نمایش چگونگی انطباق شکل طیفی کلاس صوتی با GMM..... ۵۷
شکل ۴-۷ مقایسه توزیع مدل‌های مختلف..... ۵۸
شکل ۴-۸ استفاده از مدل گویندگان دیگر برای محاسبه مدل پیش‌زمینه..... ۶۱
شکل ۴-۹ استفاده از یک مدل واحد برای نمایش مدل وانمودکنندگان..... ۶۲
شکل ۴-۱۰ آموزش UBM با استفاده از ترکیب تمام داده‌ها..... ۶۴

شکل ۴-۱۱ ترکیب زیر مدل‌های UBM برای به دست آوردن مدل نهایی..... ۶۴

شکل ۶-۱ توزیع نمرات وانمودکنندگان و FAR..... ۷۶

شکل ۶-۲ توزیع نمرات مشتری و FRR..... ۷۷

شکل ۶-۳ هم‌پوشانی توزیع نمرات کاربر با وانمودکنندگان و FAR با FRR..... ۷۷

شکل ۶-۴ منحنی DET..... ۷۸

شکل ۶-۵ تعداد انحراف استاندارد از میانگین با محاسبه درصد مساحت زیر منحنی نرمال استاندارد..... ۷۹

شکل ۶-۶ مقایسه کارایی دسته‌بندی کننده GMM با مرتبه مدل متفاوت..... ۸۳

شکل ۶-۷ مقایسه کارایی دسته‌بندی کننده GMM با تغییرات طول داده‌های تست..... ۸۴

شکل ۶-۸ مقایسه کارایی دسته‌بندی کننده GMM بر حسب تعداد کاربران موجود..... ۹۳

شکل ۶-۹ منحنی تغییرات HTER به ازاء تغییرات حد آستانه ۸۶

شکل ۶-۱۰ کاهش کارایی سیستم با افزایش میزان نویز..... ۸۷

شکل ۶-۱۱ مقایسه کارایی MFCC و HFCC در مقابل داده‌های نویزی با SNR های متفاوت..... ۸۸

شکل ۶-۱۲ مقایسه کارایی SVM با توابع هسته گوناگون..... ۹۰

شکل ۶-۱۳ مقایسه کارایی دو روش نرخ احتمال لگاریتمی و SVM در تصمیم‌گیری نمرات..... ۹۱

فهرست جدول‌ها

جدول ۶-۱ مقایسه EER و minHTER برای ارزیابی MFCC و HFCC در مقابل داده‌های نویزی با SNR های متفاوت..... ۸۹

جدول ۶-۲ مقایسه EER و minHTER برای ارزیابی روش‌های تصمیم‌گیری SVM با توابع هسته مختلف و روش نرخ احتمال لگاریتمی..... ۹۲

فهرست اختصارات

ASV	Automatic Speaker Verification
DET	Detection Error Trade-off
EER	Equal Error Rate
EM	Expectation Maximization
FA	False Acceptance
FAR	False Acceptance Rate
FR	False Rejection
FRR	False Rejection Rate
GMM	Gaussian Mixture Model
GMM-LR	Gaussian Mixture Model Likelihood Ratio
LR	Likelihood Ratio
LRR	Log Likelihood Ratios
HFCC	Human Factor Cepstral Coefficients
HMM	Hidden Markov Model
HTER	Half Total Error Rate
LPC	Linear Predictive Coding
MFCC	Mel Frequency Cepstral Coefficients
MLP	Multi-Layer Perception
MSE	Mean Square Error
PDA	Personal Digital Assistant
PIN	Personal Identification Number
SNR	Signal to Noise Ratio
SVM	Support Vector Machine
VQ	Vector Quantization
UBM	Universal Background Model

بخش نخست

بررسی منابع و پژوهشهای پیشین

فصل اول

پیش گفتار

۱-۱ مقدمه

فرض کنید در خانه و یا محل کار خود نشسته‌اید و مشغول انجام کاری می‌باشید. در این هنگام یک تماس تلفنی دریافت می‌کنید که در آن شخص تماس گیرنده خودش را معرفی می‌کند و شما به سرعت طرف مقابل را شناسایی کرده و مکالمه را ادامه می‌دهید. بدون این که متوجه شوید، با این کار شما در واقع عمل تشخیص هویت یک گوینده را به طور طبیعی انجام داده‌اید. این مثال نمونه‌ای از یک *شناسایی گوینده*[1] طبیعی است که در زندگی روزمره خود بارها آن را انجام می‌دهیم. با پیشرفت تکنولوژی و گسترش بیش از پیش رابطه انسان با ماشین، می‌توان وارد شدن سیستم‌های شناسایی صحبت را در جنبه‌های مختلف زندگی روزمره، مشاهده کرد. همچنین از آنجایی که صحبت یکی از وسایل ارتباطی ما می‌باشد، پس تکنولوژیهای مربوط به آن نیز به عنوان یک راه‌کار تقابل انسان با کامپیوتر در جوامع پیشرفته مورد استقبال قرار می‌گیرد.

یک نیاز فزآینده برای شناسایی هویت افراد در دنیای اطلاعات امروز وجود دارد که کاربردهای آن از *کارتهای اعتباری*[2] گرفته تا کنترل مرزها و *جرم‌شناسی*[3] گسترش می‌یابد. در حالت کلی یک شخص می‌تواند با استفاده از سه روش متفاوت زیر، مورد شناسایی قرار گیرد:

۱. توسط چیزی که فرد مالک آن است مثل یک کلید و یا کارت اعتباری

۲. توسط اطلاعاتی که شخص از قبل آن را می‌داند مثل رمز عبور و *شماره شناسایی هویت*[4]

۳. توسط هویتی که شخص دارد مثل اثر انگشت، صدا و مشخصات صورت

دو روش اولی متدهای تعیین هویت سنتی هستند که در طول چندین قرن مورد استفاده قرار گرفته‌اند و با این وجود دارای کاستی‌هایی نیز هستند. برای مثال کارتهای اعتباری و یا کلیدها می‌توانند گم و یا دزدیده شوند و همچنین رمز عبور و شماره شناسایی هویت به راحتی فراموش شده

---

[1] Speaker Recognition
[2] Credit Cards
[3] Forensic
[4] Personal Identification Number

و یا مورد سوء استفاده قرار گیرند. در مورد روش آخر که به عنوان شناسایی هویت *زیست‌سنجی*[1] معروف است، مشکلات ذکر شده در دو روش قبلی تا حدودی تعدیل می‌شود. در زیر بعضی از دلایل استفاده از روش زیست سنجی به طور مختصر آورده شده است.

ساختمان ظاهری، فیزیولوژی و عادات آموختاری افراد، مشخصات منحصر به فردی می‌باشند که آن‌ها را نسبت به همدیگر متمایز می‌سازند. از طرف دیگر افزایش قدرت پردازش و کاهش اندازه ریز پردازنده‌ها، انگیزه ساخت و به کار بردن روش‌های زیست‌سنجی را به ما می‌دهد. از این رو سرعت تحقیق و گسترش این روشها در سالهای اخیر به طور قابل ملاحظه‌ای افزایش یافته است. یکی از تکنیکهای شناسایی زیست‌سنجی که به عنوان یک روش مناسب مورد علاقه محققان است، روش مبتنی بر استفاده از *گفتار*[2] می‌باشد.

همان طور که گفته شد، زیست سنجی از بسیاری از جنبه‌های ظاهری افراد از قبیل الگوی عنبیه، اثر انگشت و یا گفتار به عنوان روش تعیین هویت یک شخص خاص استفاده می‌کند. هدف ما در این پایان‌نامه، تحقیق و بهبود شیوه‌های کنونی مورد استفاده برای مدل کردن، امتیازدهی و پردازش‌های ابتدایی و انتهایی سیگنال در زمینه شناسایی یک فرد با استفاده از سیگنال گفتار (شناسایی گوینده) می‌باشد. شناسایی گوینده در واقع شناسایی یک فرد بر اساس مشخصات صوتی او می‌باشد. این کار همچنین به عنوان *شناسایی صدا*[3] نیز معروف است که این نام‌گذاری اغلب با *شناسایی گفتار*[4] اشتباه گرفته می‌شود. با این وجود این دو عبارت به عنوان دو مسئله مکمل هم مورد مطالعه قرار می‌گیرند. در حالیکه شناسایی گفتار تعیین عبارت گفته شده بدون توجه به گوینده آن می‌باشد، هدف شناسایی صدا معین کردن هویت گوینده بدون توجه به عبارت ادا شده است. سیستم‌های شناسایی گوینده در یک دسته بندی کلی می‌تواند به دو زیر شاخه *تصدیق هویت گوینده*[5] و *تعیین هویت گوینده*[6] تقسیم شوند. بر خلاف سیستم تعیین هویت گوینده که در آن هویت گوینده از میان کاربرهایی که قبلاً به

---

[1] biometric
[2] Speech
[3] Voice Recognition
[4] Speech Recognition
[5] Speaker Verification
[6] Speaker Identification

سیستم معرفی شده‌اند تشخیص داده می‌شود، در سیستم تصدیق هویت گوینده، سیستم می‌تواند گوینده را به عنوان این که با هویت مرجع تطبیق دارد قبول کند و یا به عنوان یک گوینده مقلد آنرا رد کند. یک سیستم شناسایی گوینده می‌تواند در بسیاری از سیستم‌های حفاظتی از قبیل دسترسی به تلفن بانک، دسترسی به مناطق امنیتی در ساختمآن‌ها، دسترسی به *پست‌های صوتی*[1] و همچنین در زمینه های جرم‌شناسی، شناسایی گفتار، محیط‌های چند گوینده[2] (برای شناسایی یا ردیابی یک گوینده مشخص در بین چند گوینده) و ... به کار برده شود.

۱-۳ ساختار پایان‌نامه

این پایان‌نامه در سه بخش کلی سازماندهی شده است که هر کدام خود شامل چندین فصل به صورت زیر می‌باشد.

در بخش نخست یک بررسی اجمالی از منابع و پژوهش‌های پیشین ارائه شده است که از دو فصل تشکیل یافته است. فصل اول شامل معرفی کلی سیستم‌های شناسایی گوینده و اهداف و انگیزه‌های موجود در پایان‌نامه می‌باشد. فصل دوم مروری کلی بر اجزاء مختلف تشکیل دهنده یک سیستم تصدیق هویت گوینده را خواهد داشت. این فصل انواع تقسیم بندی ها و نام گذاری‌هایی که در این زمینه وجود دارد را بیان می کند و در نهایت نقاط ضعف و قوت سیستم مورد نظر و همچنین کاربردهای آن در زمینه های مختلف را مورد بحث قرار خواهد داد. در پایان مروری بر مهمترین منابع و پژوهشهای انجام یافته در این زمینه به عمل خواهد آمد.

بخش دوم مواد و روش‌های استفاده شده در این کار تحقیقاتی را بیان می‌کند. جزئیات سه فصل تشکیل دهنده این بخش به صورت زیر است. در فصل سوم روش استخراج ویژگی های منحصر به فرد گوینده از روی سیگنال گفتار مورد مطالعه قرار می‌گیرد. روش‌های مختلف مورد استفاده در این زمینه به طور خلاصه توضیح داده می‌شوند. ولی تمرکز اصلی بر روی *ضرائب کپستروم در مقیاس مِل*[3] است

---

[1] voicemail
[2] multi-speaker
[3] Mel frequency cepstral coefficients

که در این پایان‌نامه نیز از آن‌ها استفاده می‌شود. **همچنین یک ویژگی جدیدی را که در سیستم‌های شناسایی گفتار در مقابل داده های نویزی کارایی خوبی از خود نشان می دهد بررسی می‌کنیم. این ویژگی جدید که *ضرائب کپستروم فاکتورهای انسانی*[1] نامیده می‌شود برای اولین بار در این پایان‌نامه در سیستم تصدیق هویت گوینده و در مقابل داده‌های نویزی مورد استفاده قرار گرفته و کارایی آن در فصل نتایج شبیه‌سازی در مقایسه با روش قبلی مورد بررسی قرار خواهد گرفت.** فصل چهارم مروری بر روش‌های دسته‌بندی مشخصات استخراج شده در فصل قبلی را بر عهده دارد که هدف اصلی آن مطالعه وسیع دسته‌بندی کننده‌های آماری بویژه مدل مخلوط گوسی است که بدنه اصلی این پایان‌نامه را تشکیل می‌دهد. **این بخش نحوه دقیق مدل کردن مشخصات صوتی گوینده اصلی و همچنین بقیه گویندگان که به عنوان تقلید کننده مطرح می‌شوند را تحقیق می‌کند.** در فصل پنجم مرحله تصمیم گیری در رد یا قبول کردن یک گوینده مدعی مورد بررسی قرار می گیرد. در ابتدا به طور مفصل روش *نسبت احتمال لگاریتمی*[2] را که به صورت گسترده در سیستم های تصدیق هویت گوینده مبتنی بر مدل مخلوط گوسی استفاده می شود، مطرح می‌گردد. **در ادامه یک روش تصمیم گیری جدید مبتنی بر ماشین بردارهای پشتیبان که در تحقیقات اخیر مورد توجه محققان قرار گرفته، معرفی می‌شود.**

بخش سوم که شامل فصل ششم و هفتم است، نتایج شبیه سازی و مباحث پیرامون آن را تشکیل می‌دهد. در فصل ششم سیستم تصدیق هویت گوینده توسط نرم افزار مطلب پیاده‌سازی شده و نتایج شبیه سازی ارائه می‌شوند. در ابتدا جزئیات پایگاه داده مورد استفاده شده در این پایان‌نامه مورد مطالعه قرار می گیرد، سپس معیار های ارزیابی استاندارد برای سنجش کیفیت و کارایی یک سیستم تصدیق هویت گوینده مطرح می شود. سپس پیش تنظیماتی راجع به مرحله استخراج

---

[1] Human factor cepstral coefficients
[2] Log Likelihood Ratios

ویژگی‌ها، همچنین پارامترهای دسته‌بندی گوینده مدعی و مقلد انجام می گیرد. در نهایت نتایج شبیه‌سازی در مورد نشان دادن کارایی روش‌های ارائه شده آورده شده است. این نتایج در چند بخش مورد ارزیابی قرار می‌گیرند. در ابتدا نتایج شبیه‌سازی در بخش استخراج ویژگی در مورد داده‌های نویزی، دو روش مذکور را مورد مقایسه قرار می‌دهد. پس از آن انواع آزمایشات لازم برای به دست آوردن یک مدل مخلوط گوسی مناسب برای پایگاه داده مورد نظر، از قبیل تعداد توابع گوسی استفاده شده برای مدل، حجم اطلاعات صحبت در مرحله یادگیری و تست و همچنین تعداد کاربران موجود در سیستم، انجام می گیرد. سپس مقایسه‌ای بین روش تصمیم‌گیری رایج و روش تصمیم‌گیری تفکیکی جدید، که هر دو از مدل مخلوط گوسی برای دسته‌بندی کاربران و وانمودکنندگان استفاده می کنند، انجام می‌گیرد و این دو روش از نظر کارایی مورد ارزیابی واقع می‌شوند. در نهایت نتایج این شبیه‌سازی‌ها در هر سه مورد گفته شده مورد بحث قرار خواهند گرفت. در فصل ششم نتیجه گیری از کل پایان نامه به عمل می آید و پیشنهاداتی برای کارهای آینده مطرح خواهند شد.