دانشکده فنی مهندسی

پایان نامه کارشناسی ارشد

رشته مهندسی برق، گرایش بیو الکترونیک

عنوان پایان نامه :

بازشناسی مقاوم گفتار با روش دادگان مفقود با استفاده از شبکه عصبی دوسویه

استاد راهنما : دکتر منصور ولی

استاد مشاور: دکتر علی مطیع نصرآبادی

نگارش : حجت محمدنژاد

بهار۹۰

	**اظهار نامه دانشجو**	شماره: تاریخ:

اینجانب حجت محمدنژاد دانشجوی کارشناسی ارشد رشته مهندسی برق گرایش بیوالکتریک دانشکده فنی دانشگاه شاهد، گواهی می دهم که پایان نامه تدوین شده حاضر با عنوان؛ " بازشناسی مقاوم گفتار با روش دادگان مفقود با استفاده از شبکه عصبی دوسویه" به راهنمایی استاد محترم جناب آقای دکتر منصور ولی توسط شخص اینجانب انجام و صحت و اصالت مطالب تدوین شده در آن، مورد تأیید است و چنان چه هر زمان، دانشگاه کسب اطلاع کند که گزارش پایان نامه/ رساله حاضر صحت و اصالت لازم را نداشته، دانشگاه حق دارد، مدرک تحصیلی اینجانب را مسترد و ابطال نماید هم چنین اعلام می دارد در صورت بهره گیری از منابع مختلف شامل؛ گزارش های تحقیقاتی، رساله، پایان نامه، کتاب، مقالات تخصصی و غیره، به منبع مورد استفاده و پدید آورنده آن به طور دقیق ارجاع داده شده و نیز مطالب مندرج در پایان نامه حاضر تاکنون برای دریافت هیچ نوع مدرک یا امتیازی توسط اینجانب و یا سایر افراد به هیچ کجا ارایه نشده است. در تدوین متن پایان نامه حاضر، چارچوب (فرمت) مصوب تدوین گزارش های پژوهشی تحصیلات تکمیلی دانشگاه شاهد به طور کامل مراعات شده و نهایتاً این که، کلیه حقوق مادی ناشی از گزارش پایان نامه حاضر، متعلق به دانشگاه شاهد می باشد.

نام و نام خانوادگی دانشجو:.....

امضاء دانشجو:

تاریخ:

به نام خداوندگار

شکوه و زیبایی که در پدر و مادر خلاصه شد.

به نام پدر

دریای بیکرانی که چون یونسی باید در ژرفای او فرو رفت و به شکوه او رسید.

به نام مادر

گلشنی که چون پرنده‌ای باید در او پر زد و به زیباییش رسید

تقدیم به پدر و مادر عزیزم

آنان که وجودم برایشان همه رنج بود و وجودشان برایم همه مهر.

توانشان رفت تا به توانایی برسم و مویشان سپید گشت تا رویم سپید بماند.

آنان که فروغ نگاهشان،گرمی کلامشان و روشنی رویشان ...

تنها سرمایه جاودانی زندگی من است.

آنان که راستی قامتم در شکستی قامتشان تجلی یافت.

در برابر وجود گرامیشان زانوی ادب بر زمین می زنم و با دلی مملو از ...

عشق ... محبت ... خضوع

بر دستشان بوسه می زنم.

منت خدای را عزوجل که طاعتش موجب قربت است و به شکراندرش مزید نعمت.

هر نفسی که فرو می رود ممد حیات است و چون بر می‌آید مفرح ذات. پس در هر نفسی دو نعمت موجود است و بر هر نعمت شکری واجب.

باران رحمت بی حسابش همه را رسیده و خوان نعمت بی دریغش همه جا کشیده.

اگر جای تقدیر و تشکری باشد، صمیمانه‌ترین تشکراتم را نثار استاد گرانقدرم خواهم کرد که بی دریغ مرا در راستای رسالتم یاری کرد و تمام مراحل زندگی‌ام را از زمانی که با ایشان هم‌طریق شدم به‌سان یک پدرکمک-حال شد.

و تنها در یک جمله می‌تواتم بگویم :

جناب آقای دکتر منصور ولی، صمیمانه از شما سپاسگذارم

چکیده:

عملکرد سیستم‌های بازشناسی گفتار (*ASR*) زمانی که گفتار توسط نویز تخریب شده باشد، به شدت کاهش می‌یابد. روش‌های ویژگی‌های مفقود قصد دارند که این کاهش بازشناسی را با حذف مؤلفه‌هایی از نمایش زمانی- فرکانسی گفتار (اسپکتروگرام) که بیانگر نسبت سیگنال به نویز (*SNR*) پایین باشند، کاهش دهند. این روش‌ها اثر خود را در نتایج صحت بازشناسی نشان می‌دهند که در مقابل اثر نویز جمعی، مقاوم بودن بالای خود را بروز می‌دهند.

در این پایان نامه، ما از یک شیوه‌ی رایج جبران سازی دادگان که در آن عناصر مفقود، برای بدست آوردن اسپکتروگرام کامل بازسازی می‌شوند استفاده خواهیم کرد که از آن با عنوان جبران سازی مبتنی بر دادگان مفقود یاد می‌شود. در این شیوه برای تخمین مؤلفه‌های مفقود اسپکتروگرام، از همبستگی بین مؤلفه‌ها، استفاده می‌شود. در ادامه یک شیوه جدید مطرح می‌کنیم که الگوی ویژگی‌های مفقود را با دیدگاهی نو به عنوان مسئله جبران سازی دادگان مطرح می‌کند. در این روش از شبکه عصبی دوسویه بهره گرفته می‌شود که به صورت هم‌زمان بر روی دادگان تمیز و نویزی جهت بازشناسی آواهای گفتار آموزش داده می‌شود تا با انجام یک سری پردازش‌های غیر خطی و دوطرفه (جلوسو و برگشتی) بتوان از دانش نهفته در مدل، ناشی از یاد گرفتن گفتار تمیز و نویزی بهره گرفته و بردارهای بازنمایی گفتار را در جهت افزایش صحت بازشناسی آواهای گفتار بهبود بخشید. در هر دو روش ویژگی‌های کپستروم بدست آمده از اسپکتروگرام بازسازی شده، برای بازشناسی مورد استفاده قرار می‌گیرند بدون اینکه سیستم بازشناسی نیازی به اصلاح داشته باشد.

روش مبتنی بر دادگان مفقود، از دو بخش شناسایی مؤلفه‌های مفقود و اصلاح آن‌ها تشکیل شده است. روش اصلاح ویژگی مبتنی بر شبکه عصبی دوسویه، از این قاعده مستثنی بوده چرا که نیازی به شناسایی مؤلفه‌های مفقود ندارد و بازسازی را در جهت هرچه شبیه‌تر شدن تمامی مؤلفه‌ها (خواه معتبر باشد خواه نامعتبر) به مؤلفه‌های گفتار تمیز صورت می‌دهد و این یک برتری بسیار چشمگیری است که در این تحقیق حاصل شده است؛ چرا که در عمل، بحث شناسایی مؤلفه‌های مفقود، که یک بحث چالش برانگیز در تمامی روش‌های بکارگرفته شده در جهت بازشناسی مقاوم گفتار است و ارتباط مستقیمی با میزان صحت بازشناسی دارد را حذف می‌کند. ارزیابی‌هایی که در این تحقیق بر روی دو روش ذکر شده صورت گرفت، بهبود ۴/۲ درصدی بر روی صحت بازشناسی بدست آمده برای گفتار نویزی تخریب شده توسط نویز

با نسبت سیگنال به نویز $0\ dB$ با استفاده از روش اصلاح ویژگی مبتنی بر دادگان مفقود، حاصل کرد و بهمان نحـو بهبـود ۸/۵ درصدی را برای همان نرخ نویز تخریبی با استفاده از روش اصلاح ویژگی مبتنی بر شبکه عصبی دوسویه، به نمـایش گذاشت. در آخر کار با ترکیب دو روش یاد شده، توانستیم پیشرفت چشمگیری در حدود ۱۰ درصـد در رونـد بازشناسـی سیگنال‌های تخریب شده بدست آوریم.

کلمات و اصطلاحات کلیدی :

بازشناسی مقاوم گفتار، شناسایی مؤلفه‌های مفقود، روش‌های مبتنی بر دادگان مفقود، شبکه‌ی عصبی دوسویه

فهرست مطالب .....صـــفـــحه

فهرست اشکال ..... **صـفـحه**

فهرست جداول ..... **صـفـحه**

فصل اول

پیش گفتار

۱-۱ مقدمه

با رشد روزافزون استفاده از سیستم‌های بازشناسی گفتار در کاربردهای عملی و روزمره، نیاز به حفظ راندمان بازشناسی گفتار در محیط های واقعی به عنوان امری اجتناب ناپذیر مطرح می‌باشد. شرایط ایده‌آل و عاری از نویزی که در شبیه‌سازی‌های رایانه‌ای در نظر گرفته می‌شود در هیچ یک از کاربردهای واقعی صادق نیست. به عنوان مثال وجود سر و صدای محیط، انعکاس دیوارها، همهمه دیگر گویندگان و تخریب‌های ناشی از کانال‌های انتقال باعث بر هم خوردن شرایط آزمایشگاهی می‌شود.

بنابراین هنگامی که از سیستم بازشناسی گفتاری که در محیط آزمایشگاهی آموزش داده شده است، در محیط واقعی استفاده شود اغلب راندمان سیستم بازشناسی به دلیل عدم انطباق[1] دادگان آموزشی آزمایشگاه و داده جمع‌آوری شده در محیط واقعی به مقدار زیادی کاهش می‌یابد [۱]. از این رو مبحث مقاوم‌سازی[2] در برابر نویز به عنوان یکی از ضرورت‌های کاربردی و عملی از زمینه‌های فعال تحقیقاتی در سال‌های اخیر بوده است.

در سیستم‌های بازشناسی گفتار عدم تطابق بین دادگان گفتاری محیط واقعی نسبت به دادگان آموزش را می‌توان به وجود تنوعات[3] گفتار در شرایط متنوع صوتی نسبت داد. تنوعات به عواملی اطلاق می‌شود که بر الگوها تأثیر می‌گذارند و بازشناسی آن‌ها را دچار مشکل می‌کنند. برای مقاوم کردن بازشناسی خودکار الگوها نسبت به این تنوعات و نزدیک

[1] *Mismatch*
[2] *Robustness*
[3] *Variation*

کردن کیفیت بازشناسی آن ها به بازشناسی انسان، لازم است تنوعات را شناسایی کرده و آن ها را به نحوی جبران کرد. انسان در ادراک گفتار روزمره با انواع این تنوعات در سیگنال ورودی برخورد می کند و علی‌رغم وجود آن‌ها وظیفه درک را به خوبی انجام می‌دهد. به عنوان نمونه‌هایی از این تنوعات در بازشناسی گفتار می‌توان از تغییر گوینده، تغییر لهجه و لحن صدا، تغییر شدت صدا، تغییر فاصله و جهت گوینده، تغییر تأثیرات محیط مثل دیوارها و غیره، تغییر کانال انتقال، بلندگو، تغییر در سرعت بیان و ... نام برد. این عوامل باعث عدم تطبیق بین داده‌های تعلیم و دادگان تست، منجر به کاهش کیفیت بازشناسی خواهند شد [۲].

سیستم‌های بازشناسی گفتار در حقیقت، طبقه‌بندی کننده‌های الگوهای آماری هستند که بخش‌هایی از گفتار را که متعلق به یک مجموعه‌ی کلاس از صداها باشند، طبقه‌بندی می‌کنند. طبقه‌بندی مستقیماً از خود سیگنال گفتار انجام نمی‌شود، بلکه سیگنال گفتار به یک رشته از بردارهای ویژگی، تجزیه می‌شود و طبقه بندی از روی این بردارهای ویژگی انجام می‌پذیرد. بردارهای ویژگی خود به صورت‌های گوناگونی از طیف توان بخش‌های پنجره شده‌ی (فریم) سیگنال گفتار بدست می‌آیند. سیستم‌های بازشناسی گفتار سعی در یادگیری توزیع این بردارهای ویژگی، از مجموعه دادگان آموزشی گفتار دارند که هر یک از این بردارها متعلق به یک صدا می‌باشند. در طول بازشناسی، در صورتی، یک بخش از سیگنال گفتار متعلق به یک صدا، طبقه بندی خواهد شد که توزیع آن بخش از سیگنال گفتار، شباهت زیادی به بردارهای ویژگی تولید شده‌ی متعلق به آن صدا داشته باشد.

زمانی که یک سیگنال گفتار با تنوعات ایستان، تخریب شده باشد، تأثیری که نویز مخرب روی سیگنال گفتار صورت می‌دهد اینست که توزیع بردارهای ویژگی سیگنال تخریب شده، شباهت زیادی به توزیع‌هایی که توسط دادگان آموزشی یاد گرفته شده‌اند، ندارد. این نامتناسب بودن، طبقه‌بندی نامناسب و بازشناسی ضعیفی را در پی خواهد داشت [۳]. این اثر را می‌توان با آموزش سیستم بازشناسی توسط مجموعه دادگانی که دارای سطح نویز مشابه با دادگانی که قرار است بازشناسی شوند، به حداقل رساند؛ ولی حتی با وجود این شرایط، اضافه کردن نویز، اثر خود را در بالا بردن خطای تخمین طیف هر فریم از گفتار نشان خواهد داد و بنابراین تغییر پذیری ذاتی را در بردارهای ویژگی متناظر هر صدا بالا خواهد برد [۴]. در نتیجه، تفاوت توزیع‌های کلاس‌های یکسان افزایش یافته و این افزایش مغایرت، افزایش خطای طبقه بندی و بازشناسی نامناسب را نسبت به شرایطی که هر دو سیگنال گفتار آموزشی و تست از وجود نویز آزاد هستند،

شامل خواهد شد. در نهایت، زمانی که نویز تخریبی، غیرایستان باشد، حتی اگر دادگان آموزشی با سطح مشابهی از نویز در دادگان تست، تخریب شده باشند، سودمند نخواهد بود و این امر بدین علت است که اگر چه همه سطح نویز چه در دادگان آموزشی و چه در دادگان تست یکسان هستند ولی این موضوع تضمین نمی‌کند که نمونه‌های مختلف صداها در دادگان آموزشی دقیقاً توسط همان نوع نویزی تخریب شده باشد که دادگان تست با آن تخریب شده‌اند و همچنان عدم مطابقت بین توزیع‌های یاد گرفته شده در دادگان آموزشی به وسیله طبقه‌بندی کننده و توزیع‌های دادگان تست به جای خود باقی خواهد ماند.

از این رو لازم است سیستم بازشناسی گفتار به گونه ای منعطف و تطبیق پذیر طراحی گردد که امکان شناسایی گفتار نویزی و حصول به نرخ های بالای بازشناسی در شرایط محیطی متفاوت امکان پذیر باشد. چنین پروسه‌ای را سیستم بازشناسی مقاوم گفتار و مجموعه روش‌هایی را که جهت طراحی چنین سیستمی به کار می‌روند تکنیک های مقاوم سازی می‌نامند که مبحث اصلی این تحقیق را به خود اختصاص داده‌اند.

مسئله کاهش عدم مطابقت بین توزیع‌های دادگان آموزشی و تست را می‌توان در دو شیوه مطرح کرد. در شیوه اول، دادگان تست با برخی از روش‌ها به قصد اینکه به دادگان آموزشی که توزیع آن‌ها به طبقه‌بندی کننده آموزش داده شده است، شبیه شوند، تمیز می‌شوند. ما از این روش که در آن هدف جبران اثر نویز در دادگان تست می‌باشد، با عنوان روش‌های جبران سازی دادگان یاد می‌کنیم. در شیوه دوم، توزیع‌هایی که در طبقه بندی کننده برای مدل کردن کلاس-های مختلف صدا، استفاده شده است، برای اینکه با توزیع‌های دادگان تست مشابهت داشته باشند، اصلاح می‌شوند. ما از این روش که هدف آن، اصلاح مؤلفه‌های طبقه‌بندی کننده برای جبران اثر نویز است، با عنوان روش‌های جبران سازی طبقه بندی کننده یاد می‌کنیم.

نمونه‌های مختلفی از روش‌های جبران سازی دادگان و روش‌های جبران سازی طبقه بندی کننده در مقالات متعددی مطرح شده‌اند. روش‌های جبران سازی دادگان از قبیل نرمالیزه کردن وابسته به کد کپستروم[1] (*CDCN*) [۱]، سری برداری تیلور[2] (*VTS*) [۳]، تفریق طیفی[1] [۵] و فیلتر وینر[2] [۶]، قصد داشتند اثر نویز روی دادگان را به وسیله تخمین

1 codeword dependent cepstral normalization
2 vector Taylor series

زدن طیف نویز جبران کنند و سایر روش‌ها از قبیل جبران سازی کپستروم مبتنی بر تابع گوسی چند متغیره[3] (*MGCC*) [۳] و فیلتر بهینه احتمالی[4] (*POF*) [۷] سعی در تطبیق بین دادگانی که به طور هم‌زمان در شرایط آموزشی و تست ضبط شده‌اند، برای اصلاح دادگان تست، داشتند.

روش‌های جبران سازی طبقه بندی کننده از قبیل ترکیب مدل موازی[5] (*PMC*) [۸] و ترکیب مدل[6] [۹]، توزیع کلاس-های صدا را برای رسیدن به نویز جمعی، اصلاح می‌کنند و سایر روش‌ها از قبیل بیشترین احتمال خطی رگرسیون[7] (*MLLR*) [۱۰]، مشخصات توزیع‌ها را برای متناسب شدن با گفتار نویزی تست، تغییر می‌دهند. یکی از موانع تمامی این روش‌ها در این است که همگی آن‌ها، نویز مورد نظر را ایستان فرض می‌گیرند. تمامی این روش‌ها در مواجهه با نویز ایستان پایین یا متوسط موفقیت نسبتاً خوبی از خود نشان می‌دهند (گفتار نویزی با نسبت سیگنال به نویز ۱۰ *dB* یا بالاتر)؛ ولی این روش‌ها در مواجهه با نویز با سطح بالاتر، کارآمدی خود را از دست داده و حتی در برخورد با نویز غیرایستان کاملاً بی نتیجه خواهند بود. در فصل دوم، به صورت اجمالی روش‌هایی که تا کنون در زمینه بازشناسی مقاوم گفتار مورد استفاده واقع شده‌اند، شرح داده خواهد شد [۱۱].

روش‌های مطرح شده برای بازشناسی مقاوم گفتار، بر اساس مشاهده فرایند ترجیحی سیستم شنوایی انسان که مؤلفه‌های با انرژی بالای سیگنال گفتار، مؤلفه‌های ضعیف‌تر را سرکوب می‌کنند، ارائه شده‌اند. روش‌های جدید مطرح شده سعی دارند که عملکرد بازشناسی گفتار را با بازسازی مؤلفه‌های با *SNR*پایین بهبود بخشند [۱۲].

روش‌های مبتنی بر باندهای چندگانه[8] [۱۳-۱۴]، بر این حقیقت استوار بودند که باندهای فرکانسی متمایز از سیگنال گفتار ممکن است با *SNR*های مختلفی تخریب شده باشند. این روش‌ها سیگنال گفتار را به باندهای فرکانسی جداگانه تجزیه کرده و سیستم‌های بازشناسی گفتار جداگانه‌ای را برای هر باند ایجاد می‌کردند. سرانجام خروجی هر یک از این سیستم‌های بازشناسی با هم ترکیب می‌شد تا خروجی نهایی بدست آید. وزن داده شده به خروجی هر سیستم بازشناسی

---

1 spectral subtraction
2 Wiener filtering
3 multivariate Gaussian based cepstral compensation
4 probabilistic optimal filtering
5 parallel model combination
6 model composition
7 maximum likelihood linear regression
8 Multi-band based approaches

متناظر با هر باند فرکانسی، به صورت ایده‌آل به مقدار SNRدر آن باند بستگی داشت، بدین صورت که باندهای فرکانسی آلوده به نویز بالا دارای ارزش به مراتب کمتری در قبال باندهای فرکانسی که در مقابل نویز مقاوم بودند، داشتند.

روش‌های مبتنی بر ویژگی‌های مفقود [۱۵-۱۶]، بر اهمیت این حقیقت قائل هستند که نه تنها می‌توان مقدار SNR را در باندهای فرکانسی متمایز، متفاوت دانست بلکه می‌توان در زمان‌های مختلف هم، به صورت محلی و متفاوت با زمان‌های دیگر، در نظر گرفت. در این روش، سیگنال گفتار به حوزه زمانی- فرکانسی تبدیل شده و به صورت یک تصویر اسپکتروگرافیک (گرافیک طیفی) نمایش داده می‌شود که در آن محورها به ترتیب بیانگر زمان و فرکانس را بوده و مقادیر هر عنصر در شکل نمایش داده شده، بیانگر انرژی سیگنال گفتار در آن موقعیت از زمان و فرکانس می‌باشد که به آن اسپکتروگرام نیز می‌گویند.

نواحی متمایز در اسپکتروگرام، توسط سطوح مختلفی از نویز تخریب می‌شوند. در روش‌های مبتنی بر ویژگی‌های مفقود، نواحی با SNR پایین، از این نمایش پاک می‌شوند. آن دسته از روش‌هایی که بازشناسی را تنها بر پایه نواحی باقیمانده از اسپکتروگرام، که از حذف نواحی با SNR پایین بدست آمده است، انجام می‌دهند، به عنوان روش‌های اسپکتروگرام ناکامل[1] یاد می‌شوند. مزیت روش اسپکتروگرام ناکامل در اینست که این روش، هیچ فرضی درباره ایستان بودن نویز مخرب در نظر نمی‌گیرد. همچنین این روش نیازی به اطلاعات کامل از ساختار طیفی نویز نداشته و به توصیفی از نواحی نمایش زمانی- فرکانسی تنها با عنوان معتبر و نامعتبر بسنده می‌کند [۱۵]. روش اسپکتروگرام ناکامل نتایج بسیار قابل ملاحظه‌ای در صحت بازشناسی در سطوح بالایی از نویز تخریبی از خود نشان داده است [۱۷-۱۸]. تمامی روش‌های متداول اسپکتروگرام ناکامل را با عنوان روش‌های جبران سازی طبقه‌بندی کننده[2] نیز یاد می‌کنند [۱۵-۱۶, ۱۹]. این روش‌ها اثر ناکامل بودن دادگان اسپکتروگرام را در طبقه‌بندی کننده مدل کرده و طبقه‌بندی کننده را به منظور جبران سازی اصلاح می‌کنند. برای اینکه این روش‌ها قابل پیاده‌سازی باشند، طبقه‌بندی کننده باید با ویژگی‌های اسپکتروگرام، مانند ویژگی‌های طیفی یا لگاریتم طیفی، آموزش داده شود. این یک مشکل جدی برای این روش‌ها می‌باشد چرا که کاملاً واضح است که وقتی بازشناسی از روی ویژگی‌های لگاریتم طیفی صورت می‌پذیرد، صحت بازشناسی بدست آمده،

[1] *incomplete spectrogram methods*
2 Classifier Reconstruction methods

خیلی ضعیف‌تر از زمانی ست که از ویژگی‌های دیگری همچون کپستروم که خود از ویژگی‌های لگاریتم طیفی بدست می‌آیند، استفاده شود [۲۰]. صحت بازشناسی بدست آمده از ویژگی‌های کپستروم گفتار نویزی، بدون هیچ جبران سازی، اکثر اوقات بالاتر از صحت بازشناسی بدست آمده از ویژگی‌های لگاریتم طیفی است که حتی جبران سازی مبتنی بر دادگان مفقود، روی آن اعمال شده است. اما زمانی که نواحی پاک شده از اسپکتروگرام با مقادیر تخمینی بهینه‌ای از مقادیر مورد قبولشان جایگزین شوند و بازشناسی با استفاده از اسپکتروگرام کامل حاصل شود، نتایج خیلی بالاتر از زمانی خواهد بود که از روش‌های اسپکتروگرام ناکامل استفاده می‌شود. چرا که در روش اخیر ویژگی‌های لگاریتم طیفی قابلیت تبدیل به ویژگی‌های کپستروم را دارند و می‌توان بازشناسی را با استفاده از ویژگی‌های کپسترومی که از ویژگی‌های اصلاح شده استخراج شده‌اند، بدست آورد. از این روش جبران سازی به عنوان روش‌های بازسازی اسپکتروگرام[1] نیز یاد می‌شود. در فصل سوم به شرح و تفصیل در رابطه با نمونه‌ای از روش‌های بازسازی اسپکتروگرام پرداخته می‌شود.

۱-۲ تعریف مسئله و ضرورت انجام پروژه

با توجه به کاربرد وسیع سیستم‌های بازشناسی گفتار در شرایط متنوع صوتی، مقاوم سازی پروسه شناسایی نسبت به انواع تنوعات امری ضروری بوده و این مسئله ای است که در حال حاضر بیشتر پژوهشگران این شاخه تحقیقاتی را به خود مشغول کرده است. گروهی از محققان بر روی بازشناسی گفتار مقاوم نسبت به نویز محیط، گروه دیگر روی شناسایی مقاوم نسبت به تنوعات نویز جمعی و گروه‌های دیگری روی بازشناسی مقاوم نسبت به تنوعات گوینده و یا حتی تنوع نرخ صحبت مشغول بررسی و تحقیق هستند. همه این سیستم ها در یک شرایط معین و ثابت از لحاظ تنوعات کارایی خوبی نشان می‌دهند ولی با تغییر تنوعات فوق به شدت کارایی آن‌ها کاهش می‌یابد. این در حالی است که سیستم درک گفتار در انسان نسبت به همه این تنوعات مقاوم بوده و رویکرد مشابهی نشان می‌دهد؛ بنابراین به نظر می‌رسد باید به دنبال روش‌ها و ایده‌هایی بود که برای بیشتر این تنوعات قابل تعمیم باشند. هدف از پیشنهاد این پروژه یافتن روش‌هایی برای اصلاح و بهینه‌سازی بردارهای بازنمایی صوتی به گونه‌ای است که تا جای ممکن اثرات ناشی از نویز جمعی و سطح

1 Spectrogram Reconstruction methods