شناسایی گوینده وابسته به متن با استفاده از الگوریتم خوشه‌بندی K-means و روش دسته‌بندی SVM

حسین خالقی بیزکی/مجتمع دانشگاهی برق و کامپیوتر/ دانشگاه صنعتی مالک اشتر/ bizaki@gmail.com
مصطفی کاوانلویی/ کارشناس ارشد/ دانشکده مهندسی برق و الکترونیک/ دانشگاه آزاد اسلامی واحد دماوند

چکیده

هدف این مقاله توسعه یک سیستم شناسایی گوینده وابسته به متن با ارایه یک روش جدید در طبقه‌بندی گوینده می‌باشد. در این کار،یک سیستم شناسایی گوینده برای یک پایگاه داده ۱۰ نفره که در محیط عادی اتاق و با یک میکروفن معمولی هِدسِت توسط نویسنده فراهم شده، با موفقیت آزمایش شده است. ویژگی مورد استفاده، ضرایب کپسترال مبتنی بر معیار مِل(MFCC) بوده که از گفتار افراد استخراج شده و به همراه مشتقات اول و دوم بردار ویژگی ذخیره شده است. با استفاده از چندی سازی برداری(VQ) حجم داده‌ها کاهش یافته و از طبقه‌بندی‌کننده ماشین‌بردار پشتیبان(SVM) با هسته چند جمله‌ای و هسته توابع پایه شعاعی جهت طبقه‌بندی گوینده‌ها استفاده شده است. نتایج نشان می‌دهد که الگوریتم پیشنهادی در سیگنال به نویز های بالا قابلیت شناسایی قابل قبولی داشته اما با کاهش سیگنال به نویز، درصد شناسایی کاهش می‌یابد.

کلمات کلیدی: ضرایب کپسترال فرکانس مل، ماشین بردار پشتیبان، چندی‌سازی برداری، شناسایی گوینده وابسته به متن.

Text based speaker identification using K-means-clustering and SVM classification

H. Khaleghi Bizaki/ Department of Electrical and Computer Engineering/ Malek Ashtar University of Technology/ Tehran, Iran.
bizaki@gmail.com
M. Kavanloei/ MSc Student/ Science and Research Branch/ Islamic Azad University/ Damavand, Iran

ABSTRACT

The purpose of this paper is to develop a text-based speakerIdentification system by providing a new method for speaker classification. In this work, a speaker identification system has been successfully tested for a 10-person database provided by the authors in the normal room environment with a standard headset microphone.The Mel Frequency Cepstral Coefficient (MFCC) is used as a discriminating feature which is extracted from the speech of individual persons and stored with the first and second derivatives as a property vector.Using Vector Quantization (VQ), the data has been reduced and the SupportVector Machine (SVM) with the radial base functions and polynomial core to classify the speakers.The simulation results show that the proposed algorithm has acceptable detection capability in high Signal to Noise Ratio (SNR), but detection percentage decreases by decreasing the signal-to-noise ratio.

Keywords: **Mel Frequency Cepstral Coefficient, backup vector machine, SupportVector machine, Vector Quantization, Text based Speaker Identification.**

[Downloaded from kiaeee.ir on 2026-02-07]

۱- مقدمه

تمایـل به ایجاد ارتبـاط گفتاری انسـان با کامپیوتـر و عدم امنیت شناسـه‌های کاربری مختلف برای تعیین هویت افـراد مانند: کلمات عبـور و کارت‌های الکترونیکی و غیـره در مقابل جعل، محققان را به یافتن شناسـه‌های قابل اطمینان مانند شناسـه‌های بیومتریک سوق داده است. شناسه‌های بیومتریک مانند: اثر انگشت، خصوصیات چهره، عنبیه چشـم و صدا، کمتر قابل جعل هستند. بر این اساس، استخراج ویژگی‌های صوتی گوینده بر اسـاس متن موجود و سـپس تشخیص فرد مورد نظر در بین افرد مختلف بر اساس ویژگی‌های استخراج شده از اهمیت ویژه‌ای برخوردار اسـت. فرآیند شناسـایی (تشخیص) یک گوینده به‌طور عموم به‌صورت اتوماتیک توسـط ماشین و با استفاده از برخی خصوصیات صدای گوینده قابل انجام است.

تشـخیص گوینده می‌تواند به دو قسمت: شناسایی گوینده و تصدیق هویت گوینده طبقه‌بندی شـود. شناسـایی گوینده فرآیند شناسایی گوینده از میان یک پایگاه داده بوده در حالی‌که تصدیق هویت گوینده فرآیند پذیرش یا ردکردن شناسـایی یک گوینده می‌باشـد. بسیاری از روش‌های تشـخیص گوینده مانند تطبیق سـاده الگو، رویکردهای پیچـش زمانی پویـا و رویکردهای تشـخیص آماری الگـو (از قبیل شـبکه‌های عصبی و مدل‌های مخفی مارکوف)، مدل‌سـازی مخلوط گوسی، پرسپترون‌های چند لایه، توابع شعاعی پایه و الگوریتم ژنتیک برای تشخیص گوینده استفاده شده‌اند.

از طرف دیگر، روش‌های تشـخیص گوینده می‌توانند وابسـته به متن و یا مسـتقل از متن باشند. در یک سیستم مستقل از متن، بدون در نظرگرفتن آنچه که قبلا گفته شـده، تشخیص گوینده انجام می‌شود در حالی که در سیسـتم وابسـته به متن، گوینده‌ها براسـاس گفتن برخی عبـارات خاص، مانند کلمات عبور، شـماره کارت، پین‌کدها و غیره شناسایی می‌شـوند. در مقاله حاضر، تشخیص گوینده به‌صورت یک تطبیق N:1 مدل می‌شود که در آن ویژگی‌های استخراج شده یک گوینده ناشـناخته با تمام الگوهـای موجود در مدل مرجع، برای پیدا کردن بیشـترین شباهت، تطبیق داده می‌شـوند. انتخاب ویژگی‌های گوینده در طراحی یک سیستم شناسایی گوینده مهم است. مجموعه ویژگی‌هـای گوینده بایـد تغییرپذیری میان-گوینـده بالا و واریانس درون-گوینده پایینی داشـته باشـند. همچنیـن ویژگی‌های گوینده انتخاب شده به منظور کاهش افزونگی باید مستقل از یکدیگر باشند. پژوهش‌هـای زیادی در زمینه اسـتخراج ویژگی‌های صوتی مناسـب در خصوص سیسـتم‌های تشخیص گوینده انجام شده است. حسن و همکاران در[۱]یک سیسـتم شناسایی گوینده را با استفاده از ضرایب کپسترال فرکانس-مِل(Mel Frequency Cepstral Coefficient (MFCC)) توصیف و توسـعه دادند. مولفان برای آزمایش جهت نرخ شناسایی، از هر دو پنجره همینگ و مثلثی اسـتفاده کرده و نشان دادند که نتایج بـا پنجره همینـگ دارای کارآیی بهتری می‌باشـد. همچنین مولفان به‌منظور بدست آوردن نتایج رضایت‌بخش‌تر، افزایش تعداد مراکز ثقل متناسب با افزایش تعداد گویندگان را پیشنهاد کرده‌اند.

مودا و همکاران در[۲]یک سیستم تشخیص صوت با استفاده از MFCC و تکنیک پیچـش زمانـی پویـا(Dynamic Time Warping (DTW)) توصیف کردند. نتایج نشان می‌دهد که هر دو تکنیک می‌توانند به‌طور موثر برای مقاصد تشخیص صوت استفاده شوند.

الیـس در[۳] جزئیـات یک پروژه با اسـتفاده از متلـب، که طراحی یک سیسـتم تشخیص گوینده می‌باشـد، را ارایه داد. این سیستم بر تجزیه و تحلیل گام برای تشخیص یک گوینده مبتنی بوده و جزئیات مراحـل لازم برای تجزیه و تحلیل گام را در حضور نویز ارایه می‌دهد.

همچنین،یک سیستم تشـخیص اتوماتیک گوینده مبتنی برMFCC توسط ولیساوجویک[۴] ارایه شده است.

اکس‌یونگ در[۵] طراحی یک سیسـتم شناسـایی گوینده مبتنی بر تکنیکMFCC با اسـتفاده از۱۲ضریب را توصیف کرده اسـت. چهار نسـخه مختلف از سیستم با روش‌های مختلف استخراج ویژگی انجام شـده‌اند. مولفان گـزارش کرده‌اندکـه چهار روش آمـوزش مختلف کارآیی‌های مشابه در دقت و درستی را نشان می‌دهند. این پیاده‌سازی شـامل جمع‌آوری اطلاعـات، حذف سـکوت، پیش تاکیدسـیگنال، محاسبهMFCC، شیفت سیگنال، آموزش، و ارزیابی می‌باشد.

در مرجع[۶] یک روش برای تشخیص گوینده بر اساس روش ترکیبی SVM و GMM ارایه شـده است. در مرجع [۷] نیز کارایی روش‌های SVM و چند طبقه‌بندی‌کننده دیگر برای تشـخیص گوینده بررسـی شـده و یک روش ترکیبی برای بهبود کیفیت تشخیص پیشنهاد شده است.

هـدف از ایـن مقالـه، ارایه یـک سیسـتم کارآمد برای شناسـایی گوینـده وابسـته بـه متـن بـا اسـتفاده از روش‌هـایMFCC و Support Vector Machine (SVM) می‌باشد. پژوهش‌های گذشته نشان داده‌اندکهMFCCها نمایانگر جزئیات خصوصیات فردی گوینده‌ها بوده و مقاوم هستند. از طرف دیگر، SVMیک طبقه‌بندی‌کننده بسیار قوی می‌باشـد که می‌توان از آن برای تشـخیص گوینده استفاده نمود. در ساختار پیشنهادی، مراحل زیر برای طراحی سیستم تشخیص گوینده در نظر گرفته شده است:

۱- اکتساب داده‌ها ازطریق میکروفن.

۲- استخراج ویژگی.

۳- فشرده‌سـازی داده‌ها با اسـتفاده ازالگوریتـم Vector Quantization (VGLBG) Linde,Buzo and Gray

۴- تطبیق ویژگی.

با توجه به مراحل بالا، در بالاترین سـطح، تمام سیستم‌های شناسایی گوینـده دارای دو ماژول اصلی هسـتند: اسـتخراج ویژگی و تطبیق ویژگی. در ادامه این مقاله، در ابتدا، استخراج ویژگی از صدای ورودی بر اساس روشMFCC بررسی شـده سپس به ماشین بردار پشتیبان مورد استفاده برای تطبیق ویژگی پرداخته می‌شود.

۲-استخراج ویژگی و تطبیق ویژگی

۲-۱-استخراج ویژگی

واحد اسـتخراج ویژگی، یکی از واحدهای مورد نیاز بیشـتر کاربردهای بازشناسی الگو می‌باشـد. هدف از ماژول استخراج ویژگی، کاهش حجم محاسبات و حذف افزونگی‌های موجود در سیگنال گفتار با استخراج تعداد محدودی پارامتر از آن و همچنین استخراج بردارهای ویژگی صوتی، که برای مشـخص کردن خواص طیفی از زمان‌های مختلف سیگنال گفتار اسـتفاده می‌شـوند، می‌باشـد. برخی از روش‌های پر کاربرد عبارتند از: کدگـذاری پیش‌بینی خطـی (Linear Prediction Coding (LPC، ررایب کپسـترال پیشـگویی خطی -Linear Prediction Cepstral Coeffi cient(LPCC) و ضرایب کپسـتروم فرکانس-مِـل (MFCC) (مراجعه به جدول ۱). MFCCهابراسـاس تغییرات شناخته شده‌ی پهنای باند بحرانی گوش انسـان با فرکانس می‌باشند. تکنیکMFCC با استفاده از دو نـوع فیلتر ایجاد می‌شـود: فیلترهای به طول خطی و فیلترهای به طول لگاریتمی. همچنینMFCCها کمتر در معرض تغییرات شکل موج گفتار، با توجه با شرایط فیزیکی تارهای صوتی، بوده و بر همین اساس در این مقاله از این ویژگی استفاده شده است.

۲-۲-پردازنده MFCC

ضرایب کپسـترال فرکانس-مِل بهترین و شـناخته شده‌ترین ویژگی

[Downloaded from kiaeee.ir on 2026-02-07]

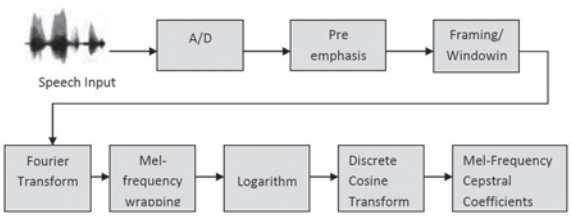
جدول ۱: روش های مختلف استخراج ویژگی

معیارها	LPC	LPCC	MFCC
وظیفه اصلی	استخراج ویژگی با تحلیل نمونه‌های پیشین	استخراج ویژگی‌ها با ترکیبLPC و تحلیل‌طیفی	استخراج ویژگی‌ها مبتنی برحوزه فرکانس با استفاده از مقیاس مل.
وابستگی گوینده	وابستگی گوینده بالا	وابستگی گوینده بالا	وابستگی گوینده متوسط(مناسب)
استحکام	ضعیف	ضعیف	خوب
انگیزه معرفی	تولید گفتار	تولید گفتار	به صورت ادراکی
فیلتر بانک	فیلترهای تمام قطب	فیلترهای تمام قطب	فیلترهای مثلثی مل
نمونه کاربردها	فشرده سازی گفتار	تشخیص گفتار و گوینده	تشخیص گفتار و گوینده

گفتار بوده که به‌طور گسترده‌تری برای هر دو هدف تشخیص گفتار و گوینده استفاده می‌شود. یک مِل،یک واحد اندازه‌گیری مبتنی بر فرکانس درک شده‌گوش انسان است. مقیاس مل تقریبا فواصل فرکانس خطی زیر ۱۰۰۰هرتز و فواصل لگاریتمی بالای ۱۰۰۰هرتز دارد. تقریبی از مل برای فرکانس می‌تواند به‌صورت زیر بیان شود[۱]:

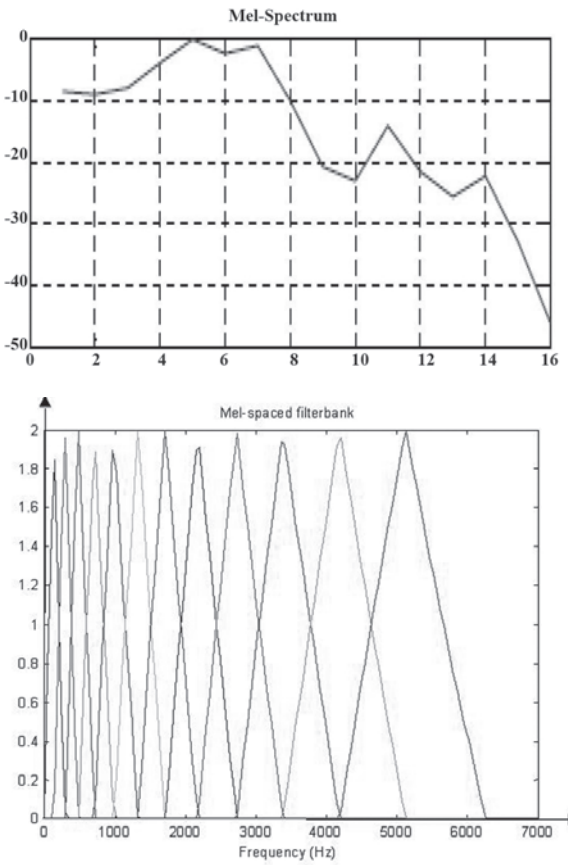
$$M = 2595\log_{10}\left(\frac{f}{700}+1\right) \quad (۱)$$

که در آن fبه فرکانس واقعی و mel(f) به فرکانس درک شده اشاره دارد. باید توجه داشت که مهم‌ترین مزیت اصلیMFCC، استحکام و مقاومت نسبت به نویز وخطاهای تخمین طیف تحت شرایط مختلف می‌باشد. با توجه به شکل (۱)، به منظوراستخراجMFCCها باید مراحل زیر صورت گیرد: تبدیل فوریه سریع، فیلترکردن وتبدیل کسینوسی از لگاریتم بردار انرژی.



شکل ۱: بلوک دیاگرام کپسترال فرکانس-مل

با توجه به شکل (۱)، سیگنال ورودی در ابتدا با نرخی بیش از دو برابر پهنای باند سیگنال، نمونه برداری میشود. نرخ‌های نمونه‌برداری رایج ۲۲۰۵۰،۱۱۰۲۵،۸۰۰۰ و ۴۴۰۰۰ نمونه در ثانیه هستند. از آنجایی تغییرات دامنه شکل موج گفتارنسبت به‌زمان، به دلیل تغییرات درمجرای صوتی و مهارت صحبت گفتار، یک فرآیند غیرایستا نمی‌باشد، با استفاده از فریم‌بندی آن را به بازه‌هایی از ۲۰میلی‌ثانیه تا۳۰ میلی‌ثانیه، که در آن سیگنال ایستان (شبه ایستان) درنظر گرفته شود، تقسیم می‌شود. پس از این که سیگنال قاب‌بندی شد، به منظورجلوگیری از مسایل ناشی از برش سیگنال، هر قاب با یک تابع پنجره (مثلا همینگ)با همپوشانی ٪۳۰ الی٪۵۰ بین قاب‌های مجاور، پنجره‌گذاری می‌شود. سپس به منظور پیداکردن مولفه‌های فرکانسی یک سیگنال نویزی و همچنین محاسبه پیچش فرکانس‌مل، از سیگنال حاصله تبدیل فوریه گسسته گرفته می‌شود. پیچش فرکانسی مل،طیف را هموار و فرکانس‌های معنی‌دار ادراکی را تایید می‌کند. در شکل (۲) فیلتر بانک مقیاس مل و طیف مل نشان داده شده است.
انجام تبدیل کسینوسی گسسته برای ناهمبستگی دامنه‌ی طیف لگاریتمی مل با ضرایب کپسترال فرکانس ملMFCC می‌باشد. در نهایت، باتبدیل لگاریتم طیف مل، سیگنال به حوزه زمان برگشت داده می‌شود. نتیجه حاصله، ضرایب کپستروم فرکانس مل (MFCC) نامیده می‌شود. ازآنجایی که ضریب طیفی مل (وبه همین ترتیب لگاریتم آن) اعدادحقیقی هستند، می‌توان با استفاده از تبدیل کسینوسی گسسته آنها را به حوزه زمان تبدیل نمود.



شکل ۲: فیلتر بانک مقیاس مل و طیف آن

۳-۲-تطبیق ویژگی

مساله تشخیص گوینده وابسته به یک زمینه بسیارگسترده علمی و مهندسی به نام تشخیص الگو، می‌باشد. هدف از تشخیص الگو، طبقه‌بندی اشیاء مدنظر در یکی از کلاس‌های از پیش تعیین شده بر اساس تطبیق ویژگی آن می‌باشد. برخی از جدیدترین روش‌های مورد استفاده در تطبیق ویژگی عبارتند از: پیچش زمانی ایستان(DTW)، مدل‌سازی مخفی مارکوف(Hidden Markov Model (HMM و چندی‌سازی برداری(VQ) می‌باشند. در این مقاله از روش VQ، به

[Downloaded from kiaeee.ir on 2026-02-07]

دلیل سهولت پیاده‌سازی و دقت بالا، استفاده شده است. روش VQ، فرآیند نقشه‌برداری بردارها از یک فضای برداری بزرگ به یک تعداد محدودی از نواحی را انجام می‌دهد. هر ناحیه با یک خوشه و مرکز آن یک کلمه کد نامیده شده است. جمع‌آوری تمام کلمات کد یک کتاب کد نامیده می‌شود[۸].

۳- تشخیص گوینده با استفاده از الگوریتم VQ

فرآیند تشخیص گوینده در دو مرحله اصلی انجام می‌شود:

- استخراج ویژگی
- تطبیق ویژگی.

۳-۱-استخراج ویژگی

هدف در استخراج ویژگی، استخراج پارامترهای اساسی برای شناسایی کردن یک سیگنال گفتار است. اگرچه چندین روش مختلف برای این کار وجود دارد، اما در این مقاله از روش MFCC استفاده شده است که خود شامل مراحل زیر می‌باشد:

الف- قاب‌بندی(تقسیم سیگنال به قاب‌ها) .

ب-پنجره‌گذاری(بدست آوردن طیف دامنه برای هر قاب).

ج-تبدیل فوریه سریع (استخراج مولفه‌های فرکانسی).

د-پیچش فرکانسی مِل (تبدیل به طیف مل).

ه-کپستروم(انجام تبدیل کسینوسی گسسته برای بدست آوردن ضرایب کپستروم).

۳-۲-تطبیق ویژگی

به‌منظور تطبیق ویژگی از کوانتیزاسیون برداری (VQ) استفاده شده است که خود شامل مراحل زیر می‌باشد:

الف- کوانتیزاسیون برداری

در این روش، در ابتدا، به منظور مشخص کردن ویژگی‌های خاص گوینده، به استخراج تعداد کمی از بردارهای ویژگی پرداخته و سپس بر اساس ویژگی‌های استخراج شده به کمک خوشه‌بندی،یک کتاب کد برای هر گوینده تشکیل می‌شود. در مرحله تشخیص، ویژگی‌های داده‌های گوینده‌ی مورد آزمایش با کتاب کد مقایسه شده و از روی میزان این اختلاف فاصله، به منظور تشخیص گوینده استفاده می‌شود. برای خوشه‌بندی از الگوریتمk-means، بر اساس کمینه کردن واریانس درون خوشه‌ها، استفاده شده است.

ب- اندازه‌گیری فاصله

به منظور اندازه‌گیری اختلاف ویژگی صوتی کاربر مورد نظر با کتاب کد از معیار کمینه فاصله اقلیدسی استفاده شده است.

۴-تشخیص گوینده با استفاده از الگوریتم SVM-VQ

در روش VQ، دسته‌بندی با استفاده از فواصل مراکز ثقل بردار ویژگی هر فرد و مقایسه آن با کل پایگاه داده، عمل شناسایی گوینده انجام می‌گیرد. به منظور بهبود الگوریتم از روش چندی‌سازی برداری جهت کاهش محاسبات استفاده شده و عملیات شناسایی گوینده در دو بخش آموزش و آزمایش به عهده ماشین بردار پشتیبان انجام شده است.

۵- شبیه‌سازی و نتایج آن

به‌منظور تشکیل پایگاه داده و اجرای سریع برنامه، یک رابط کاربری Graphical User Interface (GUI) به‌صورت شکل (۳) در نظر گرفته شده است. این رابط کاربری دارای ۱۳ گزینه است که عملکرد آنها به اختصار عبارتند از:

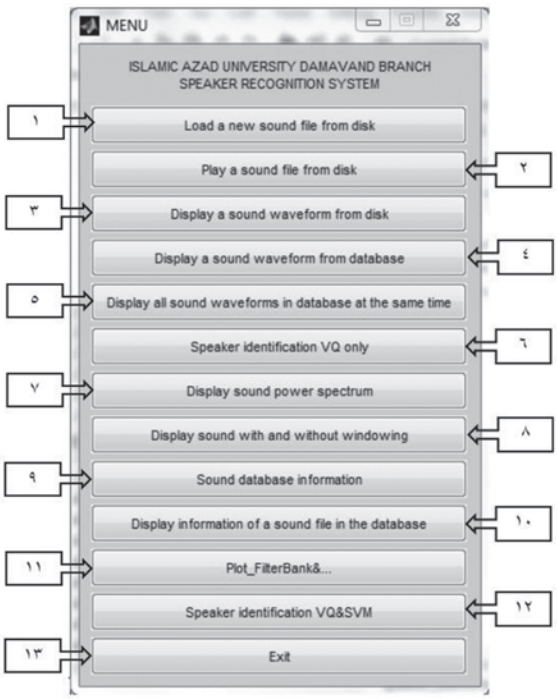
گزینه۱: خواندن یک فایل صوتی جدید از حافظه که وظیفه اضافه کردن یک فایل جدید صوتی به اندازه ۸ الی ۱۰ ثانیه را به کمک میکروفون به پایگاه داده و با درنظر گرفتن یک برچسب (ID) به عهده دارد.

گزینه۲: پخش یک فایل صوتی از دیسک سخت که با کمک یک پنجره محاوره‌ای انجام می‌گیرد.

گزینه ۳:نمایش شکل موج صوتی از دیسک که به کاربر اجازه می‌دهد تا یک فایل صوتی را از جعبه محاوره انتخاب کرده و نمایش دهد.

گزینه۴:نمایش شکل موج فایل صوتی از پایگاه داده که مشابه گزینه ۳ بوده با این تفاوت که فایل صوتی برای نمایش از صداهای پایگاه داده انتخاب می‌شود نه از روی دیسک سخت.

گزینه۵: نمایش هم‌زمان تمام شکل موج‌های صوتی موجود در پایگاه داده.



شکل ۳: GUI سیستم شناسایی گوینده

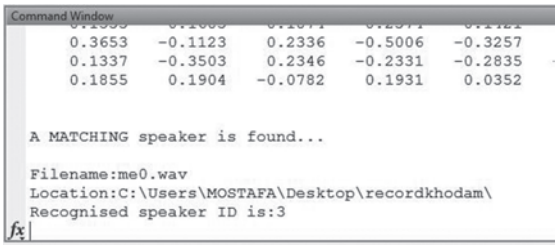
گزینه ۶: با انتخاب این گزینه تشخیص گوینده به روش VQ انجام می‌شود. این گزینه اصلی‌ترین گزینه رابط کاربری می‌باشد که فرآیند پیاده‌سازی تشخیص گوینده را انجام می‌دهد. زمانی که کاربر این گزینه را انتخاب می‌کند یک کادر محاوره‌ای مانند شکل (۴) نمایش داده می شود، سیستم بعد از اجرای برنامه به کاربر دستور ورود یک فایل صوتی را برای شناسایی می‌دهد. الگوریتم MFCC در این گزینه اجرا شده و ویژگی‌های فایل صوتی را استخراج می‌کند. سپس این ویژگی‌ها توسط الگوریتم تطبیق ویژگی کوانتیزاسیون برداری (VQ) مقایسه شده و خروجی تطبیق و شناسایی گوینده در برنامه نشان داده می‌شود.

گزینه ۷: این گزینه طیف توان فایل صوتی که توسط کاربر انتخاب شده را به‌صورت خطی و لگاریتمی نمایش می‌دهد (شکل ۵).

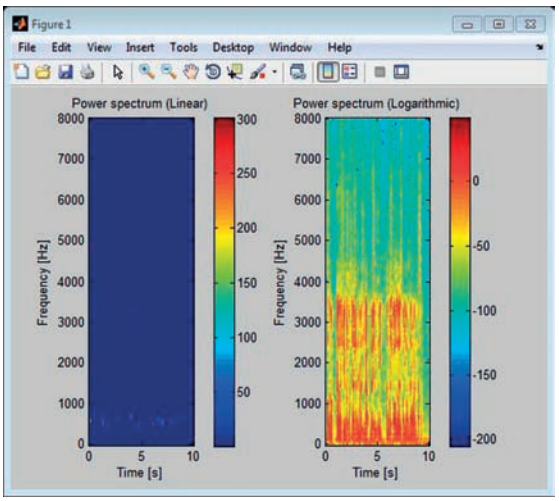
گزینه ۸: نمایش صوت بدون پنجره‌گذاری و با پنجره‌گذاری را انجام می‌دهد که این کار به‌منظور جلوگیری از مساله ناشی از برش سیگنال انجام می‌شود.

گزینه ۹: جزئیات فایل‌های صوتی ذخیره شده در ساختمان داده را نمایش می‌دهد.

[Downloaded from kiaeee.ir on 2026-02-07]



شکل ۴: کادر محاوره‌ای تشخیص گوینده



شکل ۵: نمایش طیف توان یک سیگنال صوتی

گزینه ۱۰: وظیفه نمایش اطلاعات یک فایل صوتی در پایگاه داده را به عهده دارد.

گزینه ۱۱: رسم فیلتر بانک مل و سیگنال گفتار و ضرایب کپسترال و طیف توان را انجام می‌دهد (مشابه شکل ۲).

گزینه ۱۲: شناسایی گوینده به روش SVM-VQ و رسم نتایج برای هر گوینده را به‌صورت یک نمودار میله‌ای انجام می‌دهد.

گزینه ۱۳: وظیفه خروج از برنامه را برعهده دارد.

به‌منظور بررسی سیستم تشخیص گوینده، در ابتدا، به ضبط گفتار و جمع‌آوری فایل صوتی از گوینده‌ها اقدام کرده و با یک میکروفن ساده و منویی که در GUI متلب جهت ضبط صدا بیان شد، پایگاه داده را فراهم نمودیم. برای تشکیل پایگاه، مدت زمان گفتار هر فرد به طور تقریبی ۸تا۱۰ ثانیه، تعداد افراد برابر با ۱۰ نفر شامل ۵ زن و ۵ مرد درنظر گرفته شده است. به دلیل اینکه صدای افراد و طریقه بیان جملات در شرایط مختلف کمی تفاوت دارد سعی شده جهت بررسی سیستم در شرایط مختلف، گفتارِ افراد در ده نسخه مختلف ضبط شده است (تشکیل یک پایگاه ۱۰۰ تایی). با این کار پایداری سیستم نسبت به شرایط، لهجه و تن صدای افراد، محک زده می‌شود.

۵-۱- آزمایش سیستم شناسایی گوینده با VQ

اثر تعداد مراکز ثقل (اندازه کتاب کد) مورد استفاده در الگوریتم و اثر آن بر نرخ شناسایی گوینده (درصدی از تعداد گوینده‌های شناسایی شده به کل گوینده‌های موجود در پایگاه داده) در جدول (۲) نشان داده شده است. نتایج نشان می‌دهد که زمانی که اگر تعداد مراکز ثقل بیش از ۴ و نویز کمتر از ۰٫۰۰۱ باشد، نرخ شناسایی ۱۰۰ درصد حاصل می‌شود،یعنی تمام گوینده‌های پایگاه داده به درستی شناسایی می‌شوند در حالی‌که با افزایش نویز باید تعداد مراکز ثقل را زیاد کرد تا نرخ شناسایی ۱۰۰% باقی بماند.

جدول ۲: اثر انداز کتاب کد و نویز بر روی نرخ شناسایی گوینده

تعداد مراکز ثقل (k)	نرخ شناسایی (بدون نویز)	نرخ شناسایی (واریانس‌نویز ۰.۰۰۱)	نرخ شناسایی (واریانس نویز۱. ۰)
۱	۸۰٪	۸۰٪	۶۰٪
۲	۸۰٪	۸۰٪	۳۰٪
۴	۱۰۰٪	۱۰۰٪	۵۰٪
۸	۱۰۰٪	۱۰۰٪	۷۰٪
۱۶	۱۰۰٪	۱۰۰٪	۱۰۰٪
۳۲	۱۰۰٪	۱۰۰٪	۱۰۰٪
۶۴	۱۰۰٪	۱۰۰٪	۱۰۰٪

۵-۲-آزمایش شناسایی گوینده با SVM&VQ

علاوه بر روش دسته‌بندی VQ، از روش دسته‌بندی ماشین بردار پشتیبان نیز شناسایی گوینده انجام شده است با این تفاوت که در مرحله قبل فقط با استفاده از فواصل مراکز ثقل بردار ویژگی هر فرد و مقایسه آن با کل پایگاه داده، فرد مورد نظر شناسایی‌می‌شد درحالی‌که در این قسمت، از روش چندی‌سازی برداری جهت کاهش حجم اطلاعات و همچنین کاهش محاسبات استفاده شده و عملیات شناسایی گوینده در دو بخش آموزش و آزمایش به عهده ماشین بردار پشتیبان انجام می‌گیرد.

ماشین بردار پشتیبان دارای سه نوع هسته متداول می‌باشد که با توجه به انتخاب هسته‌ها، طریقه دسته‌بندی متفاوتی داشته و در نتیجه نتایج حاصل از آن متفاوت خواهد بود. در این مقاله با استفاده از دو هسته MLP (Multilayer Perceptron) و توابع پایه شعاعی RBF (Radial Basis Function)، نرخ شناسایی گوینده و میزان تطبیق داده‌های گوینده هدف با دیگر گوینده‌ها بررسی شده است.

نرخ شناسایی و میزان تطبیق داده‌ها با انتخاب حجم کامل ضرایب کپسترال، هسته RBF و نوع خوشه‌بندیk-meansو k-meanLBG با SNR=100 (تقریبا بدون نویز) و با مراکز ثقل ۴، ۸ و ۱۶(به‌منظور شناسایی گوینده اول) در جدول (۳) نشان داده شده است. از این جدول مشاهده می‌شود که با افزایش تعداد مراکز ثقل، نرخ شناسایی کاهش می‌یابد.

همچنین، اثر تغییرات میزان ضرایب کپسترال بر نرخ شناسایی، به ازای هسته RBF و خوشه بندیk-meanLBG مرکز ثقل ۸ و SNR=100 بررسی شده و در جدول (۴) نشان داده شده است. از این جدول مشاهده می‌شودکه با کاهش تعداد ضرایب، نرخ شناسایی افزایش می‌یابد.

و بالاخره، شناسایی گوینده در حضور نویزگوسی با هسته RBF، خوشه‌بندیK-meanLBG با مرکز ثقل ۸،تعداد ضرایب ۲۰۰۰ و میزان سیگنال به نویز(SNR) متفاوت در جدول (۵) نشان داده شده است. نتایج نشان می‌دهد که با کاهش SNR، نرخ شناسایی کاهش می‌یابد.

۶-نتیجه‌گیری

در این مقاله، یک سیستم مبتنی بر متلب برای تشخیص گوینده توسعه داده شد به گونه‌ای که قادر به تشخیص گوینده مبتنی بر پایگاه داده‌اش می‌باشد. نتایج شبیه‌سازی نشان می‌دهد که الگوریتم پیشنهادی در سیگنال به نویز های بالا دارای قابلیت شناسایی قابل قبولی می‌باشد اما با کاهش سیگنال به نویز،درصد شناسایی کاهش می‌یابد.

[Downloaded from kiaeee.ir on 2026-02-07]

جدول۳: نرخ شناسایی گوینده به ازای خوشه بندی های مختلف

خوشه بندی	مرکز ثقل	درصد تشخیص
k-means	۴	۷۱.۳۸
	۸	۶۱.۳۵
	۱۶	۳۷.۱۲
k-means LBG	۴	۷۰.۳۹
	۸	۶۱.۹۵
	۱۶	۳۷.۱۹

جدول۴: نرخ شناسایی گوینده به ازای تعداد ضرایب مختلف

تعداد ضرایب	۸۰۰۰	۶۰۰۰	۴۰۰۰	۲۰۰۰	۱۰۰۰	۶۰۰
نرخ شناسایی	۶۳.۰۹	۶۴.۵۸	۶۸.۷۴	۷۳	۷۴.۸۶	۷۷.۴۵
تعداد ضرایب	۴۰۰	۲۰۰	۱۰۰	۸۰	۵۰	
نرخ شناسایی	۸۴.۸۸	۸۹.۸۲	۹۶.۷۹	۹۸.۴۶	۹۸.۷۸	

جدول۵: نرخ شناسایی گوینده اول به ازای SNRهای مختلف

SNR(dB)	۱۰۰	۸۰	۶۰	۴۰	۳۰
درصد تشخیص	۷۳	۷۲.۱۹	۶۹.۸۵	۶۳.۵۲	۵۶.۲۶
SNR(dB)	۲۵	۲۲	۲۱	۲۰	
درصد تشخیص	۵۷.۹۹	۵۴.۹۵	۵۲.۱۲	۴۹.۶۹	

سپاسگزاری

برخود لازم می‌دانیم که از زحمات جناب آقای دکتر مجتبی لطفی‌زاد کمال تشکر را داشته باشیم.

مراجع

[1] Hasan, R., Jamil, M., Rabbani, G., and Rahman, S.,"Speaker Identification Using Mel Frequency Cepstral Coefficients", 3rd International Conference on electrical & Computer Engineering, ICECE 2004, Dhaka, Bangladesh.

[2] Muda L., Begam M and Elamvazuthi I., "Voice Recognition Algorithm using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal of Computing, vol. 2, Issue 3, March 2010.

[3] Ellis., E. D. "Design of a Speaker recognition Code using MATLAB", Department of Computer and Electrical Engineering, University of Tennessee, Knoxville, 2001.

[4] Velisavljevic., V. "An automatic speaker recognition system", Mini-Project, Ecole Polytechnique Federale De Lausanne, 2003.

[5] Xiong, Y. ,"Speaker Identification System using HMM and Mel Frequency Cepstral Coefficient. Pattern Recognition and Clustering (May 2006)

[6] H.Bourouba, C.A.Korba, and Rafik Djemili, "Novel Approach in Speaker Identification using SVM and GMM", Control Engineering and Applied Informatics, Vol.15, No.3 pp. 87-95, 2013.

[7] Yussouf Nahayo, Seçkin Arı, "Performance of svm, k-nn and nbc classifiers for text-independent speaker identification with and without modelling through merging models", SAÜ Fen Bil Der 20. Cilt, 1. Sayı, s. 1-6, 2016, DOI - 10.16984/saufenbilder.00295.

[8] Nazia Hossain, Rifat Jahan, and Tanjila Tabasum Tunka, "Classified Frame-Energy Signal Using K-means Clustering", International Journal of Software Engineering & Applications (IJSEA), Vol.9, No.4, July 2018

[9] John D. Markel and Jr. Augustine H. Gray, 1976 ,"Linear Prediction of Speech", Springer Verlag, Berlin, first edition

[10] He, J., Liu, L. and Palm, G., 1999 , "A Discriminative Traning Algorithm for VQ-Based Speaker Identification", IEEE Trans on Speech and Audio Processing

[11] 14. Fang, C. (2009). "From time wraping (DTW) to Hidden markov Model (HMM)", University of Cincinatti.

[12] Linde, Y., Buzo, A., and Gray, R. (1980). "An algorithm for vector quantization design", IEEE Transactions on Communications, Vol. 28, pp. 84-95.

***

رساله دکتری مهندسی برق -کنترل

کنترل مقاوم امپدانس بازوی ربات تک رابط انعطاف‌پذیر در محیطی با سختی نامشخص

علی فیاضی

استاتید راهنما: دکتر ناصر پریز، دکتر علی کریم پور

دانشگاه فردوسی مشهد

ربات‌های انعطاف‌پذیر سبک به‌علت پاسخ سریع و کار در سرعت بالا در مقایسه با ربات‌های صلب مرسوم به‌طور گسترده‌ای در کاربردهای مختلف از جمله هوافضا، پزشکی، صنایع و غیره مورد استفاده قرار می‌گیرند. اما به‌علت قابلیت انعطاف ساختاری گسترده، مدل‌سازی و کنترل ربات‌های انعطاف‌پذیر پیچیده‌تر است. به‌علاوه برخی از پیچیدگی‌های کنترلی در این دست از ربات‌ها وجود دارد که از آن جمله می‌توان به مشخصه غیرحداقل فاز، رفتار زیرتحریک، مشخصه غیرخطی تغییر پذیر با زمان و پارامتر گسترده اشاره کرد.

رهیافت‌های کنترلی مختلفی در گذشته معرفی شده است، اما بیشتر این استراتژی‌های کنترلی به کنترل موقعیت بازوی ربات تک رابط انعطاف‌پذیر پرداخته شده است و به اثر متقابل و کنترل نیرو در برخورد با محیط توجه چندانی نشده است. علاوه بر این در تمام مطالعات انجام شده در کنترل موقعیت و نیرو تک رابط انعطاف‌پذیر، نقطه برخورد با محیط نقطه انتهایی رابط در نظر گرفته شده است و تا آنجایی‌که نویسنده اطلاع دارد، مفهوم کنترل مقاوم امپدانس ربات تک رابط انعطاف‌پذیر در محیط نامشخص با برخورد در نقطه‌ی نامعلوم میانی، به نسبت جدید و برای نخستین بار به آن پرداخته شده است.

کنترل امپدانس دارای یک ویژگی جامع است که آن را در کنترل حرکات تعاملی کارآمدتر می‌کند. در حقیقت، کنترل امپدانس با کنترل رفتار دینامیکی بین بازو و محیط به‌طور غیرمستقیم وظیفه کنترل نیرو را انجام می‌دهد. هدف نهایی این پژوهش اعمال این روش کنترلی به بازوی ربات تک رابط انعطاف‌پذیر در تعامل با محیط نامعلوم با برخورد در نقطه‌ی انتهایی و میانی بازو است. در رویکرد جدید پیشنهادی، کنترل‌کننده به‌صورت خودکار از مود کاری حرکت آزاد به مود کاری مقید تغییر وضعیت می‌دهد. بنابراین نیازی به الگوریتمی جداگانه برای تشخیص برخورد رابط با محیط نمی‌باشد. در این راستا، کنترل‌کننده امپدانس با حلقه‌ی درونی موقعیت پیشنهاد شده است. بدین معنی که در حرکت آزاد نیروی اعمالی به محیط صفر است و زاویه‌ی مرجع برای حلقه‌ی درونی موقعیت همان موقعیت مطلوب است. در حرکت مقید، مسیر مرجع برای حلقه‌ی درونی با استفاده از دینامیک امپدانس مطلوب تعیین می‌شود. استراتژی کنترل پیشنهادی، در برابر تغییرات پارامترهای محیط (نظیر سختی و ضریب میرایی)، اغتشاش نامعلوم اصطکاک کولمبی، تغییرات جرم رابط و اصطکاک ویسکوز موتور مقاوم است. همچنین، روش پیشنهادی برای هر دو حرکت آزاد و مقید معتبر می‌باشد. بخش اصلی طراحی کنترل‌کننده در استراتژی کنترل امپدانس پیشنهادی، کنترل‌کننده موقعیت حلقه‌ی داخلی است. استراتژی کنترل موقعیت حلقه‌ی داخلی براساس یک کنترل کننده مود لغزشی باسطح لغزش مرتبه کسری تعمیم یافته مجهز به رویتگر اغتشاش نامعلوم می‌باشد. کارایی و اثربخشی طرح کنترل پیشنهادی با استفاده از شبیه‌سازی عددی نشان داده شده است.

کلمات کلیدی:

کنترل مقاوم امپدانس، کنترل مود لغزشی مرتبه کسری، رویتگر اغتشاش نامعلوم، بازوی ربات تک رابط انعطاف‌پذیر، محیط نامشخص.

***

[Downloaded from kiaeee.ir on 2026-02-07]