بسم الله الرحمن الرحيم

معاونت پژوهش و فن آوری

به نام خدا

منشور اخلاق پژوهش

یاری از خداوند سبحان و اعتقاد به این که عالم محضر خداست و همواره ناظر بر اعمال انسان و به منظور پاس داشت مقام بلند دانش و پژوهش و نظریه اهمیت جایگاه دانشگاه در اعتلای فرهنگ و تمدن بشری، ما دانشجویان و اعضاء هیات علمی واحدهای دانشگاه آزاد اسلامی متعهد می گردیم اصول زیر را در انجام فعالیت های پژوهشی مد نظر قرار داده و از آن تخطی نکنیم:

۱-اصل حقیقت جویی: تلاش در راستای پی جویی حقیقت و وفاداری به آن و دوری از هرگونه پنهان سازی حقیقت.

۲-اصل رعایت حقوق: التزام به رعایت کامل حقوق پژوهشگران و پژوهیدگان (انسان، حیوان و نبات) و سایر صاحبان حق.

۳-اصل مالکیت مادی و معنوی: تعهد به رعایت کامل حقوق مادی و معنوی دانشگاه و کلیه همکاران پژوهش.

٤-اصل منافع ملی: تعهد به رعایت مصالح ملی و در نظر داشتن پیشبرد و توسعه کشور در کلیه مراحل پژوهش.

۵-اصل رعایت انصاف و امانت: تعهد به اجتناب از هرگونه جانب داری غیر علمی و حفاظت از اموال، تجهیزات و منابع در اختیار.

٦-اصل رازداری: تعهد به صیانت از اسرار و اطلاعات محرمانه افراد، سازمان ها و کشور و کلیه افراد و نهادهای مرتبط با تحقیق.

۷-اصل احترام: تعهد به رعایت حریم ها و حرمت ها در انجام تحقیقات و رعایت جانب نقد و خودداری از هرگونه حرمت شکنی.

۸-اصل ترویج: تعهد به رواج دانش و اشاعه نتایج تحقیقات و انتقال آن به همکاران علمی و دانشجویان به غیر از مواردی که منع قانونی دارد.

۹-اصل برائت: التزام به برائت جویی از هرگونه رفتار غیرحرفه ای و اعلام موضع نسبت به کسانی که حوزه علم و پژوهش را به شائبه های غیر علمی می آلایند.

دانشگاه آزاد اسلامی

واحد علوم و تحقیقات شاهرود

تعهدنامه اصالت رساله یا پایان‌نامه

اینجانب فاطمه همتی دانش آموخته مقطع کارشناسی ارشد ناپیوسته در رشته مهندسی برق گرایش الکترونیک که در تاریخ ۱۳۹۲/۱۲/۲۰ از پایان نامه خود تحت عنوان " بکارگیری اطلاعات موجود در ضرایب کپسترال دینامیکی گفتار فارسی جهت تشخیص استرس" با کسب نمره ۱۷/۷۵ و درجه خوب دفاع نموده ام بدینوسیله متعهد می‌شوم :

۱) این پایان نامه حاصل تحقیق و پژوهش انجام شده توسط اینجانب بوده و در مواردی که از دستاوردهای علمی و پژوهشی دیگران (اعم از پایان نامه، کتاب، مقاله و ....) استفاده نموده‌ام، مطابق ضوابط و رویه موجود، نام منبع مورد استفاده و سایر مشخصات آنرا در فهرست مربوطه ذکر و درج کرده‌ام.

۲) این پایان نامه قبلا برای دریافت هیچ مدرک تحصیلی (هم سطح، پایین تر یا بالاتر) در سایر دانشگاهها و موسسات آموزش عالی ارائه نشده است.

۳) چنانچه بعد از فراغت از تحصیل، قصد استفاده و هرگونه بهره برداری اعم از چاپ کتاب، ثبت اختراع و ... از این پایان نامه را داشته باشم، از حوزه معاونت پژوهشی واحد، مجوزهای مربوطه را اخذ نمایم.

٤) چنانچه در هر مقطع زمانی خلاف موارد فوق ثابت شود، عواقب ناشی از آن را می‌پذیرم و دانشگاه آزاد اسلامی واحد علوم و تحقیقات مجاز است با اینجانب مطابق ضوابط و مقررات رفتار نموده و در صورت ابطال مدرک تحصیلی‌ام هیچگونه ادعایی نخواهم داشت.

نام و نام خانوادگی:

فاطمه همتی

تاریخ و امضاء:

دانشگاه آزاد اسلامی

واحد علوم تحقیقات شاهرود

دانشکده برق گروه الکترونیک

پایان نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی برق (M.Sc)

گرایش الکترونیک

عنوان :

بکارگیری اطلاعات موجود در ضرایب کپسترال دینامیکی گفتار فارسی جهت تشخیص استرس

استاد راهنما:

دکتر حسین مروی

استاد مشاور:

دکتر علیرضا احمدی فرد

نگارش :

فاطمه همتی

زمستان ۱۳۹۲

سپاسگزاری

حمد و سپاس خداوند منان را

که توفیقی نصیب این بنده حقیر نمود تا بتوانم از زحمات و تلاش‌های استاد گران‌قدرم ، جناب آقای دکتر حسین مروی استاد راهنما و مشوق همیشگی ، جناب آقای دکتر علیرضا احمدی فرد استاد مشاور ، جناب آقای دکتر سلیمانی استاد داور ،جناب آقای دکتر حریمی مدیر محترم گروه برق و سایر عزیزانی که بنده را دلسوزانه در پیشبرد این راه یاری نمودند به استناد حدیث شریف « **من لم یشکر المخلوق ، لم یشکر الخالق** » تشکر و قدردانی نمایم .

امید است با یاری پروردگار متعال این پژوهش زمینه‌ای برای فعالیت‌های وسیع‌تر و نیل به اهداف عالی‌تر برای عمران و آبادانی کشور عزیزمان ایران اسلامی باشد.

و من الله التوفیق

تقدیم به

خدایی که آفرید

جهان را، انسان را، عقل را، علم را، معرفت را، عشق را

و به کسانی که عشقشان را در وجودم دمید:

پدر، مادر و همسر مهربانم.

گاهی بیایید و احوالشان را بپرسید...

فهرست مطالب

عنوان **صفحه**

چکیده ۱

فصل اول: استرس و تاثیر آن بر گفتار

۱-۱- مقدمه ۳
۱-۲- تعریف استرس..... ۵
۱-۲-۱- مدیریت استرس ۸
۱-۲-۲- پاسخ مغز به استرس..... ۹
۱-۳- حوزه گفتار تحت استرس ۹
۱-۳-۱- حوزه تولید..... ۱۰
۱-۳-۲- حوزه ادراک ۱۳
۱-۳-۳- حوزه سیستم‌های سخنرانی ۱۳

فصل دوم : مروری بر پیشینه تشخیص استرس از روی گفتار

۲-۱- مقدمه ۱۶
۲-۲- سیستم تشخیص استرس از روی گفتار..... ۱۶
۲-۲-۱- استخراج ویژگی ۱۶
۲-۲-۱-۱- ویژگی‌های عروضی..... ۱۷
۲-۲-۱-۱-۱- فرکانس گام ۱۸
۲-۲-۱-۱-۲- انرژی ۱۹
۲-۲-۱-۱-۳- شدت..... ۲۰

۲-۲-۱-۱-٤- مدت زمان ۲۲
۲-۲-۱-۲- ویژگی‌های طیفی..... ۲۳
۲-۲-۱-۲-۱- فرمنت‌ها ۲٤
۲-۲-۱-۲-۲- طیف دستگاه صوتی..... ۲٦
۲-۲-۱-۲-۳- ویژگی مربوط به دهانه نای ۲۹
۲-۲-۱-۲-۳-۱- طیف مربوط به دهانه نای ۲۹
۲-۲-۱-۲-٤- ضرایب مل کپستروم ۳۱
۲-۲-۱-۲-٤-۱- مل کپستروم ۳۵
۲-۲-۱-۲-٤-۲- دلتا مل و دلتا دلتا مل کپستروم..... ۳٦
۲-۲-۱-۲-٤-۳- همبستگی خودکار مل..... ۳٦
۲ -۲-۱-۲-٤-٤- ضرایب کپستروم تابع گرین‌وود..... ۳٦
۲-۲-۱-۲-٤-۵- ضرایب تابع خودهمبستگی ۳٦
۲-۲-۱-۲-۵- ضرایب پیشگویی خطی..... ۳۸
۲-۲-۱-۲-٦- ضرایب پیشگویی خطی ادراکی ۳۸
۲-۲-۱-۳- ویژگی‌های کیفیت صدا ٤۰
۲-۲-۱-۳-۱- Jitter ٤۰
۲-۲-۱-۳-۲- Shimmer ٤۰
۲-۲-۱-٤- ویژگی‌های مبتنی بر اپراتور انرژی تیگر ٤۱
۲-۲-۱-٤-۱- اپراتور انرژی تیگر..... ٤۲
۲-۲-۱-٤-۲- TEO-FM-Var ٤٤
۲-۲-۱-٤-۳- TEO-Auto-Env ٤۸
۲-۲-۱-٤-٤- TEO-CB-Auto-Env ۵۰
۲-۲-۱-۵- نرخ عبور از صفر..... ۵٤
۲-۲-۲- کاهش ویژگی ۵٤
۲-۲-۳- کلاسه‌بندی ۵۵
۲-۲-۳-۱- الگوریتم AdaBoost..... ۵۵

۲-۲-۳-۲- ماشین‌بردار پشتیبان (SVM) ۵۷
۲-۲-۳-۲-۱- ماشین‌بردار پشتیبان خطی ۵۸
۲-۲-۳-۲-۲- ماشین‌بردار پشتیبان غیرخطی ۵۹
۲-۲-۳-۳- طبقه‌بندی نزدیکترین همسایه ٦۱
۲-۲-۳-۳-۱- انتخاب k ٦۱
۲-۲-۳-۳-۲- الگوریتم دسته‌بندی Knn ٦۱
۲-۳- پایگاه داده گفتار تحت استرس ٦۳
۲-۳-۱- پایگاه داده گفتار تحت استرس واقعی و شبیه‌سازی شده (SUSAS) ٦۵
۲-۳-۲- پایگاه داده (ORI)..... ٦۸
۲-۳-۳- مشکلات پایگاه‌های داده موجود ٦۸

فصل سوم : پایگاه داده پیشنهادی و سیستم طراحی‌شده برای تشخیص استرس از روی گفتار فارسی

۳-۱- مقدمه ۷۰
۳-۲- پایگاه داده پیشنهادی ۷۰
۳-۲-۱- نحوه کار با نرم‌افزار Ulead Video Studio11 ..... ۷۱
۳-۲-۲- لیست فیلم‌های موجود..... ۷۳
۳-۳- ویژگی‌های استخراج شده از ضرایب کپسترال دینامیکی ۷۵
۳-۳-۱- ویژگی‌های زمانی ۷٦
۳-۳-۲- ویژگی‌های مبتنی بر مشتق ضرایب ۷٦
۳-۳-۳- ویژگی‌های مبتنی بر ضرایب wavelet..... ۷۷
۳-۳-۳-۱- تبدیل ویولت پیوسته ۷۷
۳-۳-۳-۲- تبدیل ویولت گسسته ۷۸

فصل چهارم : نتایج

٤-۱- مقدمه ۸۰
٤-۲- طبقه‌بندی کننده‌ها ۸۰

٤-٢-١- نحوه عملکرد الگوریتم 10 fold validation ٨١
٤ -٣- استخراج ویژگی ٨١
٤ -٤- نتایج اجرا ٨٢
٤-٤-١- نتایج اجرای برنامه توسط الگوریتم svm ٨٢
٤-٤-٢- نتایج اجرای برنامه توسط الگوریتم AdaBoost ٨٦

فصل پنجم : نتیجه‌گیری و پیشنهادات

٥-١- مقدمه ١٠١
٥-٢- جمع‌بندی ١٠١
٥-٣- پیشنهادات ١٠٢

منابع و مآخذ

فهرست منابع فارسی ١٠٣
فهرست منابع انگلیسی ١٠٦
چکیده انگلیسی ١١١

فهرست جداول

عنوان **صفحه**

جدول۱-۱- انواع امواج مغزی و کارکرد آنها ۸

جدول۱-۲- سنجش / نشانه‌های ذهنی در سخنرانی تحت استرس. اتصال بین"مشاهده / ویژگی‌ها" و این‌که آیا آن اندازه‌گیری حاکی از آن است که مولفه‌های استرس می‌تواند به‌راحتی از سوی گوینده مخابره و توسط شنونده درک‌شود..... ۱۱

جدول۲-۱- توزیع هارمونیک فرکانس گام در باند اساسی ۵۳

جدول۲-۲- معرفی چند پایگاه‌داده گفتار تحت استرس ٦٤

جدول۲-۳- پنج حوزه پایگاه‌داده SUSAS ٦٦

جدول۲-٤- خلاصه ای از همه پایگاه‌داده‌ی گفتار استرسی SUSAS..... ٦٧

جدول۳-۱- فیلم‌های ایرانی ۷٤

جدول۳-۲- فیلم‌های خارجی ۷۵

جدول۳-۳- اطلاعات مربوط به تعداد جملات پایگاه‌داده به تفکیک جنس و احساس ۷۵

جدول٤-۱- الگوریتم Svm با تابع هسته Quadratic ۸۳

جداول٤-۲- الگوریتم Svm با تابع هستهMlp ۸٤

جداول٤-۳- الگوریتم Svm با تابع هستهRbf ۸۵

جداول٤-٤- اجرای برنامه با استفاده از الگوریتم AdaBoostM1..... ۸٦

جداول٤-۵- اجرای برنامه با استفاده از الگوریتم LogitBoost ۸۸

جداول٤-٦- اجرای برنامه با استفاده از الگوریتم GentleBoost ۸۹

جداول٤-۷- اجرای برنامه با استفاده از الگوریتم RobustBoost ۹۰

جداول٤-۸- اجرای برنامه با استفاده از الگوریتم LpBoost ۹۱

جداول٤-۹- اجرای برنامه با استفاده از الگوریتم TotalBoost ۹۲

جداول٤-۱۰- اجرای برنامه با استفاده از الگوریتم RUSBoost ۹۳

جداول٤-۱۱- اجرای برنامه با استفاده از الگوریتم Subspace (طبقه‌بندی‌کننده Discriminant)..... ۹٤

جداول٤-۱۲- اجرای برنامه با استفاده از الگوریتم Subspace (طبقه‌بندی‌کننده KNN)..... ۹۵

فهرست نمودارها

عنوان **صفحه**

نمودار٤-١-میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم Svm با تابع هسته Quadratic) ۸۳

نمودار ٤-٢- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم Svm با تابع هستهMlp)..... ۸٤

نمودار ٤-٣- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم Svm با تابع هستهRbf)..... ۸۵

نمودار ٤-٤- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتمAdaBoostM1)..... ۸۷

نمودار ٤-۵- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم LogitBoost)..... ۸۸

نمودار ٤-٦- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم GentleBoost)..... ۸۹

نمودار ٤-۷- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم RobustBoost)..... ۹۰

نمودار ٤-۸- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم LpBoost.)..... ۹۱

نمودار ٤-۹- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم TotalBoost)..... ۹۲

نمودار ٤-۱۰- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم RUSBoost)..... ۹۳

نمودار٤-۱۱- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم Subspace (طبقه‌بندی‌کننده Discriminant))..... ۹٤

نمودار٤-١٢- میانگین نرخ تشخیص استرس برای همه ویژگی‌ها برای پایگاه داده پیشنهادی (با استفاده از الگوریتم Subspace (طبقه‌بندی‌کننده KNN))..... ۹۵

فهرست شکل‌ها

عنوان **صفحه**

شکل۱-۱- مدل چگونه استرس باعث تغییرات در گفتار می‌شوند. ۷
شکل۱-۲- اثر استرس و اجزای دیگر در گفتار و سیستم سخنرانی ۱۰
شکل۱-۳- اثر عوامل استرس‌زا در فرایند تولید گفتار ۱۲
شکل۲-۱- مدل کلی سیستم تشخیص استرس از روی گفتار ۱۶
شکل۲-۲- ویژگی‌های گفتار ۱۷
شکل۲-۳- توزیع نمونه‌های فرکانس گام برای بیان خنثی، لومبارد و عصبانی ۱۹
شکل۲-٤- نمونه‌هایی از اسپکتروگرام صدای بیان شده برای کلمه "break " که تحت سطوح مختلف استرس تلفظ شده ۲۰
شکل۲-۵-نمونه‌های تنوع شدت برای سبک‌های گفتار تحت استرس درشرایط خنثی، لومبارد، عصبانی.... ۲۱
شکل۲-٦- تفاوت مدت‌زمان نمونه گفتار در شرایط خنثی، لومبارد و استرس عصبانی ۲۲
شکل۲-۷- پهنای باند و مکان فورمنت‌ها..... ۲۵
شکل۲-۸- مکان فرکانس فورمنت‌ها و پهنای باند برای سبک‌های خنثی، عصبانی و لومبارد..... ۲۵
شکل۲-۹- تغییر متوسط فرکانس‌های مرکز F1 و F2 برای دو سطح نویز (گوینده مرد)..... ۲٦
شکل۲-۱۰- نمونه مربوط به طیف دستگاه صوتی برای گفتار تحت شرایط استرس عصبانی ، لومبارد و خنثی ۲۷
شکل۲-۱۱- نمونه پروفایل دستگاه آوازی برای کلمه "Freeze " هنگام شمرده صحبت‌کردن است ۲۸
شکل۲-۱۲- مدل غیرخطی انتشار صوت در طول دستگاه‌صوتی..... ۲۸
شکل۲-۱۳- نمونه طیف منبع مربوط به دهانه نای برای گفتار تحت شرایط استرس عصبانی ، لومبارد و خنثی ۳۰
شکل۲-۱٤- شیب‌های ویژه (شیب مربوط به دهانه نای)..... ۳۱
شکل۲-۱۵- بلوک دیاگرام ضرایب مل کپستروم..... ۳۲
شکل۲-۱٦- شناسایی دوره تناوب فرکانس گام از تابع همبستگی خودکار ۳۷
شکل۲-۱۷- بلوک دیاگرام PLP ۳۹

شکل۲-۱۸- شکل موج صدای گرفته‌شده از کلمه " Help " است که توسط گوینده مرد تحت شرایط عصبانی (الف) حالت طبیعی و (ب) حالت شبیه‌سازی‌شده، در مدت زمان ۱۵۰ میلی‌ثانیه به دست آمده است ٤٤
شکل۲-۱۹- شکل موج نمونه(الف)یک تک فرکانس ۱کیلوهرتز و (ب) پاسخ مدوله‌شدهAM/FM ٤٥
شکل۲-۲۰- استخراج ویژگی‌های TEO-FM-Var .شکل موج‌های بخشی از صدای /IH/؛ ستون سمت راست مربوط به شرایط استرسی و سمت چپ مربوط به شرایط خنثی است..... ٤٧
شکل۲-۲۱- استخراج ویژگی‌ها (همه شکل موج‌ها برای شکل ۲و۳ هستند و شکل ٤ برای باند دوم۱-۲ کیکوهرتز می‌باشد؛ شکل موج‌ها نشان‌دهنده‌ی یک بخش از صدای/ IH/ هستند. ستون‌های سمت چپ برای حالت خنثی و ستون‌های سمت راست برای شرایط استرسی می‌باشند)..... ٤٨
شکل۲-۲۲- استخراج ویژگی‌های TEO-CB-Auto-Env ٥٢
شکل۲-۲۳- الگوریتم ماشین بردار پشتیبان..... ٥٧
شکل۲-۲٤- ابر صفحه ای که دو کلاس را از هم جدا می‌سازد..... ٥٨
شکل۲-۲٥- ماشین بردار پشتیبان با هسته‌های : آ (خطی ب (چند جمله‌های درجه ٤ پ (شعاعی ت) سیگموید برای جداسازی داده‌های با توزیع گوسی..... ٦٠
شکل۲-۲٦- تابع چگالی توزیع نمونه‌های یک بعدی و دو بعدی ٦٢
شکل۳-۱- نرم‌افزار Ulead VideoStudio 11..... ٧١
شکل۳-۲- فضای کار در نرم‌افزار Ulead VideoStudio 11..... ٧٢
شکل۳-۳- جداسازی قسمت استرسی فیلم از بقیه فیلم..... ٧٢
شکل۳-٤- جداکردن صدا از تصویر..... ٧٣
شکل٤-۱- شکل موج‌های مربوط به حالات استرسی و بدون استرسی متعلق به یکی از شخصیت‌های پایگاه داده پیشنهادی ٩٨
شکل٤-۲- مقایسه درصد درستی ۱۱٤۱ ویژگی مرجع اسماعیلیان و ۹ ویژگی استفاده شده در پایان‌نامه بر روی پایگاه داده پیشنهادی..... ٩٩

چکیده

با توجه به گسترش شبکه‌های ارتباطی و ارتقاء روش‌های جستجو، امروزه کاربران بسیاری خواستار جستجوی اطلاعات بر اساس صوت، تصویر و یا ویدئو هستند. براین اساس و با توجه به حجم بالای اطلاعات موجود، روش‌های تحلیل و دسته‌بندی اتوماتیک صوتی از اهمیت خاصی برخوردار است. وابستگی آن به متغیرهای کیفی انسان بر پیچیدگی کار می‌افزاید. در این پایان‌نامه سعی شده‌است با آنالیز صوت، روشی برای تشخیص استرس از روی پارامترهای مبتنی بر ضرایب کپسترال دینامیکی بیان‌شود. تغییر مشخصات گویندگان، تغییرات فرهنگی، زبان و تعریف استرس به صورت دقیق و... از جمله مواردی هستند که بر پیچیدگی کار می‌افزایند. با توجه به عدم دسترسی به پایگاه داده فارسی برای تشخیص استرس ابتدا یک پایگاه داده مناسب آماده‌شده که سعی‌کردیم پارامترهای موجود در پایگاه داده‌های استاندارد در آن لحاظ شود. سپس براساس ویژگی‌های صرفا مبتنی بر ضرایب کپسترال، استرس‌دار یا بدون استرس بودن گفتار مشخص شده‌است. بدیهی است بدلیل اینکه روش فقط بر پایه ضرایب کپسترال دینامیک می‌باشد، هنگام ارزیابی درصد صحت آن بخوبی روش‌های ترکیبی یا پارامتری نخواهد بود. با این حال نتایج بدست‌آمده نشان‌داد استرس موجود در گفتار، مستقل از فرهنگ و قومیت بر ضرایب کپسترال تأثیرمعنی‌داری می‌گذارد ولذا در اکثر موارد از روی ویژگی‌های استخراجی از ضرایب کپسترال می‌توان استرس را تا حد قابل قبولی تشخیص داد و ما با استفاده از این ویژگی ها به درصد تشخیص خوبی برای تشخیص استرس رسیدیم.

کلمات کلیدی: تشخیص استرس، پایگاه داده گفتار فارسی، ویژگی‌های گفتار، ضرایب کپسترال دینامیکی.

فصل اول

استرس و تاثیر آن بر گفتار

۱-۱- مقدمه

گفتار نقش مهمی در ارتباطات انسانی دارد لذا گفتار، یک سیگنال پیچیده است بطوریکه اطلاعاتی نظیر حالات، محیط‌های صوتی، قصد، زمینه زبان ، جنبه‌های لهجه و گویش ، و بیشتر دانش زبانی فرد گوینده را ارسال می‌کند (Hansen, Patil 2007). در نتیجه از سیگنال گفتار برای انتقال اطلاعات زبانی بین افراد، و همچنین انتقال حالات هیجانی سخنران استفاده می‌شود. در ارتباطات بشر، شنوندگان فرایند تغییرات نامحسوس در گفتار را به مراتب بهتر از رابط خودکار ماشین می‌توانند اداره کنند.

پردازش صوت با هدف اصلی ارتباط بین انسان و ماشین در ۵ دهه اخیر پیشرفت زیادی داشته اما مشکل عمده باقی‌مانده این است که با وجود پیشرفت‌های زیاد، هنوز هم از داشتن یک تعامل طبیعی بین انسان و ماشین به دور هستیم چون ماشین تحت حالت عاطفی گوینده قرار نمی‌گیرد. ماشین نمی‌تواند احساسات انسان را درک و یا پاسخ‌های عاطفی تولید کند.

واژه‌ها به درستی خلق‌وخوی و قصد گوینده را درک نمی‌کنند و به این ترتیب معرفی مهارت‌های اجتماعی انسان، به ارتباط انسان و ماشین، از اهمیت خاصی برخوردار است. این‌را می‌توان با تحقیق و ایجاد روش‌های مدل‌سازی گفتار و تجزیه‌وتحلیل که شامل سیگنال، جنبه‌های زبانی و احساسی از ارتباط است ، به دست‌آورد (He, Lech and Maddage 2009).

در دامنه تعامل انسان و کامپیوتر، تشخیص احساسات گفتار مورد توجه بسیاری از محققان قرارگرفته‌است (Zhang et al. 2013). عملکرد شناخت احساسات تا حد زیادی به استخراج موفق از ویژگی‌های مستقل از گوینده‌ی مربوطه بستگی دارد. علیرغم توانایی ماشین در تبدیل گفتار به مجموعه‌ای از لغات، در بسیاری از موارد عدم تشخیص هیجان یا احساس موجود در گفتار باعث تشخیص اشتباه می‌شود (El Ayadi, Kamel and Karray 2011). در علم رباتیک یا در کلیه مواردی که نیاز به ارتباط طبیعی بین انسان و ماشین است این عدم تشخیص احساس موجود در کلام، می‌تواند منجر به اشتباه شود (Schular, Rigoll and lang 2004). در سیستم‌های داخل ماشین احساس یا هیجان موجود در گفتار می‌تواند به عنوان یک ویژگی راننده مورد استفاده قرارگیرد تا رانندگی بی‌خطر شود (France et al. 2000, 7). در گفتار درمانی یا پزشکی این

موضوع بسیار مهم است. در سیستم‌های ترجمه در بسیاری موارد معنای کلام، تابع احساس موجود در جمله است. در کابین خلبان نیز تشخیص استرس بسیار مهم است (Hansen, Cairns and Icarus 1995, 4) و سیستم‌های مبتنی بر استرس از سیستم‌های معمولی بسیار بهتر عمل می‌کنند. در سیستم‌های موبایل تشخیص استرس بسیار مهم است و هیجان به صورتی که سیستم پاسخ خود را بر اساس رنجش، ناامیدی یا خوشحالی تنظیم کند طبعاً نتیجه بسیار بهتری خواهدداد.

بخش عمده‌ای از تلاش‌های فعلی در مطالعه سخنرانی تحت استرس، به هدف تشخیص شرایط استرس برای بهبود استحکام تشخیص گفتار است (Fernandez, Picard 2003). تشخیص خودکار و طبقه‌بندی گفتار در شرایط استرس‌زا دارای برنامه‌های کاربردی در علوم سلامت رفتاری و روانی، ارتباطات انسان و ماشین، رباتیک، و پزشکی است. تکنیک‌های تطابق مدل می‌تواند بیشتر برای انطباق مدل مورد استفاده قرار گیرد به‌طوری که گفتار استرس بتواند به‌خوبی تشخیص داده‌شود. طبقه‌بندی استرس و احساسات دارای کارهای محاسباتی است که شامل دو قسمت عمده می‌باشد: استخراج و طبقه‌بندی ویژگی‌ها.

در قرن ۲۱ به طور گسترده‌ای به شناخت استرس پرداخته شده‌است(Sigmund 2012). این کار برای بعضی بخش‌ها از اهمیت ویژه‌ای برخورداراست، از جمله سیستم‌های امنیتی، تشخیص دروغ، بازی‌های ویدئویی، ربات‌های جستجو و نجات و.. (Li et al 2007).

الگوریتم‌هایی که قادر به طبقه‌بندی استرس است می‌تواند برای طبقه‌بندی سخنرانی استرسی از خنثی‌شده آن مورد استفاده قرار گیرد. درواقع، طبقه‌بندی استرس نه‌تنها می‌تواند به‌منظور بهبود استحکام سیستم‌های تشخیص گفتار مورد استفاده قرار گیرد، بلکه حالاتی نظیر مخابرات، کاربردهای نظامی، کاربردهای پزشکی، و اجرای قانون می‌توانند ازآن بهره‌مند شوند (Hansen, Patil 2007). در ارتباطات راه دور، علاوه بر پتانسیل خود که برای بهبود تلفن مبتنی بر عملکرد تشخیص گفتار است، می‌توان از طبقه‌بندی استرس برای تماس‌های اضطراری اولویت بالا به مسیر ۹۱۱ (خدمات تماس اضطراری) استفاده نمود. علاوه بر این نیز می‌تواند برای ارزیابی وضعیت عاطفی یک تماس گیرنده برای خدمات پاسخ تلفن مورد استفاده قرار گیرد. همچنین طبقه‌بندی استرس می‌تواند در تجزیه‌وتحلیل گفتار قانونی با اجرای قانون به منظور بررسی وضعیت تماس گیرنده تلفن به‌کار گرفته شود و یا به عنوان یک کمک در مصاحبه مشکوک استفاده شود.

ادغام تکنولوژی تشخیص گفتار در حال حاضر در بسیاری از ارتباطات صوتی نظامی و برنامه‌های کاربردی کنترل دیده می‌شود. از آنجایی که بسیاری از برنامه‌های کاربردی شامل محیط‌های استرس‌زا هستند (به عنوان مثال، کابین هواپیما، نظامیان حافظ صلح / تنظیم میدان جنگ)، طبقه‌بندی استرس و ارزیابی برای بهبود استحکام در این برنامه‌ها حیاتی می‌شوند.

اگر چه تحقیقات زیادی در شرایط استرس‌زا برای تشخیص گفتار انجام شده است، ولیکن فقط در حوزه طبقه‌بندی گفتار استرس محدود شده است.

اکثر مطالعات انجام شده در زمینه تجزیه‌وتحلیل استرس گوینده، با توجه به ویژگی‌های طیفی به‌دست‌آمده از یک مدل خطی تولید گفتار روی فرکانس گام[1] متمرکز شده‌اند. ویژگی‌های آکوستیک دیگر که به عنوان شاخص گفتار تحت استرس مفید می‌باشد، شامل فرکانس اساسی (F0)، مدت‌واج و شدت، ساختار منبع دهانه نای (به خصوص شیب طیفی)، و ساختار فورمنت دستگاه‌آوازی (vocal tract) است.

۱-۲- تعریف استرس

تعریف «استرس» دشوار است. یک تعریف واحد نمی‌تواند به‌طور کلی، تمام شرایط را دربرگیرد.

اکثر تعاریفی که برای مصارف عملی در نظر گرفته می‌شود ممکن است تا حدودی مبهم باشد. با این وجود، تعریف ما به جنبه‌هایی از علم زبان‌شناسی تاکید می‌کند.

از این‌رو، ما با استفاده از عبارت" گفتار تحت استرس" این مفهوم را می‌رسانیم که صحبت‌کردن تحت بعضی فشارها منجر به تغییر روند تولید گفتار می‌شود.

سخنرانی‌ای که عاری از هر گونه فشار و استرسی باشد به عنوان" گفتار در شرایط خنثی" نامیده می‌شود (Hansen, Patil 2007).

استرس به عنوان اختلال است (Zhang et al. 2009) که در حالت کلی، یک حالت روانی فیزیولوژیکی است که با فشار ذهنی، فعالیت فیزیولوژیکی ناکارآمد و کاهش عملکرد همراه است). از این‌رو، به طور معمول دربرگیرنده‌ی احساسات خاص (بعنوان مثال ترس، خشم، اضطراب ...) است. این تغییرات می‌تواند در رفتار کلام و حتی در برابر اراده فرد تاثیر گذارد. بنابراین هر گونه انحراف در گفتار، با توجه به سبک خنثی آن، به عنوان سخنرانی تحت استرس نامیده می شود.

به خوبی شناخته شده‌است استرس اشاره به وضعیت روحی سخنران دارد که ممکن است ناشی از توجه به عوامل عاطفی و خارجی اطراف آن (Womack, Hansen 1996) باشد. هنگامی که سخنران تحت شرایط

[1] pitch