

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه صداوسیما جمهوری اسلامی ایران

پایان نامه کارشناسی ارشد رشته مهندسی صدا

بازشناسی گوینده مبتنی بر روشهای ادغام اطلاعات در سطح تصمیم

دانشجو:

ناهید علینقی زاده

استاد راهنما:

دکتر علی جبار رشیدی

استاد مشاور:

دکتر معصومه شفیعیان

پاییز ۱۳۹۱

چکیده:

در میان تمام فناوریها و روشهای تشخیص هویت زیستی، بازشناسی گوینده بر مبنای اطلاعات صحبت را می‌توان طبیعی‌ترین و اقتصادی‌ترین روش برای سیستمهای ارتباط انسان-ماشین دانست. همچنین توسعه فناوری پردازش صحبت باعث تقویت بسیاری از کاربردهای بازشناسی گوینده شده است. از آنجا که یکی از مهمترین زمینه‌های تحقیقاتی فعال که در سالهای اخیر در بسیاری از کاربردها باعث بهبود عمده در بازشناسی گوینده شده است استفاده از روشها و مدل‌های ادغام اطلاعات در سطوح مختلف می‌باشد محور اصلی در این پژوهش، بهبود دقت بازشناسی گوینده با استفاده از ادغام اطلاعات در سطح تصمیم است. در این پایان نامه رویکرد استخراج ویژگیهای جدیدی بر پایه ضرایب کپسترال جهت فراهم سازی منابع تصمیم گیری مورد استفاده در ادغام تصمیم مد نظر قرار گرفته و با توجه به اینکه مشتقات هر تابع بخشی از ویژگیهای مستتر در آن را به نمایش می‌گذارد از مشتق اول و دوم ضرایب کپسترال مل-فرکانس به عنوان بردارهای ویژگی ثانویه استفاده نموده ایم. این رویکرد به مفهوم استفاده و بهره‌گیری همزمان از اطلاعات نهفته در بردار ویژگی، تغییرات (بردار سرعت) و نرخ تغییرات (بردار شتاب) ویژگی می‌باشد. پس از بازشناسی گوینده مبتنی بر این سه بردار ویژگی به صورت مجزا، جهت بهبود دقت و صحت نتایج بازشناسی و شناسایی، اقدام به طراحی چارچوب ادغام اطلاعات در سطح تصمیم نموده ایم. استفاده همزمان از این بردارهای ویژگی در بستر ادغام تصمیم تاکنون توسط محققان گزارش نشده است. استفاده از روشهای مناسب جهت خوشه‌بندی بردارهای ویژگی از جمله روش کوانتیزه کردن برداری و همچنین توابع تعیین اندازه شباهت از جمله فاصله مالهالانویس و فاصله حداکثر شباهت مبتنی بر حد آستانه از دیگر کارهای مهم انجام شده در پژوهش می‌باشد. در مرحله ادغام تصمیم، نتایج بازشناسیهای اولیه را با استفاده از روشهای ادغام تصمیم از جمله روشهای رأی‌گیری، رتبه‌بندی و روش امتیازدهی ترکیب و نتیجه را به عنوان بازشناسی نهایی استفاده ایم. نتایج نشان دهد که روشهای ادغام تصمیم باعث بهبود قابل توجه در دقت بازشناسی گوینده نسبت به حالت بدون ادغام شود. از نتایج دیگر این است که استفاده همزمان از اطلاعات مستتر در بردارهای تغییرات ضرایب کپسترال و بردارهای نرخ این تغییرات برای بازشناسی گوینده باعث بهبود کارایی سیستم بازشناسی گوینده می‌شود.

واژگان کلیدی: بازشناسی گوینده، ادغام تصمیم، استخراج ویژگی، ترکیب ویژگیها، ضرایب

کپستروم، فرکانس مل.

فهرست مطالب

صفحه	عنوان
۱	۱- فصل اول- کلیات تحقیق
۲	۱-۱- اصول بازشناسی گوینده
۵	۱-۲- مروری مختصر بر تاریخچه بازشناسی گوینده
۸	۱-۳- مسئله تحقیق و هدف آن
۸	۱-۴- مفاهیم تحقیق
۱۰	۲- فصل دوم: مبانی نظری تحقیق
۱۱	۲-۱- مقدمه
۱۲	۲-۲- بازشناسایی گوینده
۱۷	۲-۲-۱- شناسایی گوینده
۱۹	۲-۲-۲- تصدیق گوینده
۲۰	۲-۳- بخشهای اصلی سیستم بازشناسی گوینده
۲۱	۲-۴- مدل سازی آکوستیکی
۲۱	۲-۴-۱- مقدمه
۲۳	۲-۴-۲- مدل‌های مبتنی بر همپچی زمانی پویا
۲۶	۲-۴-۳- مدل‌های مبتنی بر کمی سازی برداری
۲۷	۲-۴-۴- مدل‌های مرکب گوسی
۳۱	۲-۴-۵- مدل‌های مخفی مارکف
۳۳	۲-۴-۶- حدآستانه و تصمیم های رد/قبول
۳۶	۲-۵- مروری بر تاریخچه بازشناسی گفتار و گوینده
۴۰	۲-۶- مبانی و روشهای ادغام اطلاعات و مروری بر پیشینه تحقیقات آن
۴۰	۲-۶-۱- مقدمه

۴۱ ۲-۶-۲-تعریف و مفهوم ادغام اطلاعات
۴۳ ۲-۶-۳- سطوح ادغام اطلاعات
۴۷ ۲-۶-۴- روشهای ادغام تصمیم و مرور سوابق آن
۴۷ ۲-۶-۴-۱- مقدمه
۵۱ ۲-۶-۴-۲- روشهای ادغام تصمیم
۵۱ ۲-۶-۴-۲-۱- روشهای رأی‌گیری
۵۱ ۱- ۲-۶-۴-۲-۱- روش Max- VF برای ادغام تصمیم
۵۳ ۲- ۲-۶-۴-۲-۱- روش Min-VF برای ادغام تصمیم
۵۴ ۳- ۲-۶-۴-۲-۱- روش AVF برای ادغام تصمیم
۵۳ ۴- ۲-۶-۴-۲-۱- روش اکثریت آرا
۵۵ ۵- ۲-۶-۴-۲-۱- روش رأی‌گیری موزون
۵۵ ۶- ۲-۶-۴-۲-۱- روش بهترین اکثریت برای ادغام تصمیم
۵۵ ۲-۶-۴-۲-۲- روش‌های مبتنی بر رتبه
۵۸ ۳-۶-۴-۲-۳- روش‌های مبتنی بر امتیاز
۵۹ ۴-۶-۴-۲-۴- روش استنتاج بیز
۵۹ ۵-۶-۴-۲-۵- روش دمپستر - شفر
۶۱ ۱-۵-۶-۴-۲-۵- قانون ترکیب دمپستر
۶۲ ۵-۶-۲- مروری بر تحقیقات پیشین در حوزه ادغام تصمیم
۶۶ ۳- فصل سوم: روش تحقیق
۶۷ ۳-۱- مقدمه
۶۹ ۳-۲- پردازش اولیه سیگنال صحبت
۶۹ ۳-۲-۱- فیلترهای پیش تأکید
۷۱ ۳-۲-۲- حذف نواحی سکوت از سیگنال صحبت
۷۳ ۳-۲-۲-۱- روشهای حذف سکوت

۷۴.....	۱-۲-۲-۳-نرخ عبور از صفر
۷۷.....	۲-۱-۲-۳- روش انرژی زمان کوتاه
۷۸.....	۳-۱-۲-۳- روش فاصله ماهالانویس
۸۰.....	۴-۱-۲-۳- روش مبتنی بر معیار حداکثر شباهت و حد سکوت
۸۴.....	۳-۲-۳- قاب گذاری (فریم بندی)
۸۵.....	۴-۲-۳- پنجره گذاری
۸۵.....	۱-۴-۲-۳- پنجره مستطیلی
۸۶.....	۲-۴-۲-۳- پنجره گوسی
۸۷.....	۳-۴-۲-۳- پنجره همینگ
۸۸.....	۴-۴-۲-۳- پنجره بلک من-هریس
۹۰.....	۳-۳- استخراج ویژگیها برای بازشناسی گوینده
۹۱.....	۱-۳-۳- ضرایب کپسترال فرکانس-مل
۹۱.....	۱-۱-۳-۳- مقدمه
۹۳.....	۲-۱-۳-۳- استخراج طیف سیگنال صحبت
۹۸.....	۳-۱-۳-۳- آنالیز فرکانس-مل
۱۰۵.....	۲-۳-۳- ویژگیهای حاصل از تغییرات و نرخ تغییرات MFCC
۱۰۷.....	۴-۳- طبقه بندی و تطبیق الگو (تطبیق ویژگیها)
۱۰۹.....	۱-۴-۳- ایجاد پایگاه داده گوینده
۱۱۳.....	۲-۴-۳- تطبیق الگو (تطبیق ویژگیها)
۱۱۵.....	۵-۳- سیستم بازشناسی مبتنی بر ادغام تصمیم
۱۱۸.....	۴- فصل چهارم: یافته های تحقیق
۱۱۹.....	۱-۴- مقدمه
۱۱۹.....	۲-۴- داده ها و شرایط پیاده سازی روشها
۱۲۱.....	۳-۴- پیش پردازشهای انجام شده

۱۲۱	۴-۴- نتایج مدل‌سازی گوینده در مرحله آموزش
۱۲۲	۴-۵- نتایج بازشناسی گوینده (مرحله آزمایش)
۱۲۲	۴-۵-۱- بررسی اثر تعداد خوشه‌ها در دقت روشهای بازشناسی گوینده
۱۴۳	۴-۵-۲- بررسی اثر تغییرات ابعاد فیلترهای مل و ابعاد خوشه‌ها در دقت بازشناسی گوینده
۱۵۶	۴-۵-۳- محاسبه دقت روشهای بازشناسی و استخراج اثر تغییر تعداد خوشه‌ها و تعداد فیلترها
۱۶۰	۵- فصل پنجم: نتایج و پیشنهادهای تحقیق
۱۶۱	۵-۱- مقدمه
۱۶۱	۵-۲- دستاوردها و نتیجه‌گیری
۱۶۳	۵-۳- پیشنهادهای تحقیق
۱۶۴	فهرست منابع مأخذ

فهرست جدولها

عنوان	صفحه
جدول ۱-۱ وضعیت خواص چهارگانه برای الگوهای زیستی	۲
جدول ۱-۲ تحقیقات انجام شده در حوزه بازشناسی گوینده	۳۸
جدول ۲-۲ مقایسه سطوح ادغام اطلاعات	۴۶
جدول ۲-۴ تصمیمات محلی رتبه‌بندی شده ۵ طبقه بندی کننده برای ۳ کلاس C_1, C_2, C_3	۵۷
جدول ۱-۳ مقایسه خطای روشهای حذف نواحی سکوت	۸۳
جدول ۱-۴ اطلاعات مربوط به داده های مورد استفاده	۱۲۰
جدول ۲-۴ نتایج بازشناسی صوتی برای $N=5,10,15$ و $N_f=12$	۱۲۳
جدول ۳-۴ نتایج بازشناسی صوتی برای $N=20,40,60$ و $N_f=12$	۱۲۴
جدول ۴-۴ نتایج بازشناسی صوتی برای $N=5,10,15$ و $N_f=15$	۱۲۷
جدول ۵-۴ نتایج بازشناسی صوتی برای $N=20,40,60$ و $N_f=15$	۱۲۸
جدول ۶-۴ نتایج بازشناسی صوتی برای $N=5,10,15$ و $N_f=20$	۱۳۱
جدول ۷-۴ نتایج بازشناسی صوتی برای $N=20,40,60$ و $N_f=20$	۱۳۲
جدول ۸-۴ نتایج بازشناسی صوتی برای $N=5,10,15$ و $N_f=30$	۱۳۵
جدول ۹-۴ نتایج بازشناسی صوتی برای $N=20,40,60$ و $N_f=30$	۱۳۶
جدول ۱۰-۴ نتایج بازشناسی صوتی برای $N=5,10,15,20,40,60$ و $N_f=40$	۱۳۹
جدول ۱۱-۴ نتایج بازشناسی صوتی برای $N=5,10,15,20,40,60$ و $N_f=40$	۱۴۰

فهرست شکلها

صفحه	عنوان
..... ۴	شکل ۱-۱ سیستم تصدیق گوینده
..... ۴	شکل ۲-۱ سیستم شناسایی گوینده
..... ۱۷	شکل ۱-۲ سیستم شناسایی گوینده
..... ۱۹	شکل ۲-۲ سیستم تصدیق گوینده
..... ۲۱	شکل ۳-۲ فرایند مدل سازی آکوستیکی
..... ۲۴	شکل ۴-۲ الگوریتم DTW برای همراستاسازی زمانی
..... ۳۲	شکل ۵-۲ سیستم تصدیق گوینده مبتنی بر HMM
..... ۳۴	شکل ۶-۲ استخراج آستانه با استفاده از مدل غیر واقعی
..... ۳۵	شکل ۷-۲ منحنی EER - DET نقطه تلاقی منحنی با خط $x=y$
..... ۴۴	شکل ۸-۲ ساختار ادغام در سطح سیگنال (تصویر)
..... ۴۵	شکل ۹-۲ ساختار ادغام در سطح ویژگی
..... ۴۶	شکل ۱۰-۲ ساختار ادغام در سطح تصمیم
..... ۴۸	شکل ۱۱-۲ سیستم بازشناسی گوینده موازی
..... ۶۰	شکل ۱۲-۲ مجموعه π نظریه ساده (θ)
..... ۶۸	شکل ۱-۳ ساختار سیستمهای بازشناسی گوینده اولیه (بدون ادغام)
..... ۶۸	شکل ۲-۳ ساختار سیستم بازشناسی گوینده مبتنی بر ادغام تصمیم
..... ۶۹	شکل ۳-۳ مراحل پردازش اولیه سیگنال صحبت
..... ۷۰	شکل ۴-۳ فیلتر پیش تأکید
..... ۷۱	شکل ۵-۳ نتایج انجام پیش تأکید روی سیگنال مربوط به کلمه "ایران" با دو مقدار α (۰.۶ و ۰.۸)
..... ۷۲	شکل ۶-۳ نمونه سیگنال صحبت و نواحی سکوت در آن
..... ۷۳	شکل ۷-۳ سیگنال صحبت شکل ۶-۳ با حذف نواحی سکوت

- شکل ۳-۸ سیگنال صحبت یک گوینده برای کلمه "ایران" و فیلتر ZCR ۷۵
- شکل ۳-۹ سیگنال صحبت گوینده شکل ۳-۸ برای کلمه "ایران" و فیلتر ZCR مربوطه ۷۶
- شکل ۳-۱۰ سیگنال صحبت گوینده شکل ۳-۸ برای کلمه "ایران" و فیلتر ZCR مربوطه ۷۷
- شکل ۳-۱۱ حذف نواحی سکوت با استفاده از روش STE برای سیگنال صحبت گوینده شکل ۳-۸ برای کلمه "ایران" ۷۸
- شکل ۳-۱۲ توزیع نرمال متغیر فاصله ماهالانویس ۷۹
- شکل ۳-۱۳ حذف نواحی سکوت با استفاده از روش MLM-ST برای سیگنال صحبت گوینده شکل ۳-۸ برای کلمه "ایران" ۸۲
- شکل ۳-۱۴ فریم بندی و هاپینگ در حوزه زمان ۸۴
- شکل ۳-۱۵ پنجره مستطیلی در حوزه زمان و پاسخ فرکانسی آن ۸۶
- شکل ۳-۱۶ پنجره گوسی در حوزه زمان و پاسخ فرکانسی آن ۸۶
- شکل ۳-۱۷ پنجره همینگ در حوزه زمان و پاسخ فرکانسی آن ۸۷
- شکل ۳-۱۸ مقایسه پنجره های همینگ و مستطیلی ۸۸
- شکل ۳-۱۹ پنجره بلک من-هریس در حوزه زمان و پاسخ فرکانسی آن ۸۹
- شکل ۳-۲۰ یک نمونه سیگنال و نتیجه پنجره گذاری آن با پنجره همینگ ۹۰
- شکل ۳-۲۱ ساختار پردازشها از ابتدای دریافت سیگنال صحبت تا استخراج ضرایب MFCC ۹۱
- شکل ۳-۲۲ فرمتها و پوش فرکانسی سیگنال صحبت ۹۱
- شکل ۳-۲۳ روند استخراج ضرایب MFCC ۹۲
- شکل ۳-۲۴ مفهوم تبدیل فوریه زمان-کوتاه ۹۴
- شکل ۳-۲۵ متوسط اندازه طیف حاصل از زمانهای کوتاه(فریمهای سیگنال) و پوش اندازه طیف اصلی برای ۱۰ بار تکرار کلمه "ایران" برای گوینده ۳ ۹۵
- شکل ۳-۲۶ متوسط اندازه طیف حاصل از زمانهای کوتاه(فریمهای سیگنال) و پوش اندازه طیف اصلی برای ۱۰ بار تکرار کلمه "سلام" برای گوینده ۳ ۹۶
- شکل ۳-۲۷ متوسط اندازه طیف حاصل از زمانهای کوتاه(فریمهای سیگنال) و پوش اندازه طیف اصلی برای ۱۰ بار تکرار کلمه "سلام" برای گوینده ۹۷

- شکل ۳-۲۸ مقیاس فرکانس - مل ۹۹
- شکل ۳-۲۹ بانک فیلتر مقیاس مل ۱۰۰
- شکل ۳-۳۰ ضرایب طیف مل (MFCC) ۱۰۱
- شکل ۳-۳۱ ضرایب MFCC برای یک گوینده و گفتار متفاوت ۱۰۳
- شکل ۳-۳۲ ضرایب MFCC برای چهار گوینده متفاوت و گفتار یکسان ۱۰۴
- شکل ۳-۳۳ ضرایب MFCC ۱۰۵
- شکل ۳-۳۴ بردارهای ویژگی MFCC، DMFCC و D^2 MFCC ۱۰۶
- شکل ۳-۳۵ (الف) بردارهای ویژگی (MFCC) قبل از اعمال VQ ۱۰۸
- شکل ۳-۳۶ (ب) بردارهای ویژگی (MFCC) منتخب پس از اعمال VQ ۱۰۹
- شکل ۳-۳۷ تعداد ۲ خوشه در الگوریتم K-means (تعداد مراکز=۲) ۱۱۰
- شکل ۳-۳۸ تعداد ۸ خوشه در الگوریتم K-means (تعداد مراکز=۸) ۱۱۱
- شکل ۳-۳۹ انتخاب ۱۵ بردار ویژگی مرکزی به عنوان ویژگیهای گوینده با استفاده از مدلسازی VQ ۱۱۲
- شکل ۳-۴۰ انتخاب ۵ بردار ویژگی مرکزی به عنوان ویژگیهای گوینده با استفاده از مدلسازی VQ ... ۱۱۳
- شکل ۳-۴۱ سیستم بازشناسی گوینده مبتنی بر ادغام تصمیم ۱۱۶
- شکل ۴-۱ مقایسه نتایج بازشناسی روشهای مختلف برای $N=5,10,15,20,40,60$ و $N_f=12$ ۱۲۵
- شکل ۴-۲ مقایسه دقت در هر N برای روشهای بازشناسی با $N_f=12$ ۱۲۶
- شکل ۴-۳ مقایسه نتایج بازشناسی روشهای مختلف برای $N=5,10,15,20,40,60$ و $N_f=15$ ۱۲۹
- شکل ۴-۴ مقایسه دقت در هر N برای روشهای بازشناسی با $N_f=15$ ۱۳۰
- شکل ۴-۵ مقایسه نتایج بازشناسی روشهای مختلف برای $N=5,10,15,20,40,60$ و $N_f=20$ ۱۳۳
- شکل ۴-۶ مقایسه دقت در هر N برای روشهای بازشناسی با $N_f=20$ ۱۳۴
- شکل ۴-۷ مقایسه نتایج بازشناسی روشهای مختلف برای $N=5,10,15,20,40,60$ و $N_f=30$ ۱۳۷
- شکل ۴-۸ مقایسه دقت در هر N برای روشهای بازشناسی با $N_f=30$ ۱۳۸
- شکل ۴-۹ مقایسه نتایج بازشناسی روشهای مختلف برای $N=5,10,15,20,40,60$ و $N_f=40$ ۱۴۱
- شکل ۴-۱۰ مقایسه دقت در هر N برای روشهای بازشناسی با $N_f=40$ ۱۴۲

- شکل ۴-۱۱ اثر تغییرات N(تعداد خوشه‌ها) برای روش MFCC برای بانک فیلترهای ثابت ۱۴۴
- شکل ۴-۱۲ اثر تغییرات Nf(تعداد فیلترها) برای روش MFCC برای بانک تعداد خوشه‌های ثابت ... ۱۴۵
- شکل ۴-۱۳ اثر تغییرات N(تعداد خوشه‌ها) برای روش DMFCC برای بانک فیلترهای ثابت ۱۴۶
- شکل ۴-۱۴ اثر تغییرات Nf(تعداد فیلترها) برای روش DMFCC برای بانک تعداد خوشه‌های ثابت. ۱۴۷
- شکل ۴-۱۵ اثر تغییرات N(تعداد خوشه‌ها) برای روش D^2 MFCC برای بانک فیلترهای ثابت ۱۴۸
- شکل ۴-۱۶ اثر تغییرات Nf(تعداد فیلترها) برای روش D^2 MFCC برای بانک تعداد خوشه‌های ثابت ۱۴۹
- شکل ۴-۱۷ اثر تغییرات N(تعداد خوشه‌ها) برای روش MVF برای بانک فیلترهای ثابت ۱۵۰
- شکل ۴-۱۸ اثر تغییرات Nf(تعداد فیلترها) برای روش MVF برای بانک تعداد خوشه‌های ثابت ... ۱۵۱
- شکل ۴-۱۹ اثر تغییرات N(تعداد خوشه‌ها) برای روش RANK برای بانک فیلترهای ثابت ۱۵۲
- شکل ۴-۲۰ اثر تغییرات Nf(تعداد فیلترها) برای روش RANK برای بانک تعداد خوشه‌های ثابت .. ۱۵۳
- شکل ۴-۲۱ اثر تغییرات N(تعداد خوشه‌ها) برای روش SCORE برای بانک فیلترهای ثابت ۱۵۴
- شکل ۴-۲۲ اثر تغییرات Nf(تعداد فیلترها) برای روش SCORE برای بانک تعداد خوشه‌های ثابت ۱۵۵
- شکل ۴-۲۳ مقایسه متوسط دقت روشهای بازشناسی برای تعداد فیلترهای متغیر(Nf) ۱۵۷
- شکل ۴-۲۴ مقایسه متوسط دقت روشهای بازشناسی برای تعداد خوشه‌های متغیر(N) ۱۵۸

۱- فصل اول

کلیات تحقیق

۱-۱- اصول بازشناسی گوینده

بازشناسی گوینده شاخه ای از تشخیص هویت بر پایه ویژگیهای زیستی^۱ است که به بازشناسی خودکار شناسه افراد با استفاده از مشخصات ذاتی آنها اطلاق می گردد. تشخیص هویت مبتنی بر معیارهای زیستی برای سیستمهای ارتباط انسان-ماشین در کاربردهایی که در آنها ملاحظات امنیتی وجود دارد جزو مهمترین روشها می باشد. البته غیر از صوت، الگوهای فیزیکی و رفتاری دیگری از قبیل الگوی شبکه چشم، چهره، اثرانگشت، امضاء و غیره برای تشخیص هویت مبتنی بر سنجش زیستی وجود دارد. در عمل انتخاب یک الگوی زیستی برای بازشناسی الگو حداقل بایستی خواصی همچون مقاوم بودن (در مقابل نویز و سایر شرایط مزاحم)، قابل تشخیص بین افراد مختلف (منحصر به فرد بودن)، در دسترس بودن و قابل قبول بودن را داشته باشد [1].

جدول ۱-۱ چهار خاصیت فوق را برای چند الگوی زیستی که مورد استفاده فراوان هستند را مقایسه می کند:

جدول ۱-۱ وضعیت خواص چهارگانه برای الگوهای زیستی

الگوی زیستی	شبکیه	چهره	اثرانگشت	صوت
قابلیت تمیز	بالا	بالا	بالا	قابل قبول
مقاوم بودن	بالا	بالا	قابل قبول	قابل قبول
در دسترس بودن	کم	بالا	قابل قبول	بالا
مقبولیت	قابل قبول	بالا	قابل قبول	بالا

مشخصات فوق را می توان به صورت زیر تعریف کرد:
قابلیت تمیز: یعنی تفاوت مناسب و عمده در الگو برای افراد مختلف باشد.

مقاوم بودن: قابل تکرار باشد و در تکرارها تغییرات زیادی در آن رخ ندهد.
در دسترس بودن: به سادگی قابل ارائه به یک حسگر باشد (به سادگی اندازه گیری شود)
قابل قبول بودن: به عنوان یک مشخصه مجاز توسط کاربران قابل مشاهده باشد.
قضاوت در مورد یک الگوی زیستی خوب بسیار پیچیده و وابسته به ویژگیها و شرایط کاربردها می باشد.

در میان تمام فناوریها و روشهای تشخیص هویت زیستی، بازشناسی گوینده را می توان طبیعی ترین و اقتصادی ترین روش برای سیستمهای ارتباط انسان-ماشین دانست زیرا اولاً وصول اطلاعات صحبت از دیگر الگوها راحتتر است و ثانياً صحبت رویکرد غالب و اصلی تبادل اطلاعات برای انسانهاست و می رود که رویکرد غالب برای تبادل اطلاعات انسان-ماشین باشد. همچنین توسعه فناوری پردازش صحبت باعث تقویت بسیاری از کاربردهای بازشناسی گوینده و خصوصاً حوزه های زیر شده است:

۱- کنترل دسترسی به تجهیزات فیزیکی یا شبکه های داده

۲- خرید با کارت اعتباری از طریق تلفن یا سایر تراکنشهای بانکی

۳- بازبازی اطلاعات به عنوان مثال اطلاعات مشتری برای مرکز تلفن و فهرست گذاری صوتی

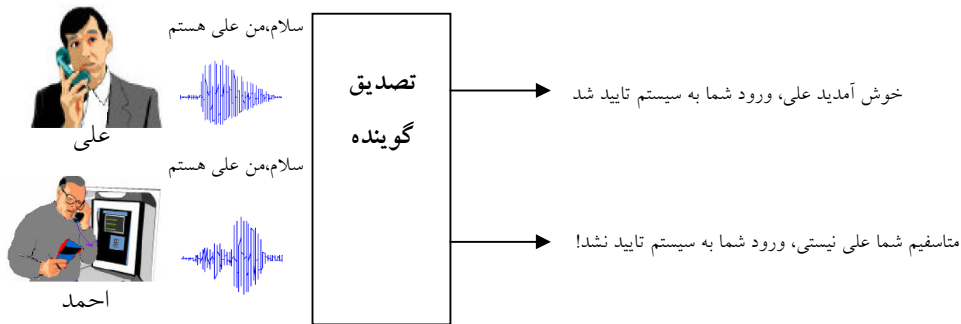
۴- پایش از دور

۵- تطبیق نمونه صوتی در محاکم قضایی

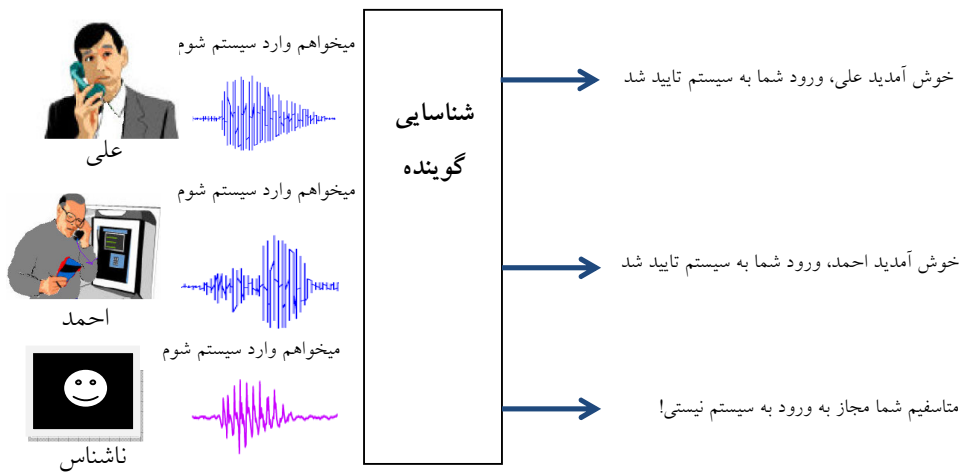
بازشناسی گوینده را می توان به دو وظیفه تصدیق^۱ و شناسایی^۲ تقسیم کرد. تصدیق عبارت است از تصمیم در مورد اینکه یک صدای نامعین مربوط به یک گوینده ادعایی هست یا نیست. فقط دو تصمیم ممکن است: یا پذیرش اینکه صدا متعلق به گوینده ادعایی است یا رد آن به عنوان یک گوینده متقلب. شناسایی عبارت است از طبقه بندی یک صوت نامعین به عنوان یکی از گویندگان ثبت شده. تعداد تصمیمات ممکن در شناسایی گوینده به تعداد گویندگان ثبت نام شده است و معمولاً کارایی سیستم با عکس تعداد گویندگان متناسب است. بنابراین وقتی تعداد گویندگان زیاد است شناسایی بسیار پیچیده تر و مشکلتر از تصدیق گوینده است. این نوع شناسایی که تشریح شد به عنوان شناسایی مجموعه-بسته^۳ نیز شناخته میشود. شناسایی مجموعه-باز^۴ متناظر با زمانی است که بتوان تصمیم گرفت که صوت نامعین ورودی به هیچکدام از گوینده های ثبت شده متعلق نباشد.

-
- 1- Verification
 - 2- Identification
 - 3- Closed-Set
 - 4- Open-Set

بنابراین تعداد تصمیمات ممکن از تعداد گویندگان ثبت شده یکی بیشتر است. شکل‌های ۱-۱ و ۲-۱ به طور کلی نحوه کار سیستم‌های تصدیق و شناسایی گوینده را نشان می‌دهند.



شکل ۱-۱ سیستم تصدیق گوینده



شکل ۲-۱ سیستم شناسایی گوینده

شناسایی مجموعه-باز ترکیب شناسایی مجموعه بسته و تصدیق گوینده است. همچنین بازشناسی گوینده را میتوان به بازشناسی وابسته به متن و بازشناسی مستقل از متن تقسیم نمود. در بازشناسی وابسته به متن، سیستم دقیقاً متن صحبت را می‌داند. در بازشناسی مستقل از متن سیستم از متن صحبت آگاهی ندارد. با داشتن اطلاعات متن صحبت، سیستم می‌تواند گوینده خاص را با اجزاء کلمه یا حروف معین شناسایی کند و بنابراین سیستم‌های وابسته به متن معمولاً کارایی بهتری نسبت به سیستم‌های مستقل از متن دارند اما در این سیستم‌ها نیاز به همکاری و تعامل زیاد گوینده می‌باشد

و بنابراین برای کاربردهایی که کنترل قوی روی ورودی کاربر وجود دارد بکار می رود. کاربرد سیستم مستقل از متن بیشتر از سیستم وابسته به متن بوده و کاربرپسندتر است اما بدون اطلاعاتی از متن صحبت رسیدن به کارایی بالا دشوار است. در کاربردهای مستقل از متن، استفاده از یک تشخیصگر متن که بتواند از متن صحبت اطلاعاتی را فراهم کند می تواند دقت بازشناسی گوینده را بهبود دهد [2]. هر چند سیستمهای مستقل از متن به عنوان بستر بهتری برای ارائه و ارزیابی فناوریها و سیستمهای بازشناسی گوینده پذیرفته شده اند، اما بسیاری از کاربردهای تجاری و صنعتی توجه بیشتری به بازشناسی وابسته به متن یا بازشناسی با متن محدود^۱ نموده اند.

۲-۱- مروری مختصر بر تاریخچه بازشناسی گوینده

پژوهش در زمینه بازشناسی گوینده حدود ۵۰ سال پیشینه دارد و همواره به عنوان حوزه ای فعال در پردازش زبان مطرح می باشد. توسعه روشها و فناوری بازشناسی گوینده همواره رابطه نزدیکی با پیشرفت در پردازش سیگنال و صحبت و فناوری رایانه داشته و دارد. بازشناسی گوینده توسط انسان به طور وسیع از دهه ۱۹۶۰ میلادی مطالعه و بررسی شد. انگیزه و هدف این مطالعات، فهم چگونگی بازشناسی گوینده ها توسط انسان و قابلیت اطمینان انسان در بازشناسی یک گوینده بود [3]. یکی از مهمترین کارهای پژوهشی که باعث گسترش تحقیقات در بازشناسی گوینده با استفاده از رایانه گردید مطرح کردن نمودار طیفی^۲ توسط کرسستا^۳ به عنوان یک ابزار شناسایی شخص بود که در سال ۱۹۶۲ در نشریه نیچر منتشر گردید.

در دهه ۱۹۷۰ توجهات به بازشناسی گوینده با استفاده از رایانه و آغاز بازشناسی خودکار گوینده^۴ (ASR) بود. سیستمهای بازشناسی گوینده در این دوره از پژوهشها محدود به تعداد محدودی گوینده (کمتر از ۲۰ نفر) بود [4]. روشها و فنون تبدیل فوریه، پیشگویی خطی^۵، آنالیز کپسترال^۶ در این دوره برای استخراج پارامترها و ویژگیها بکار برده شدند. متوسط زمان طولانی این پارامترها به عنوان مرجع گوینده استفاده می شد.

در دهه ۱۹۸۰ روشهای آماری پیچیده تر بازشناسی الگو از قبیل همپچی زمانی پویا^۷ [5] و کمی سازی برداری^۸ [6] برای سیستمهای در مقیاس بزرگ بازشناسی گوینده (بیشتر از ۱۰۰ نفر) ابداع و

1 - Text-Constraint

2 - Spectrogram

3 - L. G. Kersta

4 - Automatic Speaker Recognition(ASR)

5 - Linear Predictive

6 - Cepstral Analysis

7 - Dynamic Time Warping (DTW)

8 - Vector Quantization (VQ)

استفاده شد. همچنین استفاده توأم از ویژگیهای دینامیک و استاتیک برای بازشناسی گوینده نیز در همین دهه توسط محققان انجام پذیرفت [7].

از سالهای حدود ۱۹۹۰ در دسترس بودن پایگاه های بزرگ داده صحبت مانند YOHO مطالعات روی مدل های پیچیده تر برای بازشناسی گوینده را تقویت کرد. این مدلها شامل مدل های تصادفی^۱ (از قبیل مدل مخفی مارکف^۲ [8] و مدل ترکیبی گوس^۳)، شبکه های عصبی (مانند پرسپترون چندلایه ای^۴ [9] و توابع پایه شعاعی^۵ [10]) و ماشین های بردار پشتیبان^۶ [11] و دیگر مدلها می باشد. در بین این روش های مدلسازی، روش GMM به عنوان مؤثرترین روش در تعیین توزیع چگالی داده های صحبت شناخته می شود و روش غالب در مدلسازی برای بازشناسی گوینده است. برای استخراج ویژگیها، ضرایب کپسترال که برگرفته از مدل شنوایی انسان است یعنی ضرایب کپسترال مل-فرکانس^۷ و ضرایب دینامیک آنها به عنوان روش غالب در استخراج پارامترهای ویژگی شناخته شده می باشند. از طرفی محققان روش های مختلف نرمال سازی امتیاز^۸ برای بازشناسی مقاوم گوینده را ارائه نموده اند [12].

۳-۱- مشکلات و چالش های روشها و فناوریهای موجود بازشناسی صحبت

سیستمها و روش های موجود بازشناسی صحبت در آزمایشگاه یا بعضی کاربردهای خاص با شرایط کاری و آموزش سطح بالا و طراحی ماهرانه، خیلی خوب کار می کنند. نتایج تجربی نشان می دهد که بازشناسی خودکار گوینده در چنین محیط های ایده آلی به خوبی بازشناسی توسط انسان است. اما از منظر عملی و کاربردی، کارایی سیستم های بازشناسی گوینده فعلی در کاربردهای واقعی هنوز از درجه مقاوم بودن و قابلیت اطمینان مناسبی در مقایسه با کارایی بازشناسی توسط انسان برخوردار نیستند. اولین مسئله پیش روی سیستم های بازشناسی گوینده، بهبود سطح مقاوم بودن سیستم تحت شرایط واقعی و غیرمنطبق است. برای بازشناسی گوینده، عدم تطبیقها اصولا به دو دلیل رخ می دهد:

۱- تغییرات فردی گوینده مربوط به شیوه صحبت کردن.

-
- 1 - Stochastic Models
 - 2 - Hidden Markov Model(HMM)
 - 3 - Gaussian Mixture Model (GMM)
 - 4 - Multilayer Perceptron (MLP)
 - 5 - Radial Basis Function (RBF)
 - 6 - Support Vector Machines (SVM)
 - 7 - Mel-Frequency Cepstral Coefficients (MFCC)
 - 8 - Score Normalization

۲- تغییرات آکوستیکی محیط

سیستم صوتی انسان در گفتن یک عبارت ثابت در زمانهای مختلف نیز دارای تغییرات است و این باعث می شود که تطبیق گفته های یک فرد با ویژگیهای خودش به خوبی انجام نشود و یکی از منابع اصلی خطا در سیستمهای بازشناسی گوینده همین تغییرات ذاتی مربوط به گوینده است. تغییرات آکوستیکی محیط نیز بدلیل اعوجاجات و اختلالات غیرقابل پیش بینی مختلف در حین وصول و انتقال داده ها می باشد. به عنوان مثال در کاربردهای بازشناسی گوینده تلفنی مانند سیستمهای تراکنش تلفنی بانک، داده های صوتی ممکن است در محیطهای مختلف، با تلفنهای متفاوت و از طریق کانالهای ارتباطی مختلف دریافت شود. نويز زمينه و اعوجاج کانال و گوشی باعث تغییر در ساختار طیفی داده های صوتی می شود و ویژگیهای مستخرجه آکوستیکی (مثلا پارامترهای MFCC) نمی تواند اطلاعات گوینده را به درستی در بر داشته باشد.

برای بهبود کارایی سیستمهای بازشناسی گوینده در محیطهای واقعی نیز محققان فعالیتهای متنوعی را انجام داده اند. از جمله این کارها می توان به تحقیقات در زمینه استخراج ویژگیهای سطح بالا و جدید از قبیل ویژگیهای عروزی، لغوی و شیوه صحبت کردن و امثال آن به عنوان مکمل ویژگیهای آکوستیکی سطح پایین (مانند ضرایب کپسترال) برای بازشناسی مقاوم گوینده اشاره نمود که یکی از زمینه های تحقیقاتی بسیار فعال در این حوزه می باشد [13].

به هر حال استخراج ویژگی مؤثر و مدل سازی مناسب داده ها هنوز به خوبی کاربردی نشده و فقط بخشی از مشکلات بازشناسی را حل می کند و همیشه نمی توان از آنها استفاده نمود. یکی از زمینه هایی که در سالهای اخیر در بسیاری از کاربردها باعث بهبود عمده نتایج شده است استفاده از روشها و مدلهای ادغام اطلاعات^۱ است که در سطوح مختلف قابل کاربرد و به عنوان یکی از مهمترین زمینه های تحقیقاتی فعال مطرح است [14].

البته یکی دیگر از رویکردهای بازشناسی گوینده استفاده از الگوهای چندگانه به صورت همزمان و بهره گیری از ادغام اطلاعات چندحسگری^۲ است مثلا ادغام الگوهای چهره، اثرانگشت و صوت از جمله این رویکردهاست. پیچیدگی دریافت اطلاعات کاربر و هزینه ها در سیستم چند حسگری از جمله موانع توسعه این رویکردها می باشد.

1 - Data Fusion

2 - Multisensor Data Fusion

۳-۱- مسئله تحقیق و هدف آن

بهبود دقت بازشناسی گوینده با استفاده از ادغام اطلاعات در سطح تصمیم مورد توجه این تحقیق است. همچنین استخراج ویژگیهای جدیدی بر پایه ضرایب کپسترال جهت فراهم سازی منابع تصمیم گیری مورد استفاده در ادغام تصمیم از دیگر رویکردهای این پایان نامه می باشد. با توجه به اینکه مشتقات هر تابع بخشی از ویژگیهای مستتر در آن را به نمایش می گذارد که انتظار داریم مشتق اول و دوم ضرایب کپسترال مل-فرکانس و مشتق اول و دوم خود کپسترال مل-فرکانس به عنوان بردارهای ویژگی جدید در ترکیب با هم و با استفاده از روشهای ادغام تصمیم باعث بهبود مناسبی در دقت بازشناسی گوینده نسبت به حالت بدون ادغام شود. این رویکرد به مفهوم استفاده و بهره گیری همزمان از اطلاعات نهفته در بردار ویژگی، تغییرات (بردار سرعت) و نرخ تغییرات (بردار شتاب) ویژگی می باشد. رویکرد استفاده همزمان از این بردارهای ویژگی در بستر ادغام تصمیم تاکنون توسط محققان گزارش نشده است.

هدف این تحقیق بهبود کارایی بازشناسی گوینده با استفاده از ادغام همزمان اطلاعات حاصل از بازشناسی گوینده توسط هرکدام از بردارهای ویژگی فوق در سطح تصمیم است. در این تحقیق سعی می کنیم که به سؤالات زیر پاسخ دهیم:

آیا استفاده از اطلاعات مستتر در بردارهای تغییرات کپسترال و ضرایب کپسترال و بردارهای نرخ این تغییرات برای بازشناسی گوینده مفید است؟
چگونه به طور مؤثر می توان اطلاعات گوینده را از تغییرات و نرخ تغییرات بردارهای ویژگی کپسترال و ضرایب کپسترال استخراج کرد؟
چگونه می توان از مزایای ادغام اطلاعات در سطح تصمیم بر مبنای ادغام ویژگیها و تغییرات و نرخ تغییرات آنها برای بازشناسی گوینده بهره برد؟

۴-۱- مفاهیم تحقیق

در این پایان نامه بطور مقدماتی در خصوص مفهوم و کاربردهای بازشناسی گوینده و تاریخچه مختصر آن پرداختیم. یکی از حوزههای تحقیقاتی مورد توجه محققان، ارائه الگوهای جدید به منظور افزایش دقت عملکرد سیستمهای بازشناسی می باشد. از جمله روشها و فناوریهای مؤثر در بهبود