به نام خداوند جان و خرد

کزین برتر اندیشه برنگذرد

دانشگاه شاهرود
دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان‌نامه کارشناسی ارشد
گرایش هوش مصنوعی

عنوان

تشخیص اتوماتیک صداهای ضربه ای با استفاده از تکنیک های پردازش سیگنال

نگارش
نجمه فیاضی فر

استاد راهنما
دکتر حمید حسن‌پور

استاد مشاور
دکتر هادی گرایلو

شهریور ۱۳۹۳

دانشگاه صنعتی شاهرود

دانشکده: مهندسی کامپیوتر و فناوری اطلاعات

گروه: هوش مصنوعی

پایان‌نامه کارشناسی ارشد خانم نجمه فیاضی فر

تحت عنوان: **تشخیص اتوماتیک صداهای ضربه ای با استفاده از تکنیک های پردازش سیگنال**

در تاریخ ..... توسط کمیته تخصصی زیر جهت اخذ مدرک کارشناسی‌ارشد مورد ارزیابی و با درجه ..... مورد پذیرش قرار گرفت.

استاد راهنما	امضاء	استاد مشاور	امضاء
جناب آقای دکتر حمید حسن-پور		جناب آقای دکتر هادی گرایلو	

اساتید داور	امضاء	نماینده تحصیلات تکمیلی	امضاء
نام و نام خانوادگی:		نام و نام خانوادگی:	
نام و نام خانوادگی:			

تقدیم به پدر و مادر عزیزم

که موی سپید کشتند

تا روی سپید کردم

تشکر و قدردانی

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمت های او ندانند و کوشندگان، حق او را گزاردن نتوانند.

از راهنمایی های بی شائبه استاد فرزانه، جناب آقای دکتر حمید حسن پور که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این رساله را بر عهده گرفتند، تشکر و قدردانی می نمایم.

همچنین مشورت های گرانمایه ی جناب آقای دکتر هادی گرایلو، استادمشاور این رساله را ارج می نهم و قدردان راهنمایی های بی دریغ ایشان هستم.

لازم می دانم از اساتید بزرگوارم جناب آقای دکتر پویان، جناب آقای دکتر زاهدی و جناب آقای دکتر ابوالقاسمی به پاس زحمات گرانبهایشان در دوره کارشناسی ارشدم تشکر و قدردانی کنم.

از پدر و مادر عزیزم که همراهان همیشگی من در زندگی بوده و در تهیه و ساخت پایگاه داده در این پایان نامه مرا یاری نمودند کمال تشکر و قدردانی را دارم.

تعهد نامه

اینجانب **نجمه فیاضی فر** دانشجوی دوره کارشناسی ارشد رشته **مهندسی کامپیوتر** دانشکده **کامپیوتر و فناوری اطلاعات** دانشگاه صنعتی شاهرود نویسنده **پایان‌نامه تشخیص اتوماتیک صداهای ضربه ای با استفاده از تکنیک های پردازش سیگنال** تحت راهنمایی **دکتر حمید حسن‌پور** متعهد می‌شوم.

- 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

سیستم شناسایی و تشخیص اصوات محیط کاربردهای فراوانی در زمینه‌ی هدایت ربات، سیستم‌های نظارت پزشکی و تشخیص نفوذ در سیستم‌های امنیتی و نظارتی دارا می‌باشد. این پایان‌نامه به منظور توسعه‌ی سیستم هوشمندی انجام شده است که بتواند اصوات ضربه‌ای را از غیرضربه‌ای تمییز داده و آن‌ها را دسته‌بندی کند.

از جمله چالش‌های مطرح در روش‌های موجود در این زمینه، نیاز به پنجره‌گذاری سیگنال‌ها پیش از استخراج ویژگی و سپس استخراج ویژگی‌ها از این پنجره‌ها می‌باشد. ویژگی پیشنهاد شده در این تحقیق، مستقیما از داده‌ها استخراج می‌شود و ضرورتی برای پنجره‌گذاری سیگنال‌ها وجود ندارد. به دلیل استفاده از کل سیگنال در فرآیند استخراج ویژگی، تعداد ویژگی‌های استخراج شده کمتر از سایر روش‌ها بوده و در نتیجه حجم محاسبات سیستم پیشنهادی کم و مناسب برای کاربردهای بلادرنگ می‌باشد. همچنین این روش در مقایسه با سایر روش‌های موجود، در برابر نویز بسیار مقاوم می‌باشد.

سیستم پیشنهادی ابتدا ضربه‌ای بودن یا نبودن سیگنال ورودی را بررسی می‌کند. دو رویکرد جدید در این پایان‌نامه، برای شناسایی پیشنهاد شده است. این دو روش بر اساس تغییرات میزان انرژی سیگنال در طول زمان عمل می‌کنند. در بخش شناسایی، سیستم در شرایط بدون نویز توانسته است عملکرد بسیار مناسبی ارائه دهد. در این حالت میزان شناسایی سیستم ۱۰۰٪ گزارش شده است.

پس از شناسایی اصوات ضربه‌ای، سیستم وارد فاز دسته‌بندی می‌شود. ویژگی جدیدی که در این پایان‌نامه ارائه شده است بر مبنای تفاوت در رفتار سیگنال‌ها و نحوه‌ی رسیدن آن‌ها از مقدار کمینه به بیشینه، عمل می‌کند. در این کار از دسته‌بند k-نزدیک‌ترین همسایه استفاده شده و میزان تشخیص درست سیستم در این حالت ۹۳.۷۵٪ گزارش شده است که در مقایسه با روش‌های موجود از جمله ضرایب کپسترال فرکانس مل، نقطه‌ی تعادل طیفی و شار طیفی عملکرد بهتری داشته است.

کلمات کلیدی: اصوات ضربه‌ای، شناسایی و دسته‌بندی، استخراج ویژگی، دسته‌بند k-نزدیکترین همسایه

لیست مقالات مستخرج از پایان‌نامه

1. Najmeh Fayazi Far, Hamid HassanPour, *"A Robust Impulsive Sound Detection and Recognition System Using a Novel Statistical Feature"*, International Journal of Computer Applications, IJCA (Accepted on September 2014).

۲. نجمه فیاضی‌فر، حمید حسن‌پور،"شناسایی و دسته‌بندی اصوات ضربه‌ای جهت تشخیص نفوذ در کاربردهای نظارتی و امنیتی "، سومین همایش ملی زبان‌شناسی رایانشی، دانشگاه صنعتی شریف (ارسال شده).

فهرست مطالب

عنوان صفحه

فهرست اشکال

عنوان صفحه

فهرست جداول

عنوان صفحه

فصل اول

مقدمه

۱-۱- مقدمه

جهان، در حال حرکت به سوی استفاده‌ی گسترده از سیستم‌های هوشمندی است که همانند انسان بتواند محیط را درک نموده و تصمیمات درست و مناسبی اتخاذ نماید و به‌طور کلی، تمامی فرآیندهای انسانی را انجام داده و نقش عامل انسانی را، حذف و یا کم‌رنگ کند . از این‌رو ارائه‌ی سیستم هوشمندی که توانایی تشخیص صداهای محیط را داشته باشد و بر مبنای درک صحیح خود از اصوات موجود در محیط، پاسخ مناسبی به آن دهد بسیار ضروری به نظر می‌رسد. این سیستم در واقع همانند سیستم شنوایی انسان عمل کرده و وظیفه‌ی دریافت اطلاعات از محیط و سپس تصمیم‌گیری و پاسخ‌گویی بر مبنای اطلاعات دریافتی را بر عهده دارد.

در این فصل ابتدا کاربردهای سیستم شناسایی اصوات محیط را ذکرکرده و سپس ضرورت انجام تحقیق را بیان می‌کنیم، آن‌گاه به شرح سیستم شناسایی اصوات ضربه‌ای و نحوه‌ی عملکرد آن می‌پردازیم و چند نمونه از دست‌آوردهای این سیستم را شرح می‌دهیم. نتیجه‌گیری پایان‌بخش این فصل خواهد بود.

۱-۲- کاربردهای سیستم شناسایی اصوات محیط

تشخیص صداهای محیط شامل دامنه‌ی وسیعی است و کاربردهای متفاوتی در زمینه‌های مختلف دارد. از جمله کاربردهای آن در زمینه‌ی گفتار می‌توان به تشخیص گفتار از سایر صداهای موجود در محیط [۱]،[۲]، تشخیص گوینده [۳]،[۴]، تبدیل گفتار به نوشتار [۵]،[۶] و ترجمه‌ی همزمان دو زبان مختلف به یکدیگر [۷] اشاره نمود. بازیابی مبتنی بر محتوای اطلاعات از پایگاه‌داده‌های صوتی[۸]، برچسب‌گذاری فایل‌های صوتی [۹]، تعامل انسان و ربات [۱۰]،[۱۱]، جست‌وجو و بازیابی موسیقی در پایگاه داده‌های موسیقی[۱۲] و دسته‌بندی ادوات موسیقی[۱۳] از دیگر کابردهای این سیستم می‌باشند. سیستم شناسایی اصوات محیط، کاربردهایی در پیاده‌سازی خانه‌ی هوشمند نیز دارد[۱۴]. در بوم‌شناسی برای

تشخیص گونه‌های مختلف پرندگان و حیوانات، از سیستم هوشمند تشخیص صداهای محیط استفاده می-شود[۱۵]، [۱۶].

۱-۳- ضرورت انجام تحقیق

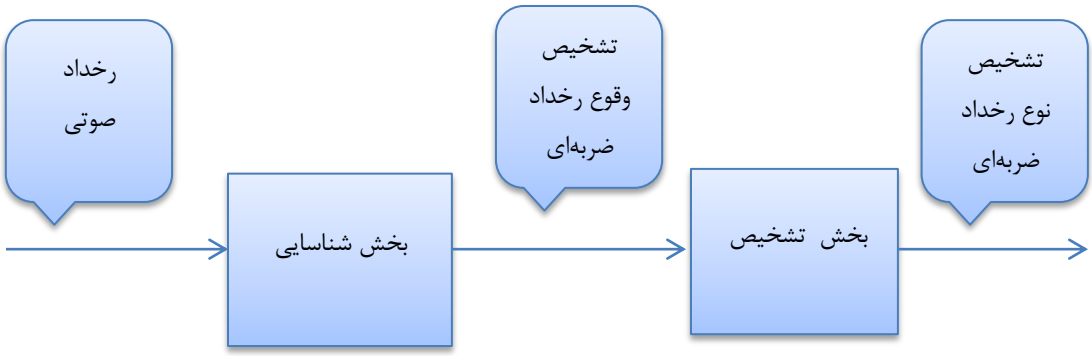
این پایان‌نامه در ابتدا با پیشنهاد شرکت ملی پالایش و پخش فرآورده‌های نفتی ایران تعریف شد. هدف آن‌ها ایجاد سیستم هوشمندی که توانایی تشخیص صدای انفجار مخزن حاوی گاز را، از سایر اصوات موجود در محیط داشته باشد، می‌باشد. سپس تصمیم بر آن شد، سیستم جامع‌تری تولید گردد که توانایی شناسایی و تشخیص انواع اصوات ضربه‌ای را داشته باشد.

سیستم هوشمند تشخیص انفجار، در جایگاه‌های سوخت‌گیری گاز طبیعی کاربرد دارد. مبنای کار این سیستم به گونه ای است که پس از وقوع انفجار در خودرویی که مشغول سوخت‌گیری می باشد، آن را تشخیص داده و کلید قطع اضطراری فعال شده و جریان سوخت را قطع می کند. در حال حاضر این عمل به صورت دستی توسط متصدی جایگاه انجام می‌شود اما با توجه به سروصدای زیادی که ناشی از خروج سوخت از نازل با فشار psi ۳۰۰۰ و پخش آن در فضای آزاد می‌باشد، متصدی جایگاه عملا قادر به قطع جریان سوخت نمی‌باشد. در نتیجه، حضور چنین سیستمی برای پیشگیری از پیامدهای بعدی، احساس می‌شود.

۱-۴- سیستم شناسایی و تشخیص اصوات ضربه‌ای

صدای ضربه‌ای با حضور ناگهانی موج فشار هوا ایجاد می‌شود و با افزایش سریع انرژی سیگنال همراه است. مدت زمان رخداد این دسته از صداها کوتاه بوده و در واقع بیانگر این نکته است که اتفاقی در محیط رخ داده است. از این افزایش ناگهانی و سریع می‌توان برای شناسایی صدای ضربه‌ای استفاده کرد. صدا-هایی همچون بسته شدن در، انفجار، رعدوبرق، فریاد زدن، زنگ تلفن و شکستن شیشه در دسته‌ی صدا-های ضربه‌ای قرار می‌گیرند. در این پایان‌نامه منظور از تشخیص صدای ضربه‌ای توانایی تمییز صدا بین

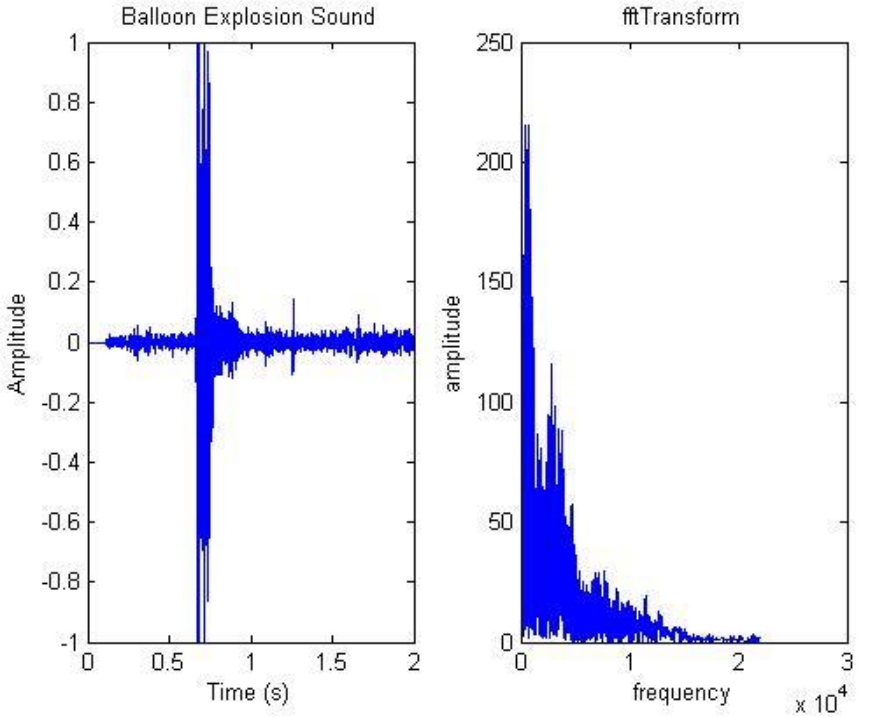
دسته‌های مختلف است در حالی که شناسایی تنها به بررسی این موضوع که رخدادی ناگهانی در محیط اتفاق افتاده است، می‌پردازد و جزئیات بیشتری راجع به نوع صدای ضربه‌ای ایجاد شده، ارائه نمی‌شود.

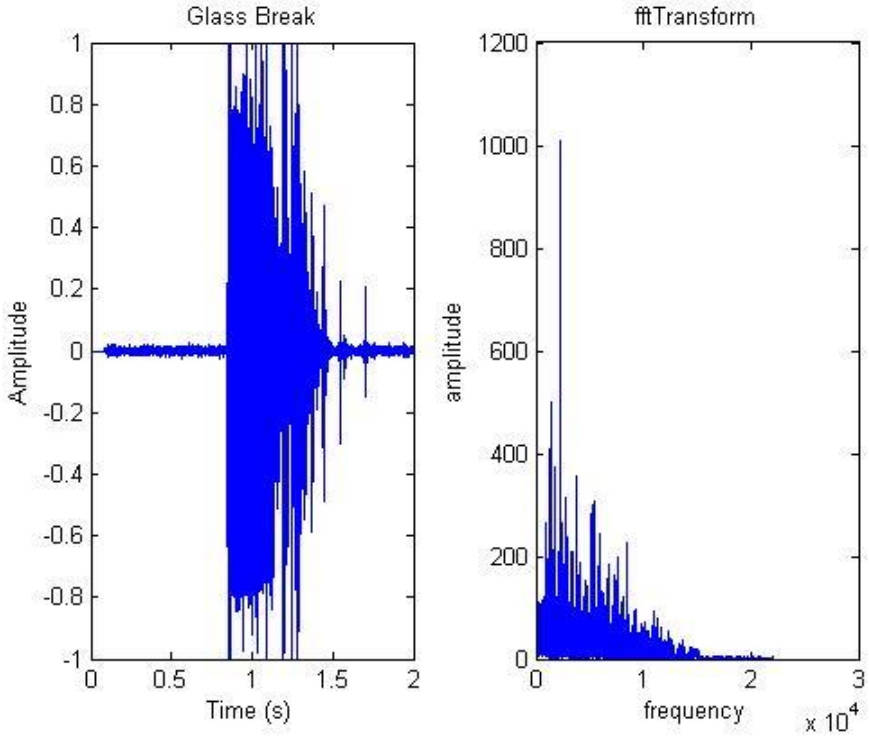


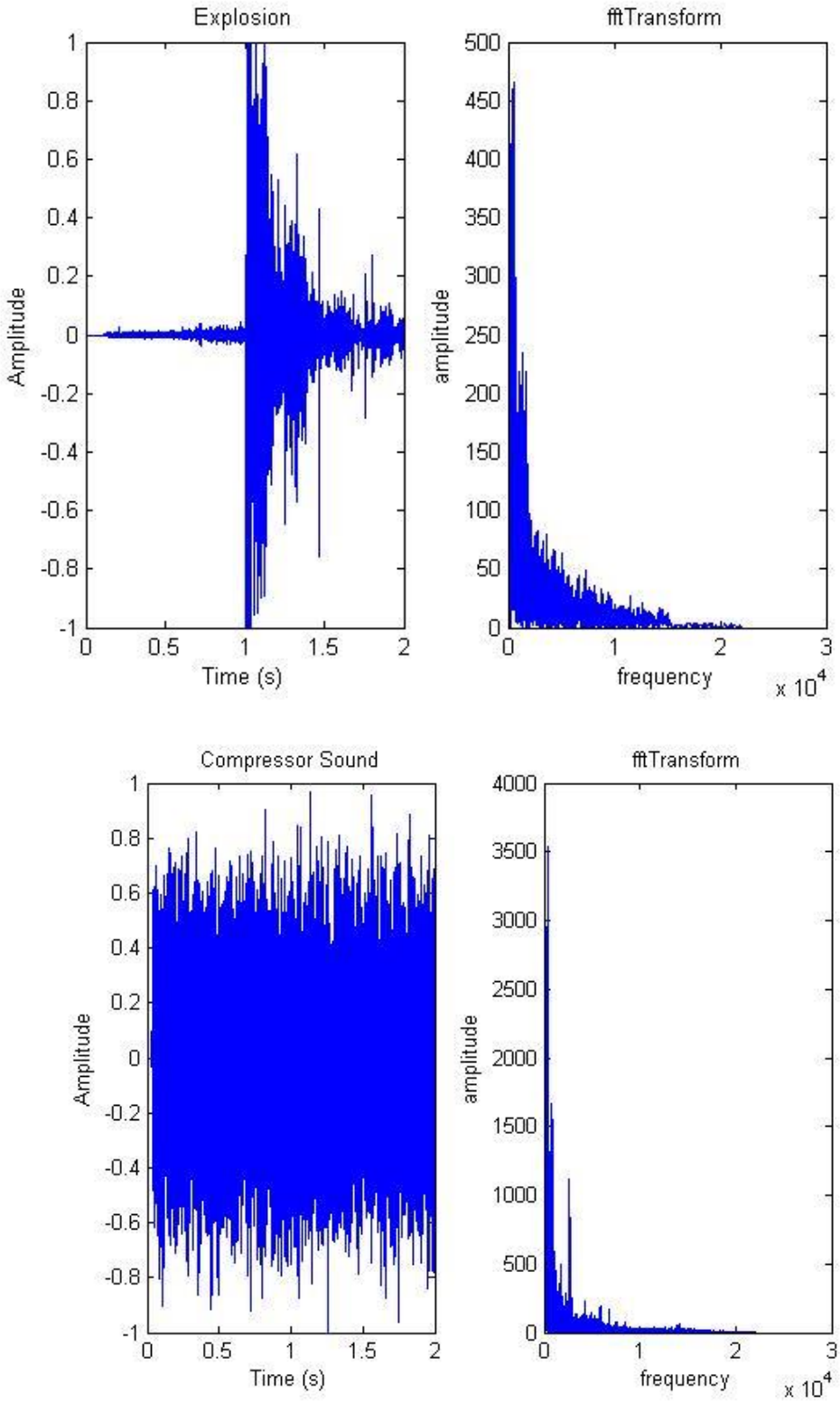
شکل ۱-۱ : سیستم شناسایی و تشخیص

در شکل ۱-۱ خروجی بخش شناسایی و تشخیص نشان داده شده‌اند. در سیستم ارائه شده ابتدا صدا-های ورودی مورد بررسی قرار می‌گیرند و در صورتی که صدایی ناگهانی باشد، بخش تشخیص شروع به کار کرده و سعی در دسته‌بندی نوع صدای ضربه‌ای می‌کند. در این حالت بخش شناسایی به صورت پیش-پردازشی برای بخش تشخیص عمل می کند و زمانی که صدای ورودی ضربه‌ای نباشد زمان محاسبات کاهش می‌یابد.

سیگنال‌های ضربه‌ای دسته‌های مختلف، از نظر ساختار و فرکانس‌های موجود در آن‌ها، با یکدیگر و با سیگنال‌های غیرضربه‌ای متفاوت می‌باشند. در شکل ۱-۲ نمونه‌ای از سیگنال‌های به کار رفته در هر دسته و تبدیل فوریه آن‌ها نشان داده شده‌اند.


Balloon Explosion Sound
Amplitude
Time (s)
fftTransform
amplitude
frequency
x 10^4


Glass Break
Amplitude
Time (s)
fftTransform
amplitude
frequency
x 10^4



شکل ۱-۲: سیگنال های دسته های مختلف به همراه تبدیل فوریه آنها

همان‌طور که در شکل۱-۲ دیده می‌شود سیگنال‌های ضربه‌ای نه تنها با سیگنال‌های غیرضربه‌ای بلکه دسته‌های مختلف اصوات ضربه‌ای، از نظر شکل ظاهری و فرکانس‌های موجود در آن‌ها با یکدیگر متفاوت می‌باشند. به طور مثال روند تغییر انرژی در سیگنال مربوط به کارکرد کمپرسور که در دسته‌ی اصوات غیرضربه‌ای قرار می‌گیرد، در طول مدت دو ثانیه، تقریبا ثابت بوده، در حالی که انرژی سیگنال‌های دسته-های دیگر به طور لحظه‌ای افزایش می‌یابد و پس از مدت بسیار کوتاهی مجددا مقدار آن کم شده و به صفر نزدیک می‌شود. بسته به این‌که سیگنال، مربوط به کدام دسته از اصوات ضربه‌ای باشد، طول مدتی که انرژی سیگنال افزایش یافته، متغییر می‌باشد. به طور مثال در دسته‌ی انفجار بادکنک، مدت زمان افزایش انرژی (رخداد ضربه) بسیار کوتاه و در حدود ۰.۱ ثانیه بوده در حالی که در دو دسته‌ی انفجار و شکستن شیشه رخداد در مدت زمان بیشتری – انفجار ۰.۸ ثانیه و شکستن شیشه ۰.۶ ثانیه- به وقوع می‌پیوندد. در سیگنال شکستن شیشه در تمام طول رخداد، انرژی سیگنال به مقدار بیشینه نزدیک بوده و پس از پایان رخداد یک مرتبه کاهش یافته اما در انفجار روند کاهش انرژی یکنواخت‌تر بوده و مقدار آن به مرور کاهش می‌یابد. علاوه بر تفاوت‌هایی که در شکل ظاهری اصوات دسته‌های مختلف ذکر شد، در تبدیل فوریه آن‌ها که نمایانگر فرکانس‌های موجود در سیگنال‌ها می‌باشد، تفاوت‌هایی به چشم می‌خورد. در دسته‌ی کارکرد کمپرسور، فرکانس‌های پایین – صفر تا ۱۰۰۰هرتز- دارای انرژی زیاد بوده و بیشترین مقدار انرژی در فرکانس ۴۵۰ هرتز حضور دارد و مقدار آن برابر ۳۵۰۰ می‌باشد. در دسته‌ی انفجار انرژی در محدوده‌ی فرکانسی بیشتری – صفر تا ۴۵۰۰ هرتز- گسترده شده ولی مقدار بیشنه انرژی، بسیار کمتر از دسته‌ی اصوات غیرضربه‌ای بوده و این مقدار تقریبا برابر ۴۵۰ می‌باشد. محدوده‌ی فرکانسی در دسته‌ی انفجار بادکنک – صفر تا ۴۰۰۰- شبیه به دسته‌ی انفجار بوده، اما مقدار بیشنه انرژی، کمتر از دسته‌ی انفجار و در حدود ۲۰۰ می‌باشد. در دسته‌ی شکستن شیشه محدوده‌ی فرکانسی بسیار گسترده‌تر از سایر

دسته‌ها- صفر تا ۹۰۰۰هرتز- بوده و مقدار بیشینه‌ی انرژی در فرکـانس ۲۵۰۰هرتـز رخ داده و مقـدار آن برابر ۱۰۰۰ می‌باشد.

همان‌طور که بیان شد، سیگنال‌های دسته‌های مختلف دارای ساختار و رفتار متفاوت در حـوزه‌ی زمـان و حوزه‌ی فرکانس می‌باشند. در فصل سوم بیان خواهد شد که چگونه از این تفاوت‌ها می‌توان بـرای دسـته-بندی اصوات بهره برد.

۱-۵- دست‌آوردهای سیستم تشخیص اصوات ضربه‌ای

این سیستم، کاربردهای فراوانی در زمینه‌های مختلف دارد. از جمله آن‌ها می‌توان به موارد زیر اشاره کرد.

- سیستم‌های امنیتی و نظارتی: به طور معمول در مکان‌هایی مانند بانک‌ها، مغازه‌ها، خانه-ها و پارکینگ‌های عمومی از این سیستم می‌توان برای تشخیص نفوذ استفاده کرد. در این حالت به راحتی می‌توان یک فعالیت غیرمعمول را شناسایی کرد. با استفاده از این سیستم حتی می‌توان صداهایی مانند شکسته شدن شیشه، بسته شدن در، شلیک و فریاد را تشخیص داد[۱۷]،[۱۸]،[۱۹].
- کمک‌رسانی به افراد ناشنوا و پیر: در این کاربرد به افرادی که مشکل شنوایی دارند کمک‌رسانی می‌شود. در محیط خانه می‌توان صداهایی همچون زنگ در یا تلفن را به صورت دیداری به این افراد اعلام کرد. در محیط بیرون نیز صدای بوق خودرو، آژیر، انفجار و سایر موقعیت‌های خطرناک اعلام می‌شوند[۲۰]،[۱۹]،[۲۱].
- کاربردهای نظارت پزشکی: در این کاربردها در محل سکونت افراد پیر، بیمار یا زنان باردار که احتمال بیهوشی در آنان وجود دارد، میکروفن‌هایی نصب می‌شود تا در صورت بروز این

حادثه و زمین افتادن افراد یا اشیا بتوان شرایط اضطراری را تشخیص داده و به این افراد کمک-رسانی شود[۲۲]،[۲۳].

- تعامل انسان و ربات و همچنین جهت‌یابی و هدایت ربات: در این کاربرد، ربات صداهای محیط را تشخیص می‌دهد و بر مبنای آنها جهت حرکت خود را تعیین می‌نماید و در صورتی که اتفاقی در محیط رخ داده باشد، از آن دوری می‌کند. در کاربردی دیگر ربات می‌تواند صدای زنگ تلفن یا در را تشخیص داده و متناسب با وظیفه‌ی خود به آنها پاسخ مناسب دهد. به طور کلی در این کاربرد همانطور که انسان صداهای محیط را مبنای تصمیم‌گیری خود قرار می‌دهد، ربات نیز قادر خواهد بود صداها را دریافت و از یکدیگر تمییز داده و شرایط محیط را تشخیص و بر مبنای آن تصمیم مناسب را اتخاذ نماید[۱۱]،[۲۱].

۱-۶- ساختار پایان‌نامه

در این فصل سیستم شناسایی و تشخیص اصوات ضربه‌ای و نحوه‌ی عملکرد آن را بررسی کردیم. در فصل دوم به مرور بر کارهایی که تاکنون در زمینه‌ی تشخیص اصوات ضربه‌ای صورت گرفته است، می-پردازیم. فصل سوم به معرفی سیستم پیشنهادی و بیان نحوه‌ی عملکرد آن اختصاص داده شده است. فصل چهارم در بردارنده‌ی نتایج و ارزیابی‌های تجربی حاصل از به کارگیری روش پیشنهادی ما برای استخراج ویژگی و مقایسه‌ی آن با کارهای گذشته است. در نهایت فصل پنجم به نتیجه‌گیری و بیان کارهای آینده اختصاص داده شده است.

۱-۷- نتیجه‌گیری

در این فصل سیستم شناسایی و تشخیص اصوات ضربه‌ای را معرفی کردیم و نحوه‌ی عملکرد آن را به طور اجمالی شرح دادیم. سپس کاربردهای متفاوت آن را در زمینه‌های مختلف مورد بررسی قرار دادیم. با

توجه به کاربردهای فراوانی که این سیستم در زمینه‌های مختلف، دارا می‌باشد، ایجاد چنین سیستمی ضروری به نظر می‌رسد.

فصل دوم

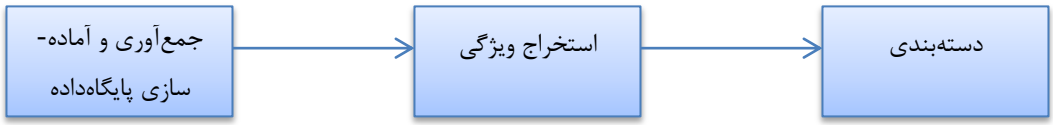
مرور کارهای پیشین

۲-۱- مقدمه

در این فصل با آگاهی از سیستم شناسایی اصوات محیط و دانستن کاربردها و ضرورت وجود چنین سیستمی، به توضیح مختصری از تحقیقات صورت گرفته در این زمینه خواهیم پرداخت. نتیجه‌گیری پایان بخش این فصل خواهد بود.

۲-۲- تشخیص اصوات ضربه‌ای

در زمینه‌ی تشخیص صداهای محیط و به طور خاص‌تر شناسایی و تشخیص صداهای ضربه‌ای کارهای متفاوت و زیادی در دهه‌ی اخیر انجام شده است. در تمامی این سیستم‌ها سه بخش اصلی وجود دارد که بدنه‌ی آن‌ها را تشکیل می‌دهند. این سه بخش در نمودار زیر نشان داده شده‌اند.



شکل۲-۱: بدنه‌ی سیستم‌های تشخیص اصوات

در زمینه‌ی تشخیص اصوات محیط، پایگاه‌داده‌ی استانداردی وجود ندارد و در هر تحقیق، نویسنده با توجه به کاربرد، پایگاه‌داده‌ای را ساخته و یا جمع‌آوری کرده است. تعداد دسته‌های موجود در هر پایگاه-داده و تعداد نمونه‌های هر یک از دسته‌ها در هر کدام از پژوهش‌ها متفاوت است.

در بخش استخراج ویژگی تحقیقات گسترده‌ای صورت گرفته است. در هر تحقیق، نویسنده با توجه به پایگاه‌داده مورد استفاده و کاربرد، ویژگی جدیدی را پیشنهاد داده است. معمولا نوآوری این پژوهش‌ها در ارائه این ویژگی نوین است که یا نرخ تشخیص صحیح سیستم را افزایش داده و یا باعث کاهش ابعاد بردار ویژگی شده است.

در بخش دسته‌بندی، با توجه به نوع ویژگی‌های استخراج شده، از دسته‌بندهای متفاوتی استفاده شده است. غالب دسته‌بندهای استفاده شده در زمینه‌ی تشخیص صداهای محیط شامل مدل مخفی مارکوف[۱]، مدل مخلوط گوسی[۲]، ماشین بردار پشتیبان[۳] و k- نزدیکترین همسایه [۴] می‌باشند.

۲-۲-۱- استخراج ویژگی

در سیستم‌های شناسایی و تشخیص اصوات محیط، بخش استخراج ویژگی نقش بسیار مهم و تعیین-کننده‌ای دارد. تاکنون روش‌های متنوع و گوناگونی برای استخراج ویژگی از صوت ارائه شده است. این ویژگی‌ها را می‌توان از جهات مختلف می‌توان دسته‌بندی کرد. در [۲۴] و [۲۵] این ویژگی‌ها با در نظر گرفتن دامنه‌ی آنها به پنج دسته‌ی ویژگی‌های حوزه‌ی زمان، حوزه‌ی فرکانس، حوزه‌ی کپسترال، حوزه‌ی فرکانس مودولاسیون[۵] و فضای فاز تقسیم می‌شوند. در [۲۶] و [۲۷] روش‌های تشخیص صدای محیط، بر مبنای ویژگی‌های استخراج شده به دو دسته تکنیک‌های ایستا و تکنیک‌های غیرایستا تقسیم می‌شوند. تکنیک‌های غیرایستا خود شامل سه دسته‌ی روش‌های مبتنی بر موجک، مبتنی بر نمایش پراکنده و مبتنی بر طیف هستند[۲۷].

ویژگی‌هایی که در کاربردهای گفتار و موسیقی استفاده می‌شوند، معمولا جز تکنیک‌های ایستای تشخیص صداهای محیط قرار می‌گیرند. این ویژگی‌ها غالبا بر مبنای خواص روان- صوتی[۶] اصوات مثل بلندی، طنین[۷] و گام[۸] می‌باشند. ویژگی‌های ذکر شده در ذیل جز این دسته قرار می‌گیرند:

[1] Hidden Markov Model (HMM)
[2] Gaussian Mixture Model (GMM)
[3] Support Vector Machine (SVM)
[4] K-Nearest -Neighbour (KNN)
[5] modulation frequency domain
[6] psychoacoustic
[7] timbre
[8] pitch

- نرخ عبور از صفر[۱]
- ضرایب کپسترال هوموورفیک[۲]
- ضرایب کپسترال فرکانس مل[۳]
- ضرایب کپسترال پیش‌بینی خطی[۴]
- ضرایب کپسترال پیش‌بینی خطی فرکانس مل [۵]
- ضرایب کپسترال فرکانس بارک [۶]
- ضرایب پیش‌بینی خطی فرکانس مل [۷]
- ویژگی پیش‌بینی خطی ادراکی[۸]
- نقطه‌ی تعادل طیف[۹]
- همواری طیفی [۱۰]
- شار طیفی[۱۱]
- نقطه قطع طیف[۱۲]
- انرژی زمان کوتاه [۱۳]

در این ادامه این فصل، پیش از مرور مقالات منتشر شده در زمینه‌ی تشخیص اصوات محیط، به توضیح و بررسی چند نمونه از ویژگی‌های فوق که نتایج حاصل از روش پیشنهادی را با آن‌ها مقایسه کرده‌ایم، می-

[1] Zero cross rate (ZCR)
[2] Homomorphic cepstral coefficients
[3] Mel frequency cepstral coefficients(MFCC)
[4] Linear prediction cepstral (LPC) coefficients
[5] Mel frequency LPC coefficients
[6] Bark frequency cepstral coefficients
[7] Bark frequency LPC coefficients
[8] Perceptual linear prediction (PLP) features
[9] Spectral Centroid
[10] Spectral Flatness
[11] Spectral Flux
[12] Spectral Roll off Point
[13] Short Time Average Energy

پردازیم. این روش‌ها شامل ضرایب کپسترال فرکانس مل، نرخ عبور از صفر، نقطه‌ی تعادل طیف، نقطه‌ی قطع طیف و شار طیفی می‌باشند.

۲-۲-۱-۱- ضرایب کپسترال فرکانس مل

ضرایب کپسترال فرکانس مل، ابتدا به دلیل استفاده در سیستم‌های تشخیص خودکار گفتار به وجود آمد، اما هم‌اکنون به یکی از تکنیک‌های استاندارد در پردازش صوت تبدیل شده است و تا به امروز یکی از پرکاربردترین و متداول‌ترین روش‌های استخراج ویژگی بوده است. همان‌طور که در فصل یک اشاره شد، به دلیل نبود پایگاه‌داده‌ی استاندارد در این زمینه، برای ارزیابی کارایی روش‌های نوین ارائه شده، مقایسه‌ای بین آن‌ها و روش‌های پایه صورت می‌گیرد. یکی از پایه‌ای ترین و اساسی‌ترین این روش‌ها،MFCC، می‌باشد. ما نیز کارایی روش پیشنهادی خود را با این ویژگی مقایسه کردیم. استخراج این ضرایب از یک سیگنال شامل چند مرحله می‌باشد که آن‌ها را به طور خلاصه به صورت زیر بیان می‌کنیم.

- ابتدا هر سیگنال را به چندین پنجره تقسیم می‌کنیم. دلیل پنجره‌گذاری سیگنال این است که می‌خواهیم سیگنال مورد نظر ما در هر پنجره، ایستا باشد و تغییرات محسوسی در آن اتفاق نیافتد.
- در این مرحله، از هر پنجره تبدیل فوریه گسسته می‌گیریم و طیف توان آن را تشکیل می‌دهیم. سپس با استفاده از فرمول زیر طیف قدرت مبتنی بر پریودگرام [۱] را برای هر پنجره $s_i(n)$ محاسبه می‌کنیم.

$$p_i(k) = \frac{1}{N}\left|S_i(k)\right|^2 \qquad (۱-۲)$$

[1] periodogram

- هم‌اکنون فیلتربانک مل را بر هر کدام از طیف‌های به دست آمده در مرحله‌ی قبل اعمال می‌کنیم و مقادیر انرژی‌ها در هر پنجره را با هم جمع می‌زنیم.
- در این مرحله از انرژی هر کدام از فیلتربانک‌ها لگاریتم می‌گیریم.
- سپس از انرژی فیلتربانک‌هاتی لگاریتمی، تبدیل کسینوسی گسسته می‌گیریم.
- در نهایت ۱۳ ضریب اول تبدیل کسینوسی را نگه می‌داریم.

با اعمال مراحل فوق بر روی سیگنال ورودی، هر سیگنال را به تعدادی پنجره تقسیم می‌کنیم و از هر یک از آن‌ها، ۱۳ ضریب استخراج می‌کنیم. بدین ترتیب بردار ویژگی که شامل، تعداد پنجره ها ×۱۳ ویژگی است، به دسته‌بند مناسب می‌دهیم. در این پایان‌نامه تعداد پنجره‌ها را تغییر دادیم و نرخ تشخیص سیستم را اندازگیری نمودیم. نتایج به دست آمده از هرکدام از حالات را در بخش ۴-۴ گزارش می‌دهیم.

۲-۲-۱-۲- نرخ عبور از صفر

این ویژگی یکی از راحت‌ترین و کم هزینه‌ترین ویژگی های به کار رفته در صوت می‌باشد. این ویژگی بیانگر تعداد دفعاتی است که یک سیگنال در حوزه‌ی زمان در مدت یک ثانیه، محور افقی را قطع می‌کند. در واقع نمایانگر تعداد دفعاتی است که سیگنال تغییر علامت داده و از منفی به مثبت یا از مثبت به منفی می‌رود. نرخ عبور از صفر به صورت زیر تعریف می‌شود.

$$ZCR = \frac{1}{T-1}\sum_{t=1}^{T-1}\prod\{S_t S_{t-1} < 0\} \qquad (۲-۲)$$

در فرمول بالا s سیگنالی با طول T می‌باشد و تابع نشانگر $\prod\{A\}$ در صورتی که A درست باشد مقدار یک و در غیر این صورت مقدار صفر دارد. نرخ عبور از صفر ویژگی بسیار پرکاربردی در دسته‌بندی گفتار و

موسیقی می‌باشد. از دیگر کاربردهای این ویژگی می‌توان به تحلیل گفتار، تشخیص آواز در موسیقی و همچنین تشخیص صداهای محیط، اشاره کرد.

در این پایان‌نامه برای استفاده از این ویژگی پس از اعمال پیش‌پردازش‌ها که در ۳-۲-۲ بیان می‌کنیم، سیگنال را به پنجره‌هایی تقسیم نموده و سپس با استفاده از فرمول ۲-۲، در هر پنجره، تعداد نقاطی که سیگنال از صفر عبور می‌کند را محاسبه کردیم. این اعداد به دست آمده را کنار یکدیگر قرار داده و بردار ویژگی را ساختیم و در نهایت با استفاده از این ویژگی‌ها و به کار گرفتن دسته‌بند مناسب داده‌ها را طبقه-بندی کردیم. در استخراج این ویژگی، دو پارامتر حضور دارند که در تعیین درصد خروجی موثرند. این دو پارامتر سایز هر پنجره و گام می‌باشند. از آنجا که فریم‌ها با یکدیگر همپوشانی دارند، تعیین اینکه با چه گامی پنجره بعدی را آغاز کنیم، مهم می‌باشد. در واقع گام بیانگر تعداد نمونه‌هایی است که بین ابتدای هر دو پنجره متوالی، وجود دارند. در بخش ۴-۴ بهترین نرخ تشخیص سیستم با به کارگیری مقدار بهینه‌ی دو پارامتر فوق گزارش شده است.

۲-۲-۱-۳- نقطه‌ی تعادل طیف

این ویژگی بیانگر نقطه‌ی تعادل، در توزیع طیف سیگنال می‌باشد. در واقع این ویژگی با میانگین وزن-دار گرفتن از فرکانس‌های موجود در سیگنال که وزن آن‌ها انرژی هر کدام از این فرکانس‌ها می‌باشد، محاسبه می‌شود. این ویژگی غالبا از هر پنجره سیگنال استخراج می‌شود و به ازای هر یک از آن‌ها، یک عدد به عنوان نقطه‌ی تعادل طیفی به دست می‌آید. این اعداد با یکدیگر تشکیل بردار ویژگی را می‌دهند. این ویژگی به ازای هر پنجره به صورت زیر تعریف می‌شود.

$$SC(t)=\frac{\sum_{k=1}^{N-1} kS_t[k]}{\sum_{k=1}^{N-1} S_t[k]} \qquad (۳-۲)$$

در فرمول ۲-۳ به ازای پنجره t ام،$S_t[k]$ انرژی فرکانس k ام و N طول تبدیل فوریه گسسته است.

در این پایان‌نامه برای استفاده از این ویژگی همانند روش قبل، ابتدا باید پیش‌پردازش‌های بیان شده در ۳-۲-۲ را بر روی داده‌ها اعمال کنیم. سپس سیگنال را به پنجره‌هایی تقسیم کرده و تبدیل فوریه هر یک را به دست آوریم. در نهایت با استفاده از فرمول ۲-۳، در هر پنجره نقطه‌ی تعادل طیف را تعیین کنیم و این اعداد به دست آمده را کنار یکدیگر قرار داده و بردار ویژگی را بسازیم. سپس با استفاده از این ویژگی و به کار گرفتن دسته‌بند مناسب به طبقه‌بندی داده‌ها بپردازیم. در استخراج این ویژگی همانند روش نرخ عبور از صفر دو پارامتر حضور دارند که در تعیین درصد خروجی موثرند. این دو پارامتر همان سایز هر پنجره و گام می‌باشند. در بخش ۴-۴ بهترین نرخ تشخیص به دست آمده با به کارگیری این ویژگی و با استفاده از بهینه‌ترین مقادیر طول پنجره و گام، گزارش شده است.

۲-۲-۱-۴- نقطه‌ی قطع طیف

این ویژگی تعیین‌کننده‌ی فرکانسی است که مقدار معینی از انرژی طیف فرکانسی در زیر آن قرار گرفته است و در واقع مرتبط با چولگی[1] طیف است و در سیگنال‌هایی که دارای محدوده‌ی فرکانسی متفاوت می‌باشند، متغیر است. برای محاسبه‌ی ویژگی نقطه‌ی قطع طیف، انرژی تمامی فرکانس‌های موجود در طیف فرکانسی تا رسیدن به یک مقدار آستانه‌ی تعیین شده با یکدیگر جمع می‌شود. فرکانسی که در آن، این مجموع به عددی بیشتر از حد آستانه‌ی تعیین شده رسید، به عنوان نقطه‌ی قطع انتخاب می‌شود. فرمولی که در محاسبه‌ی این فرکانس به کار برده می‌شود به صورت زیر تعریف شده است.

[1] Skewness

$$SRF(t) = \max\left\{ K \left| \sum_{k=0}^{K} |S_t(k)| \right|^2 < TH \sum_{k=0}^{N} |S_t(k)|^2 \right\} \quad (۴-۲)$$

در فرمول ۲-۴ حد آستانه (TH)، مقداری بین صفر و یک دارد و بسته به کاربرد معمولا بین ۰.۸۵ تا ۰.۹۵ در نظر گرفته می‌شود. در این پایان‌نامه مقدار بهینه‌ی آن با سعی و خطا مشخص شده است.

این ویژگی در تمییز صدا[1] و بی‌صدا[2] بسیار مفید است زیرا در بی‌صدا، بخش عظیمی از انرژی طیف فرکانسی در محدوده‌ی فرکانس‌های بالا قرار گرفته‌اند در حالی که در صدا بیشتر انرژی سیگنال در فرکانس‌های پایین حضور دارند. در نتیجه این ویژگی می‌تواند تمییزدهنده این دو نوع سیگنال باشد. به طور کلی اگر سیگنال‌هایی داشته باشیم که محدوده‌ی فرکانسی آن‌ها متفاوت باشد نقطه‌ی قطع می‌تواند آن‌ها را به خوبی از یکدیگر تمییز دهد.

در این پایان‌نامه این ویژگی، از داده‌های موجود در پایگاه‌داده استخراج و نتایج به دست آمده با استفاده از آن، با ویژگی پیشنهادی مقایسه شده است. برای استخراج این ویژگی ابتدا بر روی تمامی داده-های موجود در پایگاه‌داده پیش‌پردازش‌های بیان شده در ۳-۲-۲ را اعمال می‌کنیم. سپس سیگنال را به پنجره‌هایی تقسیم می‌کنیم. اینکه سیگنال را به چند پنجره تقسیم کنیم و یا میزان همپوشانی هر کدام از آن‌ها چقدر باشد، دو پارامتر مهم هستند که در میزان تشخیص سیستم موثرند. این دو پارامتر همانند روش‌های بیان شده‌ی قبل با سعی و خطا مشخص و مقدار بهینه‌ی آن‌ها تعیین شده‌اند. پس از تقسیم-بندی سیگنال ورودی به مجموعه‌ای از پنجره‌ها، از هر کدام از آن‌ها تبدیل فوریه می‌گیریم و مقدار انرژی هرکدام از فرکانس‌های موجود در هر پنجره را تا رسیدن به یک آستانه از پیش تعیین شده با هم جمع می‌کنیم. بدین ترتیب از هر پنجره یک ویژگی استخراج کرده‌ایم. ویژگی‌های به دست آمده از هر کدام از

[1] voice
[2] unvoice

پنجره‌ها را برای به دست آوردن بردار ویژگی کنار هم قرار می‌دهیم. این بردار ویژگی به دسته‌بند مناسب داده شده و عمل طبقه‌بندی را انجام می‌دهیم. پیشتر بیان شد که حین استخراج نقطه‌ی قطع، حد آستانه‌ای نیاز است که بیانگر این است که قبل از فرکانس تعیین شده چند درصد انرژی سیگنال اتفاق افتاده است. در این پایان‌نامه این حد آستانه با سعی و خطا بدین صورت تعیین می‌شود که به ازای انتخاب اعداد مختلف درصد تمییز درست کلاس‌ها بررسی و مقداری از این آستانه که بیشترین تشخیص درست با استفاده از آن به دست آمده است به عنوان مقدار بهینه انتخاب می‌شود. بهترین نتایج به دست آمده با استفاده از این ویژگی در بخش ۴-۴ آمده است.

۲-۲-۱-۵- شار طیفی

این ویژگی، نرم دو یا همان فاصله اقلیدسی دو بردار حاوی تفاضل مقدار انرژی دو پنجره متوالی در حوزه‌ی فرکانس است. شار طیفی تغییرات ناگهانی در شکل طیف را در طول زمان بیان می‌کند. سیگنال-هایی که دارای طیف با تغییرات آرام هستند، معمولا شار طیفی کمی دارند در حالی که سیگنال هایی که تغییرات ناگهانی و سریع در آن‌ها اتفاق می‌افتد، دارای شار طیفی بزرگی هستند. در واقع شار طیفی معیاری از اینکه قدرت طیف یک سیگنال با چه سرعتی تغییر می‌کند، ارائه می‌دهد و از مقایسه قدرت طیف یک پنجره و پنجره ماقبل آن به دست می‌آید. به بیان دقیق‌تر این ویژگی با محاسبه‌ی نرم دوی هر دو پنجره متوالی در حوزه‌ی فرکانس به دست می‌آید. توضیح فوق به صورت زیر فرموله شده است:

$$SF = \sum_n \left\| S[n] \right| - \left| S[n+1] \right\| \qquad (۵-۲)$$

در فرمول ۲-۵، n تعداد پنجره ها و S نمایانگر تبدیل فوریه هر پنجره می‌باشد.

در این پایان‌نامه شار طیفی را از هر کدام از سیگنال‌های موجود در پایگاه‌داده استخراج می‌کنیم. برای این کار ابتدا هر کدام از سیگنال‌ها را به پنجره‌هایی تقسیم کردیم. در تعیین سایز پنجره‌ها و گام برای انتخاب

ابتدای پنجره بعدی به طریق بیان شده در استخراج ویژگی‌های ذکرشده در بالا عمل می‌شود. پس از پنجره‌گذاری سیگنال، از هر کدام از آن‌ها به طور جداگانه تبدیل فوریه گرفته و سپس از هرکدام از بردارهای حاوی تبدیل فوریه پنجره‌های متوالی، نرم دوم گرفتیم و بدین ترتیب بردار ویژگی را ساخته می‌شود و به منظور طبقه‌بندی داده‌ها آن را به یک دسته‌بند مناسب می‌دهیم. نرخ تشخیص درست سیستم با استفاده از این ویژگی را در بخش ۴-۴ ارائه می‌کنیم.

۲-۲-۲- دسته‌بندی

به طور کلی دسته‌بندها با توجه به ساختارشان، به دو دسته‌ی یادگیری با ناظر و یادگیری بدون ناظر تقسیم می‌شوند. در یادگیری با ناظر، مجموعه‌ای از مشاهدات یا همان داده‌های آموزش، با دسته‌ای که به آن تعلق دارند، برچسب گذاری شده‌اند. سپس با استفاده از این مشاهدات برای هر دسته، مدل مربوط به آن ایجاد و داده‌های آزمایش با توجه به مدل‌های ساخته شده، دسته‌بندی می‌شوند.

در یادگیری بدون ناظر داده‌های آموزش وجود ندارد و ساخت مدل و دسته‌بندی داده‌ها تنها با استفاده از داده‌های آزمایش صورت می‌گیرد. از جمله روش‌های یادگیری بدون ناظر می‌توان به خوشه‌بندی k میانگین[1]، اشاره کرد.

الگوریتم‌های دسته‌بندی باناظر را می‌توان به سه دسته تقسیم نمود:

۱. دسته‌بندهای مبتنی بر قاعده[2]: این دسته‌بندها با استنباط یک مجموعه از قواعد از اصوات که از پیش دسته‌بندی شده‌اند، یادگیری را انجام می‌دهند. الگوریتم ریپر[3] از جمله دسته‌بندهای مبتنی بر قاعده است [۲۸]. قواعد تصمیم ممکن است به صورت درخت تصمیم مانند C4.5 در مرجع [۲۹] باشند.

[1] K-means Clustering
[2] Rule Based Classifiers
[3] Ripper

۲. دسته‌بندهای خطی[1]: در این دسته‌بندها، برای هر دسته یک نمایه محاسبه می‌شود که در واقع برداری از اوزان، براساس فرکانس و احتمالات حضور مشخصه‌ای خاص است. برای هر دسته، امتیاز بر اساس مشخصات دسته محاسبه می‌شود. از دسته‌بندهای این دسته می‌توان به دسته‌بند بیز اشاره نمود که براساس تخمین شرایط احتمالی عمل می‌کند [۳۰]. ماشین‌های بردار پشتیبان یک دسته‌بند خطی بهینه را با استفاده از انتقال به فضای مشخصه بدست می‌آورد[۳۱]، [۳۲]، [۳۳]. این گروه از دسته‌بندها همچنین شامل الگوریتم‌های یادگیری اکتشافی از هوش مصنوعی مانند پرسپترون می‌باشند که در آن اوزان از راهی پیچیده‌تر بدست می‌آیند[۳۴].

۳. دسته‌بندهای مبتنی بر مثال[2]: این دسته‌بندها یک داده‌ی جدید را با پیدا کردن K تا از داده‌های نزدیکتر به آن در مجموعه آموزشی، دسته‌بندی می‌کنند و با رأی‌گیری، آن را به دسته‌ی نزدیکتر تخصیص می‌دهند[۳۵].

۲-۲-۳- مرور بر کارهای پیشین

در این بخش مروری بر مقالاتی که در زمینه‌ی تشخیص اصوات ضربه‌ای منتشر شده‌اند، ارائه می‌کنیم. ابتدا باید یادآوری کرد به دلیل نبود پایگاه‌داده‌ی استاندارد در تحقیقات انجام شده در زمینه‌ی تشخیص صدای محیط، معمولا از ویژگی MFCC برای مقایسه‌ی روش‌های پیشنهادی استفاده می‌شود.

در [۱۹] نویسنده، از فیلتربانک دارای ۵، ۱۰ و ۲۰ ضریب که متناسب با انرژی سیگنال در باندهای فرکانسی متفاوت می‌باشد، استفاده کرده است. در واقع فضای ۰ تا ۲۰ کیلوهرتز به ۵،۱۰ یا ۲۰ باند تقسیم شده است. همچنین نویسنده بر روی ویژگی فیلتربانک دارای۲۰ ضریب، تجزیه و تحلیل مولفه‌های اصلی[3] را اعمال نموده و ابعاد ویژگی‌ها را به ۵ یا ۱۰ کاهش داده و آن‌ها را به عنوان ویژگی جدید در نظر

[1] Linear Classifiers
[2] Example Based Classifier
[3] Principal Component Analysis (PCA)

گرفته است. در این کار از دسته‌بند مدل مخلوط گوسی استفاده شده است. بهترین نرخ تشخیص سیستم، با استفاده از PCA دارای ۱۰ ضریب، به دست آمده و حدود ۹۴٪ بوده است که عملکرد بهتری نسبت به MFCC از خود ارائه داده است. قابل ذکر است که پایگاه‌داده‌ی استفاده شده در این مقاله دارای ۶ کلاس بوده است.

پایگاه‌داده‌ی استفاده شده در [۳۶] دارای ۷ کلاس از صداهای محیطی است. در این کار بخش شناسایی و تشخیص به طور مجزا عمل می‌کنند. در بخش شناسایی از سه روش همبستگی متقابل [۱]،روش کشفی مبتنی بر پیش‌بینی انرژی [۲] و فیلتر موجک [۳] استفاده می‌شود. بهترین عملکرد مربوط به فیلتر موجک بوده و ضعیف‌ترین نتایج با استفاده از پیش‌بینی انرژی به دست آمده‌است. در بخش تشخیص، ویژگی‌های ایستایی همچون MFCC و مشتق اول و دوم آن، نرخ عبور از صفر، نقطه قطع طیف، نقطه‌ی تعادل طیف، LPCC ، انرژی و ترکیب‌های متفاوت آن‌ها به دسته‌بند مدل مخلوط گوسی اعمال شده اند. کمترین نرخ خطا زمانی حاصل می‌شود که از ویژگی‌های نرخ عبور از صفر، نقطه قطع طیف، نقطه‌ی تعادل طیفی، MFCC دارای ۱۶ ضریب و انرژی در کنار هم استفاده شده و مقدار خطا ۱۵٪ بوده است.

در [۳۷] توصیف‌کننده‌های سطح پایین صوت همچون MPEG-7، نقطه‌ی تعادل طیفی، پراکندگی طیف [۴]و همواری طیفی به عنوان ویژگی به دسته‌بندی که از ترکیب دو دسته‌بند ماشین بردار پشتیبان و k- نزدیک‌ترین همسایه تشکیل شده، داده می‌شوند. آزمایشات بر روی پایگاه‌داده‌ای متشکل از ۱۲ کلاس انجام شده و بهترین درصد تشخیص ۸۵.۱٪ گزارش شده است.

[1] cross-correlation
[2] energy prediction based detection
[3] wavelet filtering based detection
[4] spectrum spread

در [۳۸] نویسنده توصیف‌کننده‌ی سطح پایین MPEG-7 را در کنار MFCC به کار برده است. توصیف‌کننده‌های MPEG-7 ابتدا توسط نسبت جداکنندگی فیشر رتبه‌بندی شده، سپس PCA بر روی ۳۰ توصیف‌کننده‌ای که بیشترین امتیاز را دارند اعمال می شود و ۱۳ ویژگی نهایی انتخاب می‌شوند. این ۱۳ ویژگی در کنار MFCC قرار می‌گیرند و بردار ویژگی نهایی ساخته می‌شود. در این پژوهش از دسته-بند مدل مخلوط گوسی استفاده شده است. پایگاه‌داده‌ی این کار شامل ۱۰ کلاس از صداهای محیطی می‌باشد. سیستم پیشنهادی عملکرد بهتری در مقایسه با زمانی که MFCC یا MPEG-7 به تنهایی به کار برده شده‌اند داشته و بهترین نرخ تشخیص این سیستم ۹۶٪ می‌باشد.

در [۱۷] ویژگی‌های LPC، MFCC دارای ۱۶ ضریب و همبستگی فایل‌های صوتی در مجموعه‌ای از غالب‌ها [1]، از پایگاه‌داده‌ای حاوی ۴ دسته، استخراج شده‌اند. زمانی که از ویژگی همبستگی استفاده شده، نمونه‌ها به کلاسی که بیشترین شباهت[2] را دارند، نسبت داده می‌شوند و زمانی که سایر ویژگی‌ها برای دسته‌بندی به کار رفته‌اند از دسته‌بند مدل مخفی مارکوف استفاده شده است. MFCC و LPC در محیط بدون نویز خوب عمل می‌کنند اما در محیط نویزدار ویژگی‌های همبستگی نتایج بهتری ارائه می‌دهند.

در [۳۹] سیستمی برای شناسایی رخدادهای شنیداری در محیط‌های واقعی ارائه شده است. رخدادها با شبکه‌ای از مدل مخفی مارکوف مدل شده‌اند و از ویژگی MFCC و مشتق اول و دوم آن استفاده شده است. دقت تشخیص به ازای ۶۱ کلاس ۲۴٪ بوده است. طبیعی است با افزایش تعداد کلاس‌ها، نرخ تشخیص کاهش یابد.

[1] correlation against audio files in a set of templates
[2] correlation

[۴۰] استفاده از ویژگی پیش‌بینی خطی بر اساس برانگیختگی کد[1] را پیشنهاد می‌دهد. نویسنده در این کار روش جدیدی برای استخراج ویژگی از جریان بیت‌های CELP ارائه نمود. چون CELP از یک دفترچه کد[2] ثابت برای برانگیختن مدل فیلتر منبع [3] ، استفاده می‌کند، عملکرد قوی‌تری نسبت به LPC دارد. با استفاده از این ویژگی‌ها و دسته‌بند بیزین عملکرد سیستم نسبت به MFCC ، ۹٪ بهبود داشته و با استفاده از ترکیب ویژگی پیشنهادی و MFCC عملکرد تا ۹۵.۲٪ ارتقا پیدا کرده است.

[۴۱] سعی کرد تغییرات زمانی در بین زیرفریم‌های یک سیگنال را با استفاده از مجموعه‌ی جدیدی از ویژگی‌ها با نام ویژگی‌های پویای طیف[4]، مد نظر قرار دهد. این ویژگی بدین صورت استخراج می‌شود: ابتدا از هر زیرفریم ویژگی‌هایی مانند MFCC، LPC و سایر ویژگی‌های مد نظر استخراج می‌شوند و درون یک بردار قرار می‌گیرند. این بردارها درون ماتریسی، کنار هم جای می‌گیرند. سپس بر روی هر سطر این ماتریس سه تبدیل فوریه، فیلتر بانک لگاریتمی و تبدیل کسینوسی به ترتیب اعمال می‌شوند و ویژگی‌های نهایی را می‌سازند. در این مقاله بیان شده است که ترکیب MFCC و مشتق اول آن بیشترین عملکرد را بین ویژگی‌های ایستا دارند و با افزودن ویژگی‌های جدید مانند نرخ عبور از صفر،LPC و انرژی باند[5] به آن‌ها عملکرد بهبود نمی‌یابد اما ویژگی‌های SDF با استفاده از دسته‌بند مدل مخلوط گوسی و ماشین بردار پشتیبان بهبود ۱۰ الی ۱۵ درصدی بین ویژگی‌های ایستا داشته است.

در [۴۲] نویسنده مجموعه‌ی جدیدی از ویژگی‌ها به نام تابع خود هم‌بسته باند باریک [6]را ارائه داد. برای محاسبه‌ی این ویژگی، ابتدا سیگنال از یک بانک فیلتر با ۴۸ باند که فرکانس مرکزی آن برابر مقیاس

[1] code excited linear prediction(CELP)
[2] code-book
[3] source-filter model
[4] spectral dynamic features(SDF)
[5] band-energy
[6] Narrow-Band Auto Correlation Function (NB-ACF)

مل [1] ، تنظیم شده است، عبور می‌کند. سپس ACF سیگنال فیلتر شده در هر باند محاسبه می‌شود. در این کار چهار مورد زیر به عنوان ویژگی در نظر گرفته شده‌اند:

- فشار صدای دریافتی در هر باند
- تاخیر اولین پیک مثبت که نشان‌دهنده‌ی فرکانس غالب در هر باند است.
- ACF نرمال‌شده‌ی اولین پیک مثبت که مرتبط با متناوب بودن سیگنال است و درکی از زیر و بمی سیگنال فیلتر شده در هر باند ارائه می‌دهد.
- طول موثر پوشش[2] ACF نرمال شده (مدت زمانی که طول می‌کشد تا ACF نرمال شده 10dB از مقدار بیشینه خود کمتر شود.)

ویژگی‌های فوق به دسته‌بند k- نزدیک‌ترین همسایه و ماشین بردار پشتیبان داده شدند. نتایج به دست آمده حاکی از آن است که این ویژگی‌ها عملکرد بهتری نسبت به MFCC و تبدیل فوریه گسسته دارند.

در [۴۳] نویسنده از شناساگر دو مرحله‌ای سطح انرژی به عنوان ویژگی استفاده کرده است. دلیل استفاده از این ویژگی این است که در حضور نویز ویژگی MFCC نمی‌تواند دسته‌بندی اصوات را با دقت مطلوب انجام دهد. در این پژوهش ابتدا سیگنال به پنجره‌هایی تقسیم می‌شود سپس با استفاده ازشناساگر انرژی در حوزه‌ی زمان[3] حضور سیگنال صوت بررسی شده و در صورت حضور صوت مجددا توسط شناساگر انرژی در حوزه‌ی فرکانس [4]وجود صوت در آن پنجره تایید می‌شود و در نهایت از پنجره حاوی صوت ضرایب MFCC استخراج می‌شوند. نتایج، بهبود ۴۱درصدی نرخ تشخیص در حضور نویز و با استفاده از این ویژگی‌ها و دسته‌بند ماشین بردار پشتیبان در مقایسه با MFCC را گزارش می‌دهند.

[1] Mel-scale
[2] envelope
[3] time-domain energy detection
[4] frequency-domain energy detection

پایگاه‌داده‌ی استفاده شده در [۴۴] شامل ۵ دسته از صداهای محیطی است. ویژگی‌های استخراج شده در این کار شامل قله‌ی طیفی [۱]، کاهش طیفی [۲]، شیب طیفی، همواری طیفی و چولگی طیفی می‌باشند. با استفاده از ویژگی‌های فوق و شبکه‌ی عصبی انتشار به عقب دقت دسته‌بندی ۹۱٪ گزارش شده است.

همانطور که پیشتر بیان شد تکنیک‌های تشخیص صداهای محیط به دو دسته ایستا و غیرایستا تقسیم می‌شوند. تکنیک‌های ایستا مورد بررسی قرار گرفت. حال به توضیح تکنیک‌های غیرایستا و مقالات مرتبط با آن می‌پردازیم. اولین دسته از این تکنیک‌ها، روش‌های مبتنی بر موجک می‌باشد. این دسته شامل روش‌هایی است که در بخش استخراج ویژگی از تبدیلاتی همچون تبدیل فوریه سریع [۳]، تبدیل فوریه زمان کوتاه [۴]، تبدیل موجک گسسته [۵] و تبدیل موجک پیوسته [۶] استفاده کرده‌اند.

در [۲۶] مقایسه‌ای بین روش‌های ایستا و روش‌های غیرایستای مبتنی بر موجک و با استفاده از دسته-بندهای مختلف ارائه شده است. در این کار روش‌های معمول مثل MFCC، LPC و PLP با روش‌هایی مثل تبدیل فوریه‌ی زمان کوتاه، تبدیل موجک گسسته، تبدیل موجک پیوسته و توزیع وینر- ویل [۷] مقایسه شدند. دسته‌بندهای چرخش پویای زمان [۸]، رقمی‌سازی بردار خطی [۹]، نگاشت‌های خود سازمان ده [۱۰]، تخمین بیشترین شباهت [۱۱]، ماشین بردار پشتیبان و مدل مخفی مارکوف مورد ارزیابی قرار گرفتند. بهترین نتیجه با استفاده از دسته‌بند DTW و با ویژگی تبدیل فوریه پیوسته به دست آمده است. در این حالت نرخ تشخیص ۷۰٪ گزارش شده است.

---

[1] spectral crest
[2] spectral decrease
[3] FFT
[4] STFT
[5] DWT
[6] CWT
[7] wignor-ville distribution(WVD)
[8] Dynamic Time Warping(DTW)
[9] LVQ
[10] self-organizing maps(SOM)
[11] maximum likelihood estimation

پایگاه‌داده‌ی [۱۸] شامل ۹ کلاس است. در این کار ویژگی‌های مبتنی بر تبدیل موجک در کنار MFCC مورد استفاده قرار گرفتند و از دسته‌بند ماشین بردار پشتیبان استفاده شد. استفاده از این دو دسته ویژگی در کنار هم نرخ تشخیص را به۹۳٪ رساند که نسبت به استفاده از MFCC یا تبدیل موجک به تنهایی نرخ تشخیص بهبود یافته است.

در [۴۵] نویسندگان از ویژگی‌های تبدیل چیرپلت گسسته[1] و تبدیل کرولت گسسته [2]به همراه ویژگی-های متداولی چون MFCC و ZCR استفاده کردند. آن‌ها گزارش دادند که نرخ تشخیص زمانی که از تمام این ویژگی‌ها در کنار یکدیگر استفاده شده به ۸۶٪ رسیده است، در صورتی که MFCC و ZCR روی همان پایگاه‌داده دقتی حدود ۷۴٪ داشته‌اند.

در [۴۶] از تبدیل موجک با موجک گاماتون[3] استفاده شده است و نتایج حاصل از آن با تبدیل موجک با استفاده از موجک دابیشز مقایسه شده است. در این پژوهش از دسته‌بند ماشین بردار پشتیبان استفاده شده است. نتایج حاکی از آن است که این موجک عملکرد بهتری نسبت به سایر موجک‌ها، هم در محیط-های بدون نویز و هم در حضور نویز داشته است. علاوه بر این ترکیب این دو دسته موجک باعث بهبود نرخ تشخیص می‌شود.

دومین دسته از تکنیک‌های غیرایستا در زمینه‌ی تشخیص صداهای محیط روش‌های مبتنی بر نمایش پراکنده می‌باشند. در ادامه مقالات منتشر شده در این زمینه را بررسی می‌کنیم.

در [۴۷] نویسنده دسته‌ای جدید از ویژگی‌ها به نام ردگیری انطباق را، برای دسته‌بندی صداهای محیط پیشنهاد داد. این روش دیکشنری حاوی اتم‌ها را برای به دست آوردن مجموعه‌ای انعطاف‌پذیر از

[1] Discrete Chirplet Transform(DChT)
[2] Discrete Curvelet Transform (DCuT)
[3] Gammatone

ویژگی‌ها بهینه می‌کند. در این مقاله از ویژگی ردگیری انطباق گابور[1] و دسته‌بندهای k- نزدیک‌ترین همسایه و مدل مخلوط گوسی استفاده شده و مقایسه‌ای بین نتایج حاصل از این روش و MFCC صورت گرفته است. ویژگی پیشنهاد شده عملکرد بهتری نسبت به MFCC دارد اما ترکیب آن‌ها با هم نرخ تشخیص را بالاتر می‌برد. لازم به ذکر است که روش پیشنهادی کلاس‌هایی را که دارای خواص ایستا هستند بسیار خوب از یکدیگر تفکیک می‌کند اما کلاس‌هایی را که خواص غیرایستا دارند، نمی‌تواند به خوبی دسته‌بندی کند.

در [۴۸] برای بهبود عملکرد ویژگی ردگیری انطباق گابور، تغییرات زیادی پیشنهاد شد. از جمله تغییرات این بود که به جای دیکشنری ثابت، یک دیکشنری وابسته به سیگنال ساخته شد و ردگیری انطباق متعامد [2] جایگزین ردگیری انطباق پایه[3] شده و در نهایت از میانگین وزن‌دار و واریانس نمونه‌ها استفاده شد. با این تغییرات، عملکرد بسیار افزایش یافت و با ترکیب این ویژگی جدید با MFCC نرخ تشخیص به ازای ۱۴ کلاس به ۹۵.۵٪ رسید.

در [۴۹] نویسنده مقایسه‌ای بین سه ویژگی موجک هار، تبدیل فوریه و ردگیری انطباق گابور ارائه داد. در این کار به جای استفاده از میانگین و انحراف معیار پارامترهای مقیاس و فرکانس اتم‌های ردگیری انطباق گابور، آن‌ها را برای ساخت یک بردار ویژگی با هم الحاق کردند. نتایج به دست آمده با استفاده از این دسته ویژگی و دسته‌بند مدل مخفی مارکوف حاکی از آن است که ویژگی‌های ردگیری انطباق گابور بهترین عملکرد را دارند.

[1] MP-Gabor
[2] orthogonal MP(OMP)
[3] basis MP(BMP)

هدف در [۱۱] دست یافتن به سیستم شنوایی ربات است که قابلیت تشخیص صداهای محیط را دارد و از آن می‌توان برای تعامل انسان و ربات استفاده کرد. در این کار از دو دسته‌بند مدل مخفی مارکوف و مدل مخلوط گوسی استفاده شده است. ویژگی‌های به کار برده شده شامل ردگیری انطباق گابور و MFCC می‌باشند و عملکرد این دو دسته ویژگی و با استفاده از دو دسته‌بند فوق بررسی شد. بهترین نتایج زمانی حاصل می‌شود که ویژگی ردگیری انطباق به مدل مخفی مارکوف اعمال شده بود. استفاده از ویژگی ردگیری انطباق و دسته‌بند مدل مخلوط گوسی نتایجی بهتر از MFCC ارائه داد.

آخرین دسته از تکنیک‌های غیرایستا در تشخیص صداهای محیط، روش‌های مبتنی بر طیف توان می-باشند. طیف توان اطلاعات مفیدی در مورد انرژی سیگنال در زمان و فرکانس محلی شده فراهم می‌کند. این طیف یک ابزار شهودی قدرتمند برای استخراج ویژگی‌های انتقال و تغییر صداهای محیط است. در این قسمت مروری بر مقالاتی که با استفاده از این ابزار به استخراج ویژگی پرداخته‌اند، ارائه می‌کنیم.

پایگاه‌داده‌ی به کار برده شده در [۵۰] شامل ۶ کلاس از صداهای ضربه‌ای می‌باشد. در این کار دو بخش شناسایی و تشخیص مستقل از هم عمل می‌کنند. در بخش شناسایی حضور یا عدم حضور صدای ضربه‌ای در سیگنال ورودی با استفاده از فیلتر میانه مورد بررسی قرار می‌گیرد و در صورت حضور، بخش تشخیص وارد عمل می‌شود. سیگنال ورودی به پنجره‌های متوالی تقسیم می‌شود. در هر پنجره انرژی N باند طیفی که محدوده 0kHz تا 20kHz را پوشش می‌دهد، محاسبه می‌شود. در این حالت هر بردار ویژگی دارای N پارامتر است که توزیع انرژی هر طیف را برای هر پنجره نشان می‌دهد. با استفاده از دسته‌بندهای مدل مخفی مارکوف و مدل مخلوط گوسی در محیط بدون نویز، نرخ تشخیص ۹۸٪ گزارش شده است.

در [۵۱] نویسنده از تحلیل زمان- فرکانس برای استخراج ویژگی استفاده کرده است. در هر ثانیه، طیف توان سیگنال با تفکیک زمانی ۵۰ میلی ثانیه محاسبه می‌شود. دامنه‌ی فرکانسی ۰ تا ۲۰ کیلوهرتز به ۵ باند ۴ کیلوهرتزی به ازای هر پنجره زمانی تقسیم می‌شود. در این تحقیق، کارایی دو دسته‌بند بیزین و شبکه عصبی مورد بررسی قرار گرفتند. تعداد دسته‌های پایگاه داده برابر با ۳ بوده است و نرخ تشخیص سیستم به ازای دسته‌بند بیزین ۹۸٪ و شبکه عصبی ۹۹٪ بوده است.

در [۵۲] یک رویکرد برای شناسایی و مدل‌سازی رخدادهای صوتی که به طور مستقیم محتوای زمانی را توصیف می‌کند، ارائه شده است. این رویکرد با استفاده از تجزیه غیر-منفی ماتریس [1]عمل می‌کند. NMF برای تجزیه کردن داده بر مبنای بخش مفید است. در واقع NMF یک الگوریتم برای توصیف داده به صورت حاصل‌ضرب مجموعه‌ای از پایه‌ها و فعالیت‌هاست که شامل مقادیری غیر صفر هستند. پایه‌ها، شامل قطعه‌های زمانی-طیفی می‌باشند. در این کار برای استخراج ویژگی، از سیگنال طیف توانی گرفته شده و سپس محور فرکانس به مقیاس فرکانس مل، تغییر داده می‌شود. سپس طیف تمام داده‌های آموزش با هم الحاق شده و NMF با ۲۰ قطعه پایه بر روی آن اعمال می‌شود. نتایج حاکی از آن است که در شرایط بدون نویز MFCC عملکرد بهتری نسبت به روش پیشنهادی دارد اما با افزایش نویز به دلیل کاهش کارایی MFCC عملکرد روش پیشنهادی بهتر می‌شود. با ترکیب MFCC و این روش می‌توان بهترین نتایج را گرفت.

در [۵۳] روشی قوی برای دسته‌بندی صداهای محیط بر مبنای متد تخصیص مجدد [2]و فیلتر گابور لگاریتمی[3] ارائه شده است. روش پیشنهادی از ۱۲ فیلتر گابور لگاریتمی که به ۳ بخش طیفی اعمال شده-اند، استفاده می‌کند. متد تخصیص مجدد نتایج دسته‌بندی را بهبود داده و عنصر اصلی در افزایش کارایی

[1] non-negative matrix factorization (NMF)
[2] reassignment
[3] log-Gabor

نسبت به روش‌های قبلی بوده است. در واقع این مقاله با ویژگی‌های قوی که توسط دسته‌بند ماشین بردار پشتیبان مورد استفاده قرار می‌گیرد و باعث شده سیستم مستقل از شرایط محیطی ضبط صدا، قوی عمل کند، سروکار دارد. در این پژوهش نرخ تشخیص به ازای ۱۰ کلاس ۹۰٪ گزارش شده است.

در [۵۴] نویسندگان از پردازش بر مبنای زیرفریم، برای محاسبه طیف سیگنال استفاده کردند. آن‌ها طیف‌نگار فوریه [۱] زیرفریم‌ها را با هم الحاق نمودند و k- نزدیک‌ترین همسایه و شبکه عصبی انتشار به جلو را برای دسته‌بندی به کار بردند. آن‌ها مطالعات بیشتری برای انتخاب پارامترهای طول طیف، طول سیگنال صوتی و نرخ نمونه‌برداری انجام دادند. ویژگی پیشنهادی با MFCC، LPC و ردگیری انطباق گابور مقایسه شد. ویژگی پیشنهادی عملکرد بهتری نسبت به MFCC و LPC دارد و نتایج آن با ردگیری انطباق گابور قابل مقایسه بود. بهترین نتایج با استفاده از دسته‌بند نزدیک‌ترین همسایه و کنار هم قرار گرفتن ویژگی‌های طیف قدرت، LPC ردگیری انطباق گابور به دست آمدند. نرخ تشخیص در این حالت به ازای ۲۰ کلاس ۹۵٪ گزارش شده است.

در [۵۵] روش‌های جدیدی برای استخراج ویژگی با استفاده از طیف توان ارائه شده است. این روش‌ها بر مبنای دامنه‌ی دیداری پیشنهاد شدند. در این مقاله سه روش برای استخراج ویژگی‌ها بیان شده است. در روش اول بر هر طیف، یک فیلتر گابور لگاریتمی اعمال می‌شود. در روش دوم بر هر طیف یک بانک فیلتر، متشکل از ۱۲فیلتر گابور لگاریتمی اعمال می‌گردد. در روش سوم هر طیف به سه بخش تقسیم شده و بر روی هر بخش آن روش دوم اجرا می‌گردد. این روش‌ها بر روی پایگاه‌داده‌ای متشکل از ۱۰ کلاس آزمایش شدند. نتایج نشان داد که روش دوم بهترین عملکرد را داشته و در این حالت با استفاده از دسته‌بند ماشین بردار پشتیبان نرخ تشخیص میانگین ۸۹٪ گزارش شده است.

[1] Fourier spectrogram

۲-۳- نتیجه‌گیری

روش‌های تشخیص صدای محیط به دو دسته ایستا و غیرایستا تقسیم می‌شوند. در بین روش‌های ایستا بهترین عملکرد را ویژگی‌های طیفی دارند که محاسبه‌ی آن‌ها راحت است؛ اما با استفاده از آن‌ها محدودیت‌هایی در مدل کردن صداهای غیرایستا وجود دارد. روش‌های غیرایستا به سه دسته‌ی ویژگی-های مبتنی بر موجک، مبتنی بر نمایش پراکنده و طیف توان تقسیم می‌شوند. عملکرد ویژگی‌های مبتنی بر موجک با ویژگی‌های ایستا قابل مقایسه است اما دو دسته‌ی دیگر ویژگی‌های غیرایستا معمولا عملکرد بهتری دارند. غالبا ویژگی‌های MFCC با یک یا دو ویژگی دیگر ترکیب می‌شوند تا عملکرد بهتری حاصل شود. با اینکه روش‌های غیرایستا عملکرد بهتری دارند اما هزینه‌ی محاسبات آن‌ها بالاست، پس استفاده از آن‌ها در بسیاری از کاربردهایی که نیاز به سرعت بالای تشخیص دارند، توصیه نمی‌شود.

فصل سوم

پیاده‌سازی سیستم پیشنهادی

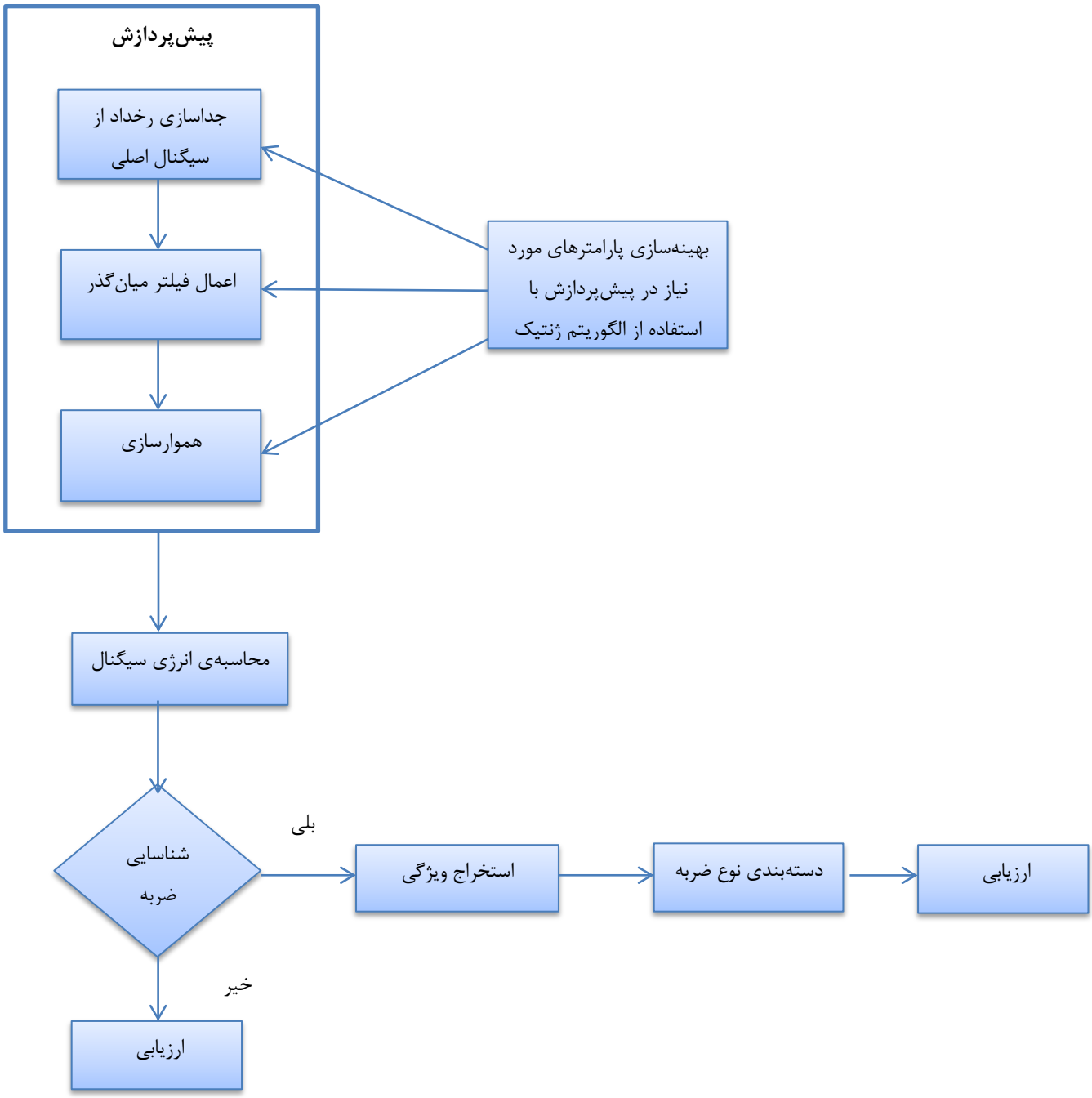
۳-۱- مقدمه

در این فصل به شرح اجزای مختلف سیستم پیشنهادی می‌پردازیم. همانطور که در فصل دوم بیان شد، هر سیستم هوشمند شناسایی اصوات محیط دارای سه بخش پیش‌پردازش، استخراج ویژگی و طبقه‌بندی می‌باشد. در ادامه‌ی این فصل هر کدام از بخش‌های فوق را با جزئیات بیشتر توضیح خواهیم داد.

۳-۲- سیستم پیشنهادی

در بخش ۱-۴ چند نمونه از سیگنال‌های به کار رفته در این پایان‌نامه و همچنین توضیحاتی راجع به تفاوت در ساختار و رفتار اصوات موجود در دسته‌های مختلف ارائه شد. در این تحقیق از تفاوت رفتار اصوات دسته‌های مختلف برای دسته‌بندی آن‌ها استفاده شده است. در واقع ویژگی جدیدی، بر مبنای تفاوت در نحوه‌ی رسیدن سیگنال‌ها از مقدار کمینه به بیشینه، معرفی شده است. سایر روش‌های موجود در استخراج ویژگی، از تعداد ویژگی‌های زیادی برای دسته‌بندی استفاده می‌کنند اما روش پیشنهادی در این پایان‌نامه تنها از ۱۰ ویژگی برای دسته‌بندی استفاده می‌کند. مهم‌ترین دلیل پیشنهاد این روش، کم بودن تعداد ویژگی‌های به کار رفته در آن و در نتیجه مناسب بودن برای کاربردهای بلادرنگ، به دلیل کم بودن بار محاسباتی و بالا بودن سرعت سیستم پیشنهادی می‌باشد. از دیگر دلایل پیشنهاد این روش مقاوم بودن آن در برابر نویز موجود در محیط می‌باشد. در این روش با اضافه شدن نویز با SNR=50، نرخ تشخیص سیستم پیشنهادی کاهش نمی‌یابد.

ساختار سیستم شناسایی و تشخیص اصوات ضربه‌ای که در این پایان‌نامه طراحی گردیده، با فلوچارت شکل ۳-۱ نشان داده شده است. در این سیستم پس از اعمال پیش‌پردازش بر روی داده‌های موجود در پایگاه‌داده، سیگنال پنجره‌گذاری شده و انرژی هرکدام از پنجره‌ها محاسبه می‌شود.



شکل۳- ۱: ساختار سیستم شناسایی اصوات ضربه‌ای

سپس با استفاده از انرژی محاسبه شده و روش ارائه شده جهت تشخیص ضربه، سیستم تصمیم‌گیری می‌کند که آیا رخداد ورودی حاوی ضربه بوده است یا خیر. در صورتی که پاسخ منفی باشد، الگوریتم خاتمه یافته و میزان دقت سیستم در شناسایی ضربه ارزیابی می‌شود. در صورتی که پاسخ مثبت باشد، از سیگنال‌ها ویژگی‌های مناسب استخراج می‌شود و نوع ضربه با استفاده از دسته‌بند مناسب تعیین شده و در نهایت میزان دقت سیستم در تشخیص نوع ضربه مورد ارزیابی قرار می‌گیرد.

۳-۲-۱ پایگاه داده

پایگاه‌داده مورد استفاده در این پایان‌نامه دارای سه کلاس از اصوات ضربه‌ای، شامل شکستن شیشه، انفجار اسپری و ترکاندن بادکنک می‌باشد. در این پایگاه داده کلاس چهارمی که حاوی اصوات غیرضربه‌ای از جمله کارکرد کمپرسور می‌باشد که در بخش شناسایی مورد استفاده قرار می‌گیرد، وجود دارد. در کنار این چهار کلاس، کلاس دیگری با عنوان کلاس متفرقه وجود دارد که خود شامل صداهایی همچون در زدن، صحبت کردن، فریاد زدن، کشیده شدن جسم روی زمین و ... است. همانطور که دیده می‌شود در کلاس متفرقه هر دو نوع صدای ضربه‌ای و غیرضربه‌ای وجود دارد و هدف از تهیه این کلاس آزمودن قدرت الگوریتم پیشنهادی در تمیز صداهای از پیش مشخص شده از سایر صداهای موجود در محیط است. در این پایان نامه نتایج در سه حالت مختلف ارائه شده است. در حالت نخست پایگاه‌داده دارای چهار کلاس بیان شده در بالاست. در حالت دوم، کلاس متفرقه در کنار چهار کلاس قبل قرار می‌گیرد و نتایج طبقه‌بندی با حضور پنج کلاس گزارش می‌شوند. در حالت سوم پایگاه‌داده تنها حاوی دو کلاس می‌باشد. درکلاس اول داده‌های انفجار و در دسته‌ی دوم سایر داده‌ها قرار می‌گیرند. هدف از به کاربردن این پایگاه-داده، بررسی توانایی سیستم پیشنهادی در تشخیص صدای انفجار از سایر اصوات موجود در محیط است.

تمامی داده‌های استفاده شده در این پایان‌نامه به صورت دیجیتال و توسط دستگاه ضبط صوت سونی مدل PX312 تهیه شده‌اند. فرکانس نمونه‌برداری ۴۴۱۰۰ هرتز بوده است و برای ذخیره‌سازی آنها از ۱۶ بیت استفاده شده است. در این حالت تمامی اطلاعات طیفی برای تشخیص و شناسایی مورد استفاده قرار می‌گیرند. این نکته در صداهای ضربه‌ای بسیار مهم است زیرا پهنای باند فرکانسی آنها به دلیل حضور ناگهانی انرژی در حوزه‌ی زمان[1]، (مانند انفجار) افزایش می‌یابد یا در مواردی مانند شکستن شیشه انرژی زیادی در فرکانس‌های بالا وجود دارد.

در ساخت پایگاه‌داده باید دقت زیادی کرد زیرا اگر سیگنال‌ها دارای کیفیت مطلوب نباشند تاثیر منفی بر نتایج شناسایی و تشخیص دارند. نکته‌ی دیگری که باید به آن توجه کرد انتخاب سیگنال‌های مطلوب از بین داده‌های ضبط شده، می‌باشد. اگر تعداد اندکی از داده‌های یک کلاس از میانگین آن کلاس دور باشند، باعث می‌شود مرزهایی که برای داده‌های آن کلاس تعیین می شود، گسترده شود و با سایر کلاس-ها تداخل پیدا کند. از طرفی اگر داده‌های یک کلاس خیلی به هم نزدیک باشند، الگوریتم پیشنهادی قوی عمل نمی‌کند، زیرا سیستم تنها با داده‌های بسیار شبیه به هم آموزش دیده و اگر داده‌ی جدیدی از همان دسته وارد شود که کمی با داده‌های آن کلاس متفاوت باشد، سیستم قادر به تشخیص آن نیست. پس انتخاب میزان گستردگی داده‌های هر کلاس نکته‌ی بسیار مهمی است که با توجه به کاربرد سیستم، به آن موضوع باید توجه کرد. پایگاه‌داده‌ی استفاده شده در این کار در جدول ۳-۱ معرفی شده است. در جدول۳-۱ دسته‌ای به نام متفرقه جود دارد. این دسته خود شامل اصوات متفاوتی است که در جدول ۳-۲ آن‌ها را معرفی می‌کنیم.

[1] sharp temporal attack

جدول۳- ۱: معرفی پایگاه‌داده و تعداد داده‌های آن به تفکیک هر دسته

کلاس‌ها	تعداد
شکستن شیشه	۳۲
انفجار اسپری	۱۸
ترکاندن بادکنک	۳۲
کارکرد کمپرسور	۲۸
دسته متفرقه	۱۶۵

جدول۳- ۲: داده‌های موجود در دسته متفرقه

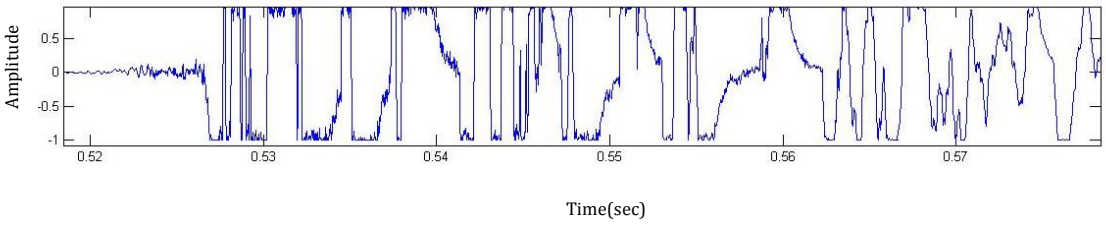
کلاس‌ها	تعداد
فریاد زدن	۱۵
صحبت کردن	۴۵
کشیدن جسم روی زمین	۱۵
در زدن	۱۵
بوق خودرو	۴۰
بسته شدن در	۳۵

با توجه به دو نکته‌ی بیان شده در بالا، حد وسطی برای گستردگی داده‌های هر کلاس در نظر گرفته شده است؛ مثلا در دسته‌ی شکستن شیشه، از شیشه‌هایی با ضخامت‌های متفاوت استفاده شده اما این تفاوت خیلی چشمگیر نیست. در کلاس بسته شدن در، جنس در یا میزان انرژی وارد شده به آن متفاوت است اما این تفاوت شامل بازه‌ی بسیار گسترده‌ای نیست. در سایر دسته‌ها نیز این نکته رعایت شده است.

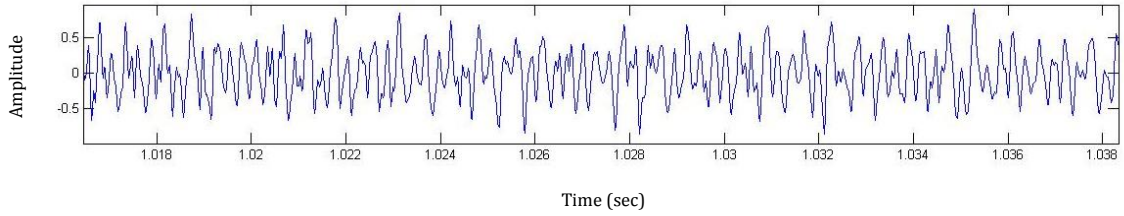
تفاوت در تعداد نمونه‌های بعضی از کلاس‌ها نه تنها خللی به سیستم وارد نمی‌کند، بلکه می‌تواند بیانگر قدرت الگوریتم پیشنهادی باشد که به تعداد نمونه‌ها در هر کلاس بستگی نداشته و احتمال رخداد تمام کلاس‌ها یکسان در نظر گرفته نمی‌شود.

۳-۲-۱-۱- چالش‌های موجود در ساخت پایگاه‌داده

چالش‌هایی که در این پایان‌نامه وجود دارد مرتبط با تهیه پایگاه‌داده‌ی مناسب است. همانطور که پیش‌تر بیان شد، تهیه و انتخاب داده‌های مناسب، تاثیر مستقیمی بر عملکرد سیستم دارد. یکی از مشکلاتی که در تهیه‌ی پایگاه‌داده وجود دارد، وجود نویز در محیط می‌باشد. نویزها شامل صداهای پس زمینه، وزش باد و نویزی که خود ضبط صوت در هنگام ضبط ایجاد می‌کند، می‌باشد. چالش دیگر در این کار انعکاس صداست. هر صدایی دارای بازتابی است. در زمان نمونه‌برداری این بازتاب نیز ضبط شده و به عنوان موج اصلی در نظر گرفته می‌شود. چالش سومی که با آن مواجه هستیم، اشباع نام دارد. زمانی که انرژی سیگنال ایجاد شده، بیشتر از قدرت ضبط صوت باشد، با آن پدیده مواجه می‌شویم. اشباع به صورت وجود بریدگی‌هایی در سیگنال خود را نشان می‌دهد و در واقع دستگاه قادر به ضبط صدا در آن بازه‌ها نمی‌باشد. در شکل ۳-۲ و شکل ۳-۳ نمونه‌ای از سیگنال اشباع شده و اشباع نشده نشان داده شده است.



شکل ۳-۲: سیگنال اشباع شده



شکل۳-۳: سیگنال اشباع نشده

۳-۲-۱-۲- رفع چالش‌های موجود در ساخت پایگاه‌داده

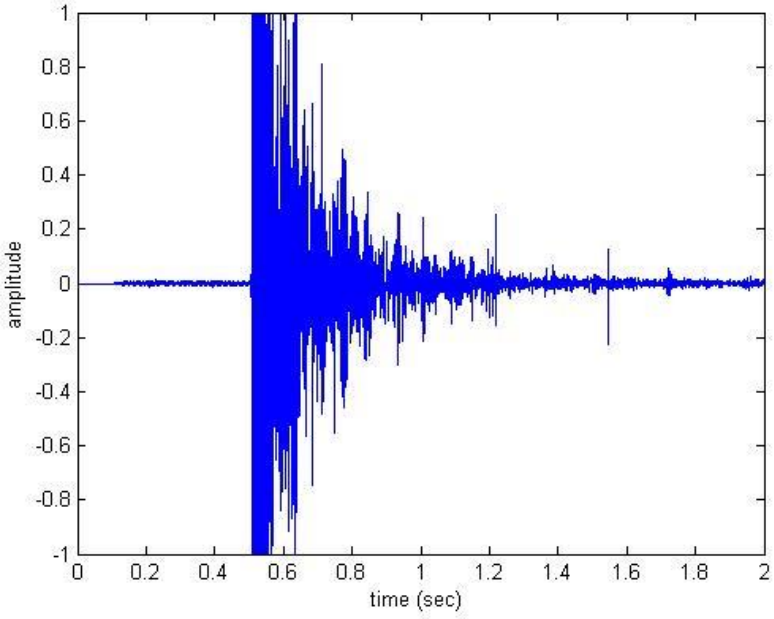
اولین و مهم‌ترین مشکلی که در ساخت پایگاه‌داده وجود دارد، حضور نویز در محیط است. برای رفع این مشکل، نمونه‌برداری را در فضای آزاد و آرام انجام دادیم. این عمل مقدار نویز در داده‌ها را کاهش داد اما علیرغم تلاش صورت گرفته در عمل باز هم مقداری نویز وجود دارد و باید از تکنیک‌های حذف نویز استفاده کرد. چالش بعدی که با آن مواجه هستیم، انعکاس است. برای کاهش این مشکل تمام نمونه-برداری‌ها در فضای آزاد و حتی المقدور به دور از اجسام منعکس‌کننده‌ی صدا، انجام شده است. آخرین مشکل اشباع نام داشت. برای رفع این موضوع باید فاصله‌ی دستگاه ضبط صوت را در هنگام نمونه‌برداری، از مکان رخداد افزایش داد. با این عمل مشکل اشباع تا حدود زیادی مرتفع می‌شود.

۳-۲-۲- پیش‌پردازش

در این بخش پیش‌پردازش‌های اعمال شده بر روی سیگنال‌ها را بررسی نموده و داده‌ها را برای استخراج ویژگی آماده می‌نماییم. داده‌ها همراه با نویزهایی هستند، همچنین انرژی سیگنال در محدوده‌ی وسیعی گسترده شده است و در بعضی نواحی آن بریدگی‌هایی وجود دارد. در این فصل تمامی عملیات انجام شده بر روی داده‌های ورودی را گام به گام تا رسیدن به داده‌هایی مناسب و قابل اعتماد بررسی می کنیم.

۳-۲-۲-۱- جداسازی رخداد از سیگنال اصلی

داده‌های موجود در پایگاه‌داده مورد استفاده شامل کلاس‌های متفاوتی از جمله انفجار می‌باشد. در شکل ۳-۴، یک نمونه از داده‌های این کلاس را نمایش می‌دهیم.

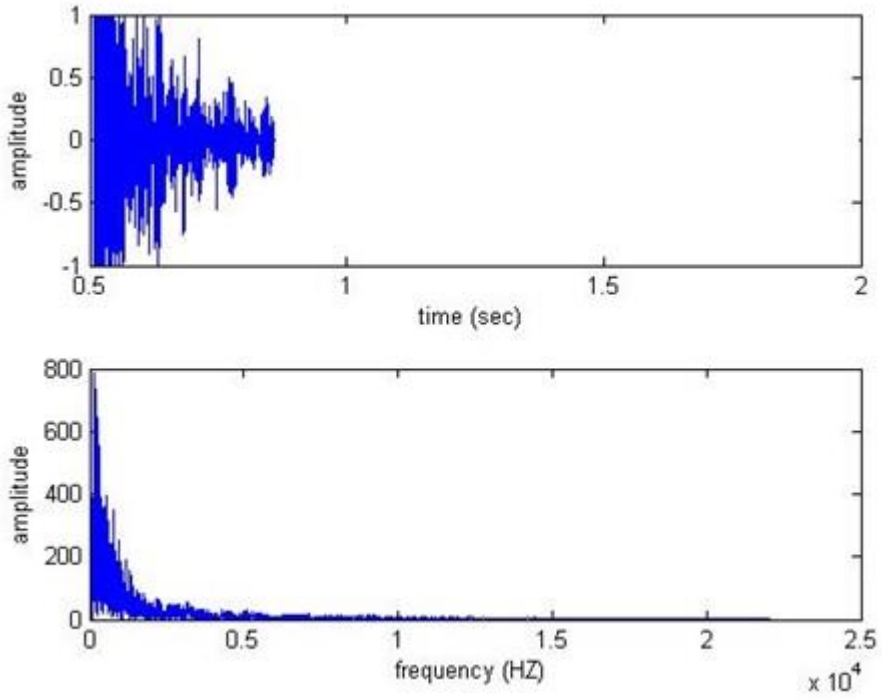


شکل ۳-۴: سیگنال انفجار در حوزه‌ی زمان

همانطور که مشاهده می‌شود رخداد مورد نظر از لحظه صفر شروع نشده است و قبل از وقوع آن، داده-ی نویز در محیط وجود داشته است. برای یکپارچه کردن تمام داده‌های موجود در پایگاه داده، از یک حد آستانه که با T1 نام‌گذاری شده، استفاده می‌شود و مقادیری از ابتدای سیگنال‌ها که اندازه‌ی عددی آن‌ها کمتر از این آستانه بوده، دور ریخته شده‌اند. از دیگر پیش‌پردازش‌هایی که روی سیگنال‌ها اعمال شده است، محلی‌سازی بخش رخداد می‌باشد. محلی‌سازی به معنای جدا کردن بخشی از سیگنال است که رخداد مورد نظر در آن اتفاق افتاده است. در مرحله‌ی قبل ابتدای رخداد تعیین شد اما برای تعیین انتهای رخداد از دو حد آستانه دیگر به طور هم‌زمان استفاده شده است. یکی از این حدود آستانه که با T2 نام-گذاری شده، برای بررسی مقدار انرژی سیگنال، مورد استفاده قرار می‌گیرد. زمانی که انرژی از این حد آستانه کمتر باشد می‌توان ادعا کرد رخداد پایان یافته است؛ اما از آنجا که ممکن است این شرط همیشه برقرار نبوده و کاهش انرژی، جزئی از ساختار رخداد مورد نظر باشد، آستانه‌ی دیگر به طور هم‌زمان مورد استفاده قرار می‌گیرد. این آستانه که با T3 نشان داده شده است، حداقل زمانی را برای طول رخداد متصور می‌شود، بدین معنی که حتی اگر انرژی سیگنال به طور لحظه‌ای از حد آستانه‌ی اولیه کمتر شد، اما طول رخداد از حداقل زمان تعیین شده کمتر بود، آن بخش از سیگنال، جزئی از رخداد تلقی می‌شود. بدین ترتیب بخشی از سیگنال که رخداد در آن اتفاق می‌افتد، جدا شده و این بخش در پردازش‌های بعدی مورد استفاده قرار می‌گیرد. مقدار بهینه این حدود آستانه با استفاده از الگوریتم ژنتیک که در ادامه فصل به توضیح آن خواهیم پرداخت، تعیین می‌شوند.

۳-۲-۲-۲- اعمال فیلتر میان‌گذر

پیش‌پردازش دیگری که روی داده‌ها اعمال می‌شود حذف فرکانس‌هایی است که در تشخیص ضروری نیستند. در شکل ۳-۵، سیگنال انفجار شکل ۳-۴، پس از جداسازی رخداد و همچنین نمایش آن در حوزه‌ی فرکانس نشان داده شده است.



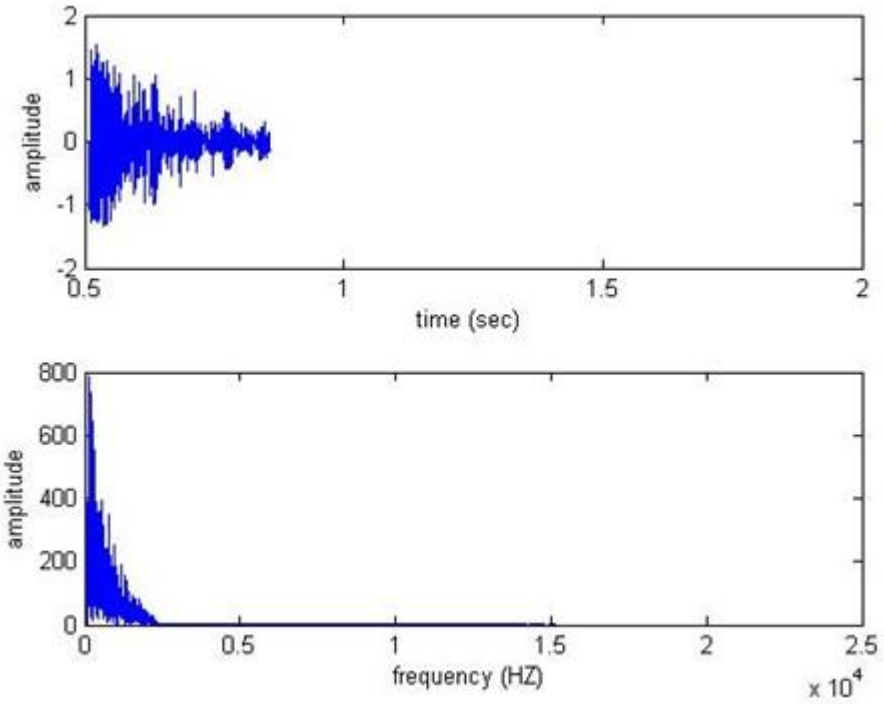
شکل۳-۵: سیگنال انفجار پس از محلی‌سازی و نمایش آن در حوزه‌ی فرکانس

همانطور که در شکل ۳-۵ ملاحظه می‌شود طیف فرکانسی شامل بازه‌ی گسترده‌ای از صفر تا ۲۰ هزار، هرتز است که بسیاری از این فرکانس‌ها زائد بوده و نه نتها جزئی از سیگنال مورد نظر نیست، بلکه ممکن است عمل تشخیص را با مشکل مواجه کند. در واقع فرکانس‌های بالاتر از ۵۰۰۰ هرتز انرژی بسیار کمی دارند و نویزهایی هستند که در محیط وجود داشته‌اند. فرکانس‌های بسیار پایین هم شامل اطلاعات ارزشمندی برای تشخیص نیستند. به منظور کاهش گستردگی طیف فرکانسی و حذف نویزهای موجود در داده‌ها و در نتیجه بهبود عملکرد تشخیص و دسته‌بندی بهتر داده‌ها، از فیلتر میان‌گذر استفاده شده است. برای بررسی این موضوع که چه محدوده‌ی فرکانسی باید از سیگنال حذف شود از چهار حد آستانه استفاده شده است. فیلتر میان‌گذر ترکیبی از دو فیلتر بالاگذر و پایین‌گذر است. نحوه‌ی عملکرد فیلتر بالا-گذر بدین صورت است که فرکانس‌های پایین‌تر از حد آستانه‌ی اولیه را به کلی حذف نموده، فرکانس‌های مابین حد آستانه‌ی اولیه و ثانویه را تضعیف نموده و فرکانس‌های بالاتر از آستانه‌ی ثانویه را عبور می‌دهد.

در این فیلتر حد آستانه‌ی اولیه با T4 و حد آستانه‌ی ثانویه با T5 نام‌گذاری شده‌اند. در فیلتر پایین‌گذر نیز حدود آستانه‌ی اولیه و ثانویه با T6 و T7 مشخص شده‌اند. مقادیر بهینه‌ی T4، T5، T6 و T7 با استفاده از الگوریتم ژنتیک تعیین شده‌اند.

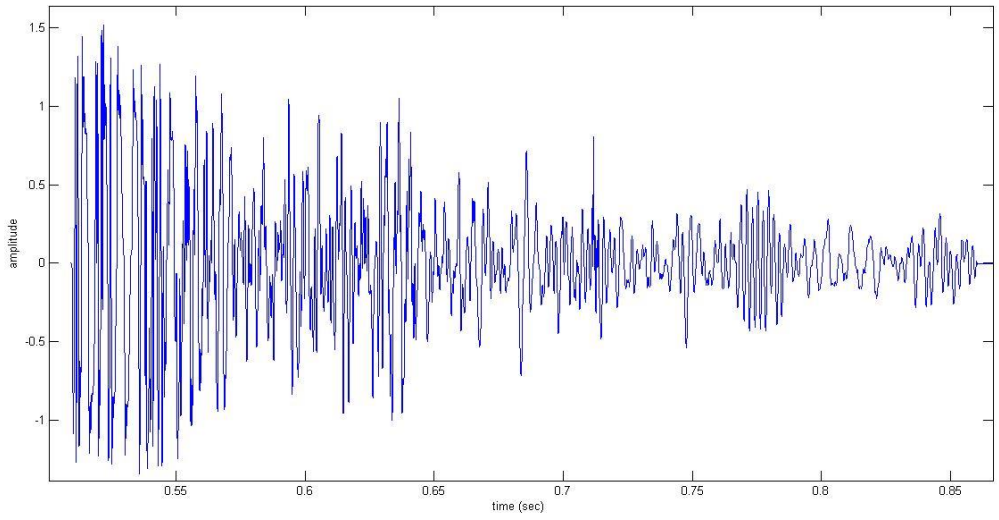
۳-۲-۲-۳- هموارسازی سیگنال

پس از اعمال فیلتر میان‌گذر سیگنال به صورت شکل ۳-۶ خواهد شد.



شکل ۳-۶: سیگنال انفجار پس از اعمال فیلتر میان‌گذر و نمایش آن در حوزه‌ی فرکانس

شکل فوق را در بازه‌ی کوچکتری مورد بررسی قرار می‌دهیم. شکل ۳-۷همان سیگنال شکل ۳-۶ می-باشد که آن را در بازه‌ی ۰.۵ تا ۰.۸۵ ثانیه نمایش داده‌ایم.



شکل۳-۷: زوم شده‌ی سیگنال انفجار پس از اعمال فیلتر میان‌گذر

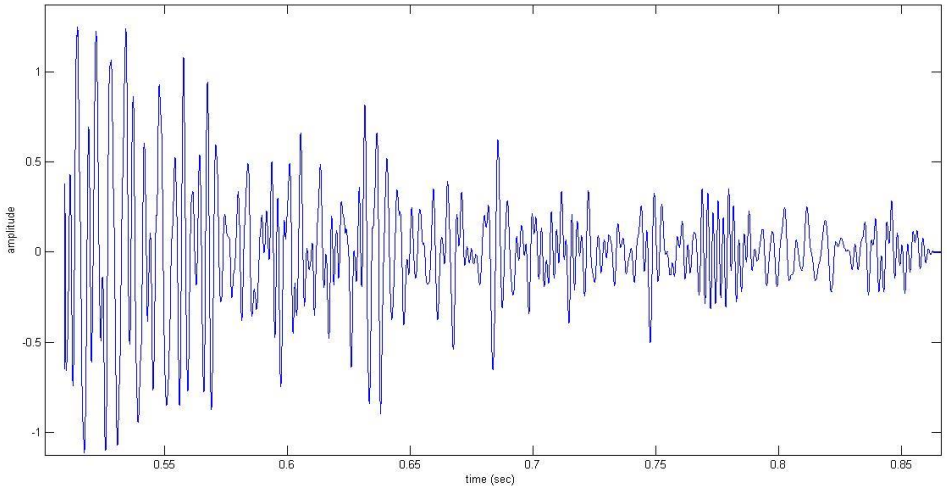
همانطور که در شکل ۳-۷ دیده می‌شود در سیگنال بریدگی‌هایی وجود دارد. این بریدگی‌ها ناشی ازحضور نویز می‌باشند. برای از بین بردن نویز و رفع بریدگی‌های موجود در سیگنال و در عین حال حفظ شکل کلی سیگنال از فیلتر سویسکی-گلی[1] استفاده شده است. فیلتر سویسکی-گلی، یک فیلتر دیجیتال است که به مجموعه‌ای از داده‌های دیجیتال، به منظور هموار کردن آن‌ها اعمال می‌شود. با این عمل نسبت سیگنال به نویز[2] افزایش می‌یابد. در این فیلتر مجموعه‌ای از نقاط همسایه (پنجره‌ای از نقاط) با یک چند جمله‌ای و با استفاده از متد حداقل مربعات خطی، تقریب زده می‌شوند. فیلتر سویسکی-گلی نویز موجود در سیگنال را تضعیف می‌کند و در نتیجه سیگنال هموارتر می‌شود و در عین حال شکل آن را حفظ نموده و ارتفاع پیک‌های سیگنال را تغییر نمی‌دهد[۵۶].

در این پایان‌نامه فیلتر سویسکی-گلی را بر روی تمام داده‌های موجود در پایگاه‌داده اعمال می‌کنیم. در این فیلتر تعیین مقدار دو پارامتر سایز پنجره‌ی تقریب و درجه‌ی چند جمله‌ای که با T8 و T9 نام‌گذاری

[1] Savitzky-Golay
[2] SNR

شده‌اند، موضوعی است که باید مورد بررسی قرار گیرند. برای تعیین مقدار بهینه این دو پارامتر از الگوریتم ژنتیک استفاده شده است.

در شکل ۳-۸ نتیجه‌ی حاصل از اعمال این فیلتر بر روی سیگنال شکل ۳-۷ نشان داده شده است.



شکل۳-۸: سیگنال انفجار پس از اعمال فیلتر سویسکی-گلی

پس از اعمال فیلتر سویسکی-گلی شکل سیگنال هموارتر و ساده‌تر شد. در این مرحله سیگنال نهایی به دست آمده، در محیط متلب پخش شد. این سیگنال شباهت قابل قبولی با سیگنال انفجار اولیه داشت. قطعا به دلیل اعمال فیلتر هموارکننده، صدای سیگنال قدری بم می‌شود اما همچنان می‌توان ادعا کرد این سیگنال مربوط به انفجار بوده و از داده‌های مربوط به سایر کلاس‌ها قابل تمییز می‌باشد.

داده‌ی به دست آمده در این مرحله مبنای ادامه‌ی کار در بخش استخراج ویژگی و دسته‌بندی می‌باشد. پس از اعمال فیلتر هموارکننده در سیگنال نهایی نمونه‌هایی ظاهر می‌شوند که مقدار آن‌ها به صفر نزدیک می‌شود. در این بخش به عنوان آخرین پیش‌پردازش اعمال شده به داده‌ها، مقادیر این نمونه‌ها را صفر می‌کنیم. دلیل اعمال چنین پیش‌پردازشی این است که در بخش استخراج ویژگی تعداد نمونه‌های صفر

می‌تواند عنصر تعیین‌کننده‌ای باشد. اینکه مقدار کدام نمونه‌ها صفر شود با استفاده از یک حد آستانه با نام T10، تعیین می‌شود. نمونه‌هایی که مقدار آن‌ها از این حد آستانه‌ی از پیش تعیین شده کمتر باشد صفر می‌شوند. مقدار بهینه‌ی این حد آستانه به همراه ۹ پارامتر دیگری که در بخش پیش‌پردازش استفاده شدند، با استفاده از الگوریتم ژنتیک تعیین شده اند. بدین ترتیب تمامی داده‌های موجود در پایگاه‌داده آماده سازی شده‌اند. در ادامه‌ی فصل از این داده‌ها برای استخراج ویژگی استفاده می‌شود.

۳-۲-۳- بهینه‌سازی پارامترهای مورد نیاز در پیش‌پردازش با استفاده از الگوریتم ژنتیک

الگوریتم ژنتیک، یکی از بهترین روش‌های موجود برای جستجوی بهینه است[۵۷]. در این پایان‌نامه از الگوریتم ژنتیک برای یافتن مقادیر بهینه‌ی برداری استفاده می‌شود که هر کدام از درایه‌های آن متناظر یکی از پارامترهای مورد نیاز در پیش‌پردازش‌ها می‌باشد. همان‌طور که در بخش ۳-۲-۲ بیان شد، در هر کدام از پیش‌پردازش‌هایی که بر روی داده‌ها اعمال می‌شود، پارامترهایی وجود دارند که باید مقادیر بهینه-ی آن‌ها تعیین شود. در این روش، مقدار بهینه‌ی ۱۰ پارامتر موجود، با توجه به نرخ دسته‌بندی درست داده‌ها، روی مجموعه‌ی آزمایش با استفاده از چهار دسته‌بند به کار برده شده در این پایان‌نامه، به طور مجزا تعیین شده‌اند.

در الگوریتم ژنتیک، یک جمعیت از افراد در محیط بقا می‌یابند. افرادی با قابلیت‌های بالاتر، شانس ترکیب و تولید مثل بیشتری دارند؛ بنابراین بعد از چند نسل، جمعیتی با کارایی بهتر به وجود می‌آید. در الگوریتم ژنتیک، هر فرد از جمعیت به صورت یک کروموزوم تعریف می‌شوند. کروموزوم‌ها در طول چندین نسل کامل‌تر می‌شوند. در هر نسل، کروموزوم‌ها ارزیابی می‌گردند و متناسب با ارزش خود، امکان بقا و تکثیر می‌یابند. تولید نسل جدید در الگوریتم ژنتیک با کمک عملگرهای جهش و ترکیب صورت می‌گیرد.

والدین برتر بر اساس تابع برازندگی انتخاب می‌شوند. در عمل ترکیب، کروموزوم‌هایی که برازندگی آن‌ها بیشتر باشد، شانس بیشتری برای انتخاب شدن دارند. این عمل به صورت اتفاقی اما بر اساس مقدار تابع برازندگی صورت می‌گیرد. مراحل تعیین مقدار بهینه‌ی پارامترها به صورت زیر خلاصه می‌شود.

۱- تشکیل جمعیت اولیه

۲- انتخاب کروموزوم‌های با مقدار شایستگی بالاتر برای تولید مثل

۳- انجام عمل ترکیب

۴- انجام عمل جهش

۵- جابه‌جایی کروموزوم‌های ایجاد شده در مرحله‌ی قبل به جای والدین در صورت داشتن مقدار شایستگی بیشتر از والدین

۶- تکرار مراحل ۲ الی آخر تا رسیدن به شرط خاتمه

شرط خاتمه، اتمام تعداد ثابتی از تکرارها می‌باشد. در این کار، این تعداد ۳۰ تکرار در نظر گرفته شده است. با این تعداد دفعات تکرار، بهترین میزان تشخیص صحیح سیستم به مقدار ثابتی می‌رسد و دیگر با افزایش تعداد دفعات تکرار نتایج بهتری حاصل نمی‌شود، در نتیجه ۳۰ تکرار شرط کافی برای خاتمه اجرای الگوریتم می‌باشد.

طول هر کروموزوم برابر تعداد پارامترهایی که قصد تعیین مقدار بهینه‌ی آن‌ها را داریم، در نظر گرفته شده است؛ بنابراین هر کروموزوم دارای ۱۰ ژن می‌باشد. تعداد جمعیت اولیه را برابر ۴۰ در نظر گرفتیم. مقدار اولیه‌ی ژن‌ها با توجه به نوع پارامتر، تعیین شده است. به طور مثال مقادیر ۴ ژن اعداد حقیقی تصادفی در بازه‌ی صفر و یک است در حالی که مقدار سایر ژن‌ها اعداد صحیح در بازه‌هایی متناسب با محدوده‌ی مجاز برای آن پارامترها تعریف شده است. تابع برازندگی مورد استفاده، بررسی صحت دسته-

بندی است. در واقع هر بار با استفاده از جمعیت موجود میزان صحت دسته‌بندی سیستم را محاسبه می-کنیم. در هر تکرار، ۴ نمونه از بهترین کروموزوم‌ها – کروموزوم‌های با شایستگی بالاتر یا به عبارتی صحت دسته‌بندی بیشتر- را انتخاب کرده و با آن‌ها عمل ترکیب و جهش را انجام می‌دهیم و مجددا میزان شایستگی آن‌ها را می‌سنجیم. در صورت بیشتر بودن این میزان، این کروموزوم‌ها جایگزین والدین خود می‌شوند و این روند در تکرارهای بعد ادامه می‌یابد تا زمان خاتمه فرا برسد.

با استفاده از الگوریتم ژنتیک مقدار بهینه‌ی پارامترهای مورد نیاز در پیش‌پردازش‌ها تعیین شد. این مقادیر تاثیر مستقیمی بر عملکرد سیستم دارند، زیرا پیش‌پردازش بخش بسیار تاثیرگذاری در سیستم شناسایی و تشخیص می‌باشد و تعیین مقادیری که این پیش‌پردازش‌ها را به نحو احسن انجام دهد بسیار ضروری و حائز اهمیت است.

۳-۲-۴- محاسبه‌ی انرژی سیگنال

این بخش در واقع نوعی استخراج ویژگی برای شناسایی رخداد ضربه‌ای از سایر اصوات موجود در محیط می‌باشد. همان‌طور که پیش‌تر بیان شد، ساختار سیستم ارائه شده بدین صورت است که ابتدا ضربه‌ای بودن رخداد ورودی مورد بررسی قرار می‌گیرد و در صورتی که ضربه‌ای باشد عملیات دسته‌بندی اجرا می‌شود؛ اما بخش شناسایی به ازای همه‌ی رخدادها اجرا می‌شود، پس این نکته که بخش شناسایی دارای بار محاسباتی اندکی باشد، بسیار حائز اهمیت است. با توجه به نکته‌ی فوق سعی شده برای بخش شناسایی از روش‌های ساده و سریع که بار محاسباتی کمی دارند، استفاده شود.

در این پایان‌نامه از بررسی انرژی پنجره‌های سیگنال، برای شناسایی رخداد ضربه‌ای استفاده کرده‌ایم. ابتدا سیگنال ورودی به پنجره‌هایی با طول مشخص و بدون همپوشانی تقسیم نموده و سپس انرژی هر کدام از این پنجره‌ها را محاسبه کردیم. مقدار انرژی این پنجره‌ها را مبنای بخش شناسایی قرار دادیم که

در ۳-۲-۵ به طور مفصل توضیحات مربوط به آن ارائه می‌شود. برای تعیین طول بهینه‌ی پنجره‌ها از سعی و خطا بهره گرفتیم و تاثیر تغییرات طول پنجره در میزان شناسایی درست سیستم را در فصل نتایج گزارش می‌دهیم. در اینجا ذکر این نکته ضروری است که طول بهینه‌ی پنجره‌ها، کسری از نرخ نمونه-برداری سیگنال می‌باشد و به صورت زیر تعریف می‌شود.

$$\text{Frame Size} = A * FS \quad (۱-۳)$$

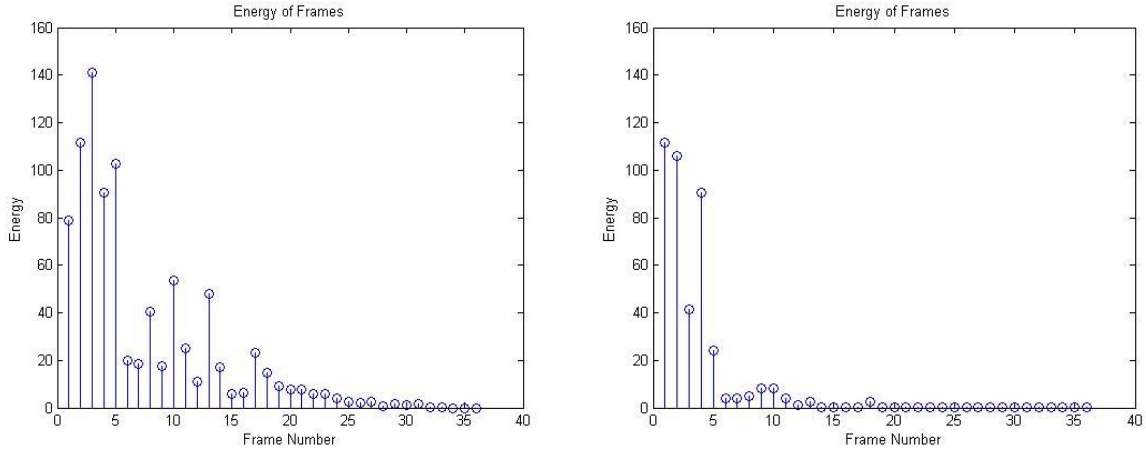
در فرمول فوق FS فرکانس نمونه‌برداری بوده که برابر ۴۴۱۰۰ هرتز می‌باشد و A ضریبی است که در بازه‌ی صفر و یک مقدار آن را تغییر دادیم تا طول بهینه‌ی پنجره‌ها را که به ازای آن سیستم بیشترین نرخ تشخیص دارد را ، تعیین کنیم.

۳-۲-۵- شناسایی رخداد ضربه‌ای

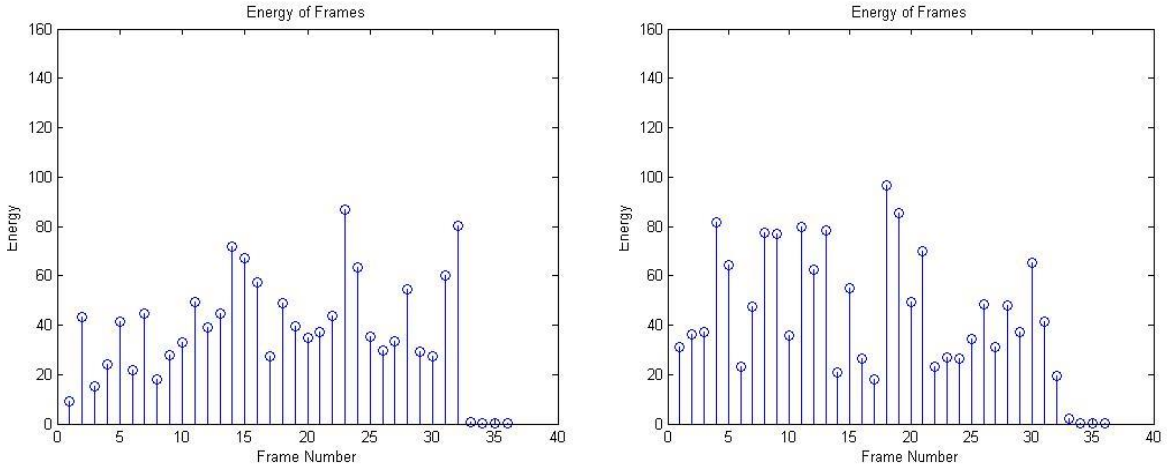
در این تحقیق دو روش نوین برای شناسایی رخداد ضربه‌ای پیشنهاد شده است.

۳-۲-۵-۱- روش پیشنهادی اول

اصوات ضربه‌ای صداهایی هستند که با افزایش ناگهانی در مقدار انرژی همراه هستند و برخلاف اصوات پریودیک که همواره یک سیکل تکراری را طی می‌کنند، یک باره اتفاق می‌افتند و پس مدت بسیار کوتاهی پایان می‌یابند. ما در این پایان‌نامه برای شناسایی و در واقع تمییز این اصوات از سایر اصوات موجود در محیط، از همین خاصیت افزایش ناگهانی انرژی استفاده کرده‌ایم. همان‌طور که در بخش ۳-۲-۴ بیان شد، پس از پنجره‌گذاری سیگنال، انرژی هر کدام از آن‌ها محاسبه شد. در شکل ۳-۹ انرژی چند نمونه از سیگنال‌هایی که مربوط به رخداد ضربه‌ای هستند و در شکل ۳-۱۰چند نمونه از سیگنال‌هایی که مربوط به رخدادی پریودیک می‌باشد را نمایش می‌دهیم.



شکل۳-۹: انرژی سیگنال‌ها با رخداد ضربه‌ای



شکل۳-۱۰: انرژی سیگنال‌ها با رخداد پریودیک

همانطور که دیده می‌شود در سیگنال‌هایی که رخداد ضربه‌ای در آن‌ها اتفاق افتاده است، عمده‌ی انرژی سیگنال در چند پنجره محدود متمرکز شده است و از آنجایی که پیش‌پردازش‌های انجام شده بر روی داده‌های پایگاه‌داده ساخته شده در این پایان‌نامه به صورتی است که رخداد در ابتدای سیگنال اتفاق می افتد، می‌توان برای تمییز رخدادهای ضربه‌ای از غیرضربه‌ای از این خاصیت که عمده‌ی انرژی سیگنال در داده‌هایی که رخداد ضربه‌ای در آن‌ها حضور دارد، در پنجره‌های نیمه‌ی اول سیگنال متمرکز شده‌اند، استفاده کرد. روند عملکرد بدین صورت است که مقدار انرژی پنجره‌های نیمه‌ی اول و نیمه‌ی دوم سیگنال

را به صورت جداگانه محاسبه می‌کنیم و سپس مقایسه‌ای بین این دو مقدار انجام می‌دهیم. در این مقایسه از یک ضریب نیز استفاده کرده‌ایم. در واقع اگر انرژی پنجره‌های نیمه‌ی اول از حاصل‌ضرب انرژی پنجره-های نیمه‌ی دوم و ضریب تعیین شده بیشتر بود، در سیگنال مورد بحث، رخدادی ضربه‌ای اتفاق افتاده است. این ضریب در فرمول (۳-۲) با B نشان داده شده و مقدار آن در بازه‌ی (۴ و ۰.۰۱) و با سعی و خطا تعیین شده است.

$$\text{First Half energy} > B * \text{Second Half Energy} \qquad (۳-۲)$$

وجود این ضریب برای بالا رفتن دقت شناسایی می‌باشد. برای تعیین مقدار این ضریب و طول پنجره‌های سیگنال، آزمایشاتی بر روی مجموعه‌ی آموزش انجام شد و پس از تعیین مقدار بهینه این پارامترها، میزان شناسایی سیستم بر روی داده‌های تست آزمایش شد. نتایج به دست آمده در فصل چهارم گزارش می‌شود.

۳-۲-۵-۲- روش پیشنهادی دوم

همانطور که در شکل‌های ۳-۹ و ۳-۱۰ دیده می‌شود، عمده‌ی انرژی سیگنال در داده‌هایی که رخداد ضربه‌ای در آن‌ها اتفاق افتاده است، تنها در چند پنجره محدود متمرکز شده است اما در داده‌هایی که رخداد ضربه‌ای وجود ندارد، انرژی سیگنال به صورت یکنواخت‌تری در سرتاسر سیگنال گسترده شده است. از این خاصیت برای تمییز اصوات ضربه‌ای و غیرضربه‌ای می‌توان استفاده کرد. در واقع با شمارش تعداد پنجره‌هایی که مقدار انرژی آن‌ها از یک حد آستانه - که ضریبی از مقدار بیشینه‌ی انرژی پنجره -هاست- بیشتر است، می‌توان ضربه‌ای بودن رخداد را تشخیص داد. این ضریب در فرمول (۳-۳) با C نشان داده شده و مقدار بهینه‌ی آن با سعی و خطا و در بازه‌ی (۱ و ۰.۰۱) تعیین شده است. پارامتر دیگری که در این روش باید مقدار آن تعیین شود، تعداد پنجره‌هایی است که مقدار انرژی آن‌ها از این حد آستانه بیشتر است. تعداد این پنجره‌ها را کسری از تعداد کل پنجره‌ها در نظر گرفتیم. مقدار این کسر را با ضریب

D، که مقدار آن در بازه‌ی (۱و ۰.۰۱) تغییر می‌کند، تعیین کردیم. روش فوق در فرمول زیر خلاصه می‌شود.

$$\text{Frame Energy Threshold} = C * \max(\text{Frame Energy}) \quad (۳-۳)$$

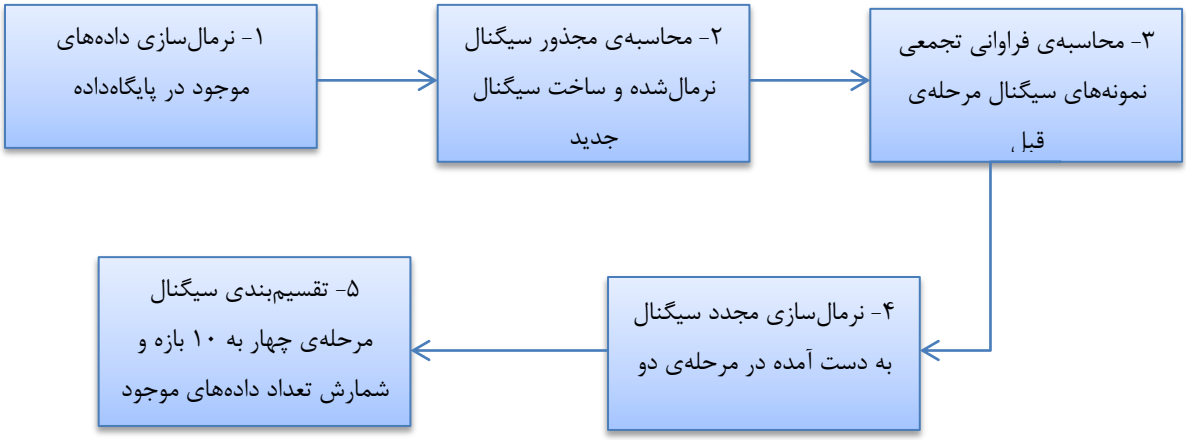
$$\text{Number of Frame's Threshold} = D * \text{Total Number of Frames} \quad (۴-۳)$$

در فصل نتایج مقدار بهینه این دو پارامتر و همچنین نرخ تشخیص درست سیستم با تغییر این دو متغییر را گزارش می‌دهیم.

مزیت این روش نسبت به روش قبل این است که این روش مستقل از پیش‌پردازش‌هایی است که بر روی داده‌ها اعمال می‌شود. همانطور که در بخش ۳-۲-۲-۱ بیان شد، مرحله‌ی اول پیش‌پردازش‌ها جداسازی بخش رخداد بود. منظور از رخداد، بخشی از سیگنال بود که رویداد مورد نظر در آن بخش اتفاق می‌افتاد. نحوه‌ی عملکرد به منظور جداسازی این بخش از سیگنال بدین صورت بود که از ابتدای سیگنال شروع می‌کردیم و زمانی که به نمونه‌ای می‌رسیدیم که مقدار انرژی آن از کسری از بیشترین انرژی سیگنال بیشتر می‌شد، آن نمونه را به عنوان ابتدای رخداد تلقی می‌کردیم و زمانی که انرژی سیگنال از کسری از بیشترین انرژی سیگنال کمتر می‌شد یا حداکثر زمانی را که برای طول رخداد در نظر گرفته بودیم، سپری می‌شد، انتهای رخداد را تعیین می‌کردیم. با این روش می‌توان تضمین نمود که در سیگنال‌های حاوی رخداد ضربه‌ای، این رخداد قطعا در ابتدای سیگنال رخ داده است و در سیگنال‌های پریودیک، رخداد در همه جای سیگنال و به طور تقریبا یک‌نواخت حضور دارد. پس با استفاده از این پیش‌شرط توانستیم رخدادهای ضربه‌ای و غیرضربه‌ای را با استفاده از روش پیشنهادی اول از یکدیگر تمیز دهیم؛ اما روش پیشنهادی دوم مستقل از مکان رخداد می‌باشد و ضربه در هر زمانی از سیگنال رخ داده باشد، قابل شناسایی می‌باشد

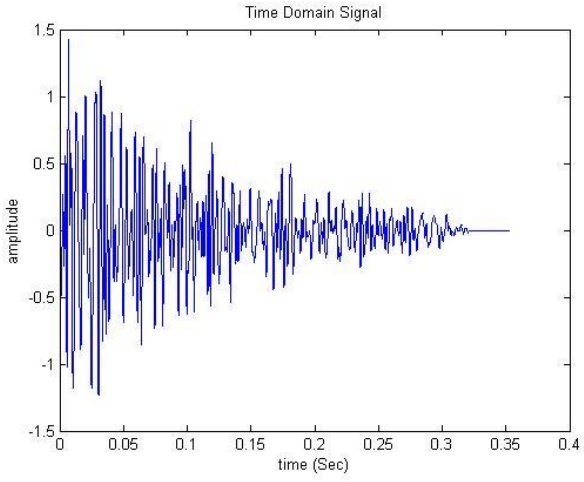
۳-۲-۶- استخراج ویژگی

در این بخش روش استخراج ویژگی پیشنهادی را توضیح می‌دهیم. روند استخراج ویژگی را در فلوچارت شکل ۳-۱۱ خلاصه می‌کنیم.

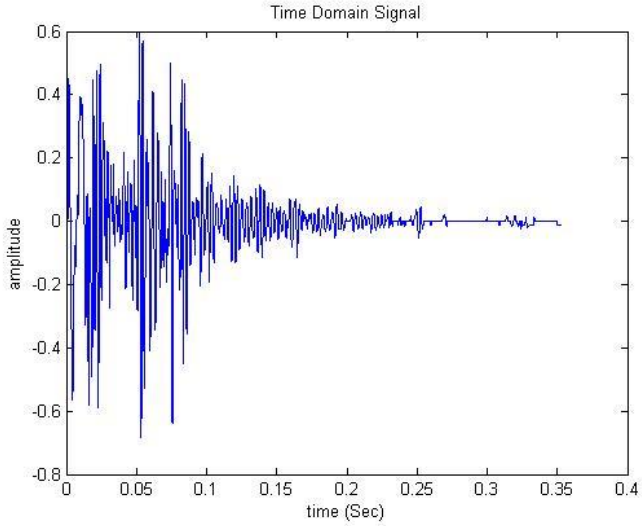


شکل۳-۱۱: مراحل استخراج ویژگی

در شکل ۳-۱۲ و ۳-۱۳ دو نمونه از سیگنال‌های انفجار موجود در پایگاه‌داده را، در حوزه‌ی زمان نشان می‌دهیم.



شکل ۳-۱۲: سیگنال انفجار شماره‌ی یک

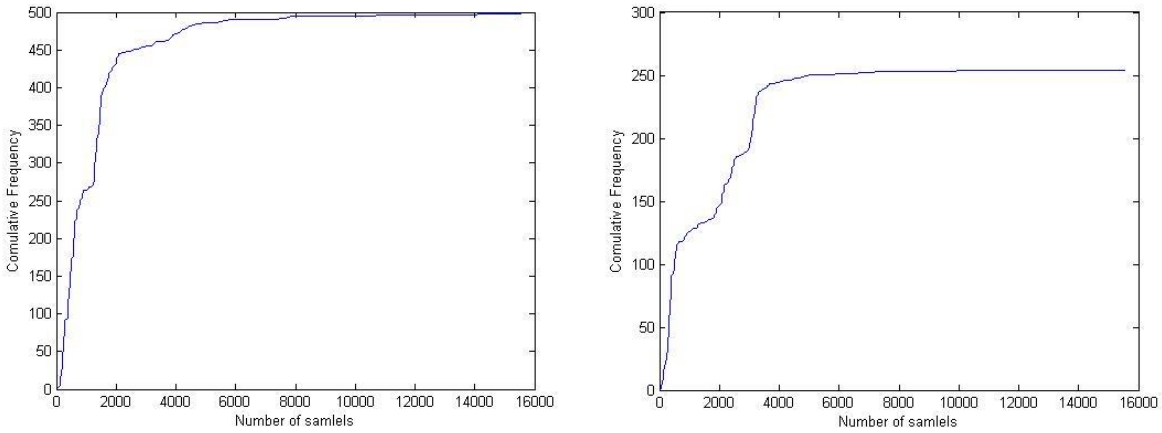


شکل۳-۱۳: سیگنال انفجار شماره‌ی دو

همانطور که در دو شکل ۳-۱۲ و ۳-۱۳ دیده می‌شود، بسته به این که قدرت انفجار چقدر باشد، مقدار دامنه‌ی سیگنال‌ها از ۰.۶ تا ۱.۵ تغییر می‌کند. به منظور یکسان‌سازی مقدار دامنه‌ی سیگنال‌ها، تمامی داده‌های موجود در پایگاه‌داده را نرمال می‌کنیم. اگر مقدار دامنه یکسان نباشد در مراحل بعدی استخراج ویژگی با مشکل مواجه خواهیم شد.

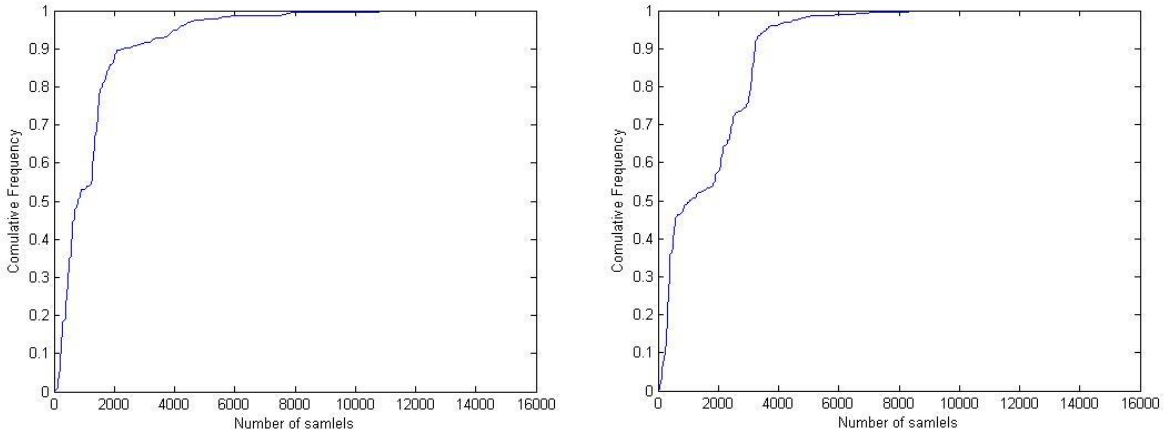
پس از نرمال‌سازی سیگنال‌ها، مجذور آن‌ها را محاسبه نمودیم. از این کار دو هدف مهم را دنبال می‌کنیم. هدف اول آن است که با به توان دو رساندن اعداد بین صفر و یک، اعداد کوچک، کوچک‌تر می‌شوند و اعداد بزرگ، بزرگ‌تر می‌شوند و به این ترتیب نوسانات ضعیف و کم اهمیت‌تر، ضعیف‌تر می‌شوند و نوسانات بزرگ‌تر، پررنگ‌تر می‌شوند. دلیل دوم آن است که بخش منفی سیگنال که نقش تعیین‌کننده‌ای در استخراج ویژگی ندارد، تغییر علامت می‌دهد و در نتیجه فرآیند استخراج ویژگی با دقت بیشتری انجام می‌شود. در این کار ما برای استخراج ویژگی، تمامی اطلاعات موجود در داده‌ها را در بازه‌ی صفر و یک محدود می‌کنیم پس حذف بخش منفی سیگنال می‌تواند ما را در استخراج ویژگی یاری کند.

پس از مجذور کردن سیگنال‌ها، ، فراوانی تجمعی آن‌ها را بدین صورت که نمونه‌ی اول را نگه داشته و نمونه‌های دوم به بعد، هرکدام مجموع نمونه‌های قبل از آن می‌باشد، محاسبه می‌کنیم. فراوانی تجمعی در واقع بیانگر نحوه‌ی تغییر رفتار سیگنال از ابتدا تا انتها می‌باشد. با محاسبه‌ی این ویژگی می‌توانیم آهنگ تغییر رفتار سیگنال را مدل کنیم. همانطور که در بخش ۱-۴ بیان شد سیگنال‌های مربوط به هر یک از دسته‌ها رفتار متفاوتی در طول زمان از خود نشان می‌دهند. در این پایان‌نامه این تفاوت را مبنای استخراج ویژگی قرار دادیم و از فراوانی تجمعی برای تمییز سیگنال‌های هر دسته بهره بردیم. در شکل ۳-۱۴ سیگنال‌های به دست آمده پس از محاسبه‌ی فراوانی تجمعی را نشان می‌دهیم.



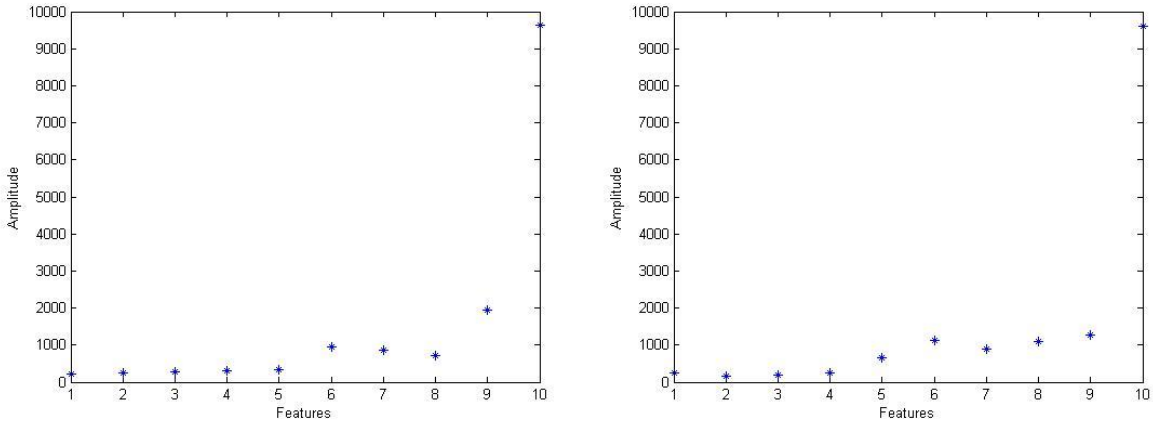
شکل۳-۱۴: - فراوانی تجمعی سیگنال شماره‌ی یک و دو

همانطور که در شکل ۳-۱۴ دیده می‌شود نقطه‌ی شروع تمامی سیگنال‌ها یکی است اما نقطه‌ی پایان آن‌ها یکسان نیست؛ مثلا نقطه‌ی پایان سیگنال شماره‌ی یک، تقریبا ۵۰۰ است در حالی که نقطه‌ی پایان سیگنال شماره‌ی دو ۲۵۰ است. برای یکسان‌سازی نقطه‌ی پایان در نمودارهای مربوط به فراوانی تجمعی سیگنال‌ها و در نتیجه داشتن معیار مناسبی برای مقایسه نحوه‌ی تغییر سیگنال‌های دسته‌های مختلف، لازم است سیگنال‌های نشان داده شده در شکل ۳-۱۴ مجددا نرمال شوند. در شکل ۳-۱۵ سیگنال‌های شکل ۳-۱۴ پس از نرمال‌سازی نشان داده شده‌اند.

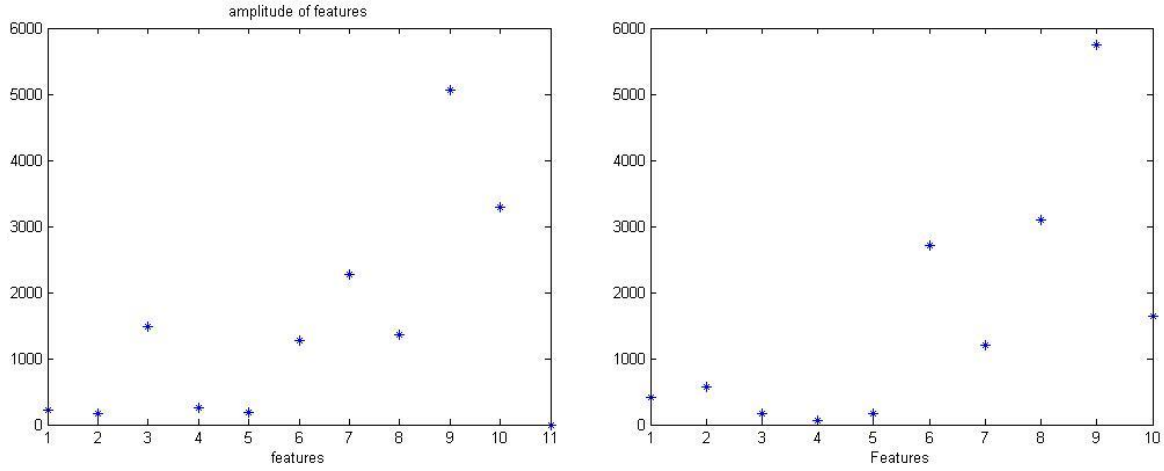


شکل۳-۱۵: نمودار فراوانی تجمعی شکل‌های ۳-۱۲ و ۳-۱۳ پس از نرمال‌سازی سیگنال‌های مربوط به آن‌ها

همانطور که در دو شکل ۳-۱۵ دیده می‌شود، توانستیم نحوه‌ی تغییر رفتار سیگنال از ابتدا تا انتها را مدل کنیم. اما از آنجا که هنوز هم مطالعه‌ی رفتار سیگنال کار راحتی نیست، بازه‌ی $[0,1]$ را به ۱۰ زیر بازه با طول ۰.۱ تقسیم می‌کنیم و تعداد نمونه‌هایی از هر سیگنال که در هر کدام از این بازه‌ها قرار می‌گیرند، شمارش کرده و آن‌ها را به عنوان ویژگی از هر سیگنال استخراج کرده و به دسته بند مناسب می‌دهیم. به این ترتیب از هر سیگنال، یک بردار ویژگی تنها با ۱۰ مولفه عددی استخراج کردیم. در شکل ۳-۱۶ دو بردار ویژگی به دست آمده از سیگنال‌های شکل ۳-۱۵ را، به صورت نموداری نشان می‌دهیم. همانطور که از شکل مشخص است، ناظر انسانی به راحتی می‌تواند تشخیص دهد این دو داده مربوط به کلاس یکسان می‌باشند. در شکل ۳-۱۷ نموداری از بردار ویژگی داده هایی که مربوط به کلاس شکستن شیشه و انفجار بادکنک است، نمایش داده‌ایم.



شکل۳-۱۶: مقادیر بردار ویژگی سیگنال شکل ۳-۱۵



الف: بادکنک ب: شکستن شیشه

شکل ۳-۱۷: مقادیر بردار ویژگی مربوط به سیگنال کلاس شکستن شیشه و بادکنک

همان‌طور که از دو شکل ۳-۱۶ و ۳-۱۷ مشخص است، ویژگی پیشنهاد شده در این تحقیق، جداکننده‌ی خوبی برای داده‌های موجود در دسته‌های مختلف است.

نوآوری این پایان‌نامه در روش نوینی می‌باشد که برای استخراج ویژگی از داده‌های صوتی پیشنهاد کرده است. این روش بدون پنجره‌گذاری سیگنال، ویژگی‌ها را مستقیما از سیگنال اصلی استخراج می-نماید و به این ترتیب تنها با ۱۰ ویژگی توانسته عمل دسته‌بندی را انجام دهد که در مقایسه با روش‌های

موجود که ابتدا سیگنال را پنجره‌گذاری نموده و سپس ویژگی‌ها را از پنجره‌ها استخراج می‌کنند، از نظر سرعت و حافظه بسیار مقرون به صرفه است.

۳-۲-۷- دسته‌بندی رخدادهای ضربه‌ای

در این پژوهش از چهار دسته‌بند k- نزدیکترین همسایه، ماشین بردار پشتیبان، بیزین و مدل مخلوط گوسی استفاده کردیم. در واقع پس از استخراج ویژگی از داده‌های موجود در پایگاه‌داده، بردار ویژگی حاوی ۱۰ مولفه را به طور جداگانه به هر کدام از دسته‌بندهای فوق دادیم و میزان تشخیص صحیح سیستم را ارزیابی نمودیم. نتایج حاصل از به کاربردن هر کدام از این دسته‌بندها را در فصل نتایج گزارش خواهیم داد.

روند کلی کار به صورت زیر خلاصه می‌شود. ابتدا داده‌ها از پایگاه داده خوانده می‌شوند. سپس مجموعه‌ای از پیش‌پردازش‌ها از جمله، جداسازی بخش رخداد، اعمال فیلتر میان‌گذر و هموارسازی بر روی داده‌ها اعمال می‌شود. سپس با استفاده از مجموعه آموزش، سیستم را آموزش می دهیم. آموزش شامل دو مرحله‌ی شناسایی اصوات ضربه‌ای از سایر اصوات و کلاس‌بندی اصوات ضربه‌ای می‌باشد. پس از آموزش سیستم و تعیین مقدار بهینه‌ی پارامترهایی که در بخش آموزش باید مقدار آن‌ها تعیین می‌شد، با استفاده از مجموعه آزمایش که شامل داده‌های دیده نشده و جدید است، کارایی سیستم را ارزیابی می-کنیم.

در دسته‌بند نزدیک‌ترین همسایه، ویژگی‌های استخراج شده از داده‌های مجموعه‌ی آموزش که کلاس آن‌ها از پیش تعیین شده است را به همراه ویژگی‌های استخراج شده از داده‌های آزمایش به عنوان ورودی به دسته‌بند می‌دهیم. دسته‌بند فاصله‌ی هر کدام از داده‌های آزمایش را از تک تک داده‌های آموزش محاسبه کرده و هر کدام از داده‌های آزمایش را به دسته‌ای که کمترین فاصله را تا داده‌های آن کلاس در

مجموعه‌ی آموزش داشته باشد، نسبت می‌دهد. در این پایان‌نامه از فاصله‌ی اقلیدسی برای تعیین فاصله‌ی داده‌های آزمایش و آموزش استفاده شده است. در این دسته‌بند پارامتری که باید مقدار آن تعیین شود، k می‌باشد. با تغییر دادن مقدار این پارامتر در بازه‌ی یک تا ده، میزان تشخیص صحیح سیستم را به ازای آن‌ها تعیین می‌کنیم و مقداری که بیشترین دقت را داشت به عنوان مقدار بهینه k در نظر می‌گیریم.

در دسته‌بند بیزین، ویژگی‌های استخراج شده از داده‌های آموزش که دسته‌ی آن‌ها از پیش تعیین شده است به همراه ویژگی‌های استخراج شده از داده‌های آزمایش را به عنوان ورودی به دسته‌بند می‌دهیم. در این دسته‌بند از توزیع نرمال گوسی برای مدل کردن داده‌های آموزش استفاده کردیم و مقدار احتمالات اولیه تمامی دسته‌ها را یکسان در نظر گرفتیم تا تاثیر متفاوت بودن تعداد نمونه‌های موجود در هر دسته را به صفر برسانیم.

در دسته‌بند مدل مخلوط گوسی، با استفاده از ویژگی‌های استخراج شده از داده‌های آموزش و به ازای هر کدام از کلاس‌ها یک نمودار گوسی که مخلوطی از چهار جز گوسی می‌باشد، ساختیم. سپس به ازای هر کدام از داده‌های آزمایش، مقدار تابع چگالی احتمال [1] را برای تمامی دسته‌ها محاسبه کردیم. مقدار این تابع به ازای هر کدام از دسته‌ها که بیشینه شد، داده‌ی آزمایش به آن دسته نسبت دادیم.

در دسته‌بند ماشین بردار پشتیبان نیز مانند دو دسته‌بند قبل، ویژگی‌های استخراج شده از داده‌های آموزش به همراه دسته‌ی آن‌ها و ویژگی‌های استخراج شده از داده‌های آزمایش را به دسته‌بند می‌دهیم. در این دسته‌بند از تابع کرنل خطی برای جداسازی داده‌های هر دسته در زمان آموزش استفاده کردیم. نتایج حاصل از به کار بردن این دسته‌بند در فصل نتایج گزارش شده است.

[1] Probability Density Function

۳-۲-۸- ارزیابی

پس از دسته‌بندی داده‌ها، با استفاده از معیارهای ارزیابی میزان کارایی روش پیشنهادی را اندازه‌گیری کرده و نتایج حاصل از آن را با روش‌های متداول استخراج ویژگی همچون ضرایب فرکانس مل مقایسه می‌کنیم. معیارهای ارزیابی و همچنین نتایج به دست آمده از روش پیشنهادی را در فصل چهارم گزارش خواهیم داد.

۳-۳- نتیجه‌گیری

در این فصل سیستم شناسایی و تشخیص اصوات ضربه‌ای را شرح دادیم. در این سیستم پس از اعمال پیش‌پردازش بر روی داده‌ها، از آن‌ها ویژگی‌های ساده، فراوانی تجمعی نمونه‌های موجود در سیگنال را استخراج کردیم و سپس با استفاده از چهار دسته‌بند بیزین، مدل مخلوط گوسی، ماشین بردار پشتیبان و k- نزدیک‌ترین همسایه عمل دسته‌بندی را انجام دادیم. این سیستم توانست با دقت خوبی عمل دسته-بندی را انجام دهد. در این پژوهش برای تعیین مقدار بهینه‌ی پارامترهای موجود در پیش‌پردازش‌ها از الگوریتم ژنتیک استفاده شد. در فصل بعد نتایج حاصل از اعمال روش پیشنهادی بر روی پایگاه‌داده ساخته شده را با سایر روش‌های موجود مقایسه می‌کنیم و نتایج حاصل از این مقایسه را ارائه می‌دهیم.

فصل چهارم

آزمایشات تجربی و ارزیابی نتایج

۴-۱- مقدمه

در این فصل ابتدا معیارهای دقت و ارزیابی عملکرد را توضیح داده و سپس نتایج حاصل از روش پیشنهادی را با روش‌های متداول استخراج ویژگی در صوت، همانند ضرایب فرکانس مل، نقطعه‌ی تعادل طیفی، شار طیفی و نرخ عبور از صفر مقایسه می‌کنیم.

۴-۲- معیارهای ارزیابی کارایی

رویکردهای ارزیابی بسیاری به منظور تعیین کارایی سیستم‌های تشخیص صدای محیط وجود دارد که معیارهای دقت[1]، فراخوانی[2] و اندازه- اف [3] از متداول‌ترین آن‌ها بوده و اغلب برای مقایسه روش‌های مختلف، مورد استفاده قرار می‌گیرند. به منظور معرفی معیارهای ارزیابی کارایی، لازم است پارامترهای به کار رفته در آن‌ها را معرفی کنیم. (جدول ۴-۱)

جدول ۴-۱: پارامترهای به کار رفته در معیارهای ارزیابی

پارامترهای به کار رفته در معیارهای ارزیابی		
	داده به دسته C_i نسبت داده شده است	داده به دسته C_i نسبت داده نشده است
داده متعلق به دسته C_i است	TP	FN
داده متعلق به دسته C_i نیست	FP	TN

[1] Precision
[2] Recall
[3] F-Measure

پس از محاسبه‌ی اجزای ساختار معیارهای ارزیابی، می‌توان معیارهای ارزیابی دقت، فراخوانی و F_1 را طبق معادله‌های (۴-۱) تا (۴-۳) به منظور بررسی و تحلیل نتایج سیستم تشخیص اصوات ضربه‌ای محاسبه نمود [۵۸].

$$\text{Precision:} \quad \pi_i = \frac{TP_i}{TP_i + FP_i} = P(\text{Relevant}|\text{Retrieved}) \qquad (۴-۱)$$

این معیار بصورت نسبت تعداد اصوات قرارگرفته در دسته‌بندی صحیح به کل اصوات قرار گرفته در همان دسته است.

$$\text{Recall:} \quad \rho_i = \frac{TP_i}{TP_i + FN_i} = P(\text{Retrieved}|\text{Relevant}) \qquad (۴-۲)$$

معیار فراخوانی به صورت نسبت تعداد اصواتی که به طور صحیح در یک دسته قرارگرفته‌اند به تعداد کل اصواتی که بایستی در آن قرار می‌گرفتند، تعریف می‌شود.

معیار F_β در واقع ترکیبی از دو معیار Precision و Recall می‌باشد. این معیار میزان تاثیر معیار دقت و معیار فراخوانی را در ارزیابی دسته‌بند در نظر می‌گیرد و از رابطه‌ی (۴-۳) بدست می‌آید.

$$F_{\beta_{Measure}} = \frac{(\beta^2 + 1) * Recall * Precision}{(\beta^2)(Recall + Precision)} \qquad (۴-۳)$$

مرسوم‌ترین مقدار β برای اررزیابی کارایی برابر ۱ است که به معنی تاثیری برابر از Precision و Recall خواهد بود. در رابطه‌ی (۴-۴) معیار F_1 تعریف شده‌است.

$$F_{1_{Measure}} = \frac{2 * Recall * Precision}{(Recall + Precision)} \qquad (۴-۴)$$

اندازه‌ی-اف کلی[1] در رابطه‌ی (۴-۵)، از مجموع نسبت حاصل‌ضرب اندازه-اف در تعداد TPهای هر دسته به تعداد کل داده‌ها به دست می‌آید.

$$TFM = \sum_{i=1}^{number\ of\ classes} TP * FM/n \qquad (۵-۴)$$

در آزمایشات انجام شده، از نرخ تشخیص سیستم [2] به منظور مقایسه نتایج استفاده شده است. این معیار به صورت میانگین فراخوانی دسته‌های مختلف تعریف شده است.

۴- ۳- آزمایشات تجربی

در این بخش آزمایشات صورت گرفته بر روی داده‌ها را بیان کرده و نتایج حاصل از آن را گزارش می-دهیم. آزمایشات در حالت‌های مختلفی صورت گرفته است. در مرحله‌ی اول تمامی داده‌های پایگاه‌داده بدون نویز هستند و نتایج حاصل از روش پیشنهادی با روش‌های بیان شده در بخش ۲-۲-۱ و با استفاده از دسته‌بندهای بیزین، مدل مخلوط گوسی، k- نزدیکترین همسایه و ماشین بردار پشتیبان، مقایسه شده-اند. آزمایشات دیگری که در این بخش صورت می‌گیرد، تعیین مقادیر بهینه‌ی پارامترهای مورد نیاز در بخش پیش‌پردازش‌هاست. از آنجا که این مقادیر با استفاده از الگوریتم ژنتیک تعیین می‌شوند و تابع هزینه در این الگوریتم نرخ تشخیص صحیح سیستم می‌باشد و به ازای دسته‌بندهای مختلف مقدار این تابع متفاوت است، مقادیر بهینه‌ی پارامترها چهار مرتبه و هر مرتبه به ازای دسته‌بند متفاوت تعیین می-

[1] Total F-Measure
[2] Recognition Rate

شوند. در مرحله‌ی دوم آزمایشات، به داده‌ها نویز با SNRهای مختلف اضافه کردیم و نرخ تشخیص صحیح سیستم به ازای روش‌های ذکر شده در مرحله‌ی قبل را با یکدیگر مقایسه کردیم. در هر دو مرحله‌ی آزمایشات پایگاه‌داده را در سه حالت بیان شده در بخش ۳-۲-۱ در نظر گرفتیم.

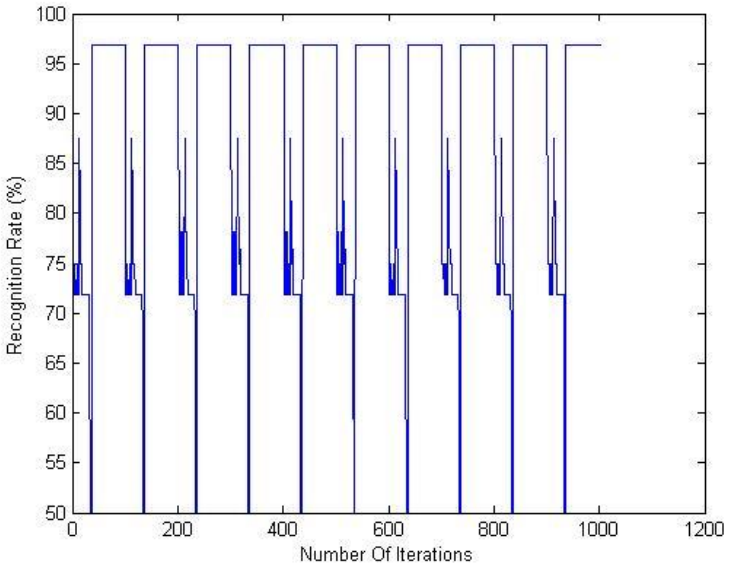
۴-۳-۱- نتایج بخش شناسایی

در این بخش نتایج به دست آمده در بخش شناسایی را ارائه می‌دهیم. همان‌طور که پیشتر بیان شد، منظور از شناسایی، دسته‌بندی داده‌ها به دو دسته اصوات ضربه‌ای و غیرضربه‌ای می‌باشد. در بخش ۳-۲-۵ دو روش جدید برای شناسایی اصوات ضربه‌ای پیشنهاد شد. در هر کدام از این دو روش پارامترهایی حضور دارند که مقادیر بهینه‌ی آن‌ها باید تعیین شود. در این بخش مقادیر این پارامترها تعیین شده و نرخ شناسایی سیستم گزارش شده است.

۴-۳-۱-۱- روش پیشنهادی اول

در این روش مقادیر بهینه‌ی دو پارامتر طول پنجره که در بخش ۳-۲-۴ از ضریب A برای نمایش آن استفاده شد و همچنین ضریب B که در بخش ۳-۲-۵ توضیحات مربوط به آن ارائه شد، باید تعیین شوند. در شکل ۴-۱ نرخ شناسایی سیستم در تکرارهای مختلف که با تغییر دو پارامتر فوق همراه می‌باشد، نشان داده شده است.

همان‌طور که دیده می‌شود با استفاده از روش پیشنهادی اول، بیشترین نرخ شناسایی سیستم ۹۶.۱۰۳٪ بوده است که این درصد در تکرارهای ۳۶ الی ۱۰۰ همچنین ۱۳۶ الی ۲۰۰ و .... اتفاق افتاده است.



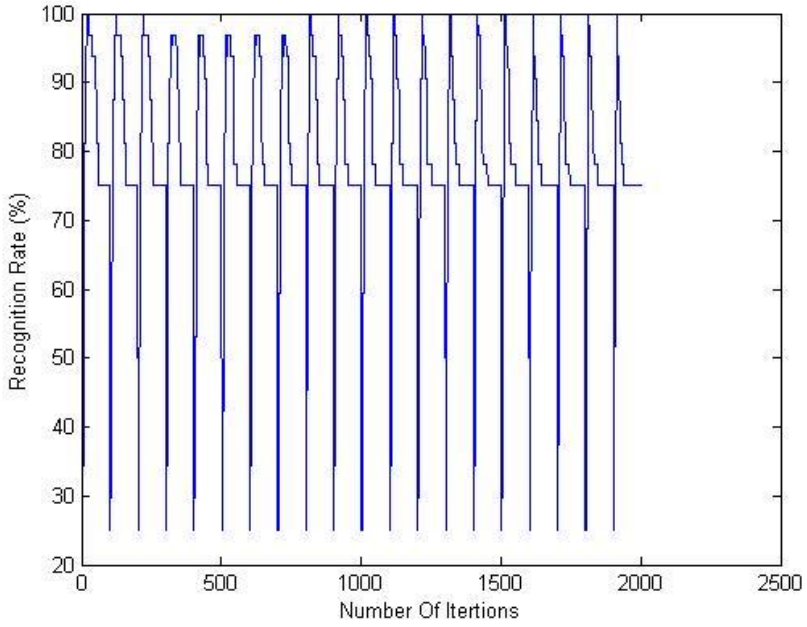
شکل۴-۱: نرخ تشخیص صحیح سیستم در تکرارهای مختلف با استفاده از روش پیشنهادی اول

در تکرار ۳۶، ضریب A برابر ۰.۰۱ بوده که با استفاده از فرمول (۳-۱) مقدار طول پنجره برابر ۴۴۱ نمونه تعیین می‌شود و مقدار ضریب B، ۱.۴۶ می‌باشد که با توجه به فرمول (۳-۲) بیانگر این نکته است که اگر مجموع انرژی پنجره‌های نیمه‌ی اول داده‌ی ورودی از ۱.۴۶ برابر مجموع انرژی پنجره‌های نیمه‌ی دوم داده‌ی مورد بحث، بیشتر باشد، رخداد ضربه‌ای اتفاق افتاده است. پس مقدار بهینه‌ی پارامتر A برابر ۰.۰۱ و B برابر ۱.۴۶ تعیین شده اند. این اعداد با استفاده از داده‌های موجود در مجموعه‌ی آموزش به دست آمده‌اند. با در نظر گرفتن این اعداد و بررسی نرخ شناسایی سیستم بر روی داده‌های موجود در مجموعه آزمایش به نرخ شناسایی ۹۶.۸۷۵٪ رسیدیم.

۴-۳-۱-۲- روش پیشنهادی دوم

در این روش علاوه بر پارامتر A در فرمول (۳-۱) که ضریب تعیین‌کننده‌ی طول پنجره‌هاست، دو پارامتر C و D نیز باید تعیین شوند. همان‌طور که در فرمول (۳-۳) و (۳-۴) آمده است، C ضریبی برای تعیین حد آستانه‌ی انرژی پنجره‌هایی است که رخداد ضربه‌ای در آن‌ها اتفاق افتاده است و D ضریبی

برای تعیین حد آستانه‌ی تعداد پنجره‌هایی است که نمایانگر رخداد ضربه‌ای هستند. در این روش نیز همانند روش پیشنهادی اول، نرخ تشخیص صحیح سیستم را بر حسب تعداد دفعات تکرار تعیین می‌کنیم. در هر تکرار مقدار سه ضریب A، C و D تغییر می‌کنند. در شکل ۴-۲ نمودار نرخ تشخیص سیستم را نشان داده‌ایم.



شکل۴-۲: نرخ تشخیص صحیح سسیستم در تکرارهای مختلف با استفاده از روش پیشنهادی دوم

همان‌طور که در شکل فوق دیده می‌شود، سیستم در چندین تکرار مختلف به تشخیص ۱۰۰٪ رسیده است. به طور مثال در تکرار ۲۲ الی ۲۵ نرخ تشخیص ۱۰۰٪ گزارش شده است. مقدار سه پارامتر فوق در تکرار ۲۲ ام را به عنوان مقادیر بهینه‌ی آن‌ها در نظر گرفته و با استفاده از آن‌ها نرخ تشخیص صحیح سیستم را بر روی مجموعه‌ی آزمایش، امتحان می‌کنیم. همان‌طور که پیشتر بیان شده بود، نرخ تشخیص گزارش شده در بالا با استفاده از مجموعه‌ی آموزش تعیین شده است. با در نظر گرفتن مقادیر A=0.01 و C=0.2 و D=0.5 - به دست آمده در تکرار ۲۲- نرخ تشخیص صحیح سیستم بر مجموعه‌ی آزمایش نیز

۱۰۰٪ گزارش شده است. این مقادیر در واقع بیانگر این هستند که با پنجره‌گذاری سیگنال‌های ورودی به پنجره‌هایی با طول ۴۴۱ نمونه (۰.۰۱×۴۴۱۰۰)، اگر تعداد پنجره‌هایی که انرژی آن‌ها از ۰.۲ بیشترین انرژی پنجره‌ها، بیشتر است، از نیمی از تعداد کل پنجره‌ها کمتر باشد، در سیگنال مورد بحث رخداد ضربه‌ای به وقوع پیوسته است.

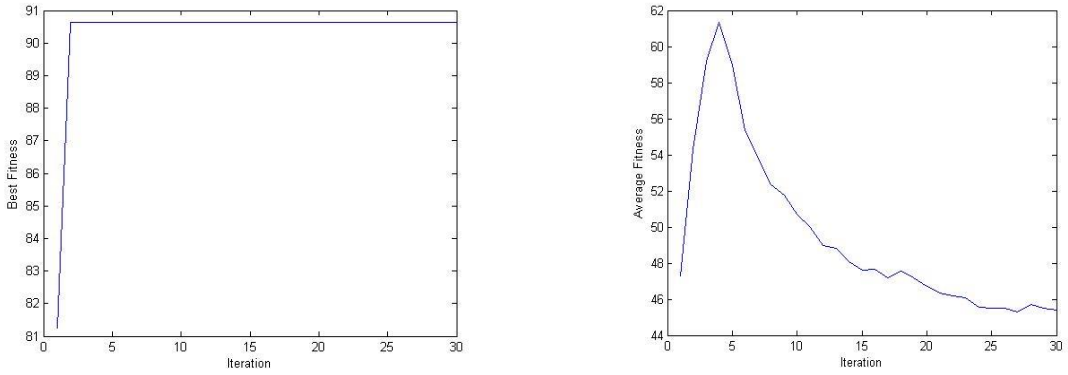
۴-۳-۱-۳- نتیجه‌گیری

در بخش ۴-۴-۱ نتایج حاصل از دو روش پیشنهادی برای دسته‌بندی اصوات محیط به دو دسته‌ی ضربه‌ای و غیرضربه‌ای را گزارش داده و مقادیر بهینه‌ی پارامترهای مورد نیاز در هر کدام از آن‌ها را تعیین نمودیم. تعیین این مقادیر با استفاده از سعی و خطا و بر روی مجموعه‌ی آموزش صورت گرفته است. پس از تعیین این مقادیر، نرخ تشخیص صحیح سیستم بر مجموعه‌ی آزمایش، مورد ارزیابی قرار گرفت. این نرخ با استفاده از روش پیشنهادی اول ۹۶.۸۷۵٪ و با استفاده از روش پیشنهادی دوم ۱۰۰٪ گزارش شده است.

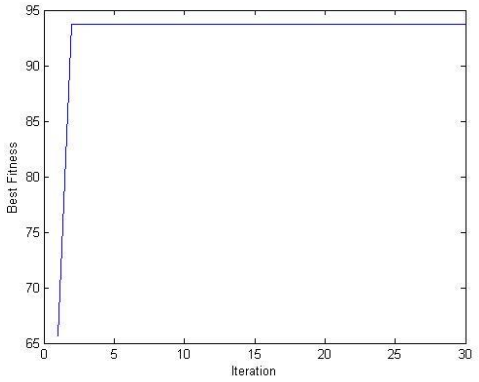
۴-۳-۲- نتایج بخش تشخیص با داده‌های بدون نویز

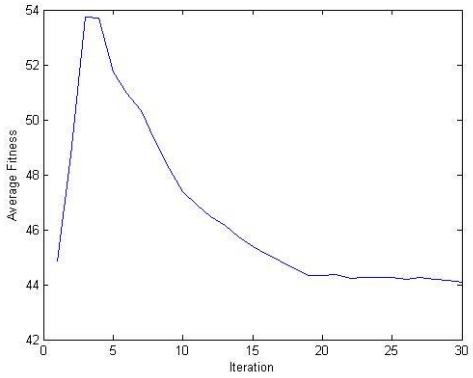
در این بخش نتایج حاصل از به کارگیری روش پیشنهادی را با سایر روش‌های متداول موجود در زمینه‌ی استخراج ویژگی از صوت، مقایسه می‌کنیم. همان‌طور که در بخش ۳-۲-۱ بیان شد پایگاه‌داده به کار رفته را در سه حالت مختلف در نظر می‌گیریم. در حالت نخست پایگاه‌داده دارای ۴ دسته، در حالت دوم همان ۴ دسته به علاوه دسته‌ی متفرقه که شامل سایر اصوات موجود در محیط است و در حالت سوم تنها دارای دو دسته انفجار و سایر اصوات می‌باشد. نتایج با استفاده از چهار دسته‌بند بیزین، مدل مخلوط گوسی، ماشین بردار پشتیبان و k- نزدیکترین همسایه گزارش شده‌اند.

پیش از مقایسه‌ی روش‌های موجود و روش پیشنهادی، نیاز است که مقادیر بهینه‌ی پارامترهای موجود در پیش‌پردازش‌ها را تعیین کنیم. برای این منظور از الگوریتم ژنتیک استفاده نمودیم. در الگوریتم ژنتیک میزان شایستگی کروموزوم‌ها برای بقا و تولید مثل با تابع برازندگی تعیین می‌شود. در این پایان‌نامه تابع برازندگی همان نرخ تشخیص صحیح سیستم می‌باشد و از آن‌جا که با استفاده از دسته‌بندهای مختلف این نرخ متفاوت می‌باشد، باید به ازای هر کدام از دسته‌بندهای به کار رفته، الگوریتم ژنتیک بهترین کروموزوم را تعیین کند. در شکل ۴-۳، ۴-۴، ۴-۵ و ۴-۶ بهترین نرخ تشخیص صحیح سیستم و میانگین آن پس از ۳۰ تکرار الگوریتم ژنتیک و با استفاده از دسته‌بند بیزین، ماشین بردار پشتیبان، k- نزدیکترین همسایه و مدل مخلوط گوسی نشان داده شده‌اند.

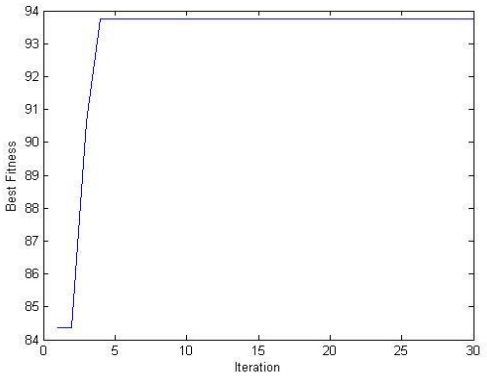


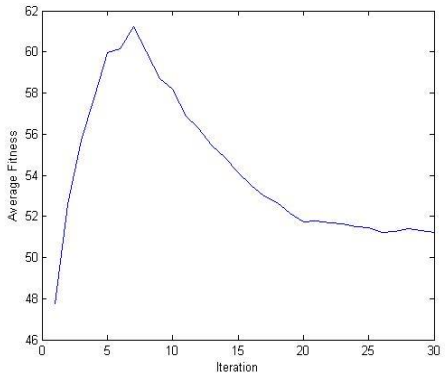
شکل۴-۳: بهترین نرخ تشخیص سیستم و میانگین آن در ۳۰ بار تکرار الگوریتم ژنتیک با استفاده از دسته‌بند بیزین



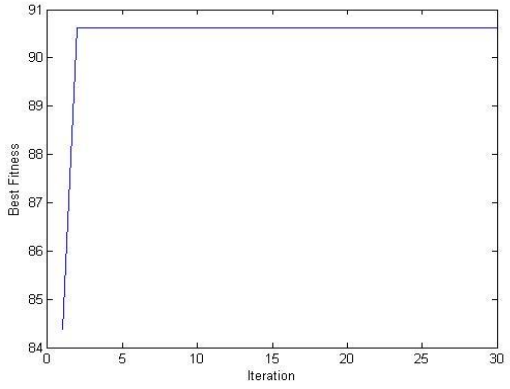


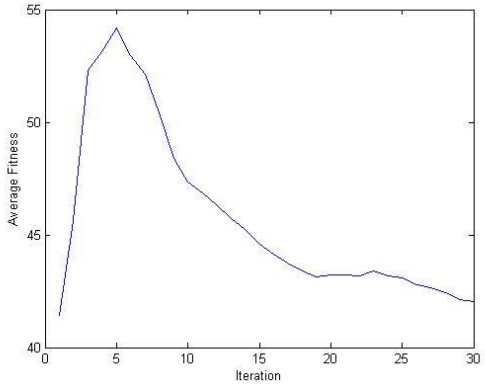
شکل ۴-۴: بهترین نرخ تشخیص سیستم و میانگین آن در ۳۰ بار تکرار الگوریتم ژنتیک با دسته‌بند ماشین بردار پشتیبان





شکل ۴-۵: بهترین نرخ تشخیص سیستم و میانگین آن در ۳۰ بار تکرار الگوریتم ژنتیک با دسته‌بند k- نزدیکترین همسایه





شکل ۴-۶: بهترین نرخ تشخیص سیستم و میانگین آن در ۳۰ بار تکرار الگوریتم ژنتیک با دسته‌بند مدل مخلوط گاوسی

همان‌طور که در شکل‌های ۴-۳، ۴-۴، ۴-۵ و ۴-۶ دیده می‌شود، بهترین میزان تشخیص صحیح سیستم پس از ۳۰ تکرار الگوریتم ژنتیک به مقداری ثابت رسید، این در حالی است که میانگین تشخیص

صحیح سیستم در این ۳۰ تکرار مقدار نزولی دارد. این موضوع نشانگر این نکته است که افزایش تعداد اجراهای الگوریتم پس از رسیدن به بالاترین درصد نشان داده شده در شکل‌های فوق الذکر، باعث بهبود میزان تشخیص صحیح سیستم نمی‌شود. بهترین نرخ تشخیص به دست آمده با استفاده از دسته‌بندهای بیزین و مدل مخلوط گوسی ۹۰.۶۲۵٪ و با استفاده از دسته‌بندهای ماشین بردار پشتیبان و k- نزدیکترین همسایه، ۹۳.۷۵٪ است. بهترین کروموزوم‌های به دست آمده توسط الگوریتم ژنتیک و با استفاده از چهار دسته‌بند ذکر شده در جدول ۴-۲ گزارش شده‌اند. توضیحات مربوط به پارامترهای T1 الی T10 موجود در جدول ۴-۲، در بخش ۳-۲-۲ ارائه شده است.

در جدول ۴-۳ دقت سیستم پیشنهادی در دسته‌بندی اصوات ضربه‌ای را در حالت چهار کلاسه و با استفاده از معیارهای موجود در روابط (۴-۱)، (۴-۲)، (۴-۴) و (۴-۵) ارائه می‌دهیم.

در جدول ۴-۴، دقت سیستم پیشنهادی با اضافه شدن کلاس متفرقه به پایگاه‌داده حاوی چهار کلاس و در جدول ۴-۵ با پایگاه‌داده دارای دو کلاس انفجار و سایر اصوات و با استفاده از معیارهای فوق‌الذکر، گزارش شده‌اند.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	10
Bayesian	0.9572	0.3500	0.5308	1609	2840	77	9	15	65	0.0225
SVM	0.5927	0.4758	0.9875	1638	2515	122	6	20	131	0.2836

جدول ۴-۲: بهترین مقادیر مورد نیاز در پیش‌پردازش‌ها، به دست آمده توسط الگوریتم ژنتیک

KNN	0.4673	0.1027	0.8808	1810	2957	74	5	25	429	0.0846
GMM	0.0916	0.2028	0.1927	1867	2827	102	7	11	135	0.1514

جدول۴- ۳: دقت سیستم پیشنهادی در حالت چهار کلاسه و با استفاده از سه دسته‌بند بیزین، ماشین بردار پشتیبان و k- نزدیکترین همسایه

	KNN			SVM			Bayesian			GMM		
	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure	Precision	recall	F1-Measure
انفجار	۱	۰.۸۷۵۰	۰۹۳۳۳	۱	۱	۱	۰.۸۸۸۹	۱	۰.۹۴۱۲	۱	۰.۷۵۰۰	۰.۸۵۷۱
کارکردکمپرسور	۰۸۸۸۹	۱	۰۹۴۱۲	۱	۱	۱	۱	۱	۱	۱	۱	۱
شکستن شیشه	۱	۰.۸۷۵۰	۰.۹۳۳۳	۰.۸۷۵۰	۰.۸۷۵۰	۰.۸۷۵۰	۰.۸۰۰۰	۱	۰.۸۸۸۹	۰.۷۲۷۳	۱	۰.۸۴۲۱
ترکیدن بادکنک	۰۸۸۸۹	۱	۰.۹۴۱۲	۰.۸۷۵۰	۰.۸۷۵۰	۰.۸۷۵۰	۱	۰.۶۲۵۰	۰.۷۶۹۲	۱	۰.۸۷۵۰	۰.۹۳۳۳
میانگین	۰.۹۴۴۴	۰.۹۳۷۵	۰.۹۳۷۳	۰.۹۳۷۵	۰.۹۳۷۵	۰.۹۳۷۵	۰.۹۲۲۲	۰.۹۰۶۳	۰.۸۹۹۸	۰.۹۳۱۸	۰.۹۰۶۳	۰.۹۰۸۱

جدول ۴-۴: دقت سیستم پیشنهادی در حالت پنج کلاسه و با استفاده از سه دسته‌بند بیزین، ماشین بردار پشتیبان و k- نزدیکترین همسایه

	KNN			SVM			Bayesian			GMM		
	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure	Precision	recall	F1-Measure	Precision	recall	F1-Measure
انفجار	۰.۸۸۸۹	۱	۰.۹۴	۰.۸۸۸۹	۱	۰.۹۴۱۲	۰.۸۷۵۰	۰.۸۷۵۰	۰.۸۷۵۰	۰.۷۷۷۸	۰.۸۷۵۰	۰.۸۲۳۵
کارکردکمپرسور	۰.۸۰۰۰	۱	۰.۸۸۸۹	۰.۶۶۶۷	۱	۰.۸۰۰۰	۱	۱	۱	۱	۱	۱
شکستن شیشه	۰.۸۷۵۰	۰.۸۷۵	۰.۸۷۵۰	۰.۸۵۷۱	۰.۷۵۰۰	۰.۸۰۰۰	۰.۸۷۵۰	۰.۸۷۵۰	۰.۸۷۵۰	۰.۷۷۷۸	۰.۸۷۵۰	۰.۸۲۳۵
ترکیدن بادکنک	۱	۰.۷۵	۰.۸۵۷۱	۰.۸۷۵۰	۰.۸۷۵۰	۰.۸۷۵۰	۰.۷۱۴۳	۰.۶۲۵۰	۰.۶۶۶۷	۰.۷۰۰۰	۰.۸۷۵۰	۰.۷۷۷۸
سایر اصوات	۰.۸۹۴۷	۰.۸۵	۰.۸۷۱۸	۱	۰.۸۰۰۰	۰.۸۸۸۹	۰.۸۵۷۱	۰.۹۰۰۰	۰.۸۷۸۰	۰.۸۷۵۰	۰.۷۰۰۰	۰.۷۷۷۸
میانگین	۰.۸۹۱۷	۰.۸۹۵۰	۰.۸۸۶۸	۰.۸۵۷۵	۰.۸۸۵۰	۰.۸۶۱۰	۰.۸۶۴۳	۰.۸۵۵۰	۰.۸۵۸۹	۰.۸۲۶۱	۰.۸۶۵۰	۰.۸۴۰۵

جدول ۴-۵: دقت سیستم پیشنهادی در حالت دو کلاسه و با استفاده از سه دسته‌بند بیزین، ماشین بردار پشتیبان و k- نزدیکترین همسایه

	KNN			SVM			Bayesian			GMM		
	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure	Precision	recall	F1-Measure
انفجار	۱	۱	۱	۱	۱	۱	۱	۰.۸۷۵۰	۰.۹۳۳۳	۰.۸۰۰۰	۱	۰.۸۵۷۱
سایر اصوات	۱	۱	۱	۱	۱	۱	۰.۹۷۷۸	۱	۰.۹۸۸۸	۱	۰۹۵۴۵	۱
میانگین	۱	۱	۱	۱	۱	۱	۰.۹۸۸۹	۰.۹۳۷۵	۰.۹۶۱۰	۰.۹۰۰۰	۰.۹۷۷۳	۰.۹۳۲۸

در بخش ۴-۲ بیان شد که نرخ تشخیص صحیح سیستم همان میانگین فراخوانی‌های دسته‌های مختلف است. در حالت چهار کلاسه، نرخ تشخیص سیستم با استفاده از دو دسته‌بند بیزین و مدل مخلوط گوسی ۹۰.۶۳٪ گزارش شده است در حالی که دسته‌بندهای k- نزدیکترین همسایه و ماشین بردار پشتیبان قدرت بیشتری از خود نشان داده‌اند و نرخ تشخیص سیستم با استفاده از آن‌ها به ۹۳.۷۵٪ افزایش یافته است. همان‌طور که پیشتر بیان شد، معیار دقت برای هر دسته، بیانگر نسبت تعداد اصوات قرارگرفته در دسته‌بندی صحیح به کل اصوات قرار گرفته در همان دسته است. معیار فراخوانی به صورت نسبت تعداد اصواتی که به طور صحیح در یک دسته قرارگرفته‌اند به تعداد کل اصواتی که بایستی در آن دسته قرار می‌گرفتند، تعریف می‌شود. معیار F1 در واقع ترکیبی از دو معیار دقت و فراخوانی می‌باشد و میزان تاثیر معیار دقت و معیار فراخوانی را یکسان در نظر می‌گیرد. پس با استفاده از این معیار می‌توان بیان کرد که سیستم داده‌های هر کلاس را با چه نرخی به دسته‌ی درست نسبت داده است. جدول ۴-۲ نشان می‌دهد که در دسته‌بند k- نزدیکترین همسایه، معیار F1 به ازای همه‌ی دسته‌ها نزدیک به هم بوده و در واقع انحراف معیار آن‌ها کم است. این موضوع بیانگر این نکته می‌باشد که با استفاده از این دسته‌بند نرخ تشخیص تمامی کلاس‌ها نزدیک به هم بوده و قدرت سیستم پیشنهادی در دسته‌بندی تمام دسته‌ها تقریبا یکسان است اما دسته‌بند ماشین بردار پشتیبان و بیزین دسته‌ی یک و دو را به خوبی از سایر دسته‌ها تفکیک می‌کنند ولی در تمییز داده‌های دسته‌ی سوم و چهارم دچار مشکل می‌شوند. دسته‌بند مدل مخلوط گوسی قدرت تفکیک‌کنندگی زیادی در دسته‌بندی داده‌های دسته‌ی دوم و چهارم دارد. در این پایان‌نامه مطلوب‌ترین دسته‌بند از جهت یکسان بودن نرخ تشخیص دسته‌های مختلف، k- نزدیکترین همسایه می‌باشد. این دسته‌بند با نرخ تشخیص سیستم ۹۳.۷۵٪ قوی تر از دو دسته‌بند بیزین و مدل مخلوط گوسی بوده و عملکرد یکسانی با ماشین بردار پشتیبان دارد. از این رو در ادامه‌ی فصل جهت مقایسه‌ی روش پیشنهادی با سایر روش‌های موجود از این دسته‌بند استفاده شده است.

جدول ۴-۳، استفاده از دسته‌بند k- نزدیکترین همسایه به عنوان دسته‌بند کارآمدتر در دسته‌بندی اصوات ضربه‌ای را تایید می‌کند. همان‌طور که ملاحظه می‌گردد با افزودن دسته‌ی سایر اصوات نرخ تشخیص سیستم کاهش می‌یابد. این امر ناشی از تنوع و گستردگی داده‌های موجود در این دسته می‌باشد. با افزودن این دسته، سیستم پیشنهادی می‌تواند تمامی اصوات موجود در محیط را، از ۴ دسته‌ی از پیش تعیین شده تمییز دهد. در واقع در حالت چهار کلاسه، چنانچه داده‌ی ورودی زمان آزمایش سیستم، به هیچ کدام از چهار دسته فوق تعلق نداشته باشد، سیستم آن را به نزدیک‌ترین دسته نسبت می‌دهد، اما با افزودن دسته‌ی پنجم این مشکل برطرف شده و سیستم توانایی تمییز اصواتی که متعلق به هیچ کدام از دسته‌های چهارگانه‌ی فوق نمی‌باشد، دارا بوده و آن‌ها را به دسته‌ی پنجم نسبت می‌دهد. گرچه نرخ تشخیص سیستم در این حالت، نسبت به حالت چهار کلاسه کاهش داشته و با استفاده از دسته‌بندهای بیزین، مدل مخلوط گوسی، ماشین بردار پشتیبان و k- نزدیکترین همسایه به ترتیب به ۸۵.۵٪، ۸۶.۵٪، ۸۸.۵٪ و ۸۹.۵٪ رسیده است، اما این کاهش قابل توجه نبوده و در برابر بهبود علمکرد قابل چشم‌پوشی می‌باشد.

در کاربردی که هدف، تشخیص داده‌های انفجار از سایر اصوات موجود در محیط است، دسته‌بند بیزین توانسته است، داده‌ها را با نرخ ۹۳.۷۵٪ از یکدیگر تفکیک دهد. دسته‌بند مدل مخلوط گوسی عملکرد بهتری داشته و نرخ تشخیص آن ۹۷.۷۳٪ گزارش شده است. در این پایگاه داده نیز دسته‌بندهای ماشین بردار پشتیبان و k- نزدیکترین همسایه همانند حالت‌های گذشته، از سایر دسته‌بندها قوی‌تر عمل کرده و داده‌ها را ۱۰۰٪ از یکدیگر تفکیک کرده‌اند.

در جدول ۴-۶ نرخ تشخیص سیستم در حالت‌های معرفی شده‌ی پایگاه‌داده و با استفاده از چهار دسته‌بند فوق‌الذکر جمع‌بندی شده است.

جدول ۴-۶: نرخ تشخیص سیستم پیشنهادی

	KNN	SVM	GMM	Bayesian
پایگاه‌داده حاوی ۵ دسته	٪۸۹.۵	٪۸۸.۵	٪۸۶.۵	٪۸۵.۵
پایگاه‌داده حاوی ۴ دسته	٪۹۳.۷۵	٪۹۳.۷۵	٪۹۰.۶۳	٪۹۰.۶۳
پایگاه‌داده حاوی ۲ دسته	٪۱۰۰	٪۱۰۰	٪۹۷.۷۳	٪۹۳.۷۵

جدول ۴-۷ نتایج روش پیشنهادی را با سایر روش‌های پایه استخراج ویژگی در صوت، مقایسه می‌کند. در این مقایسه از پایگاه‌داده یکسان استفاده شده است. در واقع روش‌های موجود بر روی پایگاه‌داده معرفی شده در بخش ۳-۲-۱، حاوی چهار دسته، پیاده‌سازی شده‌اند. در تمامی روش‌ها دسته‌بند k- نزدیکترین همسایه به کار گرفته شده است.

همان‌طور که در جدول ۴-۷ مشاهده می‌شود، نرخ تشخیص سیستم با استفاده از روش پیشنهادی از روش‌هایی همچون نرخ عبور از صفر و شار طیفی بیشتر بوده و با نقطه‌ی تعادل طیف و ضرایب فرکانس مل برابر است، اما تعداد ویژگی‌های به کار رفته در روش پیشنهادی کمتر از سایر روش‌ها می‌باشد. به طوری که گرچه نرخ تشخیص سیستم با استفاده از نقطعه قطع طیف، بیشتر است اما تعداد ویژگی‌های به کار رفته در آن سه برابر روش پیشنهادی می‌باشد.

جدول ۴-۷: مقایسه روش پیشنهادی در تشخیص اصوات ضربه‌ای با سایر روش‌های موجود با پایگاه‌داده حاوی ۴ دسته و با استفاده از دسته‌بند k- نزدیکترین همسایه

روش‌های استخراج ویژگی	تعداد ویژگی‌ها	نرخ تشخیص
نرخ عبور از صفر	۹۳	٪۸۴.۳۷۵
شار طیفی	۱۵	٪۹۰.۶۲۵
نقطه تعادل طیف	۲۵	٪۹۳.۷۵
ضرایب فرکانس مل	۴۶۸	٪۹۰.۶۲۵
ضرایب فرکانس مل	۱۸۷۲	٪۹۳.۷۵
نقطعه قطع طیف	۳۱	٪۹۶.۸۷۵
روش پیشنهادی	۱۰	٪۹۳.۷۵

در روش ضرایب فرکانس مل برای رسیدن به نرخ تشخیصی برابر با روش پیشنهادی تعداد ویژگی‌های مورد استفاده ۱۸۷۲ است که در مقایسه با ۱۰ ویژگی روش پیشنهادی بسیار زیاد می‌باشد.

از مزایای روش پیشنهادی نسبت به سایر روش‌های موجود، بالاتر بودن نرخ تشخیص سیستم، کم بودن تعداد ویژگی‌های استفاده شده و سادگی روش استخراج ویژگی می‌باشد.

۴-۴-۳- نتایج بخش تشخیص با اضافه کردن نویز به داده‌ها

در این بخش به داده‌های موجود در پایگاه‌داده چهار کلاسه، نویز سفید گوسی با SNRهای مختلف اضافه می‌کنیم و نرخ تشخیص سیستم با روش پیشنهادی و روش‌های بیان شده در بخش ۴-۴-۲ و با کمک دسته‌بند k- نزدیکترین همسایه در جدول ۴-۸ مقایسه می‌کنیم.

جدول ۴-۸: مقایسه نرخ تشخیص سیستم با روش پیشنهادی و سایر روش‌های موجود در حضور نویز با SNRهای مختلف

	۷۰	۶۰	۵۰	۴۰	۳۰	۲۰	۱۰	۰	-۱۰
نرخ عبور از صفر	٪۸۴.۳۷۵	٪۸۴.۳۷۵	٪۸۴.۳۷۵	٪۸۱.۲۵	٪۸۱.۲۵	٪۸۱.۲۵	٪۷۸.۱۲۵	٪۴۳.۷۵	٪۳۱.۲۵
شارطیفی	٪۹۰.۶۲۵	٪۸۴.۳۷۵	٪۸۱.۲۵	٪۷۸.۱۲۵	٪۷۸.۱۲۵	٪۷۵	٪۶۸.۷۵	٪۵۳.۱۲۵	٪۲۸.۱۲۵
نقطه تعادل طیف	٪۹۳.۷۵	٪۹۰.۶۲۵	٪۸۷.۵	٪۸۴.۳۷۵	٪۸۴.۳۷۵	٪۸۱.۲۵	٪۷۸.۱۲۵	٪۶۲.۵	٪۲۸.۱۲۵
ضرایب فرکانس مل	٪۹۳.۷۵	٪۹۳.۷۵	٪۹۰.۶۲۵	٪۸۷.۵	٪۸۴.۳۷۵	٪۸۱.۲۵	٪۶۲.۵	٪۴۳.۷۵	٪۲۵
نقطعه قطع طیف	٪۹۶.۸۷۵	٪۹۰.۶۲۵	٪۸۴.۳۷۵	٪۸۴.۳۷۵	٪۸۴.۳۷۵	٪۷۵	٪۶۸.۷۵	٪۲۸.۱۲۵	٪۲۵
روش پیشنهادی	٪۹۳.۷۵	٪۹۳.۷۵	٪۹۳.۷۵	٪۸۷.۵	٪۸۴.۳۷۵	٪۸۱.۲۵	٪۸۱.۲۵	٪۷۱.۸۷۵	٪۴۳.۷۵

با بررسی نتایج موجود در جدول ۴-۸ مشاهده می‌شود با افزودن نویز به داده‌ها نرخ تشخیص سیستم با استفاده از تمامی روش‌ها کاهش می‌یابد. این امر ناشی از تخریب داده‌ها توسط نویز می‌باشد. نکته‌ای که در اینجا قابل توجه است میزان کاهش قدرت سیستم با افزایش نویز می‌باشد. همان طور که در جدول ۴-۸ دیده می‌شود در تمامی روش‌های موجود با افزایش نویز نرخ تشخیص، کاهش قابل توجهی دارد اما در روش پیشنهادی این کاهش چندان قابل توجه نیست. به طور مثال روش نقطعه قطع طیف که در SNR=70 نرخ تشخیصی حدود ۳٪ بیشتر از روش پیشنهادی داشت، با افزایش قدرت نویز و رسیدن به SNR=0 تقریبا ۴۳٪ ضعیف‌تر از روش پیشنهادی عمل می‌کند. در مجموع روش پیشنهادی در تمامی SNRها عملکردی برابر و یا قوی‌تر از سایر روش‌ها داشته است. این نکته که با افزایش نویز، عملکرد سیستم پیشنهادی کاهش قابل توجهی ندارد، یکی دیگر از نقاط قوت روش پیشنهادی نسبت به سایر روش‌ها می‌باشد.

۴-۵- نتیجه‌گیری

در این فصل نتایج حاصل از روش پیشنهادی با سایر روش‌های موجود در استخراج ویژگی از صوت مقایسه شدند. سیستم پیشنهادی در شرایط بدون نویز و با استفاده از دسته‌بند k- نزدیکترین همسایه و در حالت‌های دو کلاسه، چهار کلاسه و پنج کلاسه به ترتیب نرخ تشخیص ۱۰۰٪، ۹۳.۷۵٪ و ۸۹.۵٪ دارا می‌باشد که عملکرد آن برابر و یا بهتر از سایر روش‌های موجود بوده و در عین حال از تعداد ویژگی‌های کمتر برای تشخیص استفاده کرده است. با افزودن نویز به داده‌ها نرخ تشخیص سیستم کاهش می‌یابد اما میزان کاهش در روش پیشنهادی بسیار کمتر از سایر روش‌های موجود بوده که بیانگر قدرت روش پیشنهادی می‌باشد.

فصل پنجم

نتیجه گیری و پیشنهادات

۵-۱- نتیجه‌گیری

در این پایان‌نامه، در فصل اول توضیحاتی راجع به سیستم شناسایی و تشخیص اصوات ضربه‌ای ارائه نموده و تفاوت شناسایی و تشخیص را به طور واضح بیان کردیم. سپس به کاربردهای این سیستم در زمینه‌های مختلف از جمله تشخیص نفوذ در سیستم‌های امنیتی و نظارتی، کمک به افراد پیر و ناشنوا، کاربردهای پزشکی و هدایت ربات اشاره نمودیم. از آن‌جا که این سیستم دارای کاربردهای فراوانی در زمینه‌های مختلف می‌باشد، تحقیقات متفاوت و زیادی در دهه‌ی اخیر در این زمینه انجام شده است. در فصل دوم تحقیقات انجام شده برای شناسایی و تشخیص اصوات ضربه‌ای را بررسی نمودیم. سپس در فصل سوم سیستم پیشنهادی خود را ارائه دادیم. این سیستم از دو بخش شناسایی و تشخیص ساخته شده است. در بخش شناسایی حضور یا عدم حضور صوت ضربه‌ای را با استفاده از دو تکنیک بررسی نمودیم. در تکنیک پیشنهادی اول توانستیم اصوات موجود در محیط را با نرخ تشخیص ۹۶.۸۷۵٪ به دو دسته‌ی ضربه و غیرضربه تفکیک کنیم. در روش پیشنهادی دوم، سیستم توانست حضور ضربه را با نرخ تشخیص ۱۰۰٪ شناسایی کند.

به منظور تشخیص از پایگاه‌داده‌ای متشکل از چهار دسته استفاده نمودیم. منظور از تشخیص، دسته-بندی داده‌ها به دسته‌ی درست می‌باشد. پیش از دسته‌بندی داده‌ها، از سیگنال‌ها ویژگی‌های ساده‌ای استخراج کردیم. ویژگی پیشنهاد شده در این پایان‌نامه در دسته‌ی ویژگی‌های ایستا قرار می‌گیرد. همانطور که در فصل دوم اشاره شد، از مزیت‌های این ویژگی‌ها راحتی محاسبه‌ی آن‌ها می‌باشد. برای استخراج این ویژگی ابتدا تمامی داده‌های موجود در پایگاه‌داده را نرمال می‌کنیم. سپس نمونه‌های سیگنال نرمال شده را به منظور از بین بردن بخش منفی آن و همچنین کم کردن تاثیر داده‌های کوچک، مجذور می‌کنیم. در مرحله‌ی بعد فراوانی تجمعی این سیگنال را به منظور درک نحوه‌ی تغییر رفتار سیگنال محاسبه نموده و مجددا عمل نرمال‌سازی را روی سیگنال به دست آمده، اعمال می‌کنیم. در

نهایت بازه‌ی (۱و۰) را که سیگنال نهایی در آن محدود شده است، به ۱۰ زیربازه‌ی کوچکتر تقسیم نموده و تعداد نمونه‌های موجود در هر زیربازه را به ازای تک تک سیگنال‌ها، به عنوان ویژگی انتخاب می‌نماییم. همان‌طور که ملاحظه می‌گرد محاسبه‌ی این ویژگی نه تنها از ویژگی‌های غیرایستا راحت‌تر بوده و هزینه-ی کمتری دارد، بلکه از سایر ویژگی‌های ایستا همچون MFCC که سیگنال‌ها را به منظور استخراج ویژگی، پنجره‌گذاری می‌کنند، نیز سریع‌تر می‌باشد. علاوه بر آن طول بردار ویژگی پیشنهاد شده تنها شامل ۱۰ مولفه می‌باشد که در مقایسه با سایر ویژگی‌ها بسیار کوچک‌تر می‌باشد. با استفاده از ویژگی پیشنهادی و دسته‌بند k- نزدیک‌ترین همسایه توانستیم داده‌های چهار کلاس را با نرخ تشخیص ۹۳.۷۵٪ دسته‌بندی کنیم که در مقایسه با MFCC که عملکردی حدود ۹۰٪ دارد، قوی‌تر می‌باشد. در حضور نویز روش پیشنهادی همچنان عملکرد قابل قبولی از خود ارائه می‌دهد، به طوری که با افزودن نویز سفید گوسی با SNR=50 نرخ تشخیص صحیح سیستم همچنان ۹۳.۷۵٪ بوده و در نویز با SNR=0 نرخ تشخیص حدود ۷۰٪ است که در مقایسه با MFCC که ۴۰٪ می‌باشد، عملکرد بسیار مطلوب‌تری دارد.

از مزایای روش پیشنهادی می‌توان به سریع و راحت بودن روش استخراج ویژگی، کم بودن طول بردار ویژگی و بالاتر بودن نرخ تشخیص صحیح سیستم در مقایسه با سایر روش‌های استخراج ویژگی از جمله MFCC و همچنین مقاوم بودن در برابر نویز اشاره کرد.

۵-۲- پیشنهادات

عمده‌ی پیشنهاداتی که برای ادامه کار در زمینه‌ی تشخیص اصوات ضربه‌ای می‌توان انجام داد، به ساخت پایگاه‌داده قوی‌تر منتهی می‌شود. این پیشنهادات شامل موارد زیر می‌باشد:

- جمع آوری پایگاه‌داده‌ای بزرگ‌تر که شامل دسته‌های بیشتر و تعداد نمونه‌های بیشتر از هر دسته باشد. هچنین در دسته‌ی متفرقه تمامی اصوات موجود در محیط در نظر گرفته شوند.

- لحاظ کردن تمامی نویزهای موجود در محیط از جمله وزش باد، صحبت کردن و... علاوه بر نویز سفید گوسی و بررسی نرخ تشخیص سیستم در این حالات و افزودن تکینیک‌های دیگری در کنار روش پیشنهادی تا نرخ تشخیص در این حالات نیز کاهش نیابد.
- استفاده از مجموعه‌ای از میکروفن‌ها به منظور دقت در جمع آوری پایگاه‌داده و همچنین جداسازی منبع. منظور از جداسازی منبع، تفکیک دو رخداد ضربه‌ای از یکدیگر در صورت وقوع هم‌زمان آن‌ها می‌باشد.

علاوه بر ساخت و جمع‌آوری پایگاه‌داده، در زمینه‌ی روش پیشنهادی نیز می‌توان موارد زیر را برای ادامه‌ی کار مورد توجه قرار داد:

- پنجره‌گذاری سیگنال‌ها و استخراج ویژگی‌ها از هر کدام از پنجره‌ها به منظور افزایش نرخ تشخیص سیستم در صورت وقوع چند رخداد ضربه‌ای به صورت متوالی.
- ادغام ویژگی پیشنهادی با سایر ویژگی‌های موجود به منظور افزایش نرخ تشخیص سیستم
- بررسی سایر دسته‌بندهای موجود از جمله مدل مخفی مارکوف و شبکه عصبی پرسپترون به منظور آزمایش بهبود عملکرد

فهرست مراجع

[1] D. Povey, L. Burget, M. Agarwal, P. Akyazi, F. Kai, A. Ghoshal, O. Glembek, N. Goel, M. Karafiát, A. Rastrow, The subspace Gaussian mixture model—A structured model for speech recognition, *Computer Speech & Language,* Vol. 25, No. 2, pp. 404-439, 2011.

[2] A. Graves, A.-r. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks, in *Proceeding of,* IEEE*,* pp. 6645-6649.

[3] T. Kinnunen, H. Li, An overview of text-independent speaker recognition: from features to supervectors, *Speech communication,* Vol. 52, No. 1, pp. 12-40, 2010.

[4] Z. N. Karam, W. M. Campbell, *Graph embedding for speaker recognition*, in: *Graph Embedding for Pattern Analysis*, Eds., pp. 229-260: Springer, 2013.

[5] A. Katsamanis, M. Black, P. G. Georgiou, L. Goldstein, S. Narayanan, SailAlign: Robust long speech-text alignment, in *Proceeding of*.

[6] B. Liem, H. Zhang, Y. Chen, An Iterative Dual Pathway Structure for Speech-to-Text Transcription, in *Proceeding of*.

[7] K. Wołk, K. Marasek, *Real-Time Statistical Speech Translation*, in: *New Perspectives in Information Systems and Technologies, Volume 1*, Eds., pp. 107-113: Springer, 2014.

[8] T. Virtanen, M. Helén, Probabilistic model based similarity measures for audio query-by-example, in *Proceeding of,* IEEE*,* pp. 82-85.

[9] S. Duan, J. Zhang, P. Roe, M. Towsey, A survey of tagging techniques for music, speech and environmental sound, *Artificial Intelligence Review*, pp. 1-25, 2012.

[10] S. Chu, S. Narayanan, C.-C. Kuo, M. J. Mataric, Where am I? Scene recognition for mobile robots using audio features, in *Proceeding of,* IEEE*,* pp. 885-888.

[11] N. Yamakawa, T. Takahashi, T. Kitahara, T. Ogata, H. G. Okuno, *Environmental sound recognition for robot audition using matching-pursuit*, in: *Modern Approaches in Applied Intelligence*, Eds., pp. 1-10: Springer, 2011.

[12] T. Xu, A. W. Yu, X. Liu, B. Lang, Music identification via vocabulary tree with MFCC peaks, in *Proceeding of,* ACM*,* pp. 21-26.

[13] J. G. A. Barbedo, G. Tzanetakis, Musical instrument classification using individual partials, *Audio, Speech, and Language Processing, IEEE Transactions on,* Vol. 19, No. 1, pp. 111-122, 2011.

[14] J.-C. Wang, H.-P. Lee, J.-F. Wang, C.-B. Lin, Robust environmental sound recognition for home automation, *Automation Science and Engineering, IEEE Transactions on,* Vol. 5, No. 1, pp. 25-31, 2008.

[15] F. Weninger, B. Schuller, Audio recognition in the wild: Static and dynamic classification on a real-world database of animal vocalizations, in *Proceeding of,* IEEE*,* pp. 337-340.

[16] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, K.-H. Frommolt, Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring, *Pattern Recognition Letters,* Vol. 31, No. 12, pp. 1524-1534, 2010.

[17] I. L. Freire, J. A. Apolinário Jr, Gunshot detection in noisy environments, in *Proceeding of*.
[18] A. Rabaoui, M. Davy, S. Rossignol, Z. Lachiri, N. Ellouze, Using one-class svms and wavelets for audio surveillance systems, *submitted to IEEE trans. on Information Forensic and Security*, 2007.
[19] A. D. laurent Besacier, Michael Ansorge, automatic sound recognition relying on statistical methods with application to telesurviliance, 2003.
[20] S. Oberle, A. Kaelin, Recognition of acoustical alarm signals for the profoundly deaf using hidden Markov models, in *Proceeding of,* IEEE*,* pp. 2285-2288.
[21] H. L. Xuan Guo, Jie Huang, environmental sound recognition using time freguency intersection patterns, *Awareness Science and Technology(iCAST),3rd International Conference on IEEE*, 2011.
[22] J. Beltrán-Márquez, E. Chávez, J. Favela, *Environmental sound recognition by measuring significant changes in the spectral entropy*, in: *Pattern Recognition*, Eds., pp. 334-343: Springer, 2012.
[23] M. Vacher, D. Istrate, L. Besacier, E. Castelli, J.-F. Serignat, Smart audio sensor for telemedicine, in *Proceeding of,* 15-17.
[24] D. Mitrović, M. Zeppelzauer, C. Breiteneder, Features for content-based audio retrieval, *Advances in computers,* Vol. 78, pp. 71-150, 2010.
[25] D. Mitrovic, *Discrimination and Retrieval of Environmental sounds*: na, 2005.
[26] M. Cowling, R. Sitte, Comparison of techniques for environmental sound recognition, *Pattern Recognition Letters,* Vol. 24, No. 15, pp. 2895-2907, 2003.
[27] S. Chachada, C.-C. J. Kuo, Environmental sound recognition: A survey, in *Proceeding of,* IEEE*,* pp. 1-9.
[28] W. W. Cohen, *Fast Effective Rule Induction*, in: *In Proceedings of the Twelfth International Conference on Machine Learning*, Eds., pp. 115--123: Morgan Kaufmann, 1995.
[29] W. W. Cohen, Y. Singer, A simple, fast, and effective rule learner, in *Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence*, Orlando, Florida, United States, 1999, pp. 335-342.
[30] D. D. Lewis, Naive (Bayes) at forty: The independence assumption in information retrieval, *MACHINE LEARNING: ECML-98 Lecture Notes in Computer Science*, 1998.
[31] R. Xiao, J. Wang, F. Zhang, An approach to incremental SVM learning algorithm, in *Proceeding of,* IEEE*,* pp. 268-273.
[32] M. A. Hearst, S. Dumais, E. Osman, J. Platt, B. Scholkopf, Support vector machines, *Intelligent Systems and their Applications, IEEE,* Vol. 13, No. 4, pp. 18-28, 1998.
[33] N. Cristianini, J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*: Cambridge university press, 2000.
[34] R. Hecht-Nielsen, Theory of the backpropagation neural network, in *Proceeding of,* IEEE*,* pp. 593-605.
[35] D. W. Aha, D. Kibler, M. K. Albert, Instance-based learning algorithms, *Machine learning,* Vol. 6, No. 1, pp. 37-66, 1991.

[36] M. Vacher, D. Istrate, L. Besacier, J.-F. Serignat, E. Castelli, C.-I. T. GEOD, Life Sounds Extraction and Classification in Noisy Environment, in *Proceeding of,* 77-82.
[37] J.-C. Wang, J.-F. Wang, K. W. He, C.-S. Hsu, Environmental sound classification using hybrid SVM/KNN classifier and MPEG-7 audio low-level descriptor, in *Proceeding of,* IEEE*,* pp. 1731-1735.
[38] G. Muhammad, Y. A. Alotaibi, M. Alsulaiman, M. N. Huda, Environment recognition using selected MPEG-7 audio features and Mel-Frequency Cepstral Coefficients, in *Proceeding of,* IEEE*,* pp. 11-16.
[39] A. Mesaros, T. Heittola, A. Eronen, T. Virtanen, Acoustic event detection in real life recordings, in *Proceeding of,* 1267-1271.
[40] E. Tsau, S.-H. Kim, C.-C. Kuo, Environmental sound recognition with CELP-based features, in *Proceeding of,* IEEE*,* pp. 1-4.
[41] M. Karbasi, S. Ahadi, M. Bahmanian, Environmental sound classification using spectral dynamic features, in *Proceeding of,* IEEE*,* pp. 1-5.
[42] X. Valero, F. Alías, Classification of audio scenes using Narrow-Band Autocorrelation features, in *Proceeding of,* IEEE*,* pp.
[43] X. Zhang, Y. Li, Environmental Sound Recognition Using Double-Level Energy Detection, *Journal of Signal and Information Processing,* Vol. 4, pp. 19, 2013.
[44] T. Sivaprakasam, P. Dhanalakshmi, A Robust Environmental Sound Recognition System using Frequency Domain Features, *International Journal of Computer Applications,* Vol. 80, No. 9, pp. 5-10, 2013.
[45] B.-j. Han, E. Hwang, Environmental sound classification based on feature collaboration, in *Proceeding of,* IEEE*,* pp. 542-545.
[46] X. Valero, F. Alías, Gammatone wavelet features for sound classification in surveillance applications, in *Proceeding of,* IEEE*,* pp. 1658-1662.
[47] S. Chu, S. Narayanan, C.-C. Kuo, Environmental sound recognition with time–frequency audio features, *Audio, Speech, and Language Processing, IEEE Transactions on,* Vol. 17, No. 6, pp. 1142-1158, 2009.
[48] S. Sivasankaran, K. Prabhu, Robust features for environmental sound classification, in *Proceeding of,* IEEE*,* pp. 1-6.
[49] T. K. Nobuhide Yamakawa, Toru Takahashi,Kazunori Komatani< Tetsuya Ogata, effects of modeling whitten and between frame temporal variations in power spectra on non verbal sound recognition, *ISCA*, 2010.
[50] A. Dufaux, L. Besacier, M. Ansorge, F. Pellandini, Automatic sound detection and recognition for noisy environment, in *Proceeding of,* Citeseer*,* pp.
[51] L. B. Alain Dufaux, Michael Ansorge, Fausto Pellandini, Automatic classification of wideband acoustic signals, *The Journal of the Acoustical Society of America*, 2004.
[52] C. V. Cotton, D. P. Ellis, Spectral vs. spectro-temporal features for acoustic event detection, in *Proceeding of,* IEEE*,* pp. 69-72.
[53] S. Souli, Z. Lachiri, A. Kuznietsov, *Using Three Reassigned Spectrogram Patches and Log-Gabor Filter for Audio Surveillance Application*, in: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Eds., pp. 527-534: Springer, 2013.

[54] P. Khunarsal, C. Lursinsap, T. Raicharoen, Very short time environmental sound classification based on spectrogram pattern matching, *Information Sciences,* Vol. 243, pp. 57-74, 2013.
[55] S. Souli, Z. Lachiri, Environmental Sounds Spectrogram Classification using Log-Gabor Filters and Multiclass Support Vector Machines, *arXiv preprint arXiv:1209.5756*, 2012.
[56] R. W. Schafer, What is a Savitzky-Golay filter?[lecture notes], *Signal Processing Magazine, IEEE,* Vol. 28, No. 4, pp. 111-117, 2011.
[57] D. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, MA, 1989, *NN Schraudolph and J.*. Vol. 3, pp. 1.
[58] B. Ganter, R. Wille, *Formal Concept Analysis: Mathematical Foundations*: Springer-Verlag New York, Inc., 1997.

ABSTRACT

The problem of acoustic detection and recognition is applied in robot audition, surveillance systems and security systems. This thesis addresses the problem of automatic sound detection and recognition of impulsive sounds.

Unlike other recognition techniques, which extract features form signal's frame, our proposed system extracts features from input audio signals without framing them. In this thesis a novel feature extraction method with low level feature dimension is proposed. Our proposed system has low level computational load, there for it is suitable for online applications.

The proposed system consists of detection and recognition stages. In detection stage system finds out whether a received sound is impulsive or not. For detection purpose we proposed two novel approaches based on power evolution of input audio signal. Detection rate of our approaches is 100%. When detection algorithm finds an impulsive sound, recognition stage is triggered in order to classify incoming sound. Performance of our classification method, based on signal's behavior is evaluated using KNN classifier that accuracy of classification rate is achieved 93.75%. Our proposed classification method performs well even under noise condition. The developed system has a 93.75% and 71.87% classification rate at 50dB and 0dB white Gaussian noise degradation, respectively.

Shahrood University of Technology
Faculty of Computer Engineering

Thesis Submitted in Partial Fulfillment of the Requirement for the Degree of Master of Science (M.Sc.)

Automatic Impulsive Sound Detection Based on Signal Processing Techniques

Najmeh Fayazi Far

Supervisor

Prof. Hamid Hassanpour

Associate Supervisor

Hadi Grailu

September 2014
۹۷