بسم الله الرحمن الرحيم

دانشگاه علم و صنعت ایران

دانشکده مهندسی کامپیوتر

موضوع : اعمال روشهای بهبود گفتار به عنوان پیش‌پردازش، جهت بالا بردن دقت بازشناسی گفتار فارسی

ارائه کننده : محسن رحمانی

پایان نامه برای دریافت درجه کارشناسی ارشد

در رشته مهندسی کامپیوتر

استاد راهنما : دکتر احمد اکبری

دی ماه ۱۳۸۰

تقدیم

به پدر و مادر عزیزم

چکیده

کارایی سیستمهای بازشناسی در حضور نویز کاهش می‌یابد. در این پایان‌نامه روشهای مقاوم کردن سیستم بازشناسی در برابر نویز دسته‌بندی شده و از بین آنها روشهای مبتنی بر داده بررسی می‌شود.

روشهای مختلف تفاضل طیف، جبران کپسترال، نگاشت ضرایب و روشهای مبتنی بر ویژگیهای مقاوم تحلیل شده‌اند. یک رابطه جدید برای تسطیح نویز در تفاضل طیف و یک لیفتر جبران کننده برای وزندهی ضرایب کپسترال پیشنهاد شده است. همچنین ایده تخمین ویژگیهای مقاوم از ویژگی های نویزی ارائه شده است.

نرخ بازشناسی با استفاده از پیش‌پردازنده‌های پیشنهادی افزایش داشته است. نگاشت ضرایب کپسترال نویزی با شبکه‌های عصبی نتایج قابل قبولی داشته است. بین همه روشها، نگاشت ضرایب تفاضل کپسترال میانگین با استفاده از شبکه عصبی بهترین نتیجه را دارد.

بر خود لازم میدانم از استاد راهنمای گرامی خود، جناب آقای دکتر احمد اکبری که راهنمایی‌ها و کمک‌های ارزشمند ایشان نقش بسزایی در پیشبرد پروژه داشته است، تشکر و قدردانی کنم.

همچنین از دوستان گرامیم آقایان، مهندس بابک ناصرشریف و مهندس ستار هاشمی که از رهنمودها و کمکهای آنها در این پروژه بهره بسیار بردم، صمیمانه تشکر می‌کنم.

و نیز تشکر می‌کنم از تمامی کسانی که در ضبط پایگاه صدا و دیگر مراحل انجام پروژه یاریگر من بوده‌اند.

فهرست مطالب

فهرست اشکال

عنوان **صفحه**

فهرست جداول

عنوان **صفحه**

مقدمه

هدف نهایی سیستمهای بازشناسی گفتار فراهم آوردن قابلیت درک گفتار انسان بوسیله ماشین است[۱]. از آنجا که گفتار یکی از مهمترین ابزار ارتباطی انسان با محیط اطرافش محسوب می‌شود، ایجاد توانایی در برقراری ارتباط در ماشین‌ها، تحول بزرگی در استفاده از ماشین می‌باشد.

در یک سیستم بازشناسی گفتار، پارامترهایی مختلفی تعیین کننده درجه توانایی سیستم بازشناسی هستند. این پارامترها عبارتند از: وابسته و مستقل بودن از گوینده، بازشناسی کلمات مجزا و گفتار پیوسته، اندازه کتاب لغت، محدودیتهای زبانی، گفتار مکالمه‌ای و شرایط محیطی که بازشناسی در آن انجام می‌گیرد. در این پایان‌نامه روشهای مقاوم کردن سیستم بازشناسی در محیطهای نویزی بررسی می‌شود.

سیستمهای بازشناسی گفتار اساسا کلاسه‌بند[۱]هایی هستند که قطعات گفتار را به عنوان اعضای کلاسها طبقه‌بندی می‌کنند. سیستم بازشناسی توزیع بردارهای ویژگی را که از هر واحد گفتاری حاصل می‌شود، با استفاده از مجموعه‌ای از گفتار آموزشی فرا می‌گیرد. به هنگام بازشناسی هر قطعه گفتار به عنوان یک واحد گفتاری که توزیع آن به توزیعی از بردارهای ویژگی نزدیکتر باشد، کلاسه‌بندی می‌شود.

هرگاه گفتاری که باید تشخیص داده شود، با نویز تخریب گردد کارایی سیستمهای بازشناسی به شدت کاهش می‌یابد. اثر تخریبی سیگنال گفتار آن است که توزیع بردارهای ویژگی گفتار تخریب شده، شبیه توزیعی که سیستم بازشناسی با آن آموزش دیده است، نمی‌باشد. این عدم تطابق باعث کاهش کارایی سیستمهای بازشناسی در شرایط نویزی می‌شود.

روشهای مختلفی جهت مقابله با اثر نویز بر سیستمهای بازشناسی ارائه شده است. روشهای مقابله با نویز را می‌توان به سه دسته تقسیم کرد. دسته اول روشهایی هستند که از ویژگی‌های مقاوم به نویز استفاده می‌کنند. در این روشها ویژگی‌هایی از گفتار استخراج می‌شود که در مقابل نویز حساسیت کمتری داشته باشند. دسته دوم روشهایی هستند که بر پایه تخمین گفتار تمیز عمل می‌کنند. در این روشها گفتاری که باید تشخیص داده شود با استفاده از الگوریتم‌های بهبود گفتار، بهبود می‌یابد. و دسته سوم، روشهای مبتنی بر اصلاح مدل آکوستیکی سیستم بازشناسی هستند. در این روشها سعی می‌شود سیستم بازشناسی به گونه‌ای اصلاح شود که نسبت به نویز مقاوم باشد. در این پایان‌نامه دو دسته اول بررسی شده‌اند. اهمیت روشهای این دو دسته آن است که در آنها

۱-Classifier

نیازی به تغییر دادن شیوه تطبیق الگو نیست و فرایند مقاوم‌سازی فقط در مرحله پیش‌پردازنده[۱] سیستم بازشناسی انجام می‌شود.

فصل اول این پایان‌نامه سیستمهای بازشناسی گفتار را به طور کلی بررسی می‌کند و یک سیستم بازشناسی مبتنی بر مدل مخفی مارکف را به عنوان سیستم پایه معرفی می‌کند. فصل دوم به بررسی مدل نویز، تاثیر آن بر سیگنال گفتار و نتیجه این تاثیر بر کارایی سیستم بازشناسی می‌پردازد. در فصل دوم همچنین روشهای مقاوم کردن سیستم بازشناسی در برابر نویز دسته‌بندی شده و مشخصات هر دسته توضیح داده شده است. فصل سوم این پایان‌نامه درباره ویژگی‌های مقاوم در برابر نویز بحث می‌کند. در این فصل چند ویژگی مقاوم در برابر نویز را معرفی شده و نتیجه بکارگیری بعضی از آنها گزارش شده است. در فصل چهارم، چند روش مختلف بهبود گفتار معرفی شده و در آزمایشات مختلف، نتیجه اعمال آنها بر سیگنال گفتار نویز بررسی شده‌است. فصل پنجم پایان‌نامه روشهای ارائه در فصول قبل را باهم مقایسه کرده و چند پیش‌پردازنده برای سیستم بازشناسی گفتار پایه پیشنهاد می‌دهد. فصل ششم به نتیجه گیری و تحقیقات آتی اختصاص دارد.

۱- front-end

فصل اول

سیستمهای بازشناسی گفتار

بازشناسی گفتار با توجه به کاربردهای وسیع آن در ارتباطات، تبادل اطلاعات میان انسان و ماشین و بکارگیری ماشین درترجمه مکالمات از یک زبان به زبان دیگر، مورد توجه قرار گرفته است.

در این فصل مسائل بازشناسی گفتار همچون پیوسته یا گسسته بودن، وابسته یا مستقل از گوینده بودن، اندازه کتاب لغت، محدودیتهای زبان و محیط و نوع بیان گفتار بررسی شده‌اند. همچنین اجزای یک سیستم بازشناسی و مدل مخفی مارکف به عنوان جزیی از این سیستم بازشناسی مورد بررسی قرار گرفته‌اند و در انتهای فصل یک سیستم بازشناسی پایه معرفی شده است.

۱-۱ پارامترهای بازشناسی گفتار

پارامترهایی مختلفی در یک سیستم بازشناسی گفتار موثر هستند که تعیین کننده درجه پیچیدگی سیستم می‌باشند. این پارامترها عبارتند از: وابسته و مستقل بودن از گوینده، بازشناسی کلمات مجزا و گفتار پیوسته، اندازه کتاب لغت، محدودیتهای زبانی، گفتار مکالمه‌ای و شرایط محیطی که بازشناسی در آن انجام می‌گیرد. در این بخش این پارامترها به اختصار مورد بررسی قرار می‌گیرند.

۱-۱-۱ وابسته یا مستقل از گوینده

یک سیستم وابسته به گوینده فقط برای استفاده یک گوینده طراحی می شود، درحالیکه یک سیستم مستقل از گوینده برای استفاده بوسیله هر گوینده‌ای طراحی می‌گردد. بطور معمول سیستمهای وابسته به گوینده دقیق‌تر از یک سیستم مستقل از گوینده هستند و نتایج بهتری را ارائه می‌دهند. ضعف عمده سیستم وابسته به گوینده این است که هر بار نیاز به بازشناسی گفتار گوینده جدیدی باشد، لازم است سیستم مجددا برای این گوینده آموزش ببیند. سیستم میانه سیستمهای وابسته و مستقل از گوینده، سیستم چند گوینده است که برای تعداد ثابت و کم گوینده‌ها بکار می‌رود.

۱-۱-۲ گفتار مجزا/ متصل/ پیوسته

در بازشناسی کلمات مجزا[۱]، هر کلمه بصورت جداگانه و واضح بیان می شود و سیستم بازشناسی با هر کلمه بطور مستقل سروکار دارد. در بازشناسی کلمات متصل[۲]، دنباله‌ای از کلمات برای بازشناسی مورد توجه قرار می‌گیرند، ولی کلمات جمله باید بطور مجزا و با فواصل زمانی سکوت از هم جدا شوند. در بازشناسی گفتار

---

[۱] -Isolated word recognition

[۲]- Connected word recognition