



دانشگاه تبریز
دانشکده مهندسی برق و کامپیوتر
گروه مخابرات

پایان نامه

برای دریافت درجه کارشناسی ارشد در رشته مهندسی برق
گرایش مخابرات-سیستم

عنوان

جداسازی تک‌گوشی گفتار بی‌صدا بر پایه‌ی
آنالیز ترکیب شنیداری

استاد راهنما

دکتر مسعود گراوانچی‌زاده

استاد مشاور

دکتر میرهادی سیدعربی

پژوهشگر

پریا دادور

بهمن ۱۳۹۰

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

به پاس تعبیر عظیم و انسانی شان از کلمه می‌ایثار و از خودگذشتگی،
به پاس عاطفه‌ی سرشار و گرمای امید بخش وجودشان که در این سردترین روزگار
بهترین پشتیبان است،

به پاس قلب‌های بزرگشان که فریادس است
و سرکردانی و ترس در پناهشان به شجاعت می‌گراید
و به پاس محبت‌های بی‌دینشان که هرگز فروکش نمی‌کند،

این مجموعه را به پدر و مادر عزیزم تقدیم می‌کنم.

تقدیر و شکر

الهی! ادای شکر تو را هیچ زبان نیست و دریای فضل تو را هیچ کمران نیست و سر حقیقت تو بر هیچ کس عیان نیست، بر ما بنا

رہی کہ بہتر از آن نیست.

بر خود لازم می دانم، قدردان زحمات تمام کسانی باشم کہ در انجام این پروژه مرایاری نموده اند:

از جناب آقای دکتر مسعود کراوانچی زاده، استاد فریخته، پرتلاش و دلسوزم کہ در طول سال های تحصیل اینجانب

در دانشگاه تبریز، همواره یاریم نموده و پشتیبانم بوده اند، نهایت تقدیر و شکر را دارم.

از استاد کرامتقدر، جناب آقای دکتر میرهادی سید عربی کہ زحمت مشاوره ی این پایان نامه را بر عهده گرفتند، بسیار سپاسگزارم.

در پایان، از ہمہ ی دوستان خوبم شکر می کنم و برای ہمہ ی عزیزانی کہ مراد مرا حل مختلف ہمراہی کرده اند، از خداوند بزرگ

سعادت و سلامت طلب می کنم.

این پایان نامه طی قرارداد شماره سی «۱۸۴۹۲/۵۰۰/ت» مورخ «۸۹/۱۲/۲۸» از طرف مرکز تحقیقات
مخابرات ایران مورد حمایت مالی قرار گرفته است که بدینوسیله از حمایت مالی و معنوی این مرکز قدردانی می‌کنیم.

نام خانوادگی: دادور	نام: پریا
عنوان پایان نامه: جداسازی تک‌گوشی گفتار بی‌صدا بر پایه‌ی آنالیز ترکیب شنیداری	
استاد راهنما: دکتر مسعود گراوانچی‌زاده استاد مشاور: دکتر میرهادی سیدعربی	
مقطع تحصیلی: کارشناسی ارشد	رشته: مهندسی برق دانشگاه: دانشگاه تبریز تاریخ فارغ‌التحصیلی: بهمن ۹۰ دانشکده: مهندسی برق و کامپیوتر تعداد صفحه: ۱۲۹
کلیدواژه‌ها: آنالیز ترکیب شنیداری محاسباتی، جداسازی گفتار بی‌صدا، آشکارسازی فعالیت کانالی نویز، کاهش طیفی زیرباندی مبتنی بر نسبت سیگنال به نویز، نویز رنگی.	
<p>چکیده: جداسازی تک‌گوشی گفتار از تداخل صوتی موضوع بسیار چالش‌انگیزی است. پژوهش‌های بسیاری در زمینه‌ی آنالیز ترکیب شنیداری محاسباتی (CASA) به منظور جداسازی تک‌گوشی گفتار صدادار (voiced speech) از ترکیب‌های صوتی انجام شده است. با این وجود، جداسازی گفتار بی‌صدا (unvoiced speech) به عنوان یکی از چالش‌های مهم CASA باقی مانده است. گفتار بی‌صدا به دلیل داشتن انرژی نسبتاً ضعیف و دارا نبودن ساختار هارمونیک، در برابر تداخل بسیار آسیب‌پذیر است، که این مسأله جداسازی گفتار بی‌صدا را بسیار دشوار می‌سازد. در این پایان‌نامه، سیستم جدیدی به روش کاهش طیفی زیرباندی مبتنی بر نسبت سیگنال به نویز (SNR-based SBSS) برای جداسازی گفتار بی‌صدا از تداخل غیرگفتار ارائه می‌شود. در سیستم پیشنهادی، پس از انجام آنالیز محیطی و یک عمل پیش‌پردازش، برخی ویژگی‌های مهم سیگنال ترکیب استخراج می‌شوند. سپس، جداسازی گفتار بی‌صدا در دو مرحله صورت می‌گیرد: قسمت‌بندی و گروه‌بندی. در مرحله‌ی قسمت‌بندی، ابتدا گفتار صدادار و بخش‌های متناوب سیگنال تداخل حذف می‌شوند. سپس، با استفاده از IBM صدادار، فعالیت نویز در کانال‌های فرکانسی به روش جدید آشکارسازی فعالیت کانالی نویز (CNAD) آشکار می‌شود و نسبت سیگنال به نویز سیگنال ورودی پیش‌پردازش شده تخمین زده می‌شود. آنگاه، انرژی نویز در هر کانال تخمین زده می‌شود و روش پیشنهادی SNR-based SBSS برای تولید قسمت‌های زمانی-فرکانسی در بازه‌های بی‌صدا به کار می‌رود. در مرحله‌ی بعد، قسمت‌های گفتار بی‌صدا بر اساس مشخصات فرکانسی گفتار بی‌صدا، با استفاده از یک روش آستانه‌گذاری ساده، گروه‌بندی می‌شوند. مقایسه‌ها و ارزیابی‌های اصولی با مدل (Hu & Wang (2011) نشان می‌دهند که سیستم پیشنهادی، عملکرد سیستم‌های رایج جداسازی گفتار بی‌صدا را از نظر کیفیت و قابلیت‌فهم، به میزان قابل توجهی، بهبود می‌بخشد.</p>	

فہرست مطالب

د فهرست جداول
و فهرست شکل‌ها
ی فهرست اختصارات
ک پیشگفتار

۱ فصل ۱- مقدمه‌ای بر مفاهیم آنالیز ترکیب شنیداری محاسباتی (CASA)

۲ ۱-۱- مقدمه
۳ ۲-۱- آنالیز ترکیب شنیداری (ASA)
۵ ۳-۱- آنالیز ترکیب شنیداری محاسباتی (CASA)
۵ ۱-۳-۱- CASA چیست؟
۶ ۲-۳-۱- روش‌های مختلف جداسازی در سیستم‌های CASA
۸ ۳-۳-۱- هدف CASA چیست؟
۹ ۴-۳-۱- تفاوت CASA با ICA در چیست؟
۱۰ ۵-۳-۱- کاربردهای CASA
۱۱ ۴-۱- پایه‌های سیستم‌های CASA
۱۲ ۱-۴-۱- ساختار کلی سیستم
۱۳ ۲-۴-۱- کاکلی‌گرام
۱۷ ۳-۴-۱- کریلوگرام
۱۹ ۴-۴-۱- ماسک‌های زمانی-فرکانسی
۲۲ ۵-۴-۱- بازسازی

۲۳ فصل ۲- پیشینه‌ی پژوهش در زمینه‌ی جداسازی گفتار بی‌صدا بر پایه‌ی CASA

۲۴ ۱-۲- مقدمه
۲۴ ۲-۲- چه میزان از گفتار بی‌صدا است؟
۲۵ ۱-۲-۲- تکرار نسبی
۲۶ ۲-۲-۲- مدت زمان نسبی
۲۷ ۳-۲- ویژگی‌های گفتار بی‌صدا
۲۹ ۴-۲- روش‌های جداسازی گفتار بی‌صدا بر پایه‌ی CASA
۳۰ ۱-۴-۲- جداسازی با استفاده از نشانه‌های شروع و پایان
۴۲ ۲-۴-۲- جداسازی با استفاده از CASA و روش کاهش طیفی
۴۵ ۵-۲- نتایج شبیه‌سازی جداسازی با استفاده از روش (Hu & Wang (2011

۵۴ فصل ۳- روش پیشنهادی برای جداسازی گفتار بی‌صدا بر پایه‌ی CASA
۵۵ ۱-۳- مقدمه
۵۵ ۲-۳- ساختار کلی سیستم پیشنهادی
۵۷ ۳-۳- پردازش محیطی
۶۰ ۴-۳- روش پیشنهادی برای پیش‌پردازش
۶۱ ۵-۳- استخراج ویژگی‌ها
۶۶ ۶-۳- روش پیشنهادی برای قسمت‌بندی گفتار بی‌صدا
۶۷ ۱-۶-۳- جداسازی گفتار صدادار
۶۷ ۲-۶-۳- حذف سیگنال متناوب
۶۹ ۳-۶-۳- روشی جدید برای آشکارسازی فعالیت کانالی نویز (CNAD)
۷۲ ۴-۶-۳- روشی جدید برای تخمین SNR ورودی
۷۳ ۵-۶-۳- روشی جدید برای قسمت‌بندی گفتار بی‌صدا (SNR-Based SBSS)
۷۷ ۷-۳- گروه‌بندی
۷۷ ۸-۳- بازسازی
۷۹ فصل ۴- نتایج و پیشنهادات
۸۰ ۱-۴- مقدمه
۸۰ ۲-۴- شرایط شبیه‌سازی و دادگان
۸۲ ۳-۴- معیارهای ارزیابی
۸۳ ۱-۳-۴- معیارهای ارزیابی subjective
۸۳ ۱-۱-۳-۴- نمایش زمانی
۸۴ ۲-۱-۳-۴- نمایش زمانی-فرکانسی
۸۴ ۲-۳-۴- معیارهای ارزیابی objective
۸۵ ۱-۲-۳-۴- ارزیابی کیفیت گفتار
۸۶ ۲-۲-۳-۴- ارزیابی قابلیت فهم گفتار
۸۸ ۴-۴- نتایج شبیه‌سازی الگوریتم پیشنهادی
۱۰۱ ۵-۴- مقایسه‌ی نتایج شبیه‌سازی الگوریتم پیشنهادی و الگوریتم (Hu & Wang (2011)
۱۱۳ ۶-۴- نتیجه‌گیری
۱۱۶ ۷-۴- پیشنهادات
۱۱۸ مراجع
۱۲۳ پیوست

فهرست جداول

جدول (۱-۲): بهره‌ی SNR در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم (Hu & Wang (2011) ۵۱
جدول (۲-۲): درصد تلف در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم (Hu & Wang (2011) ۵۱
جدول (۳-۲): درصد اشتباه در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم (Hu & Wang (2011) ۵۲
جدول (۴-۲): درصد خطای کل در بازه‌های بی‌صدا و صدادر برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم (Hu & Wang (2011) ۵۲
جدول (۵-۲): درصد دقت در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم (Hu & Wang (2011) ۵۳
جدول (۶-۲): درصد موفقیت بدون اشتباه در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم (Hu & Wang (2011) ۵۳
جدول (۱-۳): مقادیر تابع $f(SNR)$ در شش سطح SNR ۷۵
جدول (۱-۴): پارامترهای مهم در شبیه‌سازی الگوریتم‌های جداسازی گفتار بی‌صدا. ۸۱
جدول (۲-۴): سیگنال‌های گفتار پایگاه داده‌ی TIMIT که برای شبیه‌سازی الگوریتم‌های جداسازی گفتار بی‌صدا مورد استفاده قرار گرفته‌اند. نیمی از جملات توسط گویندگان زن و نیمی دیگر، توسط گویندگان مرد بیان شده است. ۸۲
جدول (۳-۴): حالت‌های مختلف برچسب‌گذاری واحدهای T-F در ماسک تخمینی نسبت به IBM. ۸۶
جدول (۴-۴): بهره‌ی SNR در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم پیشنهادی). ۹۷
جدول (۵-۴): درصد تلف در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم پیشنهادی). ۹۷
جدول (۶-۴): درصد اشتباه در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم پیشنهادی). ۹۸
جدول (۷-۴): درصد خطای کل در بازه‌های بی‌صدا و صدادر برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم پیشنهادی). ۹۸
جدول (۸-۴): درصد دقت در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم پیشنهادی). ۹۹
جدول (۹-۴): درصد موفقیت بدون اشتباه در بازه‌های بی‌صدا برای پنج نوع نویز مختلف در پنج سطح SNR ورودی (جداسازی توسط سیستم پیشنهادی). ۹۹

Tables of Appendix :

Tabel A-1:	SNR Gain (dB) in unvoiced intervals for 15 Noises (and their averages) and 5 input SNR levels (and their averages). Results are shown for two unvoiced speech separation systems: proposed and Hu & Wang (2011).	124
Tabel A-2:	Percentages of Miss (%) in unvoiced intervals for 15 Noises (and their averages) and 5 input SNR levels (and their averages). Results are shown for two unvoiced speech separation systems: proposed and Hu & Wang (2011).	125
Tabel A-3:	Percentages of False-Alarm (%) in unvoiced intervals for 15 Noises (and their averages) and 5 input SNR levels (and their averages). Results are shown for two unvoiced speech separation systems: proposed and Hu & Wang (2011).	126
Tabel A-4:	Percentages of Overall Error (%) in entire speech for 15 Noises (and their averages) and 5 input SNR levels (and their averages). Results are shown for two unvoiced speech separation systems: proposed and Hu & Wang (2011).	127
Tabel A-5:	Percentages of Accuracy (%) in unvoiced intervals for 15 Noises (and their averages) and 5 input SNR levels (and their averages). Results are shown for two unvoiced speech separation systems: proposed and Hu & Wang (2011).	128
Tabel A-6:	Percentages of HIT-FA (%) in unvoiced intervals for 15 Noises (and their averages) and 5 input SNR levels (and their averages). Results are shown for two unvoiced speech separation systems: proposed and Hu & Wang (2011).	129

فهرست شکل‌ها

- شکل (۱-۱): ساختار کلی سیستم‌های CASA. ۱۲
- شکل (۲-۱): ساختار داخلی گوش انسان [16]. عصب شنوایی و بخش حلزونی گوش داخلی در شکل با رنگ روشن نشان داده شده‌اند. ۱۳
- شکل (۳-۱): نمایش‌های زمانی-فرکانسی برای یکی از جملات پایگاه‌داده‌ی TIMIT بیان شده توسط گوینده‌ی زن: "We always thought we would die with our boots on." (a) طیف‌نگار، (b) کاکلی‌گرام. ۱۶
- شکل (۴-۱): (a) کریلوگرام برای مصوت /er/ با فرکانس اصلی $F_0 = 100 \text{ Hz}$ ، (b) کریلوگرام اختصاری. ۱۸
- شکل (۵-۱): جداسازی ترکیب گفتار و نویز توسط ماسک باینری ایده‌آل (IBM). (a) کاکلی‌گرام سیگنال گفتار تمیز بیان شده توسط گوینده‌ی مرد: "Ralph controlled the stopwatch from the bleachers." (b) کاکلی‌گرام سیگنال نویز صدای پرنده و جریان آب، (c) کاکلی‌گرام سیگنال ترکیب گفتار و نویز ($\text{SNR} = 0 \text{ dB}$)، (d) IBM (نقاط با رنگ روشن برچسب 1 را نشان می‌دهند)، (e) کاکلی‌گرام سیگنال گفتار جدا شده توسط IBM. ۲۱
- شکل (۱-۲): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز جمعیت در زمین بازی ($\text{SNR} = 0 \text{ dB}$)، (c) گفتار جدا شده با IBM تخمین‌زده‌شده توسط سیستم Hu & Wang (2011). (d) گفتار جدا شده توسط IBM. ۴۷
- شکل (۲-۲): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز مهمانی ($\text{SNR} = 0 \text{ dB}$)، (c) گفتار جدا شده با IBM تخمین‌زده‌شده توسط سیستم Hu & Wang (2011). (d) گفتار جدا شده توسط IBM. ۴۸
- شکل (۳-۲): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز زنگ ساعت ($\text{SNR} = 5 \text{ dB}$)، (c) گفتار جدا شده با IBM تخمین‌زده‌شده توسط سیستم Hu & Wang (2011). (d) گفتار جدا شده توسط IBM. ۴۹
- شکل (۴-۲): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز آژیر ($\text{SNR} = 5 \text{ dB}$)، (c) گفتار جدا شده با IBM تخمین‌زده‌شده توسط سیستم Hu & Wang (2011). (d) گفتار جدا شده توسط IBM. ۵۰
- شکل (۱-۳): شکل کلی سیستم پیشنهادی برای جداسازی گفتار بی‌صدا بر پایه‌ی CASA و روش کاهش طیفی زیرباندی (SBSS) مبتنی بر SNR. بلوک‌هایی که با رنگ روشن مشخص شده‌اند، مربوط به بخش‌های نوآوری در الگوریتم پیشنهادی می‌باشند. * در سیستم پیشنهادی، برای به‌دست آوردن ماسک گفتار صدادار که مورد بحث این پایان‌نامه نمی‌باشد، از ماسک IBM تولیدشده استفاده می‌شود [11]. از ماسک IBM صدادار تخمین‌زده‌شده نیز می‌توان برای جداسازی گفتار صدادار استفاده کرد که در شکل با خط-چین نشان داده شده است. ۵۶
- شکل (۲-۳): پاسخ فرکانسی یک بانک فیلتر گاماتون 64 کانالی با مراکز فرکانسی در محدوده‌ی 50 Hz تا 8000 Hz. ... ۵۹
- شکل (۳-۳): نمای کلی روش کاهش طیفی مورد استفاده در مرحله‌ی پیش‌پردازش [45]. $x(i)$ خروجی فیلتربانک گاماتون و $y(i)$ خروجی فیلتربانک گاماتون پیش‌پردازش شده در کانال i ام می‌باشد. ۶۰

- شکل (۳-۴): (a) کریلوگرام (نمایش ACF) در یک فریم مربوط به گفتار صدادر /aa/ (b) همبستگی بین کانالی، (c) کریلوگرام اختصاری. تأخیر زمانی متناظر با فله‌ی کلی SACF، $\tau = 127$ ms است. ۶۳
- شکل (۳-۵): (a) کریلوگرام پوش پاسخ (نمایش ACF) در یک فریم مربوط به گفتار صدادر /aa/ (b) همبستگی بین کانالی. ۶۴
- شکل (۳-۶): نمایش C و C_E برای جمله‌ی "We always thought we would die with our boots on." (a) همبستگی بین کانالی (C)، (b) همبستگی پوش پاسخ بین کانالی (C_E). نقاط با دامنه‌ی بالا با رنگ تیره نشان داده شده‌اند. ۶۵
- شکل (۳-۷): نمایش C و C_E برای سیگنال ترکیب با نسبت سیگنال به نویز 5 dB- از سیگنال گفتار "We always thought we would die with our boots on." و نویز صدای پرنده و جریان آب (a) همبستگی بین-کانالی (b) همبستگی پوش پاسخ بین کانالی (C_E). نقاط با دامنه بالا با رنگ تیره نشان داده شده‌اند. ۶۶
- شکل (۳-۸): طیف‌نگار سیگنال‌های مختلف نویز (a) پنکه برقی، (b) نویز سفید، (c) نویز جمعیت در زمین بازی، (d) نویز جمعیت با صدای کف زدن، (e) نویز جمعیت با صدای موسیقی، (f) باران، (g) همهمه و (h) موسیقی. ۷۰
- شکل (۳-۹): طیف‌نگار سیگنال‌های مختلف نویز (a) صدای باد، (b) نویز مهمانی، (c) صدای زنگ ساعت، (d) نویز ترافیک، (e) آژیر، (f) صدای پرنده و جریان آب و (g) زنگ تلفن. ۷۱
- شکل (۳-۱۰): تابع $f(SNR)$ بر حسب \hat{SNR} . نقاط با علامت ستاره، مقادیر به‌دست آمده برای تابع به‌صورت تجربی و منحنی خط‌چین، مقادیر درون‌یابی شده در سایر نقاط را نشان می‌دهد. ۷۵
- شکل (۴-۱): شکل موج سیگنال‌های (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز مهمانی (SNR=0 dB)، (c) گفتار جدا شده توسط IBM صدادر، (d) گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش پیشنهادی، (e) گفتار جدا شده توسط IBM. ۹۰
- شکل (۴-۲): طیف‌نگار سیگنال‌های (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز مهمانی (SNR=0 dB)، (c) گفتار جدا شده توسط IBM صدادر، (d) گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش پیشنهادی، (e) گفتار جدا شده توسط IBM. ۹۱
- شکل (۴-۳): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز جمعیت در زمین بازی (SNR=0 dB)، (c) گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش پیشنهادی، (d) گفتار جدا شده توسط IBM. ۹۲
- شکل (۴-۴): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز مهمانی (SNR=0 dB)، (c) گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش پیشنهادی، (d) گفتار جدا شده توسط IBM. ۹۳
- شکل (۴-۵): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز زنگ ساعت (SNR=5 dB)، (c) گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش پیشنهادی، (d) گفتار جدا شده توسط IBM. ۹۴
- شکل (۴-۶): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز آژیر (SNR=5 dB)، (c) گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش پیشنهادی، (d) گفتار جدا شده توسط IBM. ۹۵
- شکل (۴-۷): ماسک‌های زمانی-فرکانسی برای جداسازی ترکیب 0 dB از گفتار تمیز S3 و نویز جمعیت در زمین بازی (a) بی‌صدای تخمین‌زده‌شده توسط روش (Hu & Wang (2011)، (b) بی‌صدای تخمین‌زده‌شده توسط روش پیشنهادی، (c) بی‌صدا. ۱۰۱

- شکل (۴-۸): ماسک‌های زمانی-فرکانسی برای جداسازی ترکیب 5 dB از گفتار تمیز S3 و نویز موسیقی (a) IBM بی‌صدای تخمین‌زده‌شده توسط روش (b) IBM، Hu & Wang (2011). روش پیشنهادی، (c) IBM بی‌صدا. ۱۰۲
- شکل (۴-۹): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز جمعیت در زمین بازی (SNR= 0 dB)، (c) گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش (d) Hu & Wang (2011). گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش پیشنهادی، (e) گفتار جدا شده توسط IBM. ۱۰۴
- شکل (۴-۱۰): طیف‌نگار (A) و شکل موج (B) سیگنال‌ها در بازه‌های بی‌صدا (a) گفتار تمیز S3، (b) ترکیب گفتار و نویز موسیقی (SNR= 5 dB)، (c) گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش (d) Hu & Wang (2011). گفتار جدا شده توسط IBM تخمین‌زده‌شده توسط روش پیشنهادی، (e) گفتار جدا شده توسط IBM. ۱۰۵
- شکل (۴-۱۱): بهره‌ی SNR در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 5 سطح SNR می‌باشند و برای 15 نویز مختلف نشان داده شدند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر بهره‌ی SNR در بازه‌های بی‌صدا، حدود 2.96 dB بهبود می‌بخشد. ۱۰۷
- شکل (۴-۱۲): درصد تلف در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 5 سطح SNR می‌باشند و برای 15 نویز مختلف نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد تلف در بازه‌های بی‌صدا، حدود 8.76% بهبود می‌بخشد. ۱۰۷
- شکل (۴-۱۳): درصد اشتباه در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 5 سطح SNR می‌باشند و برای 15 نویز مختلف نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد اشتباه در بازه‌های بی‌صدا، حدود 0.63% بهبود می‌بخشد. ۱۰۸
- شکل (۴-۱۴): درصد خطای کل در بازه‌های بی‌صدا و صدادر برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 5 سطح SNR می‌باشند و برای 15 نویز مختلف نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد خطای کل در بازه‌های بی‌صدا، حدود 3.83% بهبود می‌بخشد. ۱۰۸
- شکل (۴-۱۵): درصد دقت در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 5 سطح SNR می‌باشند و برای 15 نویز مختلف نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد دقت در بازه‌های بی‌صدا، حدود 6.4% بهبود می‌بخشد. ۱۰۹
- شکل (۴-۱۶): درصد موفقیت‌بدون‌اشتباه در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 5 سطح SNR می‌باشند و برای 15 نویز مختلف نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد موفقیت‌بدون‌اشتباه در بازه‌های بی‌صدا، حدود 31.47% بهبود می‌بخشد. ۱۰۹

- شکل (۴-۱۷): بهره‌ی SNR در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 15 نویز مختلف می‌باشند و در 5 سطح SNR نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر بهره‌ی SNR در بازه‌های بی‌صدا، حدود 2.96 dB بهبود می‌بخشد. ۱۱۰
- شکل (۴-۱۸): درصد تلف در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 15 نویز مختلف می‌باشند و در 5 سطح SNR نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد تلف در بازه‌های بی‌صدا، حدود 8.76% بهبود می‌بخشد. ۱۱۰
- شکل (۴-۱۹): درصد اشتباه در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 15 نویز مختلف می‌باشند و در 5 سطح SNR نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد اشتباه در بازه‌های بی‌صدا، حدود 0.63% بهبود می‌بخشد. ۱۱۱
- شکل (۴-۲۰): درصد خطای کل در بازه‌های بی‌صدا و صدادر برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 15 نویز مختلف می‌باشند و در 5 سطح SNR نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد خطای کل، حدود 3.83% بهبود می‌بخشد. ۱۱۱
- شکل (۴-۲۱): درصد دقت در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 15 نویز مختلف می‌باشند و در 5 سطح SNR نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد دقت در بازه‌های بی‌صدا، حدود 6.4% بهبود می‌بخشد. ۱۱۲
- شکل (۴-۲۲): درصد موفقیت‌بدون‌اشتباه در بازه‌های بی‌صدا برای الگوریتم پیشنهادی و الگوریتم Hu & Wang (2011). نتایج، میانگینی از 15 نویز مختلف می‌باشند و در 5 سطح SNR نشان داده شده‌اند. به‌طور میانگین، سیستم پیشنهادی عملکرد سیستم قبلی را از نظر درصد موفقیت‌بدون‌اشتباه در بازه‌های بی‌صدا، حدود 31.47% بهبود می‌بخشد. ۱۱۲

فہرست اختصارات

ASA	Auditory Scene Analysis
CASA	Computational Auditory Scene Analysis
ERB	Equivalent Rectangular Bandwidth
IBM	Ideal Binary Mask
LC	Local Criterion
ASR	Automatic Speech Recognition
ITD	Interaural Time Difference
IID	Interaural Intensity Difference
GMM	Gaussian Mixture Model
MLP	Multi-Layer Perceptron
CNAD	Channel Noise Activity Detection
SBSS	Sub-Band Spectral Subtraction
SNR	Signal-to-Noise Ratio
ACF	Autocorrelation Function
AM	Amplitude Modulation
FM	Frequency Modulation

گفتاری که به گوش ما می‌رسد، در دنیای واقعی هرگز خالص نیست. صدای صحبت اشخاص اغلب با تداخل صداهای دیگر مانند صدای پنکه، موسیقی و یا صدای شخص دیگر به گوش می‌رسد. یک سیستم جداسازی که بتواند تداخل ترکیب شده از دیگر منابع صوتی با صدای اصلی را حذف و یا تضعیف کند، در بسیاری از کاربردها مانند تشخیص گفتار خودکار (ASR)، تشخیص گوینده در محیط‌های صوتی واقعی، بازیابی اطلاعات شنیداری^۱، برهم‌کنش صوتی رایانه و انسان^۲ و نیز طراحی وسایل کمک‌شنوایی^۳ از جمله سمعک هوشمند اهمیت فراوان دارد.

گفتار طبیعی شامل دو دسته آوا از نوع صدادار^۴ و بی‌صدا^۵ می‌باشد. گفتار بی‌صدا حدود 25% از گفتار یک زبان را تشکیل می‌دهد و به دلیل داشتن انرژی نسبتاً ضعیف و دارا نبودن ساختار هارمونیک، در برابر تداخل آسیب‌پذیرتر است. مسأله‌ی جداسازی این گفتار کمتر مورد مطالعه قرار گرفته و به‌عنوان یک چالش بزرگ باقی مانده است. در این پایان‌نامه، سیستمی برای جداسازی تک‌گوشی گفتار بی‌صدا از تداخل غیرگفتار^۶ بر پایه‌ی آنالیز ترکیب شنیداری محاسباتی^۷ (CASA) و روش کاهش طیفی زیرباندی (SBSS)^۸ مبتنی بر نسبت سیگنال به نویز (SNR) ارائه شده است. سیستم پیشنهادی، با طی مراحل

¹ Audio Information Retrieval

² Sound-Based Human Computer Interaction

³ Hearing Aid Design

⁴ Voiced

⁵ Unvoiced

⁶ Non-Speech Interference

⁷ Computational Auditory Scene Analysis (CASA)

⁸ Sub-Band Spectral Subtraction (SBSS)

اصلی سیستم‌های CASA، به تخمین ماسک باینری ایده‌ال برای جداسازی سیگنال هدف از یک ترکیب صوتی می‌پردازد.

این پایان‌نامه از فصول زیر تشکیل شده است. فصل اول به ارائه‌ی مقدمه‌ای بر مفاهیم آنالیز ترکیب شنیداری محاسباتی، به عنوان زمینه‌ی پژوهشی جدید در زمینه‌ی شنوایی ماشین^۱، اختصاص یافته است. در فصل دوم، ابتدا به بررسی ویژگی‌های گفتار بی‌صدا می‌پردازیم و سپس، پیشینه‌ای از پژوهش‌های صورت گرفته بر اساس CASA در زمینه‌ی جداسازی تک‌گوشی^۲ گفتار بی‌صدا از تداخل غیرگفتار را بیان می‌کنیم. نتایج شبیه‌سازی سیستم جداسازی یکی از پژوهش‌های قبلی در پایان این فصل آمده است. در فصل سوم، سیستم پیشنهادی معرفی می‌شود. این سیستم، مراحل اصلی سیستم-های CASA شامل پردازش محیطی^۳، استخراج ویژگی^۴، قسمت‌بندی^۵ و گروه‌بندی^۶ را در بر دارد. نحوه‌ی نحوه‌ی ارزیابی سیستم‌های جداسازی گفتار بر پایه‌ی CASA از اهمیت ویژه‌ای برخوردار است. در فصل چهارم، نتایج حاصل از شبیه‌سازی سیستم پیشنهادی، به روش‌های مختلف ارزیابی شده و نتایج حاصل از شبیه‌سازی‌ها مقایسه می‌شود. در پایان، به نتیجه‌گیری و ارائه‌ی پیشنهاد برای ارتقاء عملکرد سیستم پیشنهادی می‌پردازیم.

پریا دادور

زمستان ۱۳۹۰

¹ Machine Audition

² Monaural

³ Peripheral Processing

⁴ Feature Extraction

⁵ Segmentation

⁶ Grouping

فصل اول:

مقدمه ای بر

مفاهیم آنالیز ترکیب شنیداری محاسباتی

(CASA)

۱-۱- مقدمه

پژوهش‌های بسیاری در زمینه‌ی جداسازی منابع صوتی انجام شده است. این پژوهش‌ها شامل روش‌های بهبود کیفیت گفتار^۱، روش‌های فضایی^۲ با استفاده از آرایه‌ای از میکروفون‌ها و جداسازی کور منابع^۳ می‌باشد [1, 2]. روش‌های بهبود گفتار، کیفیت گفتار آلوده به نویز را بر اساس اطلاعات دریافتی از یک میکروفون ارتقاء می‌دهند. الگوریتم‌های ارائه شده برای این روش‌ها شامل روش کاهش طیفی^۴، فیلتر فیلتر وینر^۵، تخمین‌گر حداقل خطای مجذور میانگین^۶ و آنالیز زیر فضا^۷ می‌باشند [3]. دسته‌ی دیگری از روش‌ها که جداسازی گفتار بر پایه‌ی مدل^۸ نامیده می‌شوند، بر مدل کردن الگوهای منابع تمرکز دارند و جداسازی را به‌عنوان یک مسئله‌ی تخمین در چارچوب احتمالی فرمول‌بندی می‌کنند [4]. این سیستم‌ها با ارائه‌ی مشاهدات و استفاده از مدل‌های منابع، گفتارهای تکی را به‌طور مستقیم تخمین زده و یا یک ماسک زمانی-فرکانسی برای جداسازی هر منبع استخراج می‌کنند.

روش‌های ذکر شده با تمام گفتار آلوده به نویز سر و کار دارند و بنابراین توانایی جداسازی گفتار بی‌صدا را نیز دارند. با این وجود، روش‌های بهبود گفتار اغلب فرض‌هایی در مورد مشخصات آماری تداخل دارند که کاربرد آن‌ها را در مورد تداخل‌های کلی محدود می‌سازد. به‌عنوان مثال، اغلب فرض می‌شود نویز موجود در ترکیب ایستا^۹ است که در شرایط کلی درست نیست و تداخل می‌تواند به‌طور ناگهانی در مدت زمان کوتاهی تغییر کند. در روش‌های فضایی و جداسازی کور منابع به‌ترتیب فرض‌های ایستا بودن

¹ Speech Enhancement

² Spatial

³ Blind Source Separation

⁴ Spectral Subtraction

⁵ Wiener Filter

⁶ Minimum-Mean-Square Error (MMSE)-Based Estimator

⁷ Sub-Space Analysis

⁸ Model-Based

⁹ Stationary