



دانشگاه صنعتی شاهرود

دانشکده: برق و رباتیک

گروه: الکترونیک

پایان نامه دوره ی کارشناسی ارشد مهندسی برق - الکترونیک

کاربرد روش استخراج ویژگی RootMel جهت تخمین

سن افراد با استفاده از سیگنال گفتار

عاطفه دهقانیان

استاد راهنما:

دکتر حسین مروی

استاد مشاور:

دکتر علی سلیمانی

شهریور ۱۳۹۱

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

تقدیم به پدر و مادرم

که هستی و همه ی وجودم از آن هاست

و همسرم

به پاس قدرانی و سپاسگذاری از زحمات و

دلسوزی هایشان

تشکر و قدردانی

در ابتدا از استاد راهنمای خود، جناب آقای دکتر حسین مروی به خاطر راهنمایی و حمایتی که از من در طی انجام این تحقیق کرده اند، نهایت تشکر را می نمایم.

و نیز از تمامی دوستانم در دانشگاه صنعتی شاهرود، که با حمایت و کمک بی دریغ خود در انجام این تحقیق، مرا همراهی کردند؛ تشکر می نمایم.

از پدر و مادر و همسرم به خاطر کمک ها و دلسوزی هایشان و به خاطر حمایت های معنوی که در طی انجام این تحقیق از من داشته اند، تشکر ویژه می نمایم.

چکیده :

تخمین سن بر اساس ویژگی های گفتار انسان، یک موضوع قابل توجه در سیستم های شناسایی گفتار اتوماتیک می باشد. مطالعاتی در زمینه ی تخمین سن گوینده صورت گرفته است ولی نیاز به کار های نوین بیشتری، خصوصا برای گوینده های فارسی زبان، می باشد. در تخمین سن، مانند سایر سیستم های پردازش گفتار، با دو چالش مهم مواجه هستیم: یافتن یک روش مناسب برای استخراج ویژگی و انتخاب یک روش قابل اطمینان برای کلاسه بندی.

هدف اصلی از این تحقیق استفاده از ویژگی $\text{Root Mel Frequency Cepstral Coefficients}$ جهت بخش استخراج ویژگی در یک سیستم تخمین سن و یافتن بهترین مقدار برای داشتن درصد خطای کمتر می باشد؛ و همچنین مقایسه ی عملکرد این ویژگی با سایر ویژگی های متداول نظیر MFCC معمول، PLP و LPC نیز مورد بررسی قرار گرفته است .

برای استخراج ویژگی، کل سیگنال صوت را به کوچک ترین جزء آن، یعنی واج ها، تفکیک می کنیم و عملیات استخراج ویژگی و رده بندی را بر سیگنال مربوط به واج ها انجام می دهیم. از کلاسه بند به روش های تفکیک پذیری خطی و فواصل Mahalanobis استفاده شده است. نتیجه آزمایشات انجام شده بر پایگاه داده ی FARSDAT به کمترین ۲۸.۶۹٪ بازای ریشه ی ۰.۰۰۶ در استفاده از روش RootMFCC و تاثیر مثبت تفکیک سیگنال صوتی به واج های تشکیل دهنده ی آن، در کاهش خطا می باشد.

کلمات کلیدی : تخمین سن، گفتار، واج، استخراج ویژگی RootMel ، FARSDAT

فهرست مطالب :

- د فهرست جدول ها
- و فهرست شکل ها
- ی فهرست اصطلاحات
- ۱- فصل اول : مقدمه..... ۱
- ۱-۱ خلاصه ای از چگونگی تغییر صدای انسان با افزایش سن..... ۴
- ۱-۱-۱ مکانیزم تولید گفتار با گذشت سن..... ۵
- ۱-۱-۲ سیستم تنفسی..... ۶
- ۱-۱-۳ حنجره..... ۶
- ۱-۱-۴ سیستم فرا حنجره ای..... ۶
- ۱-۱-۵ بالارفتن سن در مردها و زن ها..... ۷
- ۲-۱ تحلیل صوتی سن..... ۷
- ۳-۱ درک انسان از سن..... ۹
- ۴-۱ روش های تکنولوژی گفتار..... ۱۰
- ۵-۱ ساختار پایان نامه..... ۱۱
- ۲- فصل دوم : تحقیقات صورت گرفته در زمینه ی تخمین سن اتوماتیک توسط گفتار..... ۱۲
- ۳- فصل سوم : مروری بر انواع روش های استخراج ویژگی..... ۴۰
- ۳-۱ مقدمه..... ۴۱
- ۳-۲ استخراج ویژگی در برابر کلاسه بندی..... ۴۱
- ۳-۳ مراحل استخراج ویژگی از سیگنال گفتار..... ۴۲
- ۳-۳-۱ شکل دهی طیفی..... ۴۲
- ۳-۳-۲ تحلیل طیفی..... ۴۳

- ۳-۳-۳ تبدیل ویژگی ۴۵
- ۳-۳-۴ دسته بندی تکنیک های استخراج ویژگی گفتار ۴۵
- ۳-۳-۵ روش های رایج استخراج ویژگی ۴۶
- ۳-۳-۵-۱ ضرایب Linear Prediction Cepstral ۴۷
- ۳-۳-۵-۲ ضرایب Perceptual Linear Prediction ۴۸
- ۳-۳-۵-۳ ضرایب Linear Frequency Cepstral ۵۱
- ۳-۳-۵-۴ ضرایب Mel Frequency Cepstral ۵۲
- ۳-۳-۶ کاهش ویژگی : ۵۳
- ۳-۳-۶-۱ LDA(linear discriminant analysis) ۵۷
- ۳-۳-۶-۲ تحلیل مؤلفه های اصلی (PCA) ۵۹
- ۳-۳-۶-۳ مقایسه ی تصویری LDA , PCA ۶۳
- ۳-۳-۷ روش های تعمیم ویژگی MFCC برای کاربردهای مختلف ۶۴
- ۳-۳-۷-۱ روش استخراج ویژگی از ضرایب قسمت بالایی خود همبستگی
- ۳-۳-۷-۲ ضرایب MFCC دو طیفی ۶۴
- ۳-۳-۷-۳ غیر حساس کردن MEL- Cepstrum نسبت به اجزای طیفی
- ۳-۳-۷-۴ نادرست (بدلی) (RootMFCC)، برای شناسایی گفتار پایدار ۹۱
- ۳-۳-۷-۴ Autocorrelation MFCC ۹۶
- ۳-۳-۷-۵ Relative MFCC برای شناسایی گفتار تلفنی پایدار ۱۰۰
- ۴ - فصل چهارم : روش پیشنهادی برای تخمین سن افراد ۱۰۶
- ۴-۱ مقدمه ۱۰۷
- ۴-۲ معرفی پایگاه داده ۱۰۸

- ۳-۴ معرفی الگوریتم پیشنهادی..... ۱۱۰
- ۴-۴ پیاده سازی الگوریتم پیشنهادی..... ۱۱۱
- ۱-۴-۴ بکارگیری روش استخراج ویژگی MFCC..... ۱۱۱
- ۲-۴-۴ بکارگیری روش استخراج ویژگی LPC..... ۱۲۰
- ۳-۴-۴ بکارگیری روش استخراج ویژگی RootMel..... ۱۲۱
- ۴-۴-۴ بکارگیری روش استخراج ویژگی PLP..... ۱۲۳
- ۵-۴-۴ اعمال روش PCA برای کاهش ابعاد ویژگی..... ۱۲۴
- ۶-۴-۴ اعمال روش LDA برای کاهش ابعاد ویژگی..... ۱۲۵
- ۷-۴-۴ بررسی تاثیر تفکیک داده ها بر اساس جنسیت بر نتایج آزمایشات... ۱۲۶
- ۵-۴ نتایج نهایی بدست آمده..... ۱۲۶
- ۶-۴ بررسی تاثیر تعداد گروه های سنی بر نتایج بدست آمده..... ۱۲۹
- ۷-۴ مقایسه نتایج آزمایش با رده بند های Linear , mahalnobis..... ۱۳۰
- ۸-۴ ارزیابی نتایج حاصل از الگوریتم های مختلف..... ۱۳۱
- ۹-۴ نتیجه گیری کلی ۱۳۲
- فهرست منابع..... ۱۳۵

فهرست جدول ها:

- جدول ۱-۱: اطلاعاتی که سیگنال گفتار حمل می کند (با توجه به نظر Fujisaki -2004)..... ۱۰
- جدول ۱-۲: نتایج بدست آمده از آزمایشات صورت گرفته در زمینه ی تخمین سن در [۹]..... ۱۳
- جدول ۲-۲: انواع ویژگی های استفاده شده در [۱۰]..... ۱۶
- جدول ۳-۲: انواع توابع kernel استفاده شده در SVM..... ۱۸
- جدول ۴-۲: انواع توابع فاصله مورد استفاده در K-NN..... ۱۸
- جدول ۵-۲: ماتریس ترکیبی بر پایگاه داده ی DES [۱۰]..... ۱۹
- جدول ۶-۲: ماتریس ترکیبی بر پایگاه داده ی ELSDSR در [۱۰]..... ۱۹
- جدول ۷-۲: ماتریس ترکیبی نسبی برای سیستم SVM در [۱۱]..... ۲۱
- جدول ۸-۲: تعداد داده های آموزشی و آزمایش برای ۳ گروه سنی در [۱۲]..... ۲۲
- جدول ۹-۲: تعداد داده های آموزشی و آزمایشی برای ۱۳ گروه سنی در [۱۲]..... ۲۳
- جدول ۱۰-۲: نتیجه ی آزمایشات برای تقسیم بندی افراد به ۳ گروه سنی در [۱۲]..... ۲۳
- جدول ۱۱-۲: نتایج کمترین نرخ خطا برای مرتبه های مختلف PLP در [۱۳]..... ۲۵
- جدول ۱۲-۲: نرخ خطای بدست آمده برای مقادیر مختلف گاما، طول قاب و تعداد ضرایب MFCC در [۱۳]..... ۲۶
- جدول ۱۳-۲: صحت پیش بینی سن و جنس بازای استفاده از رده بندهای متفاوت در [۱۴]..... ۲۸
- جدول ۱۴-۲: MAE(Mean Absolute Error)، آزمایشات با مجموعه ویژگی های متفاوت در [۸]..... ۳۲
- جدول ۱۵-۲: تعداد گروه های ویژگی مختلف، که بر اساس تحلیل MAXR انتخاب شده اند [۸]..... ۳۲
- جدول ۱۶-۲: رده بندی تخمین گروه سنی بر اساس PNN برای [۷]..... ۳۳
- جدول ۱۷-۲: رده بندی تخمین گروه سنی بر اساس GMM برای [۷]..... ۳۳
- جدول ۱۸-۲: تعداد گوینده ها در دسته های مختلف سن - جنس در [۱۵]..... ۳۶
- جدول ۱۹-۲: صحت شناسایی گروه سنی بر حسب درصد در [۱۵]..... ۳۷
- جدول ۲۰-۲: ماتریس ترکیبی نسبی برای روش مورد نظر در شناسایی ۶ گروه سن - جنسیت در [۱۵]..... ۳۷
- جدول ۲۲-۲: ماتریس ترکیبی برای ۷ دسته ی سن - جنسیت در [۱۶]..... ۳۸
- جدول ۳-۱: عملکرد تخمین فرمنت با استفاده از دو روش ذکر شده ، بر یک قاب ۳۲ میلی ثانیه ای از گفتار ساختگی نویزی در [۲۳]..... ۸۳
- جدول ۲-۳: نرخ خطای کلمه برای نویز کارخانه و f16 در [۲۵]..... ۹۵

- جدول ۳-۳: نرخ شناسایی گوینده (%) برای گفتار پاک برای مقایسه MFCC , A-MFCC در [۲۶].....۹۹
- جدول ۳-۴: نرخ شناسایی گوینده (%) برای گفتار آموزشی خراب شده با نویز F16 و کارخانه و سفید برای مقایسه MFCC , A-MFCC در [۲۶].....۹۹
- جدول ۳-۵: اندازه های مختلف SNR برای پایگاه داده در [۲۷].....۱۰۵
- جدول ۳-۶: نرخ خطای کلمه با به کارگیری انواع روش های جبران کانال در [۲۷].....۱۰۵
- جدول ۴-۱: نتایج بدست آمده برای حالت اول MFCC (خطای آزمایش بر حسب درصد).....۱۱۵
- جدول ۴-۲: نتایج بدست آمده برای حالت دوم a , b (خطای آزمایش بر حسب درصد).....۱۱۶
- جدول ۴-۳: نتایج بدست آمده برای حالت دوم c (خطای آزمایش بر حسب درصد).....۱۱۶
- جدول ۴-۴: نتایج بدست آمده برای حالت سوم (خطای آزمایش بر حسب درصد).....۱۱۸
- جدول ۴-۵: نتایج بدست آمده برای حالت چهارم (خطای آزمایش بر حسب درصد).....۱۱۹
- جدول ۴-۶: نتایج بدست آمده برای حالت اول LPC (خطای آزمایشی بر حسب درصد).....۱۲۰
- جدول ۴-۷: نتایج بدست آمده برای حالت دوم LPC (خطای آزمایشی بر حسب درصد).....۱۲۰
- جدول ۴-۸: درصد خطای آزمایش برای روش استخراج ویژگی RootMFCC بازای گامای مختلف.....۱۲۲
- جدول ۴-۹: نتایج بدست آمده برای قسمت ۴-۴-۴ PLP (خطای آزمایش بر حسب درصد).....۱۲۳
- جدول ۴-۱۰: نتایج حاصل از اعمال تابع PCA بر ماتریس ویژگی های بدست آمده.....۱۲۵
- جدول ۴-۱۱: نتایج مربوط به اعمال تابع LDA بر روش ذکر شده در قسمت ۴-۴-۱، حالت (ب a).....۱۲۶
- جدول ۴-۱۲: نتایج حاصل از تفکیک داده ها بر اساس تفکیک جنسیت.....۱۲۶
- جدول ۴-۱۳: مقایسه عملکردهای سیستم های مختلف تخمین سن با استفاده از گفتار.....۱۳۳

فهرست شکل ها:

- شکل ۱-۱ : مکانیزم تولید گفتار صدا..... ۵
- شکل ۱-۲ : نتیجه ی رده بندی سن برای پایگاه داده ی DES در [۱۰] ۱۷
- شکل ۲-۲ : نتیجه ی رده بندی سن برای پایگاه داده ی ELSDSR در [۱۰] ۱۷
- شکل ۳-۲ : مدل کردن هر گوینده با یک GMM در [۱۱] ۲۰
- شکل ۴-۲ : نمایشی از بانک فیلتر در مقیاس Mel [۱۲] ۲۲
- شکل ۵-۲ : نمودار نرخ خطا بر حسب پارامتر گاما در RBF kernel ، با استفاده از PLP با ۳ مقدار متفاوت از مرتبه (به عنوان بردار ویژگی) [۱۳] ۲۴
- شکل ۶-۲ : نمودار نرخ خطا بر حسب پارامتر گاما در RBF kernel ، با به کارگیری مرتبه ی ثابت MFCC ، با طول قاب های زمانی مختلف در [۱۳] ۲۵
- شکل ۷-۲ : نمودار نرخ خطا بر حسب پارامتر گاما در RBF kernel ، با به کارگیری مرتبه های مختلف MFCC ، با طول قاب زمانی برابر 25ms در [۱۳] ۲۶
- شکل ۸-۲ : سلسله مراتب رده بندی در [۱۴] ۲۸
- شکل ۹-۲ : مدل منحنی pitch در یک قسمت صدا در [۸] ۳۱
- شکل ۱۰-۲ : بلوک دیاگرامی از تخمین زنده ی گروه سنی در [۷] ۳۳
- شکل ۱۱-۲ : بلوک دیاگرام روش مورد نظر در فاز آموزش اولیه برای آموزش (WSNMF(weighted supervised non-negative matrix factorization) در [۱۵] ۳۵
- شکل ۱۲-۲ : بلوک دیاگرام روش مورد نظر در فاز آموزش ثانویه برای آموزش GRNN(general regression neural network) در [۱۵] ۳۶
- شکل ۱۳-۲ : بلوک دیاگرام روش مورد نظر در فاز آزمایش در [۱۵] ۳۶
- شکل ۱۴-۲ : نمایی کلی از سیستم و چگونگی ترکیب ۵ روش مختلف و به کارگیری آنها در [۱۶] ۳۹
- شکل ۱-۳ : بلوک دیاگرام نشان دهنده ی ۳ گام اصلی در استخراج ویژگی برای شناسایی گفتار..... ۴۲
- شکل ۲-۳ : دسته بندی الگوریتم های تحلیل طیفی..... ۴۳
- شکل ۳-۳ : مقایسه ی مقیاس Bark , Mel ۴۴
- شکل ۴-۳ : گام های محاسبه ی LPCC ۴۷
- شکل ۵-۳ : گام های محاسبه ی PLPCC ۴۸

- شکل ۳-۷: بلوک دیاگرام نشان دهنده ی گام های محاسبه ی MFCC ۵۲
- شکل ۳-۸: نمایشی از تاثیر کاهش ویژگی در ابعاد یک بعدی و دو بعدی..... ۵۴
- شکل ۳-۹: مسئله ی داده های ناکافی مشابه مسائل مطرح شده در curve fitting می باشد..... ۵۵
- شکل ۳-۱۰: نمایشی از عملکرد روش LDA..... ۵۷
- شکل ۳-۱۱: نمایشی از عملکرد PCA..... ۵۹
- شکل ۳-۱۲: تحلیل مقادیر ویژه ماتریس کواریانس..... ۶۲
- شکل ۳-۱۳: مقایسه عملکرد PCA , LDA و نمایش میزان تفکیک نمونه های دو کلاس..... ۶۳
- شکل ۳-۱۴: مقایسه عملکرد PCA , LDA و نمایش میزان تفکیک نمونه های دو کلاس..... ۶۳
- شکل ۳-۱۵: طیف توان و تابع خودهمبستگی برای یک قاب ۳۲ ms از گفتار 'ey' voiced در [۲۳]..... ۶۸
- شکل ۳-۱۶: طیف توان و تابع خودهمبستگی برای یک قاب ۳۲ms از گفتار 's' unvoiced در [۲۳]..... ۶۹
- شکل ۳-۱۷: تحلیل زمان کوتاه سیگنال نویز سفید اتفاقی مصنوعی با استفاده از قاب های ۳۲ms..... ۷۰
- شکل ۳-۱۸: تحلیل زمان کوتاه سیگنال نویز chrip مصنوعی با استفاده از قاب های ۳۲ms..... ۷۱
- شکل ۳-۱۹: تحلیل زمان کوتاه سیگنال نویز ضربه ای مصنوعی با استفاده از قاب های ۳۲ms..... ۷۲
- شکل ۳-۲۰: تحلیل سیگنال نویز اتومبیل با استفاده از قاب های ۳۲ میلی ثانیه ای..... ۷۳
- شکل ۳-۲۱: نمایش روش HASE (با استفاده از رشته ی خود همبستگی higher – lag ، پنجره گذاری شده با hamming) بر یک قاب ۳۲ میلی ثانیه ای از گفتار voiced یک خانم. 'ey'..... ۷۵
- شکل ۳-۲۲: تابع پنجره و طیف توان آن برای a (پنجره ی hamming و b) پنجره kaiser (۱۱.۳ =) و c) پنجره ی DDR hamming ۷۶
- شکل ۳-۲۳: نمایش روش HASE (با استفاده از رشته ی خود همبستگی higher – lag ، پنجره گذاری شده با kaiser) بر یک قاب ۳۲ میلی ثانیه ای از گفتار voiced یک خانم. 'ey'..... ۷۸
- شکل ۳-۲۴: نمایش روش HASE (با استفاده از رشته ی خود همبستگی higher – lag ، پنجره گذاری شده با DDR hamming) بر یک قاب ۳۲ میلی ثانیه ای از گفتار صدادار یک خانم. 'ey'..... ۷۹
- شکل ۳-۲۵: مقایسه روش های تخمین طیفی با به کارگیری یک قاب ۳۲ ms از گفتار صدادار ساختگی پاک. ۸۰
- شکل ۳-۲۶: مقایسه ی روش های تخمین طیفی با به کارگیری یک قاب ۳۲ میلی ثانیه ای از گفتار صدای ساختگی که با نویز سفید اتفاقی ساختگی با 10 dB SNR تخریب شده است ۸۱

- شکل ۳-۲۷: مقایسه ی روش های تخمین طیفی با به کارگیری یک قاب ۳۲ میلی ثانیه ای از گفتار صدای ساختگی که با نویز chrip ساختگی با 10 dB SNR تخریب شده است..... ۸۲
- شکل ۳-۲۸: مقایسه ی روش های تخمین طیفی با به کارگیری یک قاب ۳۲ میلی ثانیه ای از گفتار صدای ساختگی که با نویز ضربه ای اتفاقی ساختگی با 10 dB SNR تخریب شده است..... ۸۲
- شکل ۳-۲۹: مقایسه ی روش های تخمین طیفی با به کارگیری یک قاب ۳۲ میلی ثانیه ای از گفتار صدای ساختگی که با نویز واقعی اتومبیل با 10 dB SNR تخریب شده است..... ۸۳
- شکل ۳-۳۰: طیف نگاره ی گفتار واقعی (عبارت 'MAL_19Z96Z8 A' از پایگاه داده ی Aurora)..... ۸۴
- شکل ۳-۳۱: بلوک دیاگرام الگوریتم استخراج ویژگی در AMFCC..... ۸۵
- شکل ۳-۳۲: بلوک دیاگرام الگوریتم استخراج ویژگی در MFCC..... ۸۵
- شکل ۳-۳۳: بلوک دیاگرام مراحل استخراج ویژگی در روش Mfcc دوطیفی در [۲۴]..... ۸۷
- شکل ۳-۳۴: نسبت تشخیص صحیح بر حسب SNR برای روش طیفی و روش دو طیفی در زمانی که نویز سفید گوسی اضافه شود..... ۸۸
- شکل ۳-۳۵: هیستوگرام مربوط به نویز های babble، اتومبیل و کارخانه..... ۸۸
- ۳-۳۶: نسبت تشخیص صحیح بر حسب SNR برای روش طیفی (خط توپر) و روش دو طیفی (خط چین) در زمانی که نویز babble اضافه شود..... ۹۰
- ۳-۳۷: نسبت تشخیص صحیح بر حسب SNR برای روش طیفی (خط توپر) و روش دو طیفی (خط چین) در زمانی که نویز اتومبیل اضافه شود..... ۹۰
- ۳-۳۸: نسبت تشخیص صحیح بر حسب SNR برای روش طیفی (خط توپر) و روش دو طیفی (خط چین) در زمانی که نویز کارخانه اضافه شود..... ۹۰
- شکل ۳-۳۹: انرژی های بانک log mel-filter از یک گفتار پاک و نویزی (مشوش) در [۲۵]..... ۹۲
- شکل ۳-۴۰: مربع انرژی های بانک log mel-filter برای گفتار پاک و نویزی..... ۹۳
- شکل ۳-۴۱: اندازه ی تفاوت بین اولین ۱۳ ضریب DCT برای دو نمونه ی logMelFBS و اندازه ی تفاوت بین اولین ۱۳ ضریب DCT برای توان دوم logMelFBS..... ۹۳
- شکل ۳-۴۲: محاسبه ی MFCC در [۲۶]..... ۹۸
- شکل ۳-۴۳: بلوک دیاگرام پردازشگر A-MFCC در [۲۶]..... ۹۸
- شکل ۳-۴۴: نمایش اعوجاج سیگنال گفتار در [۲۷]..... ۱۰۲

- شکل ۴-۱ : معرفی اجمالی الگوریتم پیشنهادی..... ۱۱۰
- شکل ۴-۲ : ۹ تکرار مربوط به واج "ای" که توسط شخص شماره ۱۰۰ در پایگاه داده ، ادا شده است..... ۱۱۳
- شکل ۴-۳ : بلوک دیاگرام مربوط به روش حالت اول..... ۱۱۴
- شکل ۴-۴ : بلوک دیاگرام مربوط به حالت دوم ۱۱۷
- شکل ۴-۵ : بلوک دیاگرام مربوط به روش حالت سوم ۱۱۸
- شکل ۴-۶ : بلوک دیاگرام مربوط به حالت چهارم ۱۱۹
- شکل ۴-۷ : بلوک دیاگرام مربوط به محاسبه ی ضرایب ویژگی RootMFCC..... ۱۲۱
- شکل ۴-۸ : نمودار تغییرات خطای کل بر حسب تغییرات گاما ۱۲۳
- شکل ۴-۹ : نمودار مقایسه نتایج درصد خطای کل برای روش های مختلف..... ۱۲۷
- شکل ۴-۱۰ : درصد خطا برای هر ۳ گروه سنی مختلف (گروه h و گروه g و گروه w) برای همه ی روش های گفته شده ۱۲۸
- شکل ۴-۱۱ : درصد خطای کل برای تقسیم داده ها به ۲ گروه سنی ، ۳ گروه سنی و ۵ گروه سنی ۱۲۹
- شکل ۴-۱۲ : مقایسه درصد خطا در هر گروه سنی h,g,w ، برای ۴ روش نشان داده شده در شکل، با استفاده از روش mahalanobis distance..... ۱۳۰
- شکل ۴-۱۳ : مقایسه در صد خطای کل بدست آمده برای رده بند , linear و روش Mahalanobis برای ۴ روش استخراج ویژگی مذکور ۱۳۱

فهرست علائم و اختصارات:

Pitch : فرکانس صدا

Octave : فاصله ی زمانی(مدت) بین یک pitch و pitch دیگری با فرکانس دو برابر و یا نصف

Jitter : حداکثر مقدار انحراف از فرکانس پایه (F0)

Shimmer : تغییرات محلی توان

Formant: قله های موجود در طیف صوت (تجمع انرژی صوت در یک فرکانس خاص از شکل موج صوت)

: linguistic information

اطلاعات نمادین که با مجموعه ای از نمادهای مجزا و قواعدی برای ترکیب آن ها، تعریف می شوند.

: paralinguistic information

اطلاعاتی که از نوشته ها قابل استنتاج نمی باشند ولی توسط گوینده برای اصلاح و یا تکمیل اطلاعات نمادین،

اضافه می شوند.

: non-linguistic information

در رابطه با سن، جنس، ویژگی شخصی و حالات فیزیکی و احساسی گوینده و... می باشد. اگرچه گوینده برای

بیان احساسی خاص ، روش صحبت کردن خود را کنترل می کند، این ویژگی ها قابل کنترل نیستند.

ANN : artificial NEURAL NETWORK

HMM : hidden markove modle

GMM : Gussian mixture modle

KNN:K- nearest neighbor

SVM: support vector machine

SVR:support vector regression

PNN : PROBABILISTIC NEURAL NETWORK

DFT : discrete fuorier transform

KLT : Karhunen-Loeve transform

ICA : INDEPENDENT COMPONENT ANALYSIS

EM : EXPECTATION MAXIMIZATION

MAP : MAXIMUM A POSTERIORI TRAINING ALGORITHM

RBF(Radial basic function)

RASTA: Relative Spectral

MAXR : MAX-Relevance (feature selection method)

فصل اول

مقدمه

در تولید گفتار، انواع متعددی از اطلاعات نیز به طور موازی تولید می شوند، هم اطلاعات زبان شناسی و هم تعداد زیادی از صفات شخصی گوینده. که اطلاعات اخیر شامل نشانه هایی است که به ویژگی های فیزیکی دستگاه صوتی گوینده مربوط می شود. بعلاوه تعدادی از ویژگی های غیر فیزیولوژیکی گوینده نیز تاثیر خود را بر سیگنال صوتی می گذارند. مانند اطلاعاتی درباره ی حالت احساسی گوینده، ویژگی های مذهبی و اجتماعی.

تعیین کیفی دقیق این اطلاعات کار دشواری است زیرا اکثر این اطلاعات با مجموعه نشانه های یکسانی علامت گذاری می شوند. مثلا فرکانس اصلی هم به رابطه ی بین محتوی علمی _معنایی عبارت و آهنگ تلفظ بستگی دارد و هم به بسیاری فاکتورهای دیگر مانند جنس و گروه سنی و حالت احساسی گوینده .

امروزه شاهد گسترش استفاده ی افراد مختلف جامعه از کودک و خردسال تا میان سال و کهنسال و همچنین قشرهای گوناگون جامعه کودکان و دانش آموزان، دانشجویان و محققان دانشگاهی، صنعتی و کارکنان ادارات شرکتی، استفاده های خانگی تا مدیریت های شرکتی و کارخانجات بزرگ صنعتی و کشاورزی از کامپیوتر هستیم. این واقعیات به این معنی است که برای ارتباط سریع تر و آسان تر افراد از کامپیوتر، سیستم های مکالمه ای- گفتاری باید گسترش یابند تا بتوانند با کاربر ارتباط برقرار کرده و راهبرد های مکالمه ی صمیمانه با همه ی نسل ها را بهبود بخشد. اگرچه این کار با به کارگیری یک شیوه ی واحد و عمومی برای همه ی نسل ها امکان پذیر است، ولی کنترل دینامیک، انعطاف پذیر و دقیق بر ارتباط با کاربر و طرح های مکالمه ای، را می توان با تخمین اتوماتیک سن افراد و گوینده ها، محقق ساخت.

بیشتر سیستم های گفتاری - مکالمه ای رایج تنها دنباله ای از کلمات را از صدای گوینده استخراج می کنند. این کار به طور وسیعی از اطلاعات مفید دیگری که می توان از گفتار بدست آورد، چشم پوشی می کند. انسان می تواند یک سری صفات مهم را درباره ی گوینده، مانند جنسیت، سن، لهجه، نژاد، احساسات، سطح تحصیلات و حتی قد و وزن را استنباط کند. این نوع ویژگی های صوت می تواند برای طبیعی ساختن رابط های گفتاری انسان - ماشین، به طور بهینه ای مورد استفاده قرار گیرند. مثلا یک سیستم مشتری - محور با دانستن اینکه یک شخص مسن از قسمت جنوبی کشور که ناراحت و افسرده است، می تواند با ساختن یک دنباله ی مکالمه ای مناسب به لهجه ی جنوبی، مایه ی آرامش کاربر شود.

در ارتباطات گفتاری انسان - با - انسان، مخصوصا در ارتباطات شخص - با - شخص، انتظار می رود که گوینده، شیوه ی سخن گفتن خود را با توجه به خصوصیات و پاسخ های شنونده، تغییر دهد. در ارتباطات بر پایه ی گفتار، به ویژه در اولین ملاقات برای تنظیم رفتار مناسب، یک فاکتور مهم برای هر کسی سن می باشد. گفتار مردان و زنان نه تنها حامل بار معنایی عبارات گفته شده است بلکه شامل ویژگی هایی است که اطلاعاتی غیر زبانی، وابسته به گوینده را تأمین می کنند مانند جنس، سن گوینده، حالت احساسی گوینده و ... با استخراج این ویژگی ها برای هر فرد، می توانیم نحوه ی صحبت کردن خود را با طرف مقابل تنظیم کنیم.

یک کاربرد رایج تخمین سن، هر روزه در مکالمات تلفنی صورت می گیرد. هم چنین بعضی شرکت ها نیاز به سیستم تخمین سن اتوماتیک دارند تا مثلا بتوانند برای گروه های سنی مختلف از مشتریان خود، موسیقی های مناسب پخش کنند.

علاوه بر سیستم های مکالمه ای، سن به عنوان یک پارامتر کلاسه بندی مهم در بسیاری دیگر از کاربردها، در نظر گرفته می شود. مثلا اجازه دادن به گروه سنی خاصی برای دسترسی به حقوقی ویژه و یا لحاظ کردن قیمت های متفاوتی در خرید کالاها برای گروه های سنی مختلف.

نماینده های اجرای قوانین برای تشخیص هویت یک فرد، تکنیک های بیومتریک مختلفی را مد نظر قرار می دهند. مشخصه های مختلف بیومتریک، می توانند برای هویت شناسی قانونی مورد استفاده قرار گیرند. مانند الگوهای اثر انگشت، مشخصه های چهره، ترکیب هندسی دست، تغییرات امضا و الگوهای صدا. انتخاب یک روش مناسب به اطلاعات موجود و نیز به درجه اطمینان آن روش در یک کاربرد مشخص، بستگی دارد. در بعضی از جُرم ها، مدرک در دسترس ممکن است به شکل مکالمات ضبط شده باشد. الگوهای گفتاری می توانند دارای اطلاعات مهمی برای ماموران قانون باشد. برای مثال، نمونه ی گفتاری یک شخص می تواند اطلاعاتی درباره ی سن، جنس، لهجه، حالات فیزیولوژیکی باشد و یا حتی اینکه آن شخص عضو یک گروه اجتماعی یا مذهبی خاصی باشد. در نتیجه، از گفتار می توان برای شناسایی گوینده که در موارد زیادی مانند آدم دزدی، تماس های تهدید آمیز و ... مورد نیاز است، استفاده کرد.

۱-۱: خلاصه ای از چگونگی تغییر صدای انسان با افزایش سن

هر موجود زنده ای گذر عمر و افزایش سن را تجربه می کند. این یک مکانیزم پیچیده است که از جنبه های مختلفی بر یک شخص تاثیر می گذارد. در نتیجه مورد بررسی قرار دادن مفهوم تغییر سن در اکثر علوم طبیعی و قوانین بشری، گریز ناپذیر است.

بالارفتن سن در شیوه ی سخن گفتن ما تغییراتی ایجاد می کند. صدای ما و الگوهای گفتاری ما از کودکی تا پیری تغییر می کند. اگرچه اکثر تغییرات در کودکی و بلوغ اتفاق می افتد، تغییرات وابسته به سن را می توان از دوران بزرگسالی تا پیری نیز مشاهده کرد. در نتیجه، سن ما در گفتار ما منعکس می شود. بدین ترتیب سن گوینده با به کارگیری روش های متعددی، عمدتاً تحلیل صوتی و آزمایشات ادراکی، مورد مطالعه قرار می گیرد.

تغییرات وابسته به سن در گفتار بزرگسالان به طور گسترده از ۱۹۶۰ مورد مطالعه قرار گرفته است. اکثر مطالعات جنبه های صوتی و ادراکی را در نظر گرفته اند، اگرچه بعضی روش های تکنولوژی گفتار نیز دنبال شده است. بدلیل پیچیدگی فرایند افزایش سن، در بیشتر تحقیقات نیاز است تا تغییرات وابسته به سن در گفتار به خوبی درک شود.

[۱]

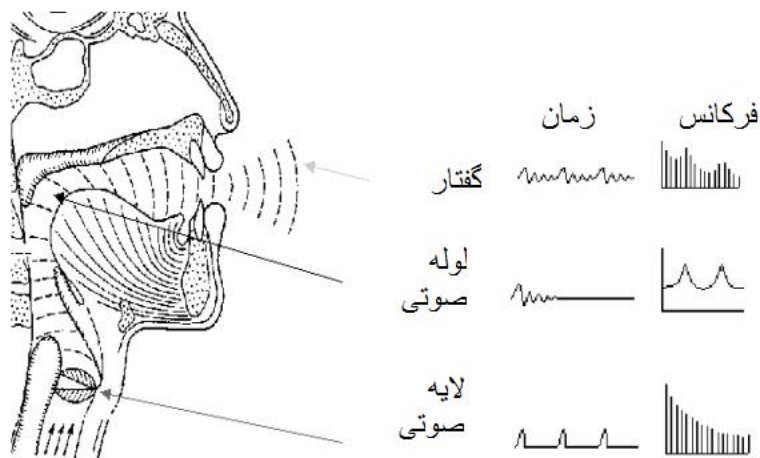
با گذشت سن مقداری تغییرات فیزیکی و هورمونی در مرد ها و زن ها رخ می دهد که می تواند بر صدای آنها تاثیر گذارد. برای مثال یک موجی از هورمون های فعال سازی در زمان بلوغ وجود می آید که بر انتقال صدا از کودکی به یک فرد بالغ تاثیر می گذارد.

در زمان بلوغ، بر اثر افزایش هورمون تستوسترون، صدای یک مرد با تقریب ۱ octave در pitch، کاهش می یابد. در خانم ها نیز در زمان بلوغ بر اثر هورمون های استروژن و پروژسترون، صدای آنها کاهش می یابد ولی با درجه ی کمتر تقریباً یک چهارم octave در pitch.

با بالا رفتن سن در مردها، مقدار هورمون تستوسترون کاهش می یابد و میانگین حجم صدا و تغییرپذیری در حجم صدای آنها افزایش می یابد. با بالا رفتن سن در خانم ها، کاهش فعالیت هورمون های جنسی، مخصوصاً در دوران یائسگی، بر لایه های صوتی خانمها و عملکرد حنجره تاثیر می گذارد. صدای خانم ها عمیق تر می شود، فرکانس آوایی حداکثر کاهش می یابد و بازه ی صوتی گسترده می شود که باعث ادا شدن pitch های پایین تر می شود.

تغییرپذیری pitch و jitter بر اثر مرور سن افزایش می یابد. [۲]

۱-۱-۱ مکانیزم تولید گفتار با گذشت سن :



شکل ۱-۱ : تولید گفتار صدا . فشار هوای شش ها تولید ارتعاش لایه های صوتی می کند ، که یک سیگنال منبع صدای ضربه ای شبه - متناوب را نتیجه می دهد . سیگنال منبع صدا باعث تحریک لوله ی صوتی ، که مانند یک بدنه ی رزونانسی عمل می کند که فرکانس های معینی را تقویت و یا تضعیف می کند و باعث تولید گفتار می شود . [۳]

مکانیزم تولید صدا در انسان را می توان به ۳ بخش تقسیم کرد: شش ها، لایه های صوتی و لوله ی صوتی. در شکل ۱-۱ مکانیزم صوتی انسان نشان داده شده است. فشار هوای شش ها باعث جریان یافتن هوا از بین حنجره، که فضای بین لایه های صوتی است، می شود . لایه های صوتی دو توده از گوشت، رباط و ماهیچه هستند که بین جلو و عقب حنجره کشیده شده و بسته به رانش و ربایش لایه های صوتی، در حالت های مختلف ارتعاشی (صوت های صدا دار) قرار می گیرند و یا اینکه اصلا ارتعاشی صورت نمی گیرد (صوت های بی صدا).

برای صوت های صدادار، لایه های صوتی به صورت شبه تناوبی باز و بسته می شوند و در نتیجه جریان هوای حنجره ای را به جریان ضربه هایی تبدیل می کند که سیگنال منبع صدا نامیده می شود. سپس سیگنال منبع صدا از لوله ی صوتی، که از حنجره آغاز و به لب ها ختم می شود، عبور می کند. لوله ی صوتی مانند یک بدنه با رزونانس ها (فرکانس های فرمنت) و ضد رزونانس ها (صفر ها) عمل می کند. لوله ی صوتی به عنوان فیلتر صوتی که طیف صوت را شکل می دهد، انجام وظیفه می کند. صداهای متنوع صوت با تنظیم شکل لوله ی صوتی و هم چنین سیگنال منبع صدا، تولید می شوند.

گفتار بر اثر ارتعاش تارهای صوتی هنگامی که نفس از شش ها بیرون داده می شوند، تولید می شود. تنوع صداهای گفتار عمدتاً به دلیل تنوع در جرم و طول تارهای صوتی و بندهای موجود در لوله ی صوتی می باشد.