بسم الله الرحمن الرحيم

دانشگاه صنعتی امیرکبیر

دانشکدهٔ مهندسی کامپیوتر و فنآوری اطلاعات

پایان‌نامه کارشناسی ارشد

رشته مهندسی کامپیوتر- هوش‌مصنوعی

مقاوم‌سازی سیستم‌های بازشناسی گفتار و گوینده در مقابل نویزهای جمع‌شونده و نویز کانال

نگارش:

امیرحسین حاج‌احمدی

استاد راهنما:

دکتر محمد مهدی همایون‌پور

استاد مشاور:

دکتر سید محمد احدی

۱۳۸۷

بسمه تعالي

دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

فرم اطلاعات پایان‌نامه
کارشناسی- ارشد و دکترا
کتابخانه مرکزی

شماره مدرک:

مشخصات دانشجو	نام خانوادگی : حاج‌احمدی	نام: امیرحسین	شماره دانشجویی: ۸۴۱۳۱۰۳۱
	دانشکده: کامپیوتر و فن‌آوری اطلاعات	رشته: هوش مصنوعی و رباتیک	گروه:
عنوان	مقاوم‌سازی سیستم‌های بازشناسی گفتار و گوینده در مقابل نویزهای جمع‌شونده و نویز کانال		
Title	Robust Speech And Speaker Recognition versus Additive and Convolutional Noises.		

استاد راهنما	نام خانوادگی: همایون پور	درجه و رتبه	استاد راهنما	نام خانوادگی:	درجه و رتبه
	نام: محمدمهدی	استادیار		نام:	
استاد مشاور	نام خانوادگی: احدی	درجه و رتبه	استاد مشاور	نام خانوادگی:	درجه و رتبه
	نام: سید محمد	استادیار		نام:	

دانشنامه	کارشناسی ○ ارشد ⊙ دکترا ○	سال تحصیلی: ۱۳۸۷
نوع پروژه	کاربردی ⊙ نظری ○ توسعه ای ○ بنیادی ○	

مشخصات ظاهری	تعداد صفحات ۱۳۶	تصویر ⊙ جدول ⊙ نمودار ⊙ نقشه ○ واژه‌نامه ○	تعداد مراجع ۶۱	ضمائم ⊙ تعداد صفحات: ۱۱
زبان متن	فارسی ⊙ انگلیسی ○	چکیده	فارسی ⊙	انگلیسی ⊙
یاداشت				
توصیفگر				
کلید واژه فارسی	*بازشناسی خودکار گفتار ، بازشناسی خودکار گوینده، مقاوم‌سازی در مقابل نویز، روش ویژگی‌های مفقود، پیک‌های طیف‌دنبالهٔ خودهمبستگی، پس‌پردازش MVA*			
Key word of English	Automatic speech recognition, Automatic speaker recognition, Noise robustness, Missing feature theory, Spectral peaks of autocorrelation sequence, MVA post processing.			

تقدیم به مادر صبورم و روح پاک پدرم

از جناب آقای دکتر همایون پور برای راهنمایی
اینجانب در انجام این پروژه و همچنین از جناب
آقای دکتر احدی به دلیل رهنمودهای اندیشمندانه
کمال تشکر را دارم.

چکیده

با توجه به اهمیت و نقش بالای مقاوم‌سازی سیستم‌های بازشناسی گفتار و گوینده در مقابل نویزهای محیطی، در این مجموعه ابتدا تأثیر استفاده از پیک‌های طیف دنبالهٔ خودهمبستگی به عنوان ویژگی‌های مقاوم برای استفاده در سیستم‌های بازشناسی گوینده مورد بررسی قرار گرفته است. سپس تأثیر روش‌های هنجارسازی ویژگی‌ها مانند روش حذف میانگین‌کپسترال، روش پس‌پردازش MVA و روش ویژگی‌های مفقود بر پیک‌های طیف دنبالهٔ خودهمبستگی برای مقاوم‌سازی در مقابل نویزهای جمع‌شونده و کانال در بازشناسی گوینده بررسی شده است. همچنین جهت استفاده از روش ویژگی‌های مفقود در سیستم‌های بازشناسی گوینده یک مدل کلی ارائه گردیده است.

در این پروژه همچنین برای افزایش کارایی روش ویژگی‌های مفقود و امکان استفاده از آن برای استخراج پیک‌های مقاوم‌تر از طیف دنبالهٔ خودهمبستگی، تغییراتی چه در تخمین ماسک و چه در بدست آوردن مدل بازسازی مناسب پیشنهاد شده است. استفاده از ماسک‌های مبتنی بر شبکه‌های عصبی که عمل تخمین ماسک را به کمک تخمین نسبت سیگنال به نویز محلی انجام می‌دهند و از قدرت تعمیم بالایی نیز برخوردار هستند و انتخاب مدل بازسازی مناسب با کمک فاصلهٔ وزن‌دار با وزن‌های حاصل از ماسک‌های نرم برای استفاده در سیستم‌های بازشناسی مبتنی بر بازسازی طیف‌نگار با روش ویژگی‌های مفقود، پیشنهاد و مورد آزمایش و بررسی قرار گرفته شده است.

آزمایش‌های انجام شده در راستای مقاوم‌سازی سیستم‌های بازشناسی گوینده، بر روی دادگان فارسدات تلفنی انجام شده است. از نتایج آزمایش‌های اولیهٔ مشخص شد که استفاده از انرژی طیفی سیگنال به عنوان یک ویژگی در سیستم‌های بازشناسی گوینده، مقاومت این سیستم‌ها را به میزان قابل توجهی کاهش می‌دهد. همچنین مشخص شد، پیک‌های طیف دنبالهٔ خودهمبستگی که در آنها از مشتق‌گیری طیفِ توان استفاده شده است، نسبت به ویژگی‌های MFCC دارای مقاومت بیشتری درمقابل به نویزهای جمع‌شونده هستند. همچنین نتایج نشان‌دهندهٔ مقاومت بالای پیک‌های طیف دنبالهٔ خودهمبستگی سیگنال‌گفتارِ استخراج شده از طیف‌نگار دنبالهٔ خودهمبستگی بازسازی شده با روش ویژگی‌های مفقود در مقابل نویزهای جمع‌شونده است. درنهایت برای سیستم‌های بازشناسی گوینده، استفاده از روش پس‌پرداز MVA را برای مقاوم‌سازی بیشتر ویژگی‌های استخراج شده در حوزهٔ کپسترال در مقابل نویزهای جمع‌شونده و کانال می‌توان پیشنهاد کرد.

در این پروژه استفاده از روش ویژگی‌های مفقود با استفاده از تخمین ماسک به کمک شبکه‌های عصبی و انتخاب مدل بازسازی مناسب با استفاده از فاصلهٔ وزن‌دار، و نیز روش پس‌پردازش MVA برای استفاده در سیستم‌های بازشناسی گفتار مقاوم در مقابل نویز نیز مورد آزمایش قرار گرفته‌اند. نتایج آزمایش‌های انجام شده جهت مقاوم‌سازی سیستم‌های بازشناسی گفتار بر روی دادگان Aurora2 نشان می‌دهند که هم روش ویژگی‌های مفقود و هم روش پس‌پردازش MVA از توانایی بالایی در مقاوم‌سازی سیستم‌های‌بازشناسی گفتار در مقابل نویزها برخوردار هستند. همچنین نتایج امیدوارکننده‌ای با استفاده از اعمال هر دو روشِ ویژگی‌های مفقود و روشِ پس‌پردازش MVA بدست آمده است.

کلمات کلیدی: *بازشناسی خودکار گفتار ، بازشناسی خودکار گوینده، مقاوم‌سازی در مقابل نویز، روش ویژگی‌های مفقود، پیک‌های طیف‌دنبالهٔ خودهمبستگی.*

فهرست مطالب

عنوان صفحه

چکیده ..... ه

۱- مقدمه ۲

۲- بازشناسی گفتار و گویندهٔ مقاوم در مقابل نویز ۷

۲-۱ مقدمه ۷

۲-۲ مختصری از تاریخچهٔ سیستم‌های بازشناسی گفتار ۱۰

۲-۳ مختصری از تاریخچهٔ سیستم‌های بازشناسی گوینده ۱۱

۲-۴ انواع نویزها و تخریب ناشی از آنها ۱۱

۲-۵ تقسیم‌بندی روش‌های مقاوم‌سازی موجود ۱۳

۲-۵-۱ استخراج ویژگی‌های مقاوم ۱٤

۲-۵-۲ هنجارسازی ویژگی‌های استخراج شده ۱۵

۲-۵-۳ ساخت مدل‌های بازشناسی مقاوم در مقابل نویز ۱٦

۲-۶ جمع‌بندی ۱۸

۳- مباحث پایه در استخراج ویژگی از سیگنال گفتار ۲۰

۳-۱ مقدمه ۲۰

۳-۲ محاسبهٔ طیف‌نگار ۲۰

۳-۳ محاسبهٔ طیف‌نگارِ مل ۲۱

۳-۴ مدل کردن تأثیر نویزهای جمع‌شونده بر ویژگی‌های طیفی ۲۲

۳-۵ محاسبهٔ ضرائب کپسترال مبتنی بر معیار مل ۲٤

۳-۶ تأثیر اعوجاج کانال بر ضرائب کپسترال ۲۶
۳-۷ تأثیر نویز جمع‌شونده بر ضرائب کپسترال ۲۷
۳-۸ جمع‌بندی ۲۸
۴- روش‌های هنجارسازی و ترکیب ویژگی‌های استخراج شده ۳۰
۴-۱ مقدمه ۳۰
۴-۲ روش هنجارسازی میانگین کپسترال (CMN) ۳۰
۴-۳ روش هنجارسازی واریانس ضرائب کپسترال ۳۱
۴-۴ فیلتر تخمینی ARMA ۳۱
۴-۵ روش پس پردازشی MVA ۳۲
۴-۶ روش ترکیب ویژگی‌ها ۳۳
۴-۷ جمع‌بندی ۳۴
۵- حوزهٔ خودهمبستگی و استخراج ویژگی‌های گفتاری مقاوم در مقابل نویز ۳۶
۵-۱ مقدمه ۳۶
۵-۲ نحوهٔ محاسبهٔ دنبالهٔ خودهمبستگی ۳۸
۵-۳ طیف‌نگار دنبالهٔ خودهمسبتگی ۳۹
۵-۴ استخراج پیک‌های طیف دنبالهٔ خودهمبستگی ۴۱
۵-۵ روش‌های استخراج پیک‌های طیف مقاوم‌تر از دنبالهٔ خودهمبستگی ۴۲
۵-۵-۱ فیلترکردن دنبالهٔ خودهمبستگی یک‌طرفه ۴۲
۵-۵-۲ محاسبهٔ مشتق طیف توان (DPS) ۴۳
۵-۵-۳ حذف تأخیرهای پایین ۴۳

۶-۵ جمع‌شونده ٤٤

۶- روش ویژگی‌های مفقود و جبران ویژگی‌های گفتاری با کمک آن ٤٦

۶-۱ مقدمه ٤٦

۶-۲ تقسیم‌بندی روش‌های مبتنی بر ویژگی‌های مفقود ٤٨

۶-۳ روش‌های تخمین ماسک ٥٠

۶-۳-۱ روش‌های مبتنی بر نسبت سیگنال به نویز محلی ٥١

۶-۳-۲ روش‌های مبتنی بر طبقه‌بندی کننده‌ها ٥٤

۶-۳-۳ روش پیشنهادی استفاده از شبکه‌های عصبی برای بدست آوردن ماسک‌های طیفی با استفاده از تخمین نسبت سیگنال به نویز ٥٦

۶-۳-۴ بردارهای ویژگی مبتنی بر طیف‌نگار پیشنهادی برای تخمین ماسک ٥٨

۶-۴ روش‌های بازسازی نواحی مفقود طیف‌نگار ٦٠

۶-۴-۱ درون‌یابی خطی ٦٠

۶-۵ روش‌های بازسازی مبتنی بر روش‌های آماری ٦٠

۶-۵-۱ استفاده از میانگین ٦١

۶-۵-۲ تخمین با حداکثر احتمال پسین (MAP) ٦٢

۶-۵-۳ تخمین کراندار حداکثر احتمال پسین ٦٤

۶-۶ استفاده از خوشه‌بندی در تخمین مقادیر مفقود ٦٧

۶-۶-۱ نحوهٔ محاسبهٔ مقادیر عضویت ٦٩

۶-۶-۲ روش پیشنهادی برای محاسبهٔ مقادیر عضویت ٧٠

۶-۷ ارائهٔ یک چارچوب کلی جهت اعمال روش ویژگی‌های مفقود بر سیستم‌های بازشناسی گوینده ٧٠

۶-۸ روش پیشنهادی برای اعمال روش ویژگی های مفقود بر ویژگی های حوزهٔ خودهمبستگی ۷۳
۶-۹ جمع‌بندی ۷۵
۷- آزمایش‌های انجام شده و نتایج آنها ۷۷
۷-۱ مقدمه ۷۷
۷-۲ مجموعه دادگان استفاده شده ۷۷
۷-۲-۱ بازشناسی گوینده (فارسدات تلفنی) ۷۸
۷-۲-۲ بازشناسی گفتار (مجموعه دادگان Aurora 2) ۸۱
۷-۳ چارچوب کلی مدل‌سازی انجام شده در سیستم‌های بازشناسی گوینده و گفتار ..۸۳
۷-۳-۱ بازشناسی گوینده ۸۳
۷-۳-۲ بازشناسی گفتار ۸۳
۷-۴ آزمایش‌های انجام‌شده در راستای مقاوم‌سازی سیستم‌های بازشناسی گوینده با استخراج ویژگی از حوزهٔ خودهمبستگی ۸٤
۷-۴-۱ تأثیر انرژی یک فریم بر مقاومت ویژگی‌ها در مقابل نویزهای جمع‌شونده ۸٤
۷-۴-۲ ویژگی‌های حوزهٔ خودهمبستگی در سیستم‌های بازشناسی گوینده ۸٦
۷-۴-۳ پیک‌های دنبالهٔ خودهمبستگی در مقایسه با ضرائب MFCC ۸۸
۷-۴-۴ روش پس‌پردازش MVA ۸۸
۷-۴-۵ روش ترکیب ویژگی‌ها ۹۰
۷-۴-۶ تأثیر اعوجاج کانال ۹۲
۷-۵ استفاده از روش دادگان مفقود در سیستم‌های بازشناسی گوینده ۹٤
۷-۵-۱ مقایسهٔ تأثیر نویز بر طیف سیگنال گفتار و طیف دنبالهٔ خودهمبستگی آن ۹٤

۷-۵-۲ مقایسهٔ روش‌های مختلف بازسازی نواحی تخریب شدهٔ طیف‌نگار ۹۷

۷-۵-۳ بررسی تأثیر تعداد خوشه‌ها بر دقت بازشناسی.....۹۹

۷-۵-۴ تخمین ماسک با استفاده از شبکه‌های عصبی ۱۰۱

۷-۵-۵ مقایسهٔ روش‌های مختلف محاسبهٔ مقادیر عضویت ۱۰۵

۷-۶ آزمایش‌های انجام شده در حوزهٔ بازشناسی گفتار ۱۱۰

۸- خلاصه و نتیجه‌گیری..... ۱۱۷

مراجع ۱۲۲

ضمیمهٔ الف: نتایج آزمایشهای انجام شده در بخش ۷-۵-۴..... ۱۲۶

ضمیمهٔ ب: نتایج آزمایشهای انجام شده به منظور بررسی تأثیر تعداد گره‌های شبکه‌های عصبی در قدرت تخمین ماسک..... ۱۳۱

ضمیمهٔ پ: نتایج آزمایشهای انجام شده به منظور بررسی حداکثر تعداد تکرارهای آموزش شبکه‌های عصبی در قدرت تخمین ماسک ۱۳۳

فهرست شکل‌ها

عنوان **صفحه**

شکل ۲-۱: تقسیم‌بندی انواع سیستم‌های بازشناسی مبتنی بر سیگنال گفتار ۹
شکل ۲-۲: مکانیزم کلی بازشناسی گفتار ۱۰
شکل ۲-۳: انواع عوامل تأثیرگذار بر کارایی سیستم‌های بازشناسی [5]. ۱۲
شکل ۲-۴: کلیات روش هنجارسازی ویژگی‌ها [14]. ۱۶
شکل ۳-۱: طیف‌نگارهای مل برای یک سیگنال گفتار ۳ ثانیه‌ای از مجموعه دادگان فارسدات تلفنی. در تمام تصاویر محور افقی شمارهٔ فریم و محور عمودی شمارهٔ فیلتر را نشان می دهد. (الف) گفتار تمیز، (ب) گفتار آغشته به نویز کارخانه۲ با SNR برابر با ۲۰ دسی بل، (ج) گفتار آغشته به نویز کارخانه۲ با SNR برابر ۱۰ دسی‌بل. (د) گفتار آغشته به نویز کارخانه۲ با SNR برابر ۰ دسی‌بل..... ۲۳
شکل ۳-۲: یک بانک فیلتر مبتنی بر معیار مل با استفاده از فیلترهای مثلثی [4]..... ۲۵
شکل ۳-۳: مراحل محاسباتی MFCC [9]..... ۲۶
شکل ۴-۱: دیاگرام بلوکی تکنیک پس پردازشی MVA [20]..... ۳۲
شکل ۴-۲: دنباله زمانی C[0]، C[1] و لگاریتم انرژی برای سیگنال گفتار خروجی از خواندن اعداد ″5376869″ که با مقادیر متفاوتی با نویز جمعی آغشته شده‌اند [20]. ۳۳
شکل ۴-۳: دیاگرام بلوکی سیستم‌های بازشناسی که از استخراج ویژگی ترکیبی استفاده می‌کنند. ۳۴
شکل ۵-۱: مراحل استخراج طیف‌نگار دنبالهٔ خودهمبستگی ۳۹
شکل ۵-۲: طیف‌نگارهای مل دنبالهٔ خودهمبستگی برای یک سیگنال گفتار ۳ ثانیه‌ای از مجموعه دادگان فارسدات تلفنی. در تمام تصاویر محور افقی شمارهٔ فریم و محور عمودی شمارهٔ فیلتر را نشان می دهد. (الف) گفتار تمیز، (ب) گفتار آغشته به نویز سفید با SNR برابر با ۳۰ دسی بل، (ج) گفتار آغشته به نویز سفید با SNR برابر ۲۰ دسی بل. (د) گفتار آغشته به نویز سفید با SNR برابر ۱۰ دسی‌بل. ۴۰
شکل ۵-۳: دیاگرام بلوکی استخراج پیک‌های طیف‌دنبالهٔ خودهمبستگی سیگنال گفتار..... ۴۱
شکل ۶-۱: دیاگرام بلوکی یک سیستم بازشناسی گفتاری که از روش ویژگی‌های مفقود مبتنی بر خوشه‌بندی استفاده می‌کند..... ۴۷
شکل ۶-۲: تقسیم‌بندی انواع روش‌های ویژگی‌های مفقود. ۴۹
شکل ۶-۳: ماسک‌های گسسته و نرم شده [36]..... ۵۳
شکل ۶-۴: مقایسهٔ دقت ماسک‌های گسسته و نرم‌شده [35]..... ۵۳
شکل ۶-۵: دیاگرام بلوکی مرحلهٔ آموزش در روش تخمین ماسک با استفاده از شبکه‌های عصبی. ۵۷

شکل ۶-۶: دیاگرام بلوکی مرحلهٔ تست در روش تخمین ماسک با استفاده از شبکه‌های عصبی. ۵۷
شکل ۶-۷: (الف) یک توزیع گوسی دو بعدی با بردار میانگین [1,1] و واریانس‌های ۱ برای مولفه‌های X و Y و کواریانس ۰.۵ بین X و Y. (ب) همان توزیع گوسی قسمت (الف) که در $X = 2$ بریده شده است. مقدار پیک در این برش در $Y = Y_1$ اتفاق افتاده است [28].....۶۳
شکل ۶-۸: تصویر توزیع شکل ۶-۷ روی محورهای $X - Y$. خط ممتد افقی مقدار مشاهده شده برای مولفهٔ X را نشان می‌دهد. خط ممتد قطری نشان دهندهٔ مکان پیک‌های توزیع بأزاء مقادیر مختلف X است. خط ممتد قطری ضخیم مقادیر تخمین MAP بأزاء مقادیر مختلف X را نشان می‌دهد [28].....۶۴
شکل ۶-۹: دو مثال از تخمین MAP کراندار. در هر دو شکل بیضی‌ها تصویر توزیع گوسی دادگان روی محورهای X-Y هستند. مولفهٔ X یک بردار مشاهده شده و با خط ممتد در طول محور X نشان داده شده است. مولفهٔ Y مشاهده نشده و بایستی تخمین زده شود. خط رگرسیون (تخمین MAP غیرکراندار) برای مقادیر مختلف X با خط قطری نشان داده شده است. (الف) کران بالا Y برابر با ۲.۵ است. تخمین MAP کراندار Y درون محدودهٔ مجاز قرار می‌گیرد و در نتیجه مقدار تخمین کراندار و غیرکراندار با یکدیگر برابر هستند. (ب) کران بالا Y برابر با ۱ است. تخمین MAP غیر کراندار بیرون ناحیهٔ مجاز قرار می‌گیرد. تخمین MAP کراندار با خط نقطه چین نشان داده شده است [28].....۶۵
شکل ۶-۱۰: مثال‌های برای تخمین MAP کراندار زمانی که بیش‌از یک مولفه بایستی تخمین زده شود. در تمام موارد بیضی ها نمایش دهندهٔ تصویر توزیع گوسی دادگان روی صفحهٔ Y-X هستند. در تمام موارد هر دو مولفه نامشخص هستند و بایستی تخمین زده شوند. (الف) تخمین غیرکراندار درون ناحیهٔ مجاز قرار می گیرد. (ب) خط قطری خط رگرسیون (MAP غیر کراندار) تخمین X نسبت به Y مربوطه را نشان می‌دهد. تخمین MAP کراندار در اشتراک خط رگرسیون و کران Y قرار می‌گیرد. (ج) خط قطری خط رگرسیون (MAP غیر کراندار) تخمین Y نسبت به X مربوطه را نشان می‌دهد. تخمین MAP کراندار در اشتراک خط رگرسیون و کران X قرار می‌گیرد. (د) تخمین MAP کراندار در هیچ یک از کرانهای مجاز قرار نمی‌گیرد. تخمین MAP کراندار در اشتراک کرانهای X و Y قرار می‌گیرد [28]..۶۶
شکل ۶-۱۱: شمایی از استفاده از چند خوشه (مدل گوسی) در مدل‌کردن [28]. ۶۸
شکل ۶-۱۲: شمای کلی مرحلهٔ بازسازی ویژگی‌ها توسط روش ویژگی‌های مفقود مبتنی بر خوشه‌بندی. ۶۹
شکل ۶-۱۳: (الف) نمونه‌ای طیف‌نگار یک سیگنال گفتار تمیز، (ب) طیف‌نگار سیگنال گفتار قسمت (الف) بعد از آغشته شدن به نویز کارخانهٔ ۲ با نسبت سیگنال به نویز ۱۰ دسیبل و (ج) ماسک نرم تخمین زده شده برای آن با استفاده از شبکه‌های عصبی..... ۷۱
شکل ۶-۱۴: شمای مرحلهٔ آموزش، برای سیستم بازشناسی گوینده‌ای که از روش ویژگی‌های مفقود مبتنی خوشه‌بندی استفاده می‌کند. ۷۳
شکل ۶-۱۵: دیاگرام بلوکی استفاده از روش ویژگی‌های مفقود در استخراج پیکهای طیف دنبالهٔ خودهمبستگی. ۷۴

شکل ۷-۱: خصوصیات طیفی نویزهای (الف) کارخانه۲، (ب) همهمه، (پ) سفید، (ت) صورتی، (ج) خودرو و (چ) هواپیما. ۷۹
شکل ۷-۲: پاسخ فرکانسی ۵ کانال مختلف که پنج اعوجاج کانال مختلف را تولید کرده‌اند. ۸۰
شکل ۷-۳: پاسخ فرکانسی فیلتر MIRS [61]. ۸۲
شکل ۷-۴: خطای بازشناسی گوینده با استفاده از تنها ضرائب MFCC و ضرائب MFCC به همراه انرژی در مقابل نویز همهمه (Babble). ۸۴
شکل ۷-۵: خطای بازشناسی گوینده با استفاده از تنها ضرائب MFCC و ضرائب MFCC به همراه انرژی در مقابل نویز هواپیما (F16). ۸۵
شکل ۷-۶:مقایسهٔ خطای بازشناسی گوینده با استفاده از ویژگیهای DAS، MFCC و MFCC نرمال شده با استفاده از هنجارسازی میانگین کپسترال بر دادگان آغشته به نویزهای (الف) همهمه، (ب) کارخانهٔ ۱، (ج) کارخانهٔ ۲ و (د) صورتی. (خطا بر حسب درصد گویندگانی است که به اشتباه تشخیص داده شده اند). ۸۹
شکل ۷-۷: مقایسهٔ دقت‌های بازشناسی ویژگی‌های MFCC، MFCC نرمال شده با MVA و ویژگی‌های DAS در مقابل نویزهای (الف) همهمه (ب) ماشین (ج) نویز سفید گوسی. ۹۰
شکل ۷-۸: نحوهٔ استخراج ویژگی‌های ترکیبی از ویژگی‌های MFCC، MFCC نرمال شده با پس‌پردازش MVA، ویژگی‌های DAS و ویژگی‌های ترکیبی در مقابل نویزهای ۹۱
شکل ۷-۹: دقت‌های بازشناسی گوینده (برحسب ٪) با استفاده از ویژگی‌های MFCC، MFCC نرمال شده با MVA، ویژگی‌های DAS و ویژگی‌های ترکیبی در مقابل نویزهای (الف) همهمه (ب) ماشین (ج) نویز سفید گوسی. ۹۳
شکل ۷-۱۰: طیف‌نگارهای مل یک سیگنال گفتار ۳ ثانیه‌ای از مجموعه دادگان فارسدات تلفنی. در تمام تصاویر محور افقی شمارهٔ فریم و محور عمودی شمارهٔ فیلتر را نشان می دهد. (الف) گفتار تمیز، (ب) گفتار آغشته به نویز کارخانه ۲ با SNR برابر با ۰ دسی‌بل، (پ) گفتار آغشته به نویز همهمه با SNR برابر ۱۰ دسی‌بل. (ت) گفتار آغشته به نویز صورتی با SNR برابر ۱۰ دسی‌بل (ث) گفتار آغشته به نویز سفید با SNR برابر ۱۰ دسی‌بل. ۹۵
شکل ۷-۱۱: طیف‌نگارهای مل دنبالهٔ خودهمبستگی برای یک سیگنال گفتار ۳ ثانیه‌ای از مجموعه دادگان فارسدات تلفنی. در تمام تصاویر محور افقی شمارهٔ فریم و محور عمودی شمارهٔ فیلتر را نشان می‌دهد. (الف) گفتار تمیز، (ب) گفتار آغشته به نویز کارخانه ۲ با SNR برابر با ۰ دسی‌بل، (پ) گفتار آغشته به نویز همهمه با SNR برابر ۱۰ دسی‌بل. (ت) گفتار آغشته به نویز صورتی با SNR برابر ۱۰ دسی‌بل (ث) گفتار آغشته به نویز سفید با SNR برابر ۱۰ دسی‌بل. ۹۶
شکل ۷-۱۲: بررسی تأثیر تعداد گره‌های شبکه‌های عصبی در قدرت تخمین ماسک طیف دنبالهٔ خودهمبستگی برای مجموعهٔ دادگان تست ماسک ۱ و ۲. ۱۰۴
شکل ۷-۱۳: بررسی تأثیر تعداد گره‌های شبکه‌های عصبی در قدرت تخمین ماسک طیف دنبالهٔ خودهمبستگی برای مجموعهٔ دادگان تست ماسک ۱ و ۲. ۱۰۵

شکل ۷-۱۴: مقایسهٔ در صد خطای بازشناسی گویندهٔ با استفاده از ویژگی های MFCC، MFCC بازسازی شده با استفاده از روش دادگان مفقود (MFCC_MD) و پیک های طیف دنبالهٔ خودهمبستگی بدست آمده از طیف نگار دنبالهٔ خودهمبستگی بازسازی شده با استفاده از روش ویژگی های مفقود (DAS_MD) برای دادگان آغشته به نویزهای (الف) کارخانه ۲، (ب) همهمه، (پ) صورتی، (ت) سفید، (ج) خودرو و (چ) هواپیما.....۱۰۹

شکل ۷-۱۵: متوسط نرخ بازشناسی لغات بر روی مجموعه دادگان Aurora2، (الف) مجموعه تست A، (ب) مجموعه تست B، (پ) مجموعه تست C. نتایج مطابق با روش های MFCC، MFCC بازسازی شده با MVA (MFCC_MVA)، MFCC بازسازی شده با روش ویژگی های مفقود (MFCC_MD)، MFCC بازسازی شده هم با MVA هم ویژگی‌های مفقود (MFCC_MD_MVA).....۱۱۵

فهرست جداول

عنوان **صفحه**

جدول ۶-۱: بردارهای ویژگی مبتنی بر طیف‌نگار پیشنهادی ۵۹
جدول ۷-۱: در صد خطای بازشناسی گوینده با استفاده از ویژگی های MFCC و MFCC به همراه انرژی در مقابل نویز همهمه (Babble). ۸۵
جدول ۷-۲: مقایسهٔ تأثیر مراحل مختلف مقاوم‌سازی پیکهای دنبالهٔ خودهمبستگی بر خطای بازشناسی گوینده برای دادگان تمیز..... ۸۶
جدول ۷-۳: مقایسهٔ تأثیر مراحل مختلف مقاوم‌سازی پیکهای دنبالهٔ خودهمبستگی بر خطای بازشناسی گوینده، برای دادگان آغشته به نویز همهمه با SNR ۱۰ دسیبل. ۸۷
جدول ۷-۴: مقایسهٔ تأثیر مراحل مختلف مقاومسازی پیکهای دنبالهٔ خودهمبستگی بر خطای بازشناسی گوینده، برای دادگان آغشته به نویز کارخانه ۲ با SNR ۱۰ دسیبل. ۸۷
جدول ۷-۵: مقایسهٔ تأثیر مراحل مختلف مقاومسازی پیکهای دنبالهٔ خودهمبستگی بر خطای بازشناسی گوینده برای دادگان آغشته به نویز سفید با SNR ۲۰ دسیبل. ۸۷
جدول ۷-۶: درصد خطای بازشناسی بدست آمده با استفاده از ویژگی های MFCC، حدف میانگین شده (MFCC_M)، واریانس نرمال شده (MFCC_V)، بازسازی شده با فیلتر ARMA (MFCC_A)، حذف میانگین با واریانس نرمال شده (MFCC_MV) و بازسازی شده با پس پردازش MVA (MFCC_MVA) بر روی دادگان فارسدات نویزی شده با پنج کانال با پاسخ‌های فرکانسی نشان‌داده شده در شکل ۷-۲.. ۹۴
جدول ۷-۷: درصد خطای بازشناسی بدست آمده با استفاده از پیک‌های طیف دنبالهٔ خودهمبستگی (AMFCC)، حدف میانگین شده (AMFCC_M)، واریانس نرمال شده (AMFCC_V)، بازسازی شده با فیلتر ARMA (AMFCC_A)، حذف میانگین با واریانس نرمال شده (AMFCC_MV) و بازسازی شده با پس پردازش MVA (AMFCC_MVA) بر روی دادگان فارسدات نویزی شده با پنج کانال با پاسخ‌های فرکانسی نشان‌داده شده در شکل ۷-۲..... ۹۴
جدول ۷-۸: بررسی تأثیر تعداد تکرار بر خطای بازشناسی در روش‌های تخمین MAP. و BMAP بر حسب در صد. ۹۷
جدول ۷-۹:مقایسهٔ در صدخطای بازشناسی گوینده با دو روش جایگذاری با میانگین (Mean) و میانگین محدود شده (Bounded Mean) برای نویز کارخانه ۲ با نسبتهای سیگنال به نویز ۰، ۱۰، ۲۰ و ۳۰ دسیبل. ۹۸
جدول ۷-۱۰: مقایسهٔ قدرت بازسازی روش‌های مختلف تخمین طیف‌نگار بر روی دادگان آغشته به نویز کارخانهٔ ۲، همهمه، صورتی و سفید با نسبت سیگنال به نویز مختلف (مقادیر برابر درصد خطای بازشناسی گوینده هستند). ۹۹
جدول ۷-۱۱: تأثیر تعداد خوشه‌ها بر خطای بازشناسی برای گفتار تست آغشته به نویز کارخانه ۲

با استفاده از بازسازی طیف‌نگار (خطا بر حسب درصد گویندگانی که اشتباه تشخیص داده شده‌اند) . ۱۰۰
جدول ۷-۱۲: تأثیر تعداد خوشه ها بر خطای بازشناسی برای گفتار تست آغشته به نویز همهمه با استفاده از بازسازی طیف‌نگار (خطا بر حسب درصد گویندگانی که اشتباه تشخیص داده شده‌اند).. ۱۰۰
جدول ۷-۱۳: تأثیر تعداد خوشه ها بر خطای بازشناسی برای گفتارِ تست آغشته به نویز صورتی با استفاده از بازسازی طیف‌نگار (خطا بر حسب درصد گویندگانی که اشتباه تشخیص داده شده‌اند).. ۱۰۰
جدول ۷-۱۴: تأثیر تعداد خوشه‌ها بر خطای بازشناسی برای گفتار تست آغشته به نویز کارخانه ۲ با استفاده از بازسازی طیف نگار دنبالهٔ خودهمبستگی(خطا بر حسب درصد گویندگانی که اشتباه تشخیص داده شده‌اند). ۱۰۰
جدول ۷-۱۵: تأثیر تعداد خوشه ها بر خطای بازشناسی برای گفتار تست آغشته به نویز همهمه با استفاده از بازسازی طیف‌نگار دنبالهٔ خودهمبستگی(خطا بر حسب درصد گویندگانی که اشتباه تشخیص داده شده‌اند). ۱۰۱
جدول ۷-۱۶: تأثیر تعداد خوشه ها بر خطای بازشناسی برای گفتار تست آغشته به نویز صورتی با استفاده از بازسازی طیف‌نگار دنبالهٔ خودهمبستگی (خطا بر حسب درصد گویندگانی که اشتباه تشخیص داده شده‌اند). ۱۰۱
جدول ۷-۱۷: مقایسهٔ در صد خطای بازشناسی بدست آمده با استفاده از ویژگی های MFCC بدست آمده از روش ویژگی‌های مفقود با روش‌های مختلف محاسبهٔ مقادیر عضویت به خوشه‌ها و ویژگی‌های MFCC خام (بدست آمده از سیگنال نویزی). ۱۰۶
جدول ۷-۱۸: مقایسهٔ در صد خطای بازشناسی بدست آمده با استفاده از پیک‌های طیف‌نگار دنبالهٔ خودهبستگی بازسازی شده با روش ویژگی‌های مفقود با فرض دانستن مقادیر واقعی تعلق به خوشه‌ها و روش پیشنهادی محاسبهٔ مقادیر تعلق به خوشه‌های مبتنی بر فاصلهٔ وزن‌دار. ۱۰۷
جدول ۷-۱۹: مقایسهٔ دقت (درصد تشخیص صحیح) بازشناسی لغات مجموعه آزمایش A از مجموعه دادگان Aurora بدست آمده با استفاده از ویژگی‌های بیان شده در ابتدای بخش ۷-۶ .... ۱۱۲
جدول ۷-۲۰: مقایسهٔ دقت (درصد تشخیص صحیح) بازشناسی لغات مجموعه آزمایش B از مجموعه دادگان Aurora بدست آمده با استفاده از ویژگی‌های بیان شده در ابتدای بخش ۷-۶ .... ۱۱۳
جدول ۷-۲۱: مقایسهٔ دقت (درصد تشخیص صحیح) بازشناسی لغات مجموعه آزمایش C از مجموعه دادگان Aurora بدست آمده با استفاده از ویژگی‌های بیان شده در ابتدای بخش ۷-۶ .... ۱۱۴

فصل اول

مقدمه

۱- مقدمه

گفتار اولین و یکی از بهترین وسایل بر قراری ارتباط بین انسان‌ها محسوب می‌شود و در برقراری ارتباط با ماشین‌ها نیز می‌تواند یکی از بهترین راه‌کارها باشد. به همین دلیل تحقیقات در شناسایی خودکار گفتار و گویندهٔ در حدود پنج دهه است که گسترش یافته است. یعنی در واقع از همان اوایل ساخت رایانه‌ها تلاش برای دستیابی به تعامل گفتاری با رایانه‌ها آغاز شده است. با توجه به هزینه کم و سهولت انتقال صوت از طریق تلفن، امروزه استفاده از آن کاربردهای بسیاری پیدا کرده‌است. مانند سیستمهای بازشناسی خودکار تلفنی و یا سیستم‌های اطلاعاتی که بر اساس پرسش و پاسخ با افراد اطلاعات خود را بروزرسانی می‌کنند، سیستم‌های تعیین هویت گوینده و غیره.

سیگنال گفتار حاوی دو سطح اطلاعاتی اصلی است. در سطح نخست، اطلاعاتِ گفتاری شامل واژه‌ها و پیام‌های بیان شده را منتقل می‌کند و در سطح دوم حاوی اطلاعات مربوط به خصوصیات گویندهٔ گفتار شامل ویژگیهای مجرای گفتاری، احساسات و غیره است.

در حالت کلی می‌توان سیستم‌های بازشناسی خودکار مبتنی بر سیگنال گفتار را به دو دستهٔ عمده تقسیم کرد:

- بازشناسی خودکار گفتار[1]
- بازشناسی خودکار گوینده[2]

اولین سیستم خودکار تشخیص گفتار در سال ۱۹۵۲ توسط آزمایشگاه Bell برای بازشناسی ارقام جدا از هم که توسط یک فرد تولید شده‌اند، طراحی شد. در تلاشهای انجام شده در طی ۳ دهه سیستم‌های بازشناسیِ گفتار به توانایی و کارایی نسبتا خوبی در تشخیص گفتار تمیز دست یافتند. ولی با توجه به کاربرد این سیستم‌ها در محیط‌های واقعی و وجود بدون تردید نویز صوتی در این محیط‌ها یافتن راهی برای بازشناسی گفتار مقاوم در مقابل نویز احساس می‌شد. در نتیجه اولین تلاشها در زمینهٔ بازشناسی گفتار مقاوم در مقابل نویز از دههٔ ۱۹۸۰ شروع شد و در دههٔ ۱۹۹۰ به نتایج نسبتاً خوبی دست پیدا کرد. در این زمان راهکارهای بسیاری برای مقاوم کردن سیستمهای بازشناسی گفتار در مقابل عدم تطابق بین داده‌های آموزشی و شرایط کاربرد پیشنهاد شد که از آن جمله می‌توان به استقلال از موزیک، میکروفون و کانال‌های انتقال، طنین اتاق و غیره اشاره کرد [1]. این تلاشها همچنان تا کنون ادامه پیدا کرده‌اند.

---

1 Automatic Speech Recognition
2 Automatic Speaker Recognition

اولین سیستم بازشناسی گوینده نیز در دهۀ ۱۹۶۰ در حدود ۱۰ سال پس از ایجاد اولین سیستم بازشناسی گفتار تولید شد. تحقیقات در زمینۀ مقاوم نمودن این نوع سیستمها نیز در دهۀ ۱۹۹۰ به اوج خود رسید و همچنان ادامه پیدا کرده است [1].

علی‌رغم بهبود بسیار ایجاد شده در کارایی سیستم‌های خودکار بازشناسی این موضوع همچنان یکی از زمینه‌های فعال تحقیقاتی محسوب می‌شود و می‌توان بیان کرد که تلاش‌های بسیاری در این زمینه انجام می‌گیرد. می‌توان انگیزۀ اصلی این تحقیق را چنین بیان کرد که *"بازشناسی مقاوم واسطۀ بین بازشناسی و کاربرد واقعی آن است"*. تا زمانی که بازشناسی مقاوم به میزان کافی توسعه پیدا نکرده باشد، سیستم‌های بازشناسی در کاربردهای واقعی به کارایی مطلوب نخواهند رسید.

مجموعۀ تلاش‌های انجام شده در راستای دستیابی به سیستم‌های بازشناسی گوینده و گفتار مقاوم در مقابل نویز را می‌توان به ۳ دستۀ زیر تقسیم کرد.

- روش‌های استخراج ویژگی‌های مقاوم
- روش‌های هنجارسازی ویژگی‌ها
- روش‌های تطبیق مدل

در دستۀ نخست هدف استخراج ویژگی‌هایی از سیگنال گفتار ورودی است که ضمن اینکه از دقت تفکیک‌پذیری بالایی برخوردار هستند، نسبت به انواع نویزها از مقاومت بالایی نیز برخوردار باشند و قدرت تفکیک‌پذیری آنها تحت تأثیر نویزها تغییر چندانی نکند. از جملۀ این ویژگی‌ها می توان به ویژگی‌های استخراج شده از حوزۀ خودهمبستگی اشاره کرد.

در دستۀ دوم سعی بر آن است تا ویژگی‌های استخراج شده از سیگنال گفتار به گونه‌ای نگاشته شوند که تأثیر نویز بر آنها تا حد امکان برطرف شود. از جملۀ رایج‌ترین این روش‌ها می‌توان به روش هنجارسازی میانگین ضرائب کپسترال، هنجارسازی واریانس، فیلتر ARMA و روش پس پردازشی MVA[1] نیز که در واقع ترکیبی از این سه روش مذکور است اشاره کرد.

در دستۀ سوم از تلاش‌های انجام شده در زمینۀ مقاوم‌سازی سیستم‌های بازشناسی از روش‌هایی برای مدل‌سازی الگوها (چه الگوهای گفتاری و چه گویندگان) استفاده می‌شود که تأثیر نویز نیز مدل شود و دقت بازشناسی نهایی تحت تأثیر نویزها تغییر چندانی پیدا نکند. از جملۀ این روش‌ها می‌توان به روش PMC[2] [41] و [42]، روش VTS[3] [42] و [43]، روش مدل اجتماع [45][1] و [46] و مدل مخفی مارکوف فاکتوریلی[2] (FHMM) [47] و [48] اشاره کرد.

---

[1] Mean Normlization, Variance normalization and ARMA Filtering
[2] Parallel Model Combination
[3] Vector Taylor Series