بسم الله الرحمن الرحيم

دانشکده مهندسی برق و رباتیک

بازشناخت مقاوم گفتارفارسی با استفاده از ضرایب مل-کپستروم بهبود یافته و شبکه عصبی

دانشجو: دانیال دارابیان

استاد راهنما:

دکتر حسین مروی

استاد مشاور:

دکتر حسین خسروی

پایان نامه جهت اخذ درجه کارشناسی ارشد

شهریور ۹۲

شماره : ۱۰۸۷/آ.ت.ب
تاریخ : ۹۲/۰۶/۲۶
ویرایش : -----

بسمه تعالی

مدیریت تحصیلات تکمیلی
فرم شماره (۶)

فرم صورتجلسه دفاع پایان نامه تحصیلی دوره کارشناسی ارشد

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عج) جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای :

دانیـــال دارابیـــان رشته : **بـــــرق** گرایش : **الکترونیک**

تحت عنوان : **بازشناخت مقاوم گفتار فارسی با استفاده از ضرایب مل – کپستروم بهبود یافته و شبکه عصبی**

که در تاریخ **۹۲/۰۶/۲۶** با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح زیر است :

قبول (با درجه : عالی) ☑ امتیاز (۱۹/۲) دفاع مجدد ☐ مردود ☐

۱ـ عالی (۲۰ ـ ۱۹)
۲ـ بسیار خوب (۱۸/۹۹ ـ ۱۸)
۳ـ خوب (۱۷/۹۹ ـ ۱۶)
۴ـ قابل قبول (۱۵/۹۹ ـ ۱۴)
۵- نمره کمتر از ۱۴ غیر قابل قبول

عضو هیأت داوران	نام ونام خانوادگی	مرتبهٔ علمی	امضـــاء
۱ـ استـاد راهنمـا	حمید سرداری	استادیار	
۲ـ استـاد مشـاور	حسین خسروی	استادیار	
۳ـ نماینده شورای تحصیلات تکمیلی	غلامرضا الهی یزدی	استادیار	
۴ـ استـاد ممتحـن	یونس زاهدی	استادیار	
۵ ـ استـاد ممتحـن	سید رضا سید قاضی	استادیار	

رئیس دانشکده :

تقدیم به پدر و مادر و همسرم که در این مسیر مرا همراه و یاور بودند.

با تشکر از

دکتر مروی به پاس راهنمایی و همراهی بی دریغشان

و دکتر خسروی که در این مسیر همراه من بودند.

تعهد نامه

اینجانب **دانیال دارابیان** دانشجوی دوره کارشناسی ارشد رشته **الکترونیک سیستم** دانشکده **مهندسی برق و رباتیک** دانشگاه صنعتی شاهرود نویسنده پایان نامه با عنوان :

بازشناخت مقاوم گفتار فارسی با استفاده از ضرایب مل-کپستروم بهبود یافته و شبکه عصبی

تحت راهنمائی آقای دکتر **حسین مروی** متعهد می شوم .

- تحقیقات در این پایان نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « Shahrood University of Technology » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان نامه ، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ

امضای دانشجو

مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .
- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نمی باشد.

چکیده:

بازشناخت اتوماتیک گفتار[1] در واقع مساله تشخیص گفتار برای یک نمونه گفتار صحبت شده توسط سخنگوی نامعلوم است. هر سیستم اتوماتیک تشخیص گفتار نیازمند استخراج ویژگی است که به وسیله-ی ویژگی استخراج شده بتواند گفتار ورودی را متمایز نماید و سپس تشخیص دهد از این دسته ویژگی ها می توان به ضرایب مل-کپستروم[2] اشاره نمود که بعدها به یکی از فراگیرترین ویژگی‌های مورد استفاده در پردازش گفتار تبدیل شدند. مشکل عمده در استفاده از این ضرایب زمانی ظاهر می شود که تشخیص گفتار در محیط‌های همراه با نویز صورت بپذیرد. عملکرد این الگوریتم در حضور نویز، با سرعت زیاد و به شدت کاهش می‌یابد.

برای دست‌یابی به یک عملکرد رضایت بخش تحت شرایط نویزی در سیستم‌های تشخیص گفتار نیاز به مصون سازی این الگوریتم در برابر نویز می‌باشد. در این پایان نامه یک روش مقاوم در برابر نویز، جهت استخراج ویژگی ضرایب مل-کپستروم معرفی شده است.

ویژگی‌های استخراج شده از سیگنال گفتار را به وسیله ی یک شبکه عصبی طبقه بندی می کنیم. در این پایان نامه از یک شبکه عصبی پرسپترون با یک لایه ورودی، دو لایه میانی یا مخفی و یک لایه خروجی جهت کلاسه بندی نتایج استفاده شده است. جهت ارزیابی الگوریتم پیشنهادی تعداد ۴۰ لغت[3] متفاوت، در هر لغت ۲۰ تکرار توسط ۲۰ گوینده ی متفاوت شامل زن و مرد بزرگسال از نمونه های صوتی پایگاه اطلاعاتی FARS-DAT استفاده شده است.

تاکنون روش های مختلفی در مصون سازی این الگوریتم به کار رفته است که هر یک به یک بلوک خاص پرداخته اند و یا بلوکی مکمل به الگوریتم پایه افزوده اند در حالی که در روش پیشنهادی علاوه بر در نظر

[1] Automatic Speech Recognition (ASR)
[2] MFCC (Mel Frequency Cepstral Coefficients)
[3] Isolated Word

گرفتن اکثر تغییرات مهم اعمال شده در سایر الگوریتم‌ها، هر یک را در مکان مناسب استفاده و نیز با روش‌های مناسب ترکیب کرده ایم و نیز بلوک‌هایی به الگوریتم پایه افزوده ایم.

نتایج آزمایش های تشخیص گفتار فارسی در نرخ سیگنال به نویزهای متفاوت، نشان گر افزایش معنادار نرخ تشخیص و نیز مصونیت بیشتر الگوریتم پیشنهادی نسبت به الگوریتم پایه و همچنین برخی از الگوریتم‌های مطرح در این خانواده همچون الگوریتم‌های :

AMFCC، GMFCC، ROOT-MFCC، CMN-SMN-MFCC و نیز **RAS-MFCC** است.

کلمات کلیدی:

بازشناخت گفتار- ضرایب مل-کپستروم- تابع خود همبستگی- فیلتربانک گوسی- تفریق میانگین گفتار- جبرانگر لگاریتم

لیست مقاله های استخراج شده :

باز شناسی مقاوم گفتار فارسی با استفاده از ضرایب مل-کپستروم بهبود یافته و شبکه عصبی (پذیرفته شده در شانزدهمین کنفرانس سیستم های هوشمند)

Improving the performance of MFCC for Persian Robust Speech Recognition

(Submmit شده در Journal of Intelligent Automation systems دانشگاه صنعتی شاهرود)

بازشناخت گفتار فارسی در محیط نویزی با استفاده از ضرایب مل-کپستروم بهبودیافته

(Submmit شده در نشریه سیستم های هوشمند در مهندسی برق دانشگاه اصفهان، علمی-پژوهشی)

فهرست مطالب

فصل اول: مقدمه

۱.۱)بیان مسئله..... ۲

۱.۲)مختصری پیرامون پیشینه ی تاریخی سیستم های اتوماتیک تشخیص گفتار..... ۳

۱.۳)چالش ها در سیستم هایASR ۵

فصل دوم: بازشناخت گفتار در محیط نویزی

۲.۱)مقدمه..... ۷

۲.۲)تغییرات نامطلوب سیگنال گفتار..... ۸

۲.۲.۱)آلودگی با نویز..... ۸

۲.۲.۲)حالات گفتار..... ۱۱

۲.۲.۳)تغییرات درونی گوینده..... ۱۱

۲.۳)آموزش سیستم های تشخیص گفتار..... ۱۱

۲.۴)روش های مصون سازی سیستم های اتوماتیک تشخیص گفتار..... ۱۳

۲.۵)طبقه بندها..... ۱۵

۲.۵.۱)شبکه عصبی..... ۱۵

۲.۵.۲)مدل مخفی مارکوف.....۱۷

۲.۵.۳)مدل مخلوط گوسی.....۱۸

فصل سوم: روش های متداول استخراج ویژگی

۳.۱)مقدمه.....۲۳

۳.۲)ضرایب کپسترال پیش گوی خطی.....۲۴

۳.۳)ضرایب مل-کپستروم.....۲۷

۳.۳.۱)فریم بندی و فیلتر پیش تاکید.....۲۸

۳.۳.۲)پنجره گذاری.....۲۹

۳.۳.۳)فیلتربانک.....۲۹

۳.۳.۴)لگاریتم و تبدیل کسینوسی.....۳۲

۳.۴)پیش گوی خطی مبتنی بر درک انسان.....۳۳

۳.۴.۱)آنالیز طیفی.....۳۴

۳.۴.۲)آنالیز بحرانی.....۳۴

۳.۴.۳)پیش تاکید مربوط به یکسان سازی بلندی صدا.....۳۵

۳.۴.۴)قانون توان شدت بلندی صدا.....۳۵

۳.۴.۵)مدل کردن AR ۳۵

۳.۴.۶)آنالیز کپسترال..... ۳۵

۳.۵)ضرایب موجک..... ۳۶

فصل چهارم: مروری بر کارهای انجام شده

۴.۱)مقدمه..... ۳۹

۴.۲)مرور و طبقه بندی روش های استفاده شده در بهبود الگوریتم ضرایب مل-کپستروم..... ۴۰

۴.۳)ضرایب مل-کپستروم بهبود یافته بر اساس افزودن بلوک مکمل به الگوریتم پایه..... ۴۳

۴.۴)ضرایب مل-کپستروم بهبود یافته بر اساس تغییر در بلوک های پایه ی الگوریتم..... ۴۸

۴.۵)ضرایب مل-کپستروم بهبود یافته بر اساس تغییر در پیاده سازی سخت افزاری این الگوریتم..... ۵۵

فصل پنجم: روش پیشنهادی

۵.۱)مقدمه..... ۵۷

۵.۲)بررسی روش پیشنهادی..... ۵۸

۵.۲.۱)حذف سکوت از گفتار..... ۶۰

۵.۲.۲)تفریق میانگین گفتار در حوزه ی زمان..... ۶۱

۵.۲.۳)فیلتر پیش تاکید..... ۶۲

۵.۲.۴)فریم بندی.....۶۳

۵.۲.۵)عبور از پنجره ی بهبود یافته.....۶۳

۵.۲.۶)اتوکورلیشن و حذف ضرایب مرتبه پایین.....۶۴

۵.۲.۷)عبور از فیلتر بالاگذر.....۶۵

۵.۲.۸)تبدیل فوریه و تفریق میانگین گفتار در حوزه ی فرکانس.....۶۶

۵.۲.۹)فیلتربانک گوسی.....۶۶

۵.۲.۱۰)به توان رساندن خروجی هریک از کانال ها.....۶۸

۵.۲.۱۱)بلوک CLMN ۶۹

۵.۲.۱۲) فیلتر RASTA.....۷۱

۵.۲.۱۳)بلوک ریشه گیری (ROOT).....۷۱

۵.۳)ارزیابی الگوریتم پیشنهادی.....۷۲

۵.۴)دیتا بیس.....۷۳

فصل ششم: نتایج آزمایش های انجام شده

۶.۱)مقدمه.....۷۵

۶.۲)نتایج آزمایش های انجام شده برای تعیین توان مورد نیاز در بلوک توان رسانی.....۷۶

۶.۳)مرتبه مورد نیاز در حذف ضرایب اتوکورلیشن.....۷۸

۶.۴)مرتبه مورد نیاز در بلوک ریشه گیری.....۷۹

۶.۵)ارزیابی نهایی روش پیشنهادی.....۸۰

۶.۶)نتیجه گیری و پیشنهادات.....۸۶

مراجع.....۸۸

فهرست شکل ها:

فصل اول:

(شکل ۱-۱) بلوک دیاگرام پایه در یک سیستم ASR.....۴

(شکل ۲-۱) شبکه عصبی پرسپترون دولایه.....۱۶

فصل دوم:

(شکل ۲-۲) روند آموزش و بازشناسی در مدل مخفی مارکوف.....۱۷

(شکل ۲-۳) مدل مخلوط گوسی.....۱۹

فصل سوم:

(شکل ۳-۱) بلوک دیاگرام پایه ی محاسبه ضرایب مل-کپستروم.....۲۸

(شکل ۳-۲) دو نوع پیاده سازی فیلتر بانک مل.....۳۰

(شکل ۳-۳)بلوک دیاگرام پایه ی محاسبه ی ضرایب پیش گوی مبتنی بر درک انسان.....۳۳

فصل چهارم:

(شکل ۴-۱) الگوریتم حاصل از حذف ضرایب مرتبه پایین.....۴۵

(شکل ۴-۲) بلوک دیاگرام الگوریتم DRHOASS-MFCC.....۴۶

(شکل ۴-۳) بلوک دیاگرام الگوریتم ROOT-MFCC.....۴۸

(شکل ۴-۴) بلوک دیاگرام الگوریتم GMFCC..... ۵۰

(شکل ۴-۵) فیلتربانک گوسی..... ۵۱

(شکل ۴-۶) فیلتر بانک با توزیع معکوس مثلثی..... ۵۲

(شکل ۴-۷) فیلتر بانک با توزیع معکوس گوسی..... ۵۲

(شکل ۴-۸) پنجره همینگ و همینگ بهبود یافته..... ۵۴

(شکل ۴-۹) بلوک دیاگرام الگوریتم MFCCE..... ۵۵

فصل پنجم:

(شکل ۵-۱) بلوک دیاگرام الگوریتم پیشنهادی..... ۵۹

(شکل ۶-۱) مقایسه ی عملکرد روش پیشنهادی با الگوریتم پایه و نیز ROOT-MFCC..... ۸۳

(شکل ۶-۲) مقایسه ی عملکرد روش پیشنهادی با الگوریتم پایه و نیز GMFCC..... ۸۴

(شکل ۶-۳) مقایسه ی عملکرد روش پیشنهادی با الگوریتم پایه و نیز CMN-SMN MFCC ۸۴

(شکل ۶-۴) مقایسه ی عملکرد روش پیشنهادی با الگوریتم پایه و نیز AMFCC..... ۸۵

(شکل ۶-۵) مقایسه ی عملکرد روش پیشنهادی با الگوریتم پایه و نیز RAS-MFCC..... ۸۵

(شکل ۶-۶) مقایسه ی عملکرد روش پیشنهادی با سایر روش های مطرح در خانواده ی الگوریتم های ویژه ی ضرایب مل با استفاده از درصد میانگین تشخیص گفتار نویزی..... ۸۶

فهرست جدول ها :

(جدول ۴-۱) سیر روش های ایجاد بهبود در الگوریتم ضرایب ملکپستروم.....۴۲

(جدول ۴-۲) مقایسه ی پارامترهای پنجره ی بهبود یافته با پنجره ی همینگ.....۵۴

(جدول ۶-۱) مقایسه و تعیین مقدار عددی توان در مرحله ی توان رسانی در نرخ سیگنال به نویزهای متفاوت.....۷۷

(جدول ۶-۲) مقایسه و تعیین مقدار عددی آستانه ی حذف در مرحله حذف ضرایب مرتبه پایین خودهمبستگی در نرخ سیگنال به نویز های متفاوت.....۷۸

(جدول ۶-۳) مقایسه ی درصد تشخیص گفتار در مقادیر متفاوت ریشه (α) در نرخ های متفاوت سیگنال به نویز.....۷۹

(جدول ۶-۴) مقایسه ی درصد تشخیص گفتار در روش پیشنهادی با نرخ تشخیص گفتار در روش های متداول از خانواده ی ضرایب مل-کپستروم.....۸۱

فهرست علائم و اختصارات

ASR	**Automatic Speech Recognition**
AR	**Autocorrelation Regressive**
AMFCC	**Autocorrelation Mel Frequency Cepstral Coefficient**
CMN-SMN MFCC	**Cepstral mean normalization Spectral Mean Normalization Mel Frequency Cepstral Coefficient**
DRHOASS- MFCC	**Differentiated Higher Order Relative Autocorrelation Sequence**
GMFCC	**Gaussian Mel Frequency Cepstral Coefficient**
GMM	**Gaussian Mixture Model**
HMM	**Hidden Morkov Model**
IMFCC	**Inverse Mel Frequency Cepstral Coefficient**
IGMFCC	**Inverse Gaussian Mel Frequency Cepstral Coefficient**
LPC	**Linear Prediction Cepstral**
LPCC	**Linear Prediction Cepstral Coefficient**
MLP	**Multy Layer Perceptron**
NN	**Neural Network**
PLP	**Perceptual Linear Prediction**
WGN	**White Gaussian Noise**

فصل اول

مقدمه

فصل اول: مقدمه

۱.۱) بیان مسئله:

یکی از زیرشاخه‌های علم پردازش سیگنال پردازش گفتار است که در روند پیشرفت‌های صورت گرفته در زمینه‌ی پردازش سیگنال، بسیار مورد توجه قرار گرفته است و تحقیقات گسترده ای را متوجه خود ساخته است. پردازش گفتار شامل سه بخش اصلی می باشد: تبدیل متن به گفتار ، بازشناسی گفتار و بهسازی گفتار.

بازشناسی گفتار توسط سیستم های اتوماتیک تشخیص گفتار(ASR) انجام می گیرد. هدف از انجام این پایان‌نامه ایجاد بهبود در عملکرد این سیستم با تکیه بر بهبود یکی از پرکاربردترین الگوریتم های استخراج ویژگی در حوزه‌ی پردازش گفتار می باشد .

هر سیستم تشخیص گفتار نیازمند استخراج ویژگی است که بوسیله‌ی آن بتواند گفتار ورودی را به طور صحیح تشخیص دهد. مقاوم سازی این ویژگی ها و بطور خاص ویژگی ضرایب مل-کپستروم، به عنوان رایج‌ترین آن‌ها در حوزه‌ی پردازش گفتار، در برابر نویز از اهمیت بسزایی در بهبود نرخ تشخیص، در سیستم‌های اتوماتیک تشخیص گفتار برخوردار است.