دانشکده : برق و رباتیک

گروه : الکترونیک

یک روش استخراج ویژگی برای شناسایی گوینده بر مبنای توزیع ویگنر

دانشجو : جلیل قاسمی

استاد راهنما

دکتر حسین مروی

استاد مشاور

دکتر امید رضا معروضی

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

بهمن ۱۳۹۰

شماره :

بسمه تعالی

تاریخ :

مدیریت تحصیلات تکمیلی

فرم صورتجلسه دفاع پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای جلیل قاسمی رشته برق گرایش الکترونیک تحت عنوان

یک روش استخراج ویژگی برای شناسایی گوینده بر مبنای توزیع ویگنر

که در تاریخ ۹۰/۱۱/۱۲ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح زیر است :

قبول (با درجه : ..... امتیاز .....) ☐ دفاع مجدد ☐ مردود ☐

۱ـ عالی (۲۰ ـ ۱۹) ۲ـ بسیار خوب (۱۸/۹۹ ـ ۱۸)

۳ـ خوب (۱۷/۹۹ ـ ۱۶) ۴ـ قابل قبول (۱۵/۹۹ ـ ۱۴)

۵- نمره کمتر از ۱۴ غیر قابل قبول

a عضو هیأت داوران	نام ونام خانوادگی	مرتبهٔ علمی	امضاء
۱ـ استادراهنما			
۲ـ استاد مشاور			
۳ـ نماینده شورای تحصیلات تکمیلی			
۴ـ استاد ممتحن			
۵ ـ استاد ممتحن			

تأیید رئیس دانشکده :

تقدیم به

پدر و مادر بزرگوارم

و

همسرمهربانم

سپاس بی کران خداوند متعال که مرا فرصت اندیشیدن داد.

بر خود لازم می‌دانم که از استاد گرانقدرم جناب آقای دکتر حسین مروی

و تمامی اساتید گرامی که مرا در انجام این تحقیق یاری نمودند

کمال تشکر و قدردانی را داشته باشم

تعهد نامه

اینجانب جلیل قاسمی دانشجوی دوره کارشناسی ارشد رشته برق- الکترونیک دانشکده برق و رباتیک دانشگاه صنعتی شاهرود نویسنده پایان نامه یک روش استخراج ویژگی برای شناسایی گوینده بر مبنای توزیع ویگنرتحت راهنمائی دکتر حسین مروی متعهد می شوم.

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری ، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده:

با پیشرفت روز افزون علم و تکنولوژی نیاز بشر به سیستم‌های هوشمند در تمامی زمینه‌ها مشهود می‌باشد. با توجه به اینکه صدای هر فرد مانند شناسه‌های دیگری چون اثر انگشت، خصوصیات چهره، عنبیه چشم و ... منحصر به فرد می‌باشد به همین دلیل از آن در سیستم‌های هوشمند بازشناسی گوینده استفاده می‌شود. اما یکی از مهمترین مسائل مطرح شده در زمینه سیستم‌های بازشناسی گوینده تأثیر نویز بر سیگنال گوینده و کاهش دقت بازشناسی است. لذا بازشناسی مقاوم گوینده در مقابل نویز یکی از موارد مورد مطالعه محققان می‌باشد که در سالهای اخیر تحقیقات زیادی در این زمینه انجام شده و روشهای مختلفی ارایه شده است.

هدف از انجام این تحقیق استفاده از توزیع زمان- فرکانس ویگنر- ویل برای استخراج ویژگی از سیگنال گفتار می‌باشد. در این تحقیق از ترکیب تابع توزیع ویگنر- ویل و تبدیل هیلبرت و ضرایب MFCC برای استخراج ویژگی استفاده شده است. در این روش از سیگنال گفتار تبدیل هیلبرت گرفته شده و پس از آن، تبدیل ویگنر- ویل گرفته می‌شود. سیگنال خروجی از تبدیل ویگنر- ویل از فیلتر بانک مل عبور داده می‌شود و پس از لگاریتم گیری، از آن تبدیل فوریه کسینوسی گرفته می‌شود. خروجی بدست آمده از این سیستم پیشنهادی بعنوان یک ویژگی استخراج شده از سیگنال گفتار جهت تشخیص گوینده بکار می‌رود. به کمک این ویژگی‌ها برای هر گوینده یک مدل مخلوط گوسی(GMM) بدست آورده شده است. نتایج حاصله در این روش با ضرایب MFCC و PLP مقایسه شده است که جهت پیاده سازی روش‌های پیشنهادی از دیتا بیس TIMIT استفاده شده است. نتایج گویای این مطلب است که روش پیشنهادی در سیگنال به نویزهای پایین جواب بهتری را از ضرایب MFCC و PLP ارایه می‌دهد.

کلید واژه : بازشناسی گوینده و گفتار، توزیع ویگنر- ویل، استخراج ویژگی

فهرست مطالب **صفحه**

فصل اول: سیستم‌های بازشناسی گوینده و گفتار

۱-۱- مقدمه ۲

۱-۲- تاریخچه ۳

۱-۳- سیستم های تشخیص گفتار یا بازشناسی گفتار ۳

۱-۴- سیستم های بازشناسی گوینده ٤

۱-۴-۱- سیستم های تأیید هویت گوینده ٤

۱-۴-۲- سیستم های بازشناسی هویت گوینده ٤

۱-۴-۳- سیستم های وابسته به متن ٦

۱-۴-۴- سیستم های مستقل از متن ٦

۱-۵- ساختار پایان نامه ۷

فصل دوم : روشهای متداول استخراج ویژگی

۲-۱- مقدمه ۱۰

۲-۲- ویژگیهای مورد استفاده در تحلیل گفتار ۱۱

۲-۳- مراحل استخراج ویژگی ۱۲

۲-۳-۱- پیش تأکید ۱۲

۲-۳-۲- پنجره بندی ۱۲

۲-۳-۳- فیلتر کردن ۱۳

۲-۴- روشهای آنالیز زمانی ۱٤

۲-۴-۱- انرژی ۱٤

۲-۴-۲- تعداد عبور از صفر ۱٤
۲-۴-۳- سیگنال به نویز ۱٥
۲-۵- روشهای آنالیز طیفی ۱٦
۲-۵-۱- روش کپستروم ۱۷
۲-۵-۱-۱- کپستروم حقیقی ۱۷
۲-۵-۱-۲- کپستروم مختلط ۱۸
۲-۵-۱-۳- کپستروم ریشه ۱۸
۲-۵-۲- ضرایب پیشگویی خطی یا LPC ۱۸
۲-۵-۳- ضرایب LPC-CEPSTRUM ۲۱
۲-۵-۴- ضرایب کپسترال مبتنی بر معیار مل (MFCC) ۲۱
۲-۵-۴-۱- استخراج ویژگی با استفاده از طیف های دنباله خود همبستگی ۲٦
۲-۵-۵- پیشگویی خطی ادراکی ۲۷
۲-۵-۶- تحلیل PLP-RASTA ۳۱
۲-۵-۷- ترکیب MFCC و PCA برای استخراج ویژگی ۳۳
۲-۵-۸- استخراج ویژگی با استفاده از MFCC وزن دهی شده ۳۳
۲-۵-۹- استخراج ویژگی با استفاده از MVDR و DCT ۳۴

فصل سوم : توزیع ویگنر- ویل

۳-۱- مقدمه ۳٦
۳-۲- سیگنالهای ایستان و غیر ایستان ۳۷
۳-۳- مشکل تبدیل فوریه در آنالیز سیگنالهای غیر ایستان ۳۸
۳-۴- نمایش زمان – فرکانس ۳۹

۳-۴-۱- فرم دو جملهای ٤٣
۳-۴-۲- فرم خطی ٤٣
۳-۵- توزیع ویگنر - ویل ٤٤
۳-۵-۱- ویژگی های توزیع ویگنر ویل ٤٥
۳-۵-۲- ترم تداخلی ٤٨
۳-۵-۳- نمونه برداری از تابع توزیع ویگنر- ویل : سیگنال تحلیلی ٥١
۳-۵-۴- توزیع ویگنر ویل هموار شده در فرکانس ٥٣
۳-۵-۵- توزیع ویگنر- ویل هموار شده در زمان و فرکانس ٥٤
۳-۶- استفاده از WVD در تحلیل سیگنال گفتار..... ۵۷

فصل چهارم : مدل مخلوط گوسی

۴-۱- مقدمه..... ٦٠
۴-۲- انگیزه استفاده از مدل مخلوط گوسی ٦١
۴-۳ مدل مخلوط گوسی و فرمول بندی آن ٦١
۴-۴- الگوریتم EM ٦٣
۴-۵- کاربرد GMM در تشخیص گوینده..... ٦٦

فصل پنجم : روش پیشنهادی و پیاده سازی الگوریتم

۵-۱- مقدمه..... ۶۸
۵-۲- روش پیشنهادی..... ۶۹
۵-۲-۱- پیش پردازش..... ۷۰
۵-۲-۲- پنجره گذاری ۷۰
۵-۲-۳- تبدیل هیلبرت ۷۰

۵-۲-۴- تبدیل ویگنر- ویل ۷۲
۵-۲-۵- فیلتر بانک مل ۷۶
۵-۲-۶- لگاریتم گیری ۷۶
۵-۲-۷- تبدیل فوریه کسینوسی ۷۶
۵-۳- دیتا بیس ۷۷
۵-۴- مشخصات سیستم ۸۰
۵-۵- سنجش کارایی ۸۱
۵-۵- نتایج ۸۳
۵-۵-۱- نتایج حاصل از ضرایب MFCC ۸۳
۵-۵-۲- نتایج حاصل از ضرایب PLP ۸۴
۵-۵-۳- نتایج حاصل از ضرایب MFCC – WVD - S ۸۵
۵-۵-۴- نتایج حاصل از ضرایب MFCC –PWVD – S ۸۶
۵-۶- نتیجه گیری ۸۷
۵-۷- کارهای آینده ۸۸
مراجع ۸۹

فهرست اشکال **صفحه**

شکل (۱-۱) : ساختار پایه از سیستم های بازشناسی گوینده..... ۵

شکل (۲-۱): بازه‌ای ایستان از یک سیگنال گفتار..... ۱۱

شکل (۲-۲): حوزه فرکانس و زمان پنجره همینگ..... ۱۳

شکل (۲-۳): متوسط عبور از صفر یک جمله..... ۱۵

شکل (۲-۴): مراحل کپستروم حقیقی..... ۱۸

شکل (۲-۵): پوش طیف سیگنال گفتار با استفاده از LPC..... ۲۰

شکل (۲-۶): بلوک دیاگرام روش MFCC..... ۲۱

شکل (۲-۷): فیلتر بانک مل..... ۲۳

شکل (۲-۸): بلوک دیاگرام روش استخراج گفتار در حوزه خود همبستگی..... ۲۶

شکل (۲-۹) بلوک دیاگرام PLP..... ۲۷

شکل (۲-۱۰) مقایسه MFCC و PLP..... ۳۰

شکل (۲-۱۱): پاسخ فرکانسی فیلتر رستا..... ۳۱

شکل (۲-۱۲): مراحل استخراج ضرایب PLP- RASTA..... ۳۲

شکل (۲-۱۳): مراحل استخراج ضرایب با استفاده از MFCC و PCA ۳۳

شکل (۲-۱۴): مراحل استخراج ضرایب با استفاده از DCT و MVDR ۳۴

شکل (۳-۱): سیگنال ایستان..... ۳۷

شکل (۳-۲): سیگنال غیر ایستان..... ۳۷

شکل (۳-۳): تبدیل فوریه شکل(۳-۱)..... ۳۸

شکل (۳-۴): تبدیل فوریه شکل(۳-۲)..... ۳۹

شکل (۳-۵): سیگنال chirp.....۴۴

شکل (۳-۶): توزیع ویگنر- ویل سیگنال chirp.....۴۵

شکل (۳-۷): نمایش ترم تداخلی در سیگنال.....۴۹

شکل (۳-۸): نمایش ترم تداخلی در سیگنال.....۴۹

شکل (۳-۹): ساختار ترم تداخلی بین دو جزء گوسی با موقعیت های متفاوت.....۵۰

شکل (۳-۱۰): سیگنالی با دو جزء گوسی.....۵۲

شکل (۳-۱۱): توزیع ویگنر- ویل سیگنال غیر تحلیلی متشکل از دو گوسی.....۵۲

شکل (۳-۱۲): توزیع ویگنر- ویل سیگنال تحلیلی متشکل از دو گوسی.....۵۲

شکل (۳-۱۳): توزیع WVD سیگنالی متشکل از یک گوسی و یک سینوسی.....۵۵

شکل (۳-۱۴): توزیع PWVD سیگنالی متشکل از یک گوسی و یک سینوسی.....۵۵

شکل (۳-۱۵): توزیع SPWVD سیگنالی متشکل از یک گوسی و یک سینوسی.....۵۶

شکل (۳-۱۶): سیگنال زمانی واج /i/ و نمودار SPWVD ۵۷

شکل (۳-۱۷): سیگنال زمانی واج /ɑ/ و نمودار SPWVD ۵۸

شکل (۴-۱): نمایشی از مدل مخلوط گوسی با دو بعد و دو گوسی.....۶۰

شکل (۴-۲): ساختار مدل مخلوط گوسی.....۶۱

شکل (۵-۱): بلوک دیاگرام روش پیشنهادی.....۶۹

شکل (۵-۲): مشخصه فرکانسی تابع توزیع ویگنر- ویل با استفاده از تبدیل هیلبرت.....۷۱

شکل (۵-۳): مشخصه فرکانسی تابع توزیع ویگنر- ویل بدون استفاده از تبدیل هیلبرت.....۷۱

شکل (۵-۴): نمایشی سه بعدی از تبدیل ویگنر- ویل یک پنجره از سیگنال گفتار.....۷۲

شکل (۵-۵): نمایشی سه بعدی از تبدیل ویگنر- ویل هموار شده یک پنجره از سیگنال گفتار.....۷۳

شکل (۵-۶): مشخصه فرکانسی حاشیه‌ای روی یک فریم WVD.....۷۴

شکل (۵-۷): مشخصه فرکانسی حاشیه‌ای روی یک فریم PWVD.....۷۵

فهرست جداول صفحه

جدول (۳-۱): مقایسه چند تابع توزیع.....۴۱

جدول (۵-۱): توزیع گوینده ها برای هر ایالت.....۷۷

جدول (۵-۲): هشت ایالت گوینده ها.....۷۸

جدول (۵-۳): تقسیم بندی جملات بر حسب SX و SA و SI.....۷۸

جدول (۵-۴): تعداد زن و مرد در هر لهجه.....۷۹

جدول (۵-۵): نتایج کلاسه بندی حاصل از چهار روش متفاوت در روش ارزشیابی اول.....۸۷

جدول(۵-۶): نتایج کلاسه بندی حاصل از چهار روش متفاوت در روش ارزشیابی دوم.....۸۸

فهرست نمودار صفحه

نمودار (۵-۱): نتایج حاصل از ضرایب MFCC.....۸۳

نمودار (۵-۲): نتایج حاصل از ضرایب PLP.....۸۴

نمودار (۵-۳): نتایج حاصل از ضرایب MFCC – WVD – S.....۸۵

نمودار (۵-۵): نتایج حاصل از ضرایب MFCC –PWVD –S.....۸۶

فصـــــــل اول

سیـــستم های

بازشناسی گوینـــده و گفتـــار

۱-۱- مقدمه

دیر زمانی است که استفاده از شناسه‌های کاربری مختلف برای تعیین هویت افراد مورد استفاده قرار می‌گیرد. از مرسوم‌ترین این شناسه‌ها می‌توان به شماره شناسنامه، کد ملی، نام و نام خانوادگی و ... اشاره کرد. یکی از اشکالات عمده این شناسه‌ها ضعفی است که در مقابل جعل آنها قرار دارد که این موضوع خود امنیت این شناسه‌ها را زیر سؤال برده است و دانشمندان را به یافتن شناسه‌های قابل اطمینانی مانند شناسه‌های بیومتریک[1] سوق داده است. شناسه‌های بیومتریک مانند اثر انگشت، خصوصیات چهره، عنبیه چشم، ویژگیهای صدا و ... کمتر قابل جعل هستند و قابلیت اطمینان بیشتری دارند.

در اغلب کاربردها، مانند سیستم‌های امنیتی، سیستم‌های الکترونیکی، جرم شناسی و ... شناسایی فرد گوینده حائز اهمیت می‌باشد. حال با توجه به اینکه از لحاظ علمی ویژگی‌های گفتار، منحصر بفرد می‌باشد، در سیستم‌های تشخیص گوینده[2] از ویژگی‌های صوتی صدای افراد جهت بازشناسی آنها استفاده می‌شود. الگوهای صوتی افراد به دو دسته تقسیم بندی می‌شود. دسته اول شامل ساختار اندامهای صوتی، اندازه، شکل گلو، دهان و ویژگی‌های تارهای صوتی است و دیگری الگوهای رفتاری آموخته شامل تحصیلات، موقعیت اجتماعی و سبک سخن گفتن است که غالباً در سیستم‌های بازشناسی گوینده امروزی از ویژگی‌های دسته اول استفاده می‌شود[۱۰].

قابلیت بالای این سیستم‌ها و نیز دسترسی ساده از طریق میکروفن، احساس نیاز به طراحی اینگونه سیستم‌ها را جهت جانشینی امضاء، اثر انگشت، کارت شناسایی، رمز عبور و مانند اینها پدید آورده است.

[1] Biometrik
[2] Speaker Recognition

۱-۲- تاریخچه

اولین تلاشها برای ایجاد و ساخت سیستم‌های پردازش گفتار در اوایل دهه ۱۹۵۰ شروع شد. تقریبا ۱۰ سال پس از آن در اوایل دهه ۱۹۶۰ اولین سیستم بازشناسی گوینده و سیستم تأیید هویت گوینده[1] اختراع شد. آقای Pruzanky اولین کسی بود که در آزمایشگاه بل با استفاده از فیلتر بانک و ارتباط بین دو اسپکتوگرام برای اندازه گیری شباهت بین آنها تحقیقات خود را شروع کرد. آقای Li و همکارانش از جداسازهای خطی برای جدا سازی این طیف‌ها استفاده کردند و همچنین در تگزاس آقای Doddington از آنالیز طیف به جای روش فیلتر بانک استفاده کرد. تحقیقات در مورد مقاوم سازی[2] سیستم‌های بازشناسی گوینده و گفتار در مقابل انواع نویزها و شرایط محیطی در دهه ۱۹۹۰ به اوج خود رسید و هم اکنون محققان زیادی بر روی مقاوم‌سازی سیستم‌های بازشناسی گوینده و گفتار فعالیت دارند[۴].

۱-۳- سیستم های تشخیص گفتار یا بازشناسی گفتار[3]

به سیستم‌هایی که اطلاعات گفتاری را دریافت کرده و متن بیانی گوینده را استخراج می‌کنند یا به بیان دیگر گفتار شخص گوینده را تشخیص می‌دهند، سیستم‌های تشخیص گفتار یا بازشناسی گفتار می‌گویند[۴]. کاربردهای این نوع سیستم‌ها بسیار متنوع بوده، که ارتباط با بانک اطلاعاتی از راه دور، ثبت نام دانشجویان از طریق تلفن، اعلام شماره شناسایی شخصی، شماره حساب بانکی، شماره عضویت برای کاربران یک سیستم خدمات رسانی و در کمک به افراد ناشنوا و ... نمونه‌هایی از کاربردهای این نوع سیستم‌ها به حساب می‌آیند.

[1] Speaker Identification
[2] Robustness
[3] Speech Recognition

۱-۴- سیستم های بازشناسی گوینده

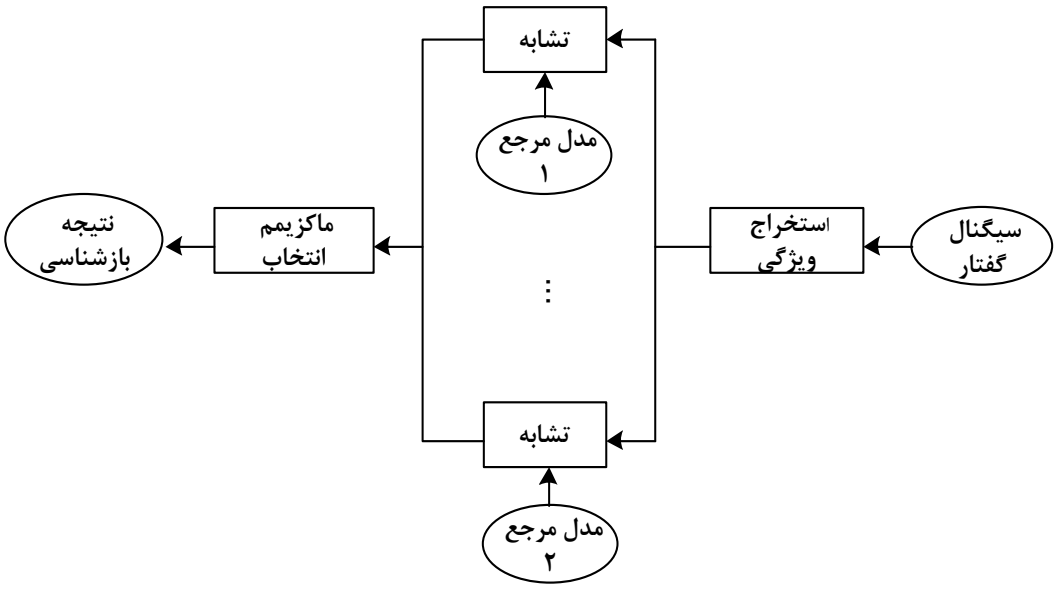
یک سیستم تشخیص گوینده سیستمی است که شخص را از صدای آن تشخیص می‌دهد. سیستم‌های بازشناسی گوینده از لحاظ روش استفاده به دو دسته سیستم‌های تایید هویت گوینده و سیستم های بازشناسی هویت گوینده تقسیم‌بندی می‌شوند. در یک تقسیم‌بندی دیگر سیستم‌های تشخیص گوینده به سیستم‌های وابسته به متن و مستقل از متن تقسم‌بندی می‌شوند که در ادامه توضیح داده می‌شوند[۴].

۱-۴-۱- سیستم های تأیید هویت گوینده

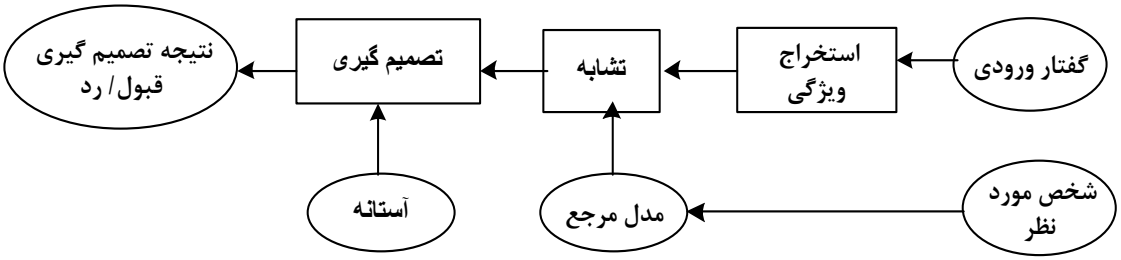
در یک سیستم تایید هویت گوینده شخص عموماً با انتخاب یا وارد کردن نام یکی از کاربران خاص سیستم ادعا می‌کند که او همان کاربر ثبت شده‌ی سیستم است. در این حالت سیستم وظیفه دارد ویژگی‌های صوتی شخص مدعی را با ویژگی‌های صوتی ذخیره شده‌ی کاربر ثبت شده‌ی مورد ادعا مقایسه نموده و با استفاده از نتیجه‌ی بدست آمده ادعای شخص را بپذیرد یا رد کند[۴].

۱-۴-۲- سیستم های بازشناسی هویت گوینده

در یک سیستم بازشناسی هویت گوینده، شخص صحبت کننده ادعای هویت یک کاربر خاص ثبت شده را نمی‌نماید و این سیستم است که وظیفه دارد او را در میان کاربران ثبت شده‌ی سیستم بازشناسی نماید و یا تشخیص دهد که ویژگی‌های صوتی او با هیچ یک از کاربران ثبت شده‌ی سیستم همخوانی ندارد[۴]. در شکل(۱-۱) شمای کلی این نوع سیستم‌ها به نمایش گذاشته شده است.



(الف) : تشخیص گوینده



(ب) : تأیید گوینده

شکل(۱-۱) : ساختار پایه از سیستم های بازشناسی گوینده[۱۱].

۱-۴-۳- سیستم های وابسته به متن

در سیستم‌های وابسته به متن، در مرحله تست، گوینده باید از همان جمله‌ای استفاده کند که در مرحله آموزش استفاده کرده است. به طور کلی این سیستم‌ها کاربرد زیادی ندارند[۴] و [۱۱].

۱-۴-۴- سیستم های مستقل از متن

در سیستم‌های مستقل از متن، در مرحله تست، لزومی ندارد که شخص از همان جمله‌ای استفاده کند که در مرحله آموزش استفاده کرده است. بنابراین همانطور که مشخص است کاربرد سیستم‌های مستقل از متن گسترده تر از سیستم‌های وابسته به متن است[۴] و [۱۱].

دقت تمامی سیستم‌های خودکار بازشناسی گوینده و گفتار در هنگام استفاده در محیط نویزی کاهش می‌یابد. دلیل این افت دقت تفاوت بین داده‌های آموزش و داده‌های تست این سیستم‌ها می‌باشد[۱]. با توجه به وجود نویز در اکثر محیط‌های واقعی، مقاوم‌سازی سیستم‌های خودکار بازشناسی گفتار و گوینده در مقابل نویزها را می‌توان واسطه بین بازشناسی خودکار و کاربرد واقعی آن دانست[۱] و[۴].

مجموعه تلاش‌های انجام شده برای مقاوم‌سازی سیستم‌های باز شناسی گفتار و گوینده را می‌توان به سه دسته تقسیم کرد. در دسته اول که استخراج ویژگی‌های مقاوم[1] نامیده می‌شود، سعی شده است تا مستقیما از سیگنال گفتار ویژگی‌های مقاوم در مقابل نویزها استخراج شود. در این مرحله باید از سیگنال گفتار ویژگی‌هایی انتخاب شود که تأثیر پذیری آنها در مقابل نویز کم باشد. در دسته دوم که نرمال سازی ویژگی‌ها[2] نامیده می‌شوند، سعی شده است تاثیر نویز تا حد امکان از ویژگی‌های استخراج شده حذف گردد. در دسته‌ای دیگر که تطبیق مدل[3] نامیده می‌شود، مقاوم‌سازی در مرحله

[1] Robust Feature Extraction
[2] Feature Normalization
[3] Model Adaptation