دانشگاه صنعتی امیرکبیر

(پلی تکنیک تهران)

دانشکده: مهندسی پزشکی

پایان نامه کارشناسی ارشد

رشته تحصیلی: بیوالکتریک

عنوان

بهبود کیفیت سیگنال گفتار آغشته به نویز و اعوجاج توسط شبکه‌های عصبی

استاد راهنما

دکتر سیدعلی سیدصالحی

دانشجو

لوییزا دهیادگاری

بسمه تعالی

شماره:

تاریخ :

معاونت پژوهشی
فرم پروژه تحصیلات تکمیلی ۷

فرم اطلاعات پایان نامه
کارشناسی ارشد و دکترا

۱-مشخصات دانشجو

نام و نام خانوادگی: لوییزا دهیادگاری ☑ دانشجوی آزاد ☐ بورس ☐ معا

شماره دانشجویی : ۸۲۱۳۳۲۲۳ دانشکده : مهندسی پزشکی رشته تحصیلی : بیوالکتریک

نام و نام خانوادگی استاد راهنما : دکتر سیدعلی سیدصالحی

عنوان به فارسی : بهبود کیفیت سیگنال گفتار آغشته به نویز و اعوجاج توسط شبکه‌های عصبی

عنوان به انگلیسی : Quality improvement of speech signal due to noise and disturbance using neural networks

نوع پروژه: کاربردی ☑ بنیا ☐ توسعه ای ☑ ای ☐

تاریخ شروع : ۱۳۸۳/۱۱/۱ ۵ تاریخ خاتمه : ۱۳۸۴/۱۲/۲۶ تعداد واحد : ۹

سازمان تأمین کننده اعتبار : مرکز تحقیقات مخابرات ایران

واژه های کلید به فارسی :

شبکه های عصبی بازگشتی ، مقاوم بودن ، بازشناسی گفتار ، جاذب ، تنوع ، دینامیک‌های غیرخطی

واژه های کلیدی به انگلیسی :

Recurrent Neural Network, Robustness, Speech Recognition, Attractor, Variability, Nonlinear Dynamics

نظرها و پیشنهادها به منظور بهبود فعالیت های پژوهشی دانشگاه :

استاد راهنما : دکتر سیدعلی سیدصالحی

دانشجو : لوییزا دهیادگاری

امضاء استاد راهنما : تاریخ : ۱۳۸۴/۱۲/۱۰

نسخه ۱ : معاونت پژوهشی

نسخه ۲ : کتابخانه و به انضام دو جلد پایان نامه به منظور تسویه حساب با کتابخانه و مرکز اسناد و مدارک علمی

چکیده :

انسان در ادراک گفتار روزمره با انواع تنوعات در سیگنال ورودی برخورد می‌کند و علیرغم آنها وظیفه درک به خوبی انجام می‌شود. به عنوان نمونه‌هائی از این تنوعات در بازشناسی گفتار می‌توان از نویزهای مختلف مانند نویزهای جمعی یا نویز کانال نام برد که به صورت ایستان و یا غیرایستان به سیگنال گفتار اضافه می‌شوند. مشاهده می‌شود که کارائی سیستمهای بازشناسی گفتار به عنوان مدلهائی از درک گفتار در انسان با تغییر تنوعات بشدت افت پیدا می‌کند.

در این پروژه از توانائی‌های شبکه‌های عصبی با اتصالات بازگشتی برای کاهش میزان نویز ، اعوجاج و تنوعات ناخواسته از سیگنال گفتار استفاده می‌شود. ساختارهای مختلف شبکه‌های عصبی بازگشتی که به منظور بازشناسی گفتار در سطح بازشناسی آوا طراحی و پیاده سازی شده‌اند، برای حذف نویز از سیگنال گفتار مورد بررسی قرار می‌گیرند و نتایج بدست آمده از آنها با یک شبکه عصبی ساده که اتصالات بازگشتی در آن ملحوظ نشده است، مقایسه می‌شوند.

در آزمایشات اولیه نحوه عملکرد شبکه‌های بازگشتی و نحوه به قعر رفتن الگوهای نویزی با استفاده از چند نمونه ساده بررسی شده‌اند. در آزمایشات بعد ساختارهای شبکه عصبی بازگشتی به منظور بازشناسی گفتار در سطح بازشناسی آوا طراحی و پیاده سازی شده است که با هدف بازشناسی سیگنال گفتار نویزی مورد بررسی قرار می‌گیرد. ساختار این شبکه در طی آزمایشات مختلف بررسی و به تدریج کامل می‌گردد. در انتها ساختاری از شبکه عصبی بازگشتی طراحی شد که می‌تواند با استفاده از دور زدن در شبکه و به قعر رفتن الگوها، نمونه‌های نویزی شده را از روی الگوهای تمیزی که به شبکه تعلیم داده شده است بازیابی کند. در طی آزمایشات از دادگان صحبت یک نفر از گویندگان استفاده شد و پس از به دست آوردن یک ساختار نهائی از شبکه عصبی عملکرد شبکه در مورد دادگان زیاد و افراد مختلف نیز مورد ارزیابی قرار گرفت.

بهترین مدل از شبکه‌های عصبی بازگشتی توانسته است درصد صحت بازشناسی سیگنال نویزی با نویز صفر دسی‌بل را برای دادگان تعلیم ده جمله از یک نفر ۲۰٪ و برای ۴۰۰ جمله از نفرات زیاد ۲۱٪ نسبت به یک شبکه ساده که در آن اتصالات بازگشتی ملحوظ نشده است و خاصیت حذف نویز را ندارد، بهبود دهد.

فصل اول

مقدمه

۱-۱- مقدمه

درک انسان از گفتار در رویاروئی با تغییرات وسیع چه در آزمایشگاه و چه در حالت طبیعی مقاوم است. به عبارتی انسان در بازشناسی الگوها بسیار مقاوم عمل می‌کند. به عنوان مثال تصویر چهره یک فرد را علیرغم تغییر ابعاد آن، جهت چهره و ... به خوبی تشخیص می‌دهد و سیگنال گفتار را علیرغم آنکه تحت تأثیر عوامل مختلف تغییر یافته باشد به خوبی تشخیص می‌دهد. این عوامل تأثیر گذار روی سیگنال عوامل متعددی می‌توانند باشند و تأثیر خاص خود را روی سیگنال هر یک به صورت خطی یا غیر خطی ایجاد می‌نمایند. انسان در ادراک گفتار روزمره با انواع تنوعات در سیگنال ورودی برخورد می‌کند و علیرغم آنها وظیفه درک به خوبی انجام می‌شود. به عنوان نمونه هائی از این تنوعات در بازشناسی گفتار می‌توان از تغییر گوینده، تغییر بلندی صدا، تغییر تأثیرات محیط مثل دیوارها و ... ، تغییر در کانال انتقال مثل بلندگو، تلفنی، لهجه، تغییر میکروفون تغییر در سرعت بیان و ... نام برد [۲].

برای بازشناسی گفتار در محیطی که شامل صداهای مختلف است انسان باید دو مسئله ادراکی را حل کند. اول تداخل صداهائی که به گوش می‌رسند که باید منبع هر یک را به صورت متمایز درک کند، یا به عبارت دیگر نسبت به تنوعات گفتار مقاوم باشد. و دوم بازشناسی گفتار وقتی که نشانه‌های صوتی در بین سایر صداها و منابع نویز مفقود شده و یا تضعیف می‌شوند [۱۸].

پیشرفتهای اخیر در بازشناسی گفتار تحسین برانگیز بوده است و سیستم های تشخیص گفتار در کنترل و هماهنگ کردن شرایط یادگیری و شرایط بازشناخت خوب عمل کرده‌اند. اما فناوری بازشناخت در موارد نویزی شکست خورده است و بازشناسی مقاوم گفتار در محیطهای‌صوتی نویزی هنوز یک مسئله بزرگ حل نشده باقی مانده است [۱۸]. بازشناسی گفتار در حضور نویزهای مختلف به دلیل عدم هماهنگی بین مدلهای صوتی و داده‌هائی که توسط نویز تخریب شده‌اند، یک کار دشوار می‌باشد. روشهای مرسوم برای بهبود مقاوم‌سازی بازشناسی گفتار سعی دارند که این عدم هماهنگی را با استفاده از روشهای مختلف کاهش دهند. یکی از این روشها که در بازشناسی مقاوم گفتار نویزی استفاده می‌شود، شبکه‌های عصبی بازگشتی می‌باشد. در استفاده از شبکه‌های عصبی بازگشتی برای بازشناسی مقاوم گفتار نویزی سعی می‌شود از مزایای این شبکه‌ها مانند تعلیم گسسته، توانائی محاسبه ثابتهای زمانی و امکان ترکیب طبقه‌بندی و مقداردهی، بهره‌برداری شود [۱۳].

رویاروئی با نویز جمعی یک ایده رایج در بازشناسی گفتار مقاوم در برابر نویز است. روشهای مختلفی برای بهبود مقاوم سازی بازشناسی گفتار پیشنهاد شده است که از آن جمله می‌توان به تعلیم موقعیتهای مختلف، بازیابی گفتار و بهنگام کردن مدلهای آماری برای واحدهای گفتار نام برد. بازیابی گفتار به تمیز کردن گفتار نویزی با کم کردن نویز تخمینی از سیگنال گفتار نویزی کمک می‌کند. در این تحقیق سعی می‌شود با تعلیم سیگنال تمیز به شبکه عصبی بازگشتی به کاهش نویز سیگنالهای گفتار نویزی کمک شود [۲۰].

هدف این پروژه استفاده از توانائی‌های شبکه‌های عصبی با اتصالات بازگشتی در کاهش میزان نویز ، اعوجاج و تنوعات ناخواسته از سیگنال گفتار است. ساختار شبکه عصبی بازگشتی بگونه‌ای است که در آن حالت شبکه در هر زمان به ورودی‌های قبلی وابسته است. همچنین یک شبکه عصبی بازگشتی قابلیت ذخیره تغییرات زمانی الگوهای گفتار را برای تمایز خوب و مناسب دارد. اتصالات بازگشتی مثل یک حافظه کوتاه مدت عمل می‌کنند که در موارد بازشناسی مفید است.

در این شبکه‌ها در اثر دور زدن، حالت شبکه تغییر تدریجی پیدا می‌کند و بهنگام می‌شود. تغییرات تدریجی شبکه در برخی شرایط به سمت کاهش انرژی پیش می‌رود تا اینکه سرانجام شبکه به یک حالت تعادل متناظر با یک کمینه محلی انرژی یا قعر می‌رسد. اگر این کمینه های انرژی متناظر با مجموعه ای از الگوهای تعلیم از پیش تعیین شده باشند ، شبکه عصبی به صورت یک حافظه عمل خواهد کرد . به عبارت دیگر اگر از نسخه اعوجاج یافته یک الگوی ذخیره شده شروع کنیم ، شبکه قادر است بردار حالت مربوط به الگوی ذخیره شده را به یاد آورد. چنین توانائی در تصحیح خطا در روشهای بازشناسی الگوها و بازیابی اطلاعات بکار می رود [۳].

در واقع می‌توان این مسیر را مقدمه‌ای برای استفاده از شبکه‌های عصبی برای تصمیم‌گیری بهینه دانست. به این ترتیب امید می‌رود که بتوان سیستم‌های بازشناسی گفتار را نسبت به موارد فوق مقاوم ساخت. این مدل می‌تواند به عنوان فیلتر غیر خطی پردازش اولیه قبل از سیستم‌های بازشناس بکار رود و در حذف نویز و تنوعات ناخواسته کمک کند. این تنوعات از قبل برای مدل تعریف نمی‌شوند و در شبکه عصبی حتی الامکان سعی می‌شود که سیگنالهای بهنجار (نرمال) به عنوان قعرهای بستر جذبها شکل بگیرند. و با حرکت به سمت قعر امکان حذف تنوعات ناخواسته از سیگنال گفتار فراهم آید.

در ادامه این رساله ابتدا به بررسی اصول بازشناسی گفتار نویزی می‌پردازیم و بعد از آن مروری خواهیم داشت بر تحقیقاتی که در این زمینه انجام شده‌اند. سپس به معرفی دادگان استفاده شده در آزمایش و آزمایشات اولیه انجام شده می‌پردازیم. در فصول بعد آزمایشات انجام شده در این پروژه را بیان و نتایج آنها را ارزیابی می‌کنیم. ساختارهای شبکه‌های عصبی بازگشتی که به منظور کاهش نویز سیگنال گفتار طرح و بررسی شده‌اند و نتایج آزمایشاتی که بر روی این شبکه‌ها انجام شده است به ترتیب در فصول بعد مطرح می‌شوند. در انتها نیز به بررسی نتایج و یافته‌های این تحقیق می‌پردازیم و راهکارهائی برای بهتر کردن نتایج، در آینده بیان خواهیم کرد.

فصل دوم

نگرشی بر بازشناسی مقاوم گفتار نویزی

۲-۱- مقدمه

با رشد روزافزون استفاده از سیستم‌های گفتار در کاربردهای عملی و روزمره نیاز به حفظ راندمان بازشناسی گفتار در محیطهای واقعی به عنوان امری اجتناب ناپذیر مطرح گردیده است. شرایط ایده‌ال و عاری از نویزی که در کارها و شبیه‌سازیهای کامپیوتری در نظر گرفته می‌شود در بسیاری از کاربردهای واقعی به صورت جدی نقض می‌شود. به عنوان مثال وجود نویز سر و صدای محیط، انعکاس دیوار و تغییر در کانالهای انتقال باعث بر هم خوردن شرایط آزمایشگاهی می‌شود.

بنابراین هنگامی گه از سیستم بازشناسی گفتار که در محیط آزمایشگاهی آموزش داده شده است، در محیط واقعی استفاده می‌شود اغلب راندمان سیستم بازشناسی به دلیل عدم انطباق دادگان آموزشی آزمایشگاه و داده آموزشی جمع آوری شده در محیط واقعی به مقدار زیادی کاهش می‌یابد. از این رو مبحث مقاوم‌سازی در برابر نویز به عنوان یکی از ضرورتهای کاربردی و عملی از زمینه های فعال تحقیقاتی در سالهای اخیر بوده است.

۲-۲- مفاهیم بازشناسی گفتار

۲-۲-۱- محیط صوتی

محیط صوتی نه تنها بر تولید گفتار تأثیر می‌گذارد بلکه ممکن است سیگنال گفتار را نیز آلوده کند. اگر نسبت سیگنال به نویز در سیگنال نویزی پائین باشد راندمان سیستم‌های بازشناسی گفتار به مقدار زیادی کاهش می‌یابد. در حالی که سیستم شنوائی انسان تا حدود زیادی سیگنال گفتار نویزی را درک می‌کند.

در محیطهای نویزی، دره‌های طیف نسبت به پیکهای طیف بیشتر تحت تأثیر نویز قرار می‌گیرند. علاوه بر آن نویز سفید گائوسین اضافه شده به سیگنال باعث کاهش نرم بردارهای کپسترال[1] بدست آمده از پیش‌بینی خطی[2] می‌گردد. همچنین نویز غیر ایستان و دیگر تغییرات ایجاد شده به وسیله محیط صوتی، هنوز به طور کامل و جامع در سیستم‌های بازشناسی گفتار مورد بررسی قرار نگرفته است [۱].

۲-۲-۲- تغییرات گوینده

این تغییرا ت می‌تواند به وسیله عواملی از قبیل سبک صحبت کردن، نرخ صحبت کردن، کیفیت صدا، استرس ایجاد شده بر اثر شرایط محیط و هیجان بوجود آید. استرس و خصوصاً اثر لمبارد[3] (یعنی تغییر گفتار توسط گوینده به دلیل نویز موجود در محیط) در چندین مطالعه مورد بررسی قرار گرفته‌اند [۲ ، ۱۴]. از مطالعات انجام گرفته مشخص می‌شود که شیب طیفی، توزیع انرژی، فرکانس‌های فرمانت و نرم ضرایب کپسترال به علت اثر لمبارد تغییر می‌کند. به هر حال وجود تفاوت بین گفتار تولید شده در شرایط بدون نویز و گفتار تولید شده در شرایط نویزی قابلیت بازشناسی گفتار را مشکل می‌سازد.

۲-۲-۳- نویز

[1] Cepstral
[2] Linear Prediction
[3] Lombard

هر عاملی که باعث تغییر مشخصات سیگنال گفتار، قبل از دریافت توسط گیرنده گفتار شود را به عنوان نویز تعریف می‌کنیم. که این نویز می‌تواند بعلت اثرات محیط صوتی بر روی سیگنال گفتار و یا گوینده باشد.

عموماً نویزها به دو دسته کلی ایستان[1] و غیرایستان[2] تقسیم بندی می‌شوند. نویز ایستان نویزی است که چگالی طیف توان آن در طول زمان ثابت است، ولی نویز غیرایستان دارای خواص آماری متغیر در طول زمان است. نویز ایستان از قبیل نویز ایجاد شده توسط فن کامپیوتر و یا دستگاههای تهویه می‌باشد. نویز غیرایستان از قبیل بهم خوردن درب، صدای رادیو، تلویزیون و صدای گوینده‌های دیگر است.

از لحاظ اثرگذاری نویز بر روی سیگنال، نویزها به دو دسته نویز جمعی و اعوجاج کانال تقسیم‌بندی می‌شوند. اثر نویز جمعی بر روی سیگنال گفتار به صورت اضافه شدن نویز به سیگنال در حوزه زمان است و در حقیقت سیگنال نویز با سیگنال گفتار جمع می‌شود. در حالی که اعوجاج کانال و مشخصه فرکانسی آن در سیگنال گفتار در حوزه فرکانس ضرب می‌شود و در حوزه زمان در سینال گفتار کانوالو می‌شود.

در سیستم‌های بازشناسی گفتار از نویزهای مختلفی جهت تست و یا آموزش این سیستم‌ها استفاده می‌شود. اگر سیستم بازشناسی گفتار توسط گفتار تمیز آموزش داده شود سیستم را آموزش تمیز می‌نامند. و اگر بازشناسی با گفتار نویزی در نسبتهای سیگنال به نویز[3] مختلف آموزش داده شود، به آن سیستم آموزش چند حالته گفته می شود. معمولاً در سیستم‌های بازشناسی جهت تست سیستم از نویزهای کارخانه، ماشین و هواپیما بیشتر استفاده می‌شود. سیگنال نویز مشخصات طیفی سیگنال گفتار را تغییر می‌دهد که بسته به نوع نویز این تغییرات متفاوت است.

۲-۲-۳-۱- میکروفون و کانال انتقال

[1] Estationary
[2] Non Estationary
[3] Signal to Noise Ratio (SNR)

بعد از مرحله تولید گفتار و قبل از رسیدن گفتار به شنونده، سیگنال گفتار ممکن است توسط میکروفون و کانال انتقال دچار اغتشاش شود. این اغتشاشات ایجاد شده در سیگنال گفتار عموماً به صورت نویز کانوالوی هستند.

۲-۲-۳-۲- نویز جمعی

اگر نویزی که بر سیگنال گفتار تأثیر می‌گذارد در حوزه طیف با سیگنال گفتار جمع شود، این نویز را نویز جمعی می‌نامند. سیگنالی که با میکروفون صحبت نزدیک بدست آید دارای مقدار کمی نویز جمعی و پژواک می‌باشد و سیگنال گفتار حاصل از میکروفونی که به دهان گوینده نزدیک نیست ممکن است مقدار زیادی نویز جمعی و پژواک داشته باشد.

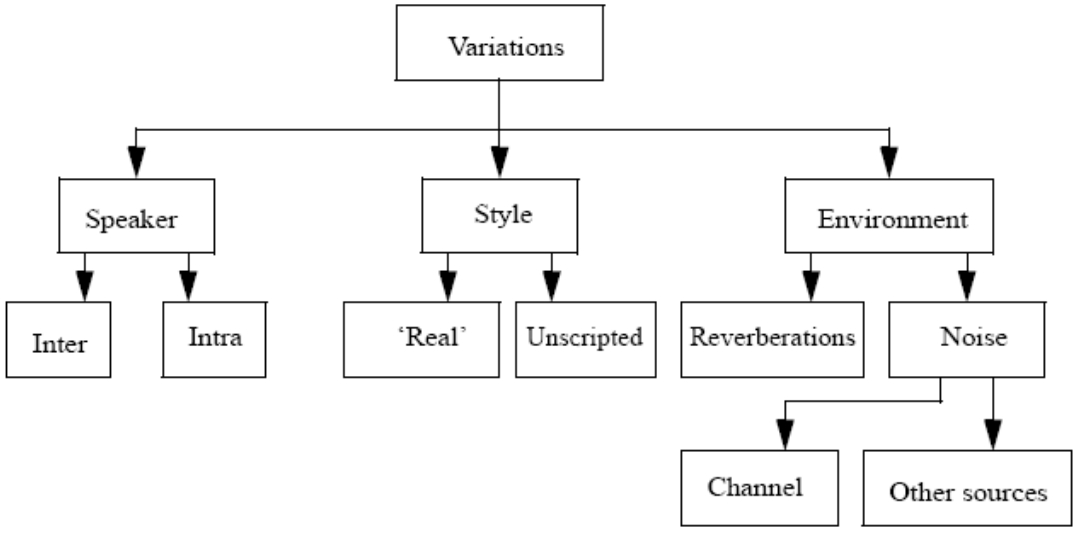
رویاروئی با نویز جمعی یک ایده رایج در بازشناسی گفتار مقاوم است. نویز جمعی می‌تواند به صورت ایستان یا غیرایستان باشد. انسان توانائی بازشناسی گفتاری را که با هر یک از این دو نوع نویز تخریب شده باشد دارد [۱۰ ، ۱۲].

۲-۳- منابع تنوعات در سیستم‌های بازشناسی گفتار

کارائی یک سیستم بازشناسی گفتار وابسته به شرایطی است که تحت آن ارزیابی می‌شود. اغلب سیستم‌های بازشناس گفتار می‌توانند تحت شرایط مناسب به دقتی همانند دقت انسان برسند. برای مثال می‌توان به دقت بازشناسی کلمه در حدود ۹۷٪ در پایگاه داده *DARPA* اشاره کرد [۱۲]. ولی هنگامی که بین شرایط تعلیم و تست عدم انطباق وجود دارد سخت است که شرایط مقاوم را بدست آوریم. در کاربردهای ساده مثل بازشناسی ارقام تلفنی ۹۹٪ دقت را در شرایط تمیز داشته‌ایم در حالی که این درصد وقتی که تغییرات کانال تلفن در دست نیستند به ۴۹٪ کاهش پیدا کرده است. عدم انطباق ممکن است به دلایل مختلف بوجود آید. برای مثال وقتی که سیستم با گفتار تمیز تعلیم می‌بیند و گفتار ورودی در طی بازشناسی تخریب می‌شود. یا وقتی که سیستم با سایر انواع تنوعات

مثل تغییرات صوتی، فونتیکی و تغییرات دیگر یا تنوعاتی که از اختلاف در لهجه‌ها ایجاد می‌شود، مواجه می‌شود.

در مورد عدم انطباقهای شناخته شده سیستم‌های بازشناس گفتار می‌توانند طراحی شوند و با یک پایگاه داده بزرگ گفتار که انواع مختلف تنوعات گفتار را با هم بکار می‌گیرد تعلیم ببینند و به طور قابل قبولی در طی کاربردهای واقعی عمل کنند. ولی این سیستم‌ها به دلیل آن که فقط برای بازشناسی گفتار تعلیم دیده‌اند و برای بازیابی گفتار تعلیم ندیده‌اند، هزینه بالائی را در بر دارند. بازشناسی قابل توجه بخاطر تأثیرات مختلف و غیر قابل پیش بینی مثل تغییرات محیط سخت است. در مجموع سیستم شنوائی انسان نسبت به همه انواع این تنوعات حتی در نسبتهای سیگنال به نویز خیلی کم مقاوم است [۱۴]. فاکتورهای اصلی که شرایط سیستم بازشناسی گفتار را در شرایط تست تخریب می‌کند در شکل ۲-۱ آمده است.



شکل ۲-۱- منابع تنوعات در سیستم‌های بازشناسی مقاوم گفتار[۱۳]

- تغییرات گوینده: این تغییرات منجر به تغییراتی در سیستم های صوتی بین افراد می‌شود. لهجه نیز نقش مهمی در این مقوله بازی می‌کند.
- تغییرات درون گوینده: این تغییرات شامل فاکتورهائی مثل مشرب گوینده، سلامت فیزیکی، حالت احساسی و ... در جلسات مختلف می‌باشد. سیستم های بازشناسی گفتاری که برای

ارتباط با وابستگی‌های به گوینده طراحی شده‌اند بازشناسی خیلی ناچیزی برای درصد کوچکی از گوینده‌ها دارند. علت اصلی اختلافات بین گوینده‌ها اختلاف بین آناتومی، شرایط صوتی، سن و لهجه است و می تواند با طراحی سیستم‌های بهنگام کننده گوینده کاهش یابد.

- تأثیر لمبارد: دیده می‌شود که در حضور نویز زمینه زیاد گوینده‌ها گفتار خود را تغییر می‌دهند تا برای سایرین بهتر قابل درک باشد.
- وابستگی به متن و زمینه و روش گفتگو: این عوامل تغییرات مهمی بین شرایط بازشناسی و تعلیم ایجاد می‌کنند. دیده شده است که دقت بازشناسی یک سیستم که با یک گفتار خاص برای هدف بازشناسی تعلیم دیده است، وقتی که استفاده از گفتارهای ناگهانی منجر به تنوعاتی در ریت گفتار، نحوه گفتارو کرولیشن‌ها می‌شود خیلی کاهش می‌یابد. در کنار این تغییرات، تغییر در حوزه (به عنوان مثال مدلهای زبانی تعلیم دیده شده در محل کار مشابه مدلهای زبانی تعلیم دیده در مشاورات یا مذاکرات است) بخاطر عدم انطباق مدل زبانی و خروج واژگان یک زبان باعث کاهش دقت می‌شود.
- تغییرا ت محیط: این تغییرات بخاطر نویز جمعی، نویز کانوال شده، انعکاسها و ... ایجاد می‌شوند و باعث ایجاد یک تخریب قابل توجه در شرایط گفتار می‌شوند. نویز جمعی در اثر آلوده شدن سیگنال گفتار با سایر منابع صدا ایجاد می‌شود. این منابع می‌تواند شامل سایر گوینده‌ها یا نویزی باشد که از منابع زمینه مثل ترافیک، موزیک، صدای کارخانه و ... بدست می آید.

نویز جمعی از طریق شیفت دادن میانگین طیفی و کوچک کردن واریانس توزیع روی سیگنال تأثیر می‌گذارد. تغییر در کانال تلفن، تغییر در میکروفون یا اضافه شدن یک گوینده باعث تخریبهای کانال در سیگنال گفتار می‌شود. این تخریبها یک تغییر جمعی در دامنه لگاریتمی ایجاد می‌کنند. و نیز یک سطح آستانه متغیر با زمان ایجاد می‌کنند. انعکاس‌ها که نمونه‌های تغییر یافته

گفتار هستند در اثر انعکاس از اشیا یک اتاق ایجاد می‌شوند و باعث تغییر کانوال شده در طیف گفتار می‌شوند.

تعلیم یک سیستم بازشناس با همه انواع ممکن نویزکه ممکن است در طی بازشناسی اتفاق بیافتد شرایط بازشناسی را بهبود می‌دهد. گرچه این روش معایبی دارد که از آن جمله نیاز به داشتن یک پایگاه داده بزرگ است.

خصوصیات کانال ارتباطی و نویز زمینه مهمترین فاکتورها در مدلسازی گفتار نویزی می‌باشند. به طور کلی کانال ارتباطی رفتارهای غیرخطی دارد که می تواند دلایل مختلفی داشته باشد. به عنوان مثال تقویت نویز منجر به کنترل اتوماتیک بهره و نویز غیرخطی منجر به سوئیچ کردن روی شبکه های تلفنی می‌شود. گرچه برای سادگی تأثیرات کانال را خطی در نظر می‌گیرند. گفتار نویزی می‌تواند به عنوان کانولوشن خطی یک سیگنال گفتار تمیز $x(t)$ با پاسخ ضربه کانال $h(t)$ مدل شود.

$$y(t)=h(t).x(t)+n(t) \qquad (۱-۲)$$

که $n(t)$ مشخص کننده تخریب کانال و نویز است. گرچه با در نظر گرفتن اینکه کانال اغلب برای گفتار تمیز طراحی می‌شود تأکید روشهای ASR اغلب بر روی نویزهای زمینه جمعی است. طیف قدرت کوتاه مدت گفتار نویزی طبق رابطه زیر می باشد:

$$P_y(f)=P_x(f)+P_n(f) \qquad (۲-۲)$$

که $P_x(f)$ و $P_n(f)$ تخمین طیف کوتاه مدت گفتار تمیز و نویز مربوطه را مشخص می‌کنند.

۲-۴- بازشناسی مقاوم گفتار نویزی

نویز زمینه جمعی که گفتار را تخریب می‌کند می‌تواند به صورت ایستان یا غیرایستان باشد. در مورد نویز ایستان توزیع بردارهای الگو در گفتار تخریب شده (نویزی) شبیه به توزیع یاد گرفته

شده از روی داده های تعلیمی نیست. این عدم انطباق باعث بازشناسی ضعیف و طبقه بندی نادرست می‌شود [۱۵].

تعلیم سیستم بازشناس با گفتاری که سطح یکسانی از نویز را با گفتار مورد بازشناسی دارد، می‌تواند بازشناسی را بهبود دهد. گرچه یک سیستم بازشناس گفتار که روی یک نویز تعلیم می‌بیند نمی‌تواند دقت بازشناسی را برای انواع نویز تضمین کند. یک راه برای حل این مسئله تعلیم چند موقعیتی[1] است، که در آن داده های تعلیم با انواع نویزها مخلوط می شوند و نسبتهای سیگنال به نویز مختلف به شبکه کمک می‌کند که برای نویزهای ناشناخته در طی بازشناسی سازمان‌دهی شود.

معایب تعلیم سیستم بازشناس با نویزهای جمعی این است که باعث افزایش خطا در تخمین طیف هر فریم گفتار می‌شود. و بنابراین تنوعات ذاتی بردارهای الگوی انواع صداها را افزایش می‌دهد. به عنوان نتیجه واریانس توزیع کلاسهای مختلف صدا افزایش می‌یابد و باعث افزایش خطای طبقه‌بندی می‌شود و بنابراین طبقه‌بندی ناصحیح را در طی موقعیتهائی که گفتارهای تعلیم و تست نویزی نیستند افزایش می‌دهد [۱۶].

وقتی که گفتار با نویز غیرایستان تخریب می‌شود تعلیم سیستم با گفتار تخریب شده با سطح نویز یکسان با گفتار تست کافی نیست. در این حالت سطح نویز در داده‌های تعلیم و تست مشخص نیست. انسان می‌تواند گفتاری را که با نویز ایستان و غیرایستان تخریب شده است تشخیص دهد [۱۲،۱۱]. برای بدست آوردن سطوح کارائی انسان لازم است که سیستم‌های بازشناس گفتار به اندازه کافی برای بدست آوردن انواع مختلف تنوعات مقاوم باشند.

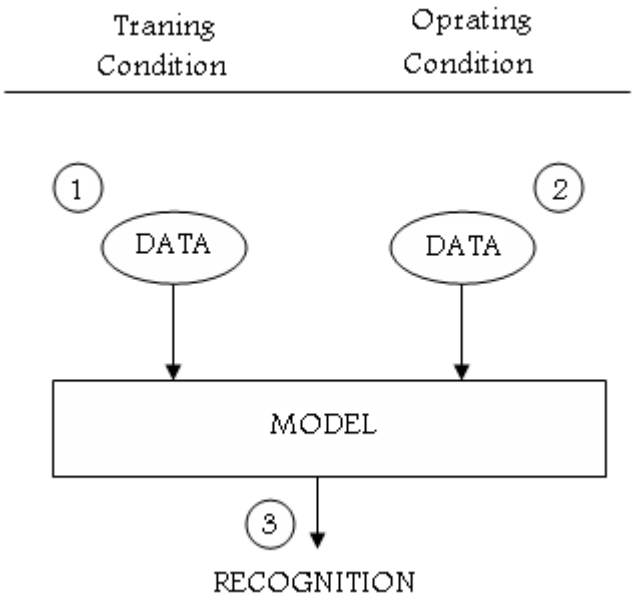
روشهای مختلفی برای بدست آوردن مقاوم‌سازی لازم در سیستم‌های بازشناس گفتار وجود دارد که می‌تواند در مقوله‌های زیر طبقه بندی شود:

[1] *multiconditional*

- استخراج ویژگی های مقاوم[1]
- بهسازی گفتار[2]
- جبران سازی مدل برای نویز[3]

۲-۴-۱- استخراج ویژگی های مقاوم

استخراج ویزگی های مقاوم را می‌توانیم توسط دیاگرام شکل ۲-۲ نمایش دهیم.



شکل ۲-۲- دیاگرام کلی استخراج ویژگی‌های مقاوم در برابر نویز[۱]

با توجه به اینکه پارامترهائی که جهت بازشناسی گفتار استفاده می‌شوند نسبت به اغتشاشات حساس می‌باشند به همین دلیل بررسی اثر نویز بر روی پارامترها مسئله مهمی است که در این دسته قرار می‌گیرد. هدف یافتن پارامترهای خاص و یا معیار شباهت مقاوم در برابر تغییرات حاصل از نویز بر روی گفتار است. از جمله روشهائی که در این دسته قرار می‌گیرند استفاده از بردارهای کپسترال

[1] Robust Speech Feature extraction
[2] Speech Enhansement
[3] Model base Compensation for noise

نرمالیزه شده، حذف تغییرات آرام سیگنال توسط فیلتر *RASTA*[1]، *CMN*[2]، مشتق زمانی، *SMC*[3] و *OSALPC*[4] می‌باشد.

روشهای بازشناسی مقاوم در این زمینه ساده هستند. این روشها بر روی تأثیرات نویز و استخراج الگوهای مقاوم در برابر نویز از گفتار نویزی متمرکز است. انتخاب مناسب نوع پارامترها به افزایش مقاوم سازی کمک می‌کند. به عنوان مثال نشان داده شده است که ضرائب کپسترال نسبت به ضرائب طیفی مقاوم‌ترند. چون از تغییرات کوچک طیفی بین صداهایی که متعلق به یک کلاس هستند جلوگیری می‌کنند. استفاده از رفتارهای سیستم شنوائی در استخراج الگو توانائی بازشناسی مقاوم گفتار را بیشتر بهبود می‌دهد. برای مثال مل-کپستروم و پیشگوئی خطی ادراکی[5] روشهای شناخته شده در این زمینه هستند [۱۴].

۲-۴-۱-۱- تجزیه و تحلیل اجزای اصلی

وابستگی‌های خطی بین مجموعه متغیرها می‌تواند با استفاده از انتقال *PCA* از بین برود. این روشها نه تنها می‌توانند برای غیر وابسته کردن مجموعه هائی از سطوح انرژی یک کپستروم استفاده شود بلکه برای ترکیب مجموعه پارامترها مثل الگوهای استاتیک و دینامیک کاربرد دارند.

PCA می‌تواند برای مدلسازی سه بعدی نواحی محلی استفاده شود. مطالعات اخیر برتری انتقال *PCA* را در *DCT* به عنوان بازنمائی در بازشناسی گفتار مقاوم چند بانده نشان داده است. کاربرد دوگانه *PCA* با یک عمل وزندار که به عنوان تجزیه و تحلیل گسسته خطی شناخته شده است می تواند اطلاعات گسسته لازم برای متمایز کردن صداهای گفتار را بکار گیرد و یک مجموعه از پارامترها که به آن ضرائب *IMELDA* گفته می‌شود و به صورت مناسب برای محاسبات فاصله ایوکلیدین وزندار می‌شوند را بسازد. مدارکی وجود دارد که در این محاسبات سیگنالهای تخریب شده

[1] RelAtive SpeTrAl
[2] Cepstral Mean Normalization
[3] Short-time Modified Coherence
[4] One Sided Autocorrellation LPC
[5] Perceptual Linear Prediction

می‌توانند مقاوم‌سازی را در حالت تخریب شده بهبود دهند، در حالی که آسیبی به سیگنالهای تخریب نشده وارد نمی‌شود [۱۸].

۲-۴-۱-۲- نرمالیزاسیون بردار الگو

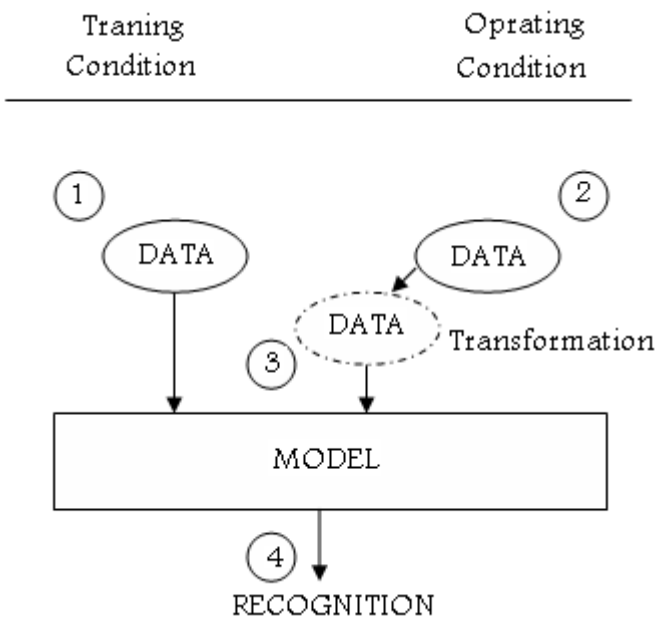
نرمالیزاسیون بردار الگو یا روشهای جبران تخریبها برای برگرداندن تأثیرات فیلترسازی خطی ناشناخته از منابعی مثل انعکاس اتاق و تغییر شکل طیفی با میکروفون استفاده می‌شوند. نرمالیزاسیون می‌تواند در دامنه طیفی یا در دامنه کپستروم انجام شود. در سایر موارد روش کلی این است که توان میانگین کل عبارت را از هر جزئی از بردار الگو کم کنند و سپس بر انحراف استاندارد تقسیم کنند.

هدف انتقال میانگین تشکیل فیلتر بالاگذر است در حالی که تقسیم بر انحراف معیار استاندارد برای کنترل اتوماتیک بهره *(AGC)* انجام می‌شود تا پارامترهای آمارای را بدون در نظر گرفتن شرایط نویزی یکسان نگه دارد.

روشهای سیستم بازشناسی گفتار می‌تواند با بکار گرفتن نرمالیزاسیون بردار الگوی تقسیم‌بندی شده بهبود یابد. در نرمالیزاسیون بردار الگوی چند قسمتی، خروجی‌های استخراج کننده الگو با یک رنج عددی و بدون در نظر گرفتن شرایط نویزی با در نظر گرفتن میانگین و انحراف معیار استاندارد یک قسمت مورد توجه تقویت می‌شوند.

۲-۴-۲- بازیابی گفتار

بازیابی گفتار بخاطر نیازبه بهبود سیستم‌های ارتباطی گفتار در شرایط نویزی مورد توجه قرار گرفته است. این روش یک پردازش برای کاهش نویز گفتار نویزی است. این پردازش بطور کلی شامل کم‌کردن نویز تخمینی از سیگنال گفتار نویزی است. هدف بهبود شرایط درک گفتار یا افزایش قابل فهم بودن آن است. دیاگرام کلی این روش در شکل ۲-۳ آمده است.



شکل ۲-۳- دیاگرام کلی روش بهسازی گفتار[۱]

قابل فهم بودن گفتار بازیابی شده می‌تواند در طی بازشناسی گفتار اندازه‌گیری شود. گرچه هنوز ثابت نشده است که روشهای پیش پردازش بازیابی گفتار در بهبود ریت بازشناسی انسان موفق بوده اند. اما تحقیقات نشان داده است که دقت انسان می‌تواند با کاهش شدید کیفیت گفتار بهبود یابد. به عبارت دیگر بهبود وضوح (قابل درک بودن) به صورت واضح در سیستم‌های بازشناسی گفتاری تشریح شده است که از سطوح کم نویز به صورت جدی تأثیر گرفته‌اند. در این مورد پیش پردازش بازیابی گفتار می‌تواند به مقدار زیادی وضعیت سیستم بازشناسی گفتار را در محیطهای نویزی بهبود دهد.

روشهای بازیابی گفتار همیشه توانائی کاهش ریت خطا در سیستم‌های بازشناسی مقاوم گفتار را ندارند. چون تغییر شکلهای تولید شده در طیف گفتار یک نتیجه از بازیابی انجام شده با استفاده از تخمین نویز نسبت به مقدار واقعی نویز است[۱].

۲-۴-۲-۱- تفریق طیف