دانشگاه تبریز
دانشکده مهندسی برق و کامپیوتر
گروه مخابرات

پایان‌نامه

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


بسم الله الرحمن الرحيم

به پاس تعبیر عظیم و انسانی شان از کلمه ی ایثار و از خودگذشتگی،

به پاس عاطفه ی سرشار و گرمای امید بخش وجودشان که در این سردترین روزگاران

بهترین پشتیبان است،

به پاس قلب های بزرگشان که فریادرس است

و سرگردانی و ترس در پناهشان به شجاعت می گراید

و به پاس محبت های بی دریغشان که هرگز فروکش نمی کند،

این مجموعه را به پدر و مادر عزیزم تقدیم می کنم.

تقدیر و تشکر

الهی! ادای شکر تو را هیچ زبان نیست و دریای فضل تو را هیچ کران نیست و سرّ حقیقت تو بر هیچ کس عیان نیست، بر ما بنما

رهی که بهتر از آن نیست.

بر خود لازم می دانم، قدردان زحمات تمام کسانی باشم که در انجام این پروژه مرا یاری نموده اند:

از جناب آقای دکتر مسعود کراوانچی زاده، استاد فرهیخته، پر تلاش و دلسوزم که در طول سال های تحصیل اینجانب

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

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

در پایان، از همه ی دوستان خوبم تشکر می کنم و برای همه ی عزیزانی که مرا در مراحل مختلف همراهی کرده اند، از خداوند بزرگ

سعادت و سلامت طلب می کنم.

این پایان نامه طی قرارداد شماره ی «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 می کنیم.

نام خانوادگی: دادور **نام:** پریا

عنوان پایان‌نامه: جداسازی تک‌گوشی گفتار بی‌صدا بر پایه‌ی آنالیز ترکیب شنیداری

استاد راهنما : دکتر مسعود گراوانچی‌زاده

استاد مشاور: دکترمیرهادی سیدعربی

مقطع تحصیلی: کارشناسی ارشد **رشته:** مهندسی برق **گرایش:** مخابرات

دانشگاه: دانشگاه تبریز **دانشکده:** مهندسی برق و کامپیوتر

تاریخ فارغ‌التحصیلی: بهمن ۹۰ **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.

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 هارمونیکی، 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 پیش‌پردازش‌شده 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 به‌کار 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.

فهرست مطالب

فهرست جداول

Tabels of Appendix:

فهرست شکل‌ها

فهرست 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

پیشگفتار

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[1]، برهم‌کنش صوتی رایانه و انسان[2] و نیز طراحی وسایل کمک‌شنوایی[3] از جمله سمعک هوشمند اهمیت فراوان دارد.

گفتار طبیعی شامل دو دسته آوا از نوع صدادار[4] و بی‌صدا[5] 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[6] بر پایه‌ی آنالیز ترکیب شنیداری محاسباتی[7](CASA) و روش کاهش طیفی زیرباندی (SBSS)[8] 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 برمفاهیم آنالیز ترکیب شنیداری محاسباتی، به عنوان زمینه‌ی پژوهشی جدید در زمینه‌ی شنوایی ماشین[1]، 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[2] 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[3]، استخراج ویژگی[4]، قسمت‌بندی[5] و گروه‌بندی[6] 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 از روش‌ها که جداسازی گفتار بر پایه‌ی مدل[۸] نامیده می‌شوند، بر مدل کردن الگوهای منابع تمرکز دارند 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