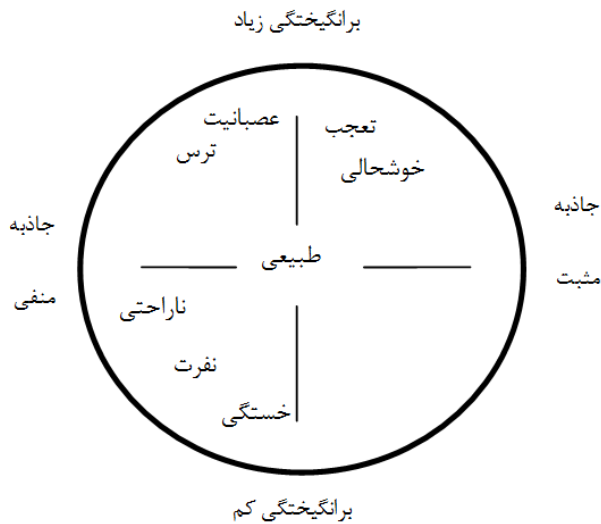


تشخیص احساس از روی گفتار با استفاده از طبقه‌بند مبتنی بر مدل و ویژگی‌های دینامیکی غیر خطی

علی حریمی، علی‌رضا احمدی فرد، علی شهزادی و خشایار یغمایی



شکل ۱: مدل دوبعدی برانگیختگی - جاذبه احساس.

[۵]. اغلب ویژگی‌هایی که برای بازشناسی احساس از روی گفتار به کار گرفته می‌شوند می‌توانند در دو گروه ویژگی‌های عروزی و ویژگی‌های طیفی دسته‌بندی شوند. ویژگی‌های عروزی بیشتر با لحن و ریتم گفتار در ارتباط هستند. این ویژگی‌ها معمولاً از مشخصات آماری منحنی‌های فرکانس گام و انرژی محاسبه می‌شوند [۶] و با دربرداشتن اطلاعات مهم احساسی، پرکاربردترین ویژگی‌های این حوزه به شمار می‌آیند [۵] و [۶]. ویژگی‌های طیفی که از طیف سیگنال به دست می‌آیند در سال‌های اخیر جایگاه ویژه‌ای پیدا کرده‌اند. این ویژگی‌ها به عنوان مکمل ویژگی‌های عروزی نقش به‌سزایی در افزایش راندمان تشخیص احساس از روی گفتار داشته‌اند [۶].

اغلب ویژگی‌های یادشده با برانگیختگی احساس ارتباط دارند [۷]. لذا در جداسازی احساس‌های با سطح برانگیختگی یکسان مانند عصبانیت و خوشحالی ناکارآمد می‌باشند [۶] و [۷]. شکل ۱ موقعیت احساس‌های مختلف را در مدل دوبعدی برانگیختگی - جاذبه نشان می‌دهد [۸]. بر اساس پژوهش‌های گذشته تداخل احساس‌های عصبانیت و خوشحالی و همچنین خستگی و عادی بخش مهمی از خطای سیستم‌های بازشناسی احساس از روی گفتار را به خود اختصاص داده است [۵] تا [۷]. بنابراین به نظر می‌رسد استخراج ویژگی‌هایی که بتواند این دو احساس را از یکدیگر جدا کند مسئله مهمی باشد.

در فرایند تولید گفتار شارش غیر خطی و دینامیکی هوا در فضای سه‌بعدی وجود دارد که این مسئله در مدل خطی منبع-فیلتر متداولی که برای استخراج ویژگی‌های عروزی و طیفی استفاده می‌شود در نظر گرفته نمی‌شود [۹] تا [۱۱]. به منظور در نظر گرفتن این اثر غیر خطی، دینامیک غیر خطی ابزارهای مختلفی از جمله فضای فاز بازسازی‌شده را پیشنهاد کرده است. در دهه اخیر، فضای فاز بازسازی‌شده برای حوزه‌های مختلف

چکیده: با توجه به پیشرفت‌های صورت‌گرفته در زمینه رباتیک و تعامل انسان و ماشین، تشخیص احساس از روی گفتار اهمیت ویژه‌ای پیدا کرده است. در این مقاله یک طبقه‌بند مبتنی بر مدل احساسی برانگیختگی - جاذبه، برای بازشناسی احساس از روی گفتار استفاده شده است. در این روش، در مرحله اول نمونه‌ها با استفاده از ویژگی‌های متداول عروزی و طیفی بر مبنای سطح برانگیختگی طبقه‌بندی می‌شوند. سپس احساس‌های با سطح برانگیختگی یکسان با استفاده از ویژگی‌های پیشنهادی دینامیکی غیر خطی از یکدیگر جدا می‌شوند. ویژگی‌های دینامیکی غیر خطی از روی مشخصات هندسی فضای فاز بازسازی شده سیگنال گفتار استخراج می‌شوند. بدین منظور چهار منحنی توصیفگر برای بازنمایی مشخصات هندسی فضای فاز بازسازی شده محاسبه می‌شوند. سپس مؤلفه‌های مهم تبدیل کسینوسی گسسته این منحنی‌ها به عنوان ویژگی‌های دینامیکی غیر خطی مورد استفاده قرار می‌گیرند. روش پیشنهادی بر روی پایگاه داده برلین با استفاده از تکنیک ۱۰ تکه برابر ارزیابی شده و نرخ بازشناسی ۹۶٫۳۵٪ و ۸۷٫۱۸٪ برای زنان و مردان به دست آمد. با توجه به تعداد نمونه‌ها در هر گروه جنسیتی، متوسط نرخ بازشناسی ۹۲٫۳۴٪ برای سیستم پیشنهادی به دست آمد.

کلیدواژه: بازشناسی احساس از روی گفتار، احساس‌های با جاذبه یکسان، طبقه‌بند متوالی، ویژگی‌های دینامیکی غیر خطی، فضای فاز بازسازی شده.

۱- مقدمه

پیشرفت‌های اخیر در زمینه رباتیک محققان را بر آن داشته است تا در زمینه بازشناسی احساس از روی سیگنال‌های بیولوژیک و به ویژه گفتار تحقیقات مهمی انجام دهند. بازشناسی احساس از روی گفتار کاربردهای مختلفی در سیستم‌های تعاملی انسان و ماشین دارد، به طور مثال می‌تواند عملکرد سیستم‌های تشخیص گفتار را بهبود بخشد [۱]. همچنین در زمینه آموزش، بازی‌های کامپیوتری، پزشکی، روان‌شناسی و خودروهای هوشمند کاربردهای متنوعی دارد [۲] تا [۴].

تشخیص احساس از روی گفتار را می‌توان به عنوان یک مسئله تشخیص الگو در سه زیربخش استخراج ویژگی، انتخاب ویژگی و طبقه‌بندی بررسی نمود. با وجود تلاش‌های گسترده، استخراج ویژگی‌های مؤثر هنوز به عنوان یکی از چالش‌های مهم این حوزه به شمار می‌رود

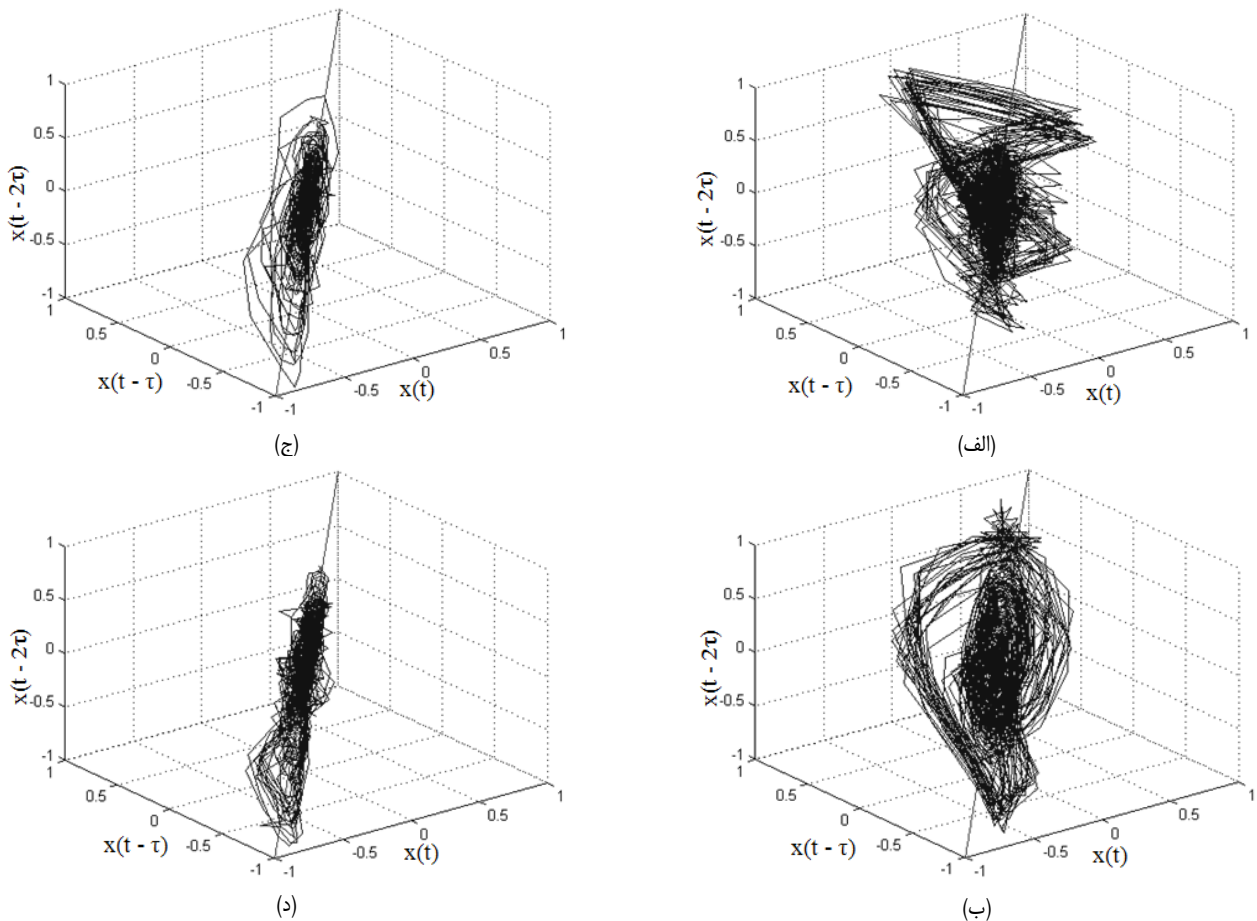
این مقاله در تاریخ ۱۷ اسفند ماه ۱۳۹۱ دریافت و در تاریخ ۱ شهریور ماه ۱۳۹۳ بازنگری شد.

علی حریمی، دانشکده برق و کامپیوتر، دانشگاه آزاد اسلامی، واحد شاهرود، شاهرود، (email: a.harimi@gmail.com).

علی‌رضا احمدی فرد، دانشکده مهندسی برق و رباتیک، دانشگاه صنعتی شاهرود، شاهرود، (email: ahmadyard@shahroodut.ac.ir).

علی شهزادی، دانشکده برق و کامپیوتر، دانشگاه سمنان، سمنان، (email: shahzadi@iust.ac.ir).

خشایار یغمایی، دانشکده برق و کامپیوتر، دانشگاه سمنان، سمنان.



شکل ۲: فضای فاز بازسازی شده برای بخش کوچکی از یک جمله با احساس‌های (الف) عصبانیت، (ب) خوشحالی، (ج) خستگی و (د) عادی.

در این مدل سیگنال تولیدشده توسط یک سیستم به عنوان یک سری زمانی در نظر گرفته می‌شود. بر اساس نظریه تاکن [۹]، برای سری زمانی $x_n, n = 1, 2, \dots, N$ بردار زیر تولید می‌شود

$$\bar{x}_n = [x_n, x_{n+\tau}, x_{n+2\tau}, \dots, x_{n+(d-1)\tau}] \quad (۱)$$

\bar{x}_n یک بردار سطری است که موقعیت یک نقطه در فضای فاز بازسازی شده را نشان می‌دهد و d ابعاد فضای فاز و τ تأخیر می‌باشد. این بردارها به صورت ستونی به هم پیوند داده می‌شوند تا یک ماتریس توصیفگر دینامیک سیستم به صورت زیر تشکیل شود

$$X = \begin{bmatrix} x_1 & x_{1+\tau} & \dots & x_{1+(d-1)\tau} \\ x_2 & x_{2+\tau} & \dots & x_{2+(d-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n+\tau} & \dots & x_{n+(d-1)\tau} \end{bmatrix} \quad (۲)$$

در این تحقیق به پیشنهاد [۱۳] و [۱۵]، τ و d برابر یک و سه تنظیم شده‌اند. نتایج آزمایش‌های ما نیز مناسب بودن این مقادیر را تأیید می‌کند. منحنی فضای فاز بازسازی شده برای یک بخش کوچک با صدای "ein" از نمونه‌های پایگاه داده آلمانی که توسط یک زن در چهار احساس عصبانیت، خوشحالی، خستگی و عادی بیان شده است در شکل ۲-الف تا ۲-د نشان داده شده است.

همان طور که در شکل ۲ مشاهده می‌شود، شکل هندسی فضای فاز بازسازی شده با احساس گفتار مرتبط است. این مسئله الهام‌بخش استخراج ویژگی‌های دینامیکی غیر خطی از فضای فاز بازسازی شده برای جداسازی احساس‌های عصبانیت و خوشحالی و همچنین خستگی و عادی می‌باشد.

پردازش گفتار از جمله تشخیص گفتار [۱۲] و [۱۳]، ارتقای گفتار [۱۴] و تشخیص خواب آلودگی از روی گفتار [۱۵] مورد استفاده قرار گرفته است اما بر اساس بررسی‌های ما تاکنون تحقیقات مهمی در کاربرد این ابزار در بازشناسی احساس از روی گفتار صورت نگرفته است.

مهم‌ترین کار صورت گرفته در این مقاله استفاده از ویژگی‌های دینامیکی غیر خطی مستخرج از مشخصات هندسی فضای فاز بازسازی شده سیگنال گفتار به منظور جداسازی احساس‌های عصبانیت و خوشحالی و همچنین خستگی از عادی می‌باشد. این طبقه‌بند به صورت متوالی با یک طبقه‌بند کلی که احساس‌ها را بر اساس سطح برانگیختگی به پنج گروه عصبانیت-خوشحالی، خستگی-عادی، انزجار، ترس و ناراحتی با استفاده از ویژگی‌های متداول عروزی و طیفی جدا می‌کند بسته شده است. این تکنیک باعث کاهش چشم‌گیر تداخل بین احساس‌های عصبانیت و خوشحالی و همچنین خستگی و عادی و در نتیجه افزایش دقت کلی طبقه‌بند می‌شود.

در ادامه این مقاله، در بخش ۲ فضای فاز بازسازی شده بیان می‌شود. در بخش ۳، ویژگی‌های پیشنهادی دینامیکی غیر خطی را در کنار ویژگی‌های متداول عروزی و طیفی معرفی می‌کنیم. بخش ۴ پایگاه داده مورد استفاده را معرفی می‌نماید. نتایج در بخش ۵ بیان شده و مورد بحث قرار می‌گیرند و نهایتاً در بخش ۶ به جمع‌بندی و نتیجه‌گیری خواهیم پرداخت.

۲- فضای فاز بازسازی شده

فضای فاز بازسازی شده که در دینامیک غیر خطی مطرح شده است، یک مدل پیشگویی قدرتمند برای سیستم‌های دینامیک آشوبگر می‌باشد.

۳-۲ ویژگی‌های عروزی

ویژگی‌های عروزی از پرکاربردترین ویژگی‌ها در حوزه بازشناسی احساس از روی گفتار به شمار می‌روند [۵] و [۶]. به منظور استخراج این ویژگی‌ها، مشخصات آماری منحنی‌های فرکانس گام و انرژی مورد استفاده قرار می‌گیرند. در اینجا ۲۰ تابع آماری شامل مقادیر مینیمم، ماکسیمم، رنج، میانگین، میانه، میانگین ترمیم‌شده ۱۰٪، میانگین ترمیم‌شده ۲۵٪، صدک‌های یکم، پنجم، دهم، بیست و پنجم، هفتاد و پنجم، نودم، نود و پنجم و نود و نهم، دامنه میانبرها، انحراف معیار، واریانس، ناهمواری و درجه اوج نمودار آماری مورد استفاده قرار می‌گیرند [۱۵]. همچنین معمول است که این توابع به مشتقات اول و دوم منحنی‌های مورد نظر نیز اعمال شوند [۶]. بنابراین در مجموع ۶۰ ویژگی مبتنی بر فرکانس گام و ۶۰ ویژگی مبتنی بر انرژی استخراج می‌شوند.

همچنین نرخ عبور از صفر و عملگر انرژی تیگر [۱۸] نیز در این تحقیق مورد استفاده قرار می‌گیرند. این ویژگی‌ها مستقیماً جزء ویژگی‌های عروزی به حساب نمی‌آیند اما در این تحقیق این ویژگی‌ها را در کنار ویژگی‌های عروزی ارزیابی می‌کنیم. تیگر حاوی اطلاعاتی از شارش غیر خطی هوا در سیستم تولید گفتار است [۱۹] و برای سیگنال x_n به صورت زیر محاسبه می‌شود

$$TEO(x_n) = x_n^2 - x_{n-1}x_{n+1} \quad (۷)$$

به منظور استخراج ویژگی‌های مبتنی بر نرخ عبور از صفر و تیگر، ۲۰ تابع آماری به منحنی نرخ عبور از صفر و منحنی تیگر و همچنین مشتقات اول و دوم آنها اعمال می‌شود. بدین ترتیب نهایتاً، ۲۴۰ ویژگی عروزی خواهیم داشت.

۳-۳ ویژگی‌های طیفی

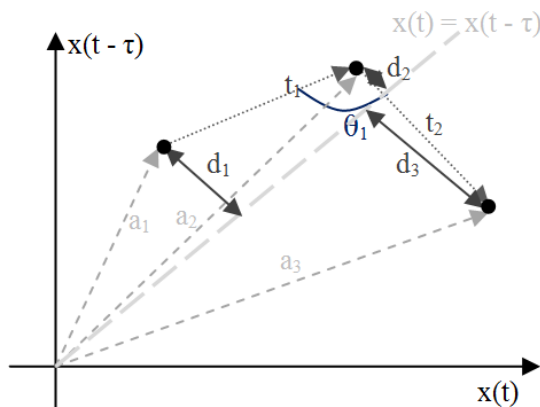
دو نوع ویژگی طیفی در این مقاله استفاده شده است: ضرایب مل کپستروم (MFCC) و فرمت‌ها که به عنوان پرکاربردترین ویژگی‌های طیفی در حوزه تشخیص احساس از روی گفتار معرفی شده‌اند [۲۰] تا [۲۳]. سیگنال گفتار ابتدا با استفاده از یک فیلتر پیش‌تأکید بالاگذر با ضریب ۰/۹۷ فیلتر می‌شود. سپس ۱۲ ضریب اول MFCC و ۴ فرمت اول از پنجره‌های همپینگ سیگنال گفتار به طول ۲۰ میلی‌ثانیه و همپوشانی ۱۰ میلی‌ثانیه استخراج می‌شوند. سپس ۲۰ تابع آماری بیان‌شده در بخش قبل به این منحنی‌ها و مشتقات اول و دوم آنها اعمال می‌شوند. در مجموع ۹۶۰ ویژگی طیفی از سیگنال گفتار استخراج می‌شود.

۴- پایگاه داده گفتار احساسی

پایگاه داده احساسی آلمانی برلین، پایگاه داده‌ای شناخته‌شده است که به صورت رایگان از طریق اینترنت در اختیار عموم قرار دارد [۲۴]. بسیاری از تحقیقات روش‌های پیشنهادی خود را با این پایگاه داده ارزیابی نموده‌اند [۶] و [۲۵] تا [۲۸]. این پایگاه داده شامل ۵۳۵ جمله با ۱۰ محتوای مختلف است که توسط ۱۰ گوینده (۵ زن و ۵ مرد) در ۷ احساس مختلف بیان شده‌اند. جدول ۱ مشخصات این پایگاه داده را از نظر توزیع جملات نشان می‌دهد.

۵- نتایج

در این تحقیق فرض شده است که ابتدا یک طبقه‌بند تفکیک جنسیتی با دقت کامل مشابه آنچه در [۲۹] عنوان شده است در اولین طبقه به کار گرفته شود. بنابراین سیستم برای زنان و مردان کاملاً به صورت جداگانه



شکل ۳: نمایش سه نقطه متوالی در فضای فاز سه‌بعدی و چگونگی تشکیل منحنی‌های توصیفگر.

۳-۳ استخراج ویژگی

در این بخش ویژگی‌های غیر خطی دینامیکی مستخرج از فضای فاز بازسازی شده با جزئیات بیان می‌شوند. ویژگی‌های عروزی و طیفی استفاده‌شده در این مقاله نیز توصیف می‌شوند.

۳-۱ ویژگی‌های دینامیکی غیر خطی

در این روش مشخصات هندسی فضای فاز بازسازی شده با چهار منحنی توصیفگر بیان می‌شوند. اولین منحنی توصیفگر فاصله نقاط تا مرکز می‌باشد [۱۵]

$$\bar{a} = [a_1, a_2, \dots, a_N] \quad (۳)$$

که در آن N تعداد نقاط در فضای فاز بازسازی شده است و a_i فاصله i امین نقطه تا مرکز می‌باشد.

دومین منحنی فاصله بین نقاط متوالی است [۱۵]

$$\bar{t} = [t_1, t_2, \dots, t_{N-1}] \quad (۴)$$

سومین منحنی زاویه بین خطوط متوالی متصل کننده نقاط است [۱۵]

$$\bar{\theta} = [\theta_1, \theta_2, \dots, \theta_{N-1}] \quad (۵)$$

و نهایتاً چهارمین منحنی فاصله نقاط تا خط همانی می‌باشد

$$\bar{d} = [d_1, d_2, \dots, d_N] \quad (۶)$$

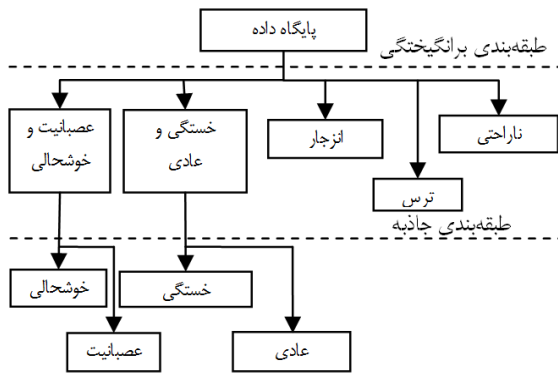
شکل ۳ چگونگی تشکیل این منحنی‌ها را برای سه نقطه متوالی در فضای فاز بازسازی شده دوبعدی نشان می‌دهد.

سپس تبدیل کسینوسی گسسته برای فشرده‌سازی اطلاعات منحنی‌ها استفاده می‌شود به این صورت که ۲۰۰ ضریب اول هر منحنی ویژگی‌های دینامیکی غیر خطی پیشنهادی را تشکیل می‌دهند. بنابراین مجموعاً ۸۰۰ ویژگی دینامیکی غیر خطی از هر نمونه صوتی استخراج می‌شود.

در این تحقیق ویژگی‌های دینامیکی غیر خطی از سیگنال گفتار اصلی، بخش‌های آوایی سیگنال گفتار، بخش‌های غیر آوایی گفتار، سیگنال تحریک حلقی، منحنی فرکانسی گام و منحنی انرژی استخراج می‌شود. بخش‌های آوایی و غیر آوایی گفتار با آستانه‌گذاری بر روی منحنی‌های انرژی و نرخ عبور از صفر محاسبه می‌شوند. سیگنال تحریک حلقی با استفاده از فیلتر معکوس [۱۶] و فرکانس گام نیز با استفاده از روش خودهمبستگی محاسبه می‌شود [۱۷]. در نتیجه مجموعاً $4800 = 6 \times 800$ ویژگی دینامیکی غیر خطی از هر نمونه به دست می‌آید.

جدول ۱: تعداد جملات در پایگاه داده برلین.

احساس	زن	مرد
عصبانیت	۶۷	۶۰
خوشحالی	۴۴	۲۷
خستگی	۴۶	۳۵
عادی	۴۰	۳۹
انزجار	۳۵	۱۱
ترس	۳۲	۳۷
ناراحتی	۳۷	۲۵
همه	۳۰۱	۲۳۴



شکل ۴: ساختار طبقه‌بند پیشنهادی.

جدول ۲: ماتریس تداخل طبقه‌بندی احساس بر مبنای سطح برانگیختگی (زنان).

احساس	عصبانیت و خوشحالی	عادی و خستگی	انزجار	ترس	ناراحتی	نرخ تشخیص
عصبانیت و خوشحالی	۱۱۰	۰	۱	۰	۰	%۹۹,۱۰
عادی و خستگی	۰	۸۵	۱	۰	۰	%۹۸,۸۴
انزجار	۰	۰	۳۴	۱	۰	%۹۷,۱۴
ترس	۳	۰	۱	۲۷	۰	%۸۴,۳۸
ناراحتی	۰	۲	۰	۰	۳۵	%۹۴,۵۹
درستی	%۹۷,۳۵	%۹۷,۷۰	%۹۱,۸۹	%۹۶,۴۳	%۹۷,۲۲	

دقت کل: %۹۶,۶۸

جدول ۳: ماتریس تداخل طبقه‌بندی احساس بر مبنای سطح برانگیختگی (مردان).

احساس	عصبانیت و خوشحالی	عادی و خستگی	انزجار	ترس	ناراحتی	نرخ تشخیص
عصبانیت و خوشحالی	۷۹	۰	۰	۸	۰	%۹۰,۸۰
عادی و خستگی	۱	۶۹	۰	۰	۴	%۹۳,۲۴
انزجار	۱	۲	۷	۱	۰	%۶۳,۶۴
ترس	۷	۱	۰	۲۹	۰	%۷۸,۳۸
ناراحتی	۰	۳	۰	۰	۲۲	%۸۸,۰۰
درستی	%۸۹,۷۷	%۹۲,۰۰	%۱۰۰	%۷۶,۳۲	%۸۴,۶۲	

دقت کل: %۸۸,۰۳

طراحی می‌شود. ویژگی‌ها قبل از اعمال به طبقه‌بند ماشین‌های بردار پشتیبان به صورت خطی به بازه $[-1, 1]$ نرمال می‌شوند [۶]. به منظور اجتناب از پیامدهای اثرات منفی ابعاد بالا در مسئله [۳۰]، یک روش انتخاب ویژگی مبتنی بر فیلتر با معیار فیشر استفاده می‌شود. این معیار با اندازه‌گیری شباهت درون‌کلاسی و اختلاف میان‌کلاسی به هر ویژگی رتبه‌ای اختصاص می‌دهد [۶]

فرایند ارزیابی بررسی شوند. نتیجه از میانگین ۱۰ آزمایش محاسبه می‌شود. شکل ۴ طبقه‌بند استفاده‌شده در این مقاله را نشان می‌دهد. همان طور که در شکل مشاهده می‌شود، ابتدا احساس‌ها بر مبنای سطح برانگیختگی با استفاده از یک طبقه‌بند از هم جدا می‌شوند. سپس احساس‌های با سطح برانگیختگی یکسان با استفاده از دو طبقه‌بند متوالی از هم جدا می‌شوند. به منظور ارزیابی روش پیشنهادی از منحنی دقت که در آن دقت طبقه‌بند به صورت تابعی از N_{top} ویژگی انتخاب‌شده با معیار فیشر نشان داده می‌شود، استفاده می‌کنیم.

$$FDR(u) = \frac{2}{C(C-1)} \sum_{c_1} \sum_{c_2} \frac{(\mu_{c_1,u} - \mu_{c_2,u})^2}{\sigma_{c_1,u}^2 + \sigma_{c_2,u}^2} \quad (۸)$$

$1 \leq c_1 < c_2 \leq C$

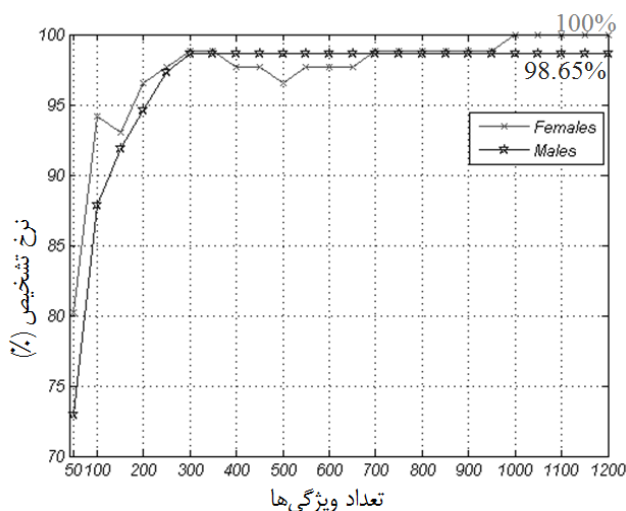
۱-۵ طبقه‌بندی بر مبنای سطح برانگیختگی

در روش پیشنهادی، نمونه‌ها ابتدا با استفاده از ویژگی‌های متداول عروسی و طیفی که ارتباط عمیقی با سطح برانگیختگی گفتار دارند [۶] و [۷] به ۵ کلاس تقسیم‌بندی می‌شوند: (۱) عصبانیت و خوشحالی، (۲) خستگی و عادی، (۳) انزجار، (۴) ترس و (۵) ناراحتی. شکل ۵ منحنی‌های دقت را برای ۵ گروه مربوط برای زنان و مردان نشان می‌دهد.

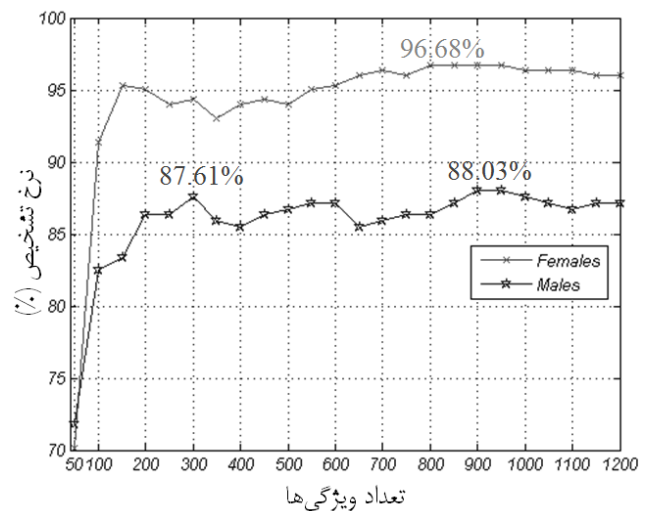
در این رابطه $\mu_{c_i,u}$ و $\sigma_{c_i,u}^2$ مقادیر میانگین و واریانس را برای u امین ویژگی در i امین کلاس نمایش می‌دهند. C نیز تعداد کلاس‌ها را نشان می‌دهد. ویژگی‌ها با معیار فیشر کمتر با انتخاب یک مقدار آستانه مناسب از لیست حذف می‌شوند.

بر اساس شکل ۵، بیشترین نرخ تشخیص %۹۶,۶۸ و %۸۸,۰۳ برای زنان و مردان با استفاده از ۸۰۰ و ۹۰۰ ویژگی قابل دستیابی است. نتایج طبقه‌بندی به صورت ماتریس‌های تداخل برای زنان و مردان در جداول ۲ و ۳ قابل مشاهده است. در این جداول ستون سمت چپ احساس واقعی و

با توجه به تعداد اندک نمونه‌ها در پایگاه داده برلین [۳۱] برای ارزیابی از تکنیک ۱۰ تکه برابر استفاده شده است. در این تکنیک پایگاه داده به ۱۰ قسمت مساوی بدون همپوشانی تقسیم‌بندی می‌شود. در هر آزمایش ۹ قسمت برای آموزش و یک قسمت برای ارزیابی در نظر گرفته می‌شود. آزمایش ۱۰ مرتبه تکرار می‌شود به طوری که همه تکه‌ها یک بار در



شکل ۷: منحنی دقت برای طبقه‌بندی احساس‌های خستگی و عادی.



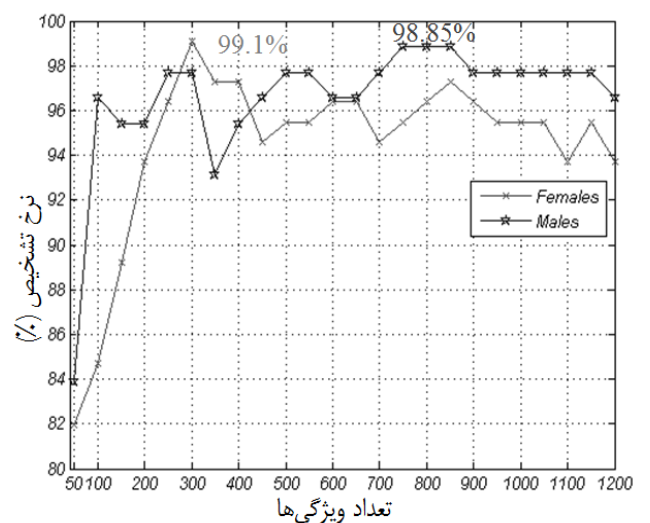
شکل ۵: منحنی دقت برای طبقه‌بندی بر اساس سطح برانگیختگی.

جدول ۴: ماتریس تداخل برای طبقه‌بندی عصبانیت در مقابل خوشحالی (زنان).

احساس	عصبانیت	خوشحالی	نرخ تشخیص (%)
عصبانیت	۶۷	۰	۱۰۰
خوشحالی	۱	۴۳	۹۷٫۷۳
درستی (%)	۹۸٫۵۳	۱۰۰	
دقت کل: ۹۹٫۱٪			

جدول ۵: ماتریس تداخل برای طبقه‌بندی عصبانیت در مقابل خوشحالی (مردان).

احساس	عصبانیت	خوشحالی	نرخ تشخیص (%)
عصبانیت	۶۰	۰	۱۰۰
خوشحالی	۱	۲۶	۹۶٫۳
درستی (%)	۹۸٫۳۶	۱۰۰	
دقت کل: ۹۸٫۸۵٪			



شکل ۶: منحنی دقت برای طبقه‌بندی عصبانیت و خوشحالی.

همان طور که در شکل ۶ مشاهده می‌شود، نرخ تشخیص متوسط ۹۹٫۱٪ و ۹۸٫۸۵٪ برای زنان و مردان با استفاده از ۳۰۰ و ۷۵۰ ویژگی دینامیکی غیر خطی قابل دستیابی است. با توجه به جداول ۴ و ۵، برای زنان و مردان، فقط یک نمونه خوشحال به اشتباه عصبانی تشخیص داده شده است. این نتایج حاکی از کارآمد بودن ویژگی‌های پیشنهادی برای جداسازی احساس‌های عصبانیت و خوشحالی که سطح برانگیختگی یکسانی دارند می‌باشد. شایان ذکر است پیش از این تداخل این دو احساس با استفاده از ویژگی‌های متداول طیفی و عروسی بخش مهمی از خطای کل سیستم را تشکیل می‌داد [۶]، [۷]، [۲۷] و [۳۲].

۳-۵ خستگی و عادی

ویژگی‌های دینامیکی غیر خطی به منظور طبقه‌بندی احساس‌های خستگی و عادی نیز استفاده شدند. شکل ۷ منحنی دقت و جداول ۶ و ۷ ماتریس‌های تداخل را نشان می‌دهند.

همان طور که در شکل ۷ نشان داده شده است، طبقه‌بندی خستگی و عادی برای زنان به طور شگفت‌انگیزی با استفاده از ۱۰۰۰ ویژگی دینامیکی غیر خطی با دقت کامل انجام شده است. نتایج جدول ۶ هم تأکیدی بر این مسئله است. برای مردان نیز این طبقه‌بندی با دقت ۹۸٫۶۵٪ با استفاده از ۳۰۰ ویژگی انجام شده است. با دقت در جدول ۷، برای مردان فقط یک نمونه خستگی اشتباهاً به عنوان عادی شناخته شده است. این نتایج حاکی از موفقیت ویژگی‌های دینامیکی غیر خطی در

سطر بالایی احساس تشخیص داده شده را نشان می‌دهد. ستون نرخ تشخیص، دقت تشخیص برای هر کلاس را نشان می‌دهد. این عدد از تقسیم تعداد نمونه‌های درست تشخیص داده شده به تعداد کل نمونه‌ها قابل محاسبه است. درستی هر کلاس از تقسیم تعداد نمونه‌های درست تشخیص داده شده در هر کلاس به تعداد کل نمونه‌هایی که به آن کلاس اختصاص داده شده‌اند به دست می‌آید.

همان طور که در جدول ۲ مشاهده می‌شود، برای زنان ویژگی‌های عروسی و طیفی به کار رفته برای طبقه‌بندی احساس‌ها بر مبنای برانگیختگی بسیار خوب عمل کرده‌اند. همچنین بر اساس جدول ۳ تداخل احساس ترس با دو احساس عصبانیت و خوشحالی منشأ بخش مهمی از خطای طبقه‌بندی در این مرحله می‌باشد. هر چند که به نظر می‌رسد ویژگی‌های طیفی و عروسی برای جداسازی احساس‌ها بر اساس سطح برانگیختگی کفایت کنند، این ویژگی‌ها برای جداسازی احساس‌های با سطح برانگیختگی یکسان مناسب نمی‌باشند [۶] و [۷].

۲-۵ عصبانیت و خوشحالی

به منظور جداسازی عصبانیت در مقابل خوشحالی، فقط از ویژگی‌های دینامیکی غیر خطی استفاده می‌کنیم. شکل ۶ منحنی دقت را برای زنان و مردان نشان می‌دهد. ماتریس‌های تداخل برای بهترین نتایج به دست آمده برای زنان و مردان نیز در جداول ۴ و ۵ نشان داده شده است.

جدول ۶: ماتریس تداخل طبقه‌بندی خستگی و عادی (زنان).

نرخ تشخیص (%)	عادی	خستگی	احساس
۱۰۰	۰	۴۶	خستگی
۱۰۰	۴۰	۰	عادی
درستی (%)	۱۰۰	۱۰۰	
دقت کل: ۱۰۰٪			

جدول ۷: ماتریس تداخل طبقه‌بندی خستگی و عادی (مردان).

نرخ تشخیص (%)	عادی	خستگی	احساس
۹۷/۱۴	۱	۳۴	خستگی
۱۰۰	۳۹	۰	عادی
درستی (%)	۹۷/۵	۱۰۰	
دقت کل: ۹۸/۶۵٪			

جدول ۸: ماتریس تداخل برای ۷ احساس (زنان).

نرخ تشخیص (%)	ناراحتی	ترس	انزجار	عادی	خستگی	خوشحالی	عصبانیت	احساس
۹۸/۵۱	۰	۰	۱	۰	۰	۰	۶۶	عصبانیت
۹۷/۷۳	۰	۰	۰	۰	۰	۴۳	۱	خوشحالی
۹۷/۸۳	۰	۰	۱	۰	۴۵	۰	۰	خستگی
۱۰۰	۰	۰	۰	۴۰	۰	۰	۰	عادی
۹۷/۱۴	۰	۱	۳۴	۰	۰	۰	۰	انزجار
۸۴/۳۸	۱	۲۷	۱	۰	۰	۰	۳	ترس
۹۴/۵۹	۳۵	۰	۰	۱	۱	۰	۰	ناراحتی
درستی (%)	۹۷/۲۲	۹۶/۴۳	۹۱/۸۹	۹۷/۵۶	۹۷/۸۳	۱۰۰	۹۴/۲۹	
دقت کل: ۹۶/۳۵٪								

جدول ۹: ماتریس تداخل برای ۷ احساس (مردان).

نرخ تشخیص (%)	ناراحتی	ترس	انزجار	عادی	خستگی	خوشحالی	عصبانیت	احساس
۸۸/۳۳	۰	۷	۰	۰	۰	۰	۵۳	عصبانیت
۹۲/۵۹	۰	۱	۰	۰	۰	۲۵	۱	خوشحالی
۸۸/۵۷	۳	۰	۰	۱	۳۱	۰	۰	خستگی
۹۴/۸۷	۱	۰	۰	۳۷	۰	۱	۰	عادی
۶۳/۶۴	۰	۱	۷	۲	۰	۰	۱	انزجار
۷۸/۳۸	۰	۲۹	۰	۱	۰	۲	۵	ترس
۸۸/۰۰	۲۲	۰	۰	۱	۲	۰	۰	ناراحتی
درستی (%)	۸۴/۶۲	۷۶/۳۲	۱۰۰	۸۸/۱۰	۹۳/۹۴	۸۹/۲۹	۸۸/۳۸	
دقت کل: ۸۷/۱۸٪								

ماشین‌های بردار پشتیبان و ارزیابی ۱۰ تکه برابر برای جداسازی ۷ احساس دقت ۸۵/۶٪ گزارش شده است. در [۳۴] نیز دقت ۸۵/۵٪ به دست آمده است.

جداسازی احساس‌های خستگی و عادی می‌باشد که بخش مهمی از خطای طبقه‌بندی را در سیستم‌های متداول به خود اختصاص می‌دهند [۶]، [۲۷] و [۳۲].

۵-۴ طبقه‌بند نهایی

نتایج عددی طبقه‌بندی ۷ احساس برای زنان و مردان در جداول ۸ و ۹ نشان داده شده است. همان‌طور که در این جداول مشاهده می‌شود، متوسط نرخ تشخیص ۹۶/۳۵٪ و ۸۷/۱۸٪ برای زنان و مردان قابل دستیابی است. نرخ تشخیص بیشتر زنان نسبت به مردان به دلیل ادراک و بروز بیشتر احساسات توسط زنان است [۳۳].

با در نظر گرفتن ۳۰۱ جمله زنان و ۲۳۴ جمله مردان، متوسط نرخ تشخیص ۹۲/۳۴٪ قابل دستیابی است.

مقایسه عددی نتایج با کارهایی که روش خود را با پایگاه داده برلین ارزیابی نموده‌اند نیز می‌تواند در این زمینه مفید باشد. هر چند که اعداد به دلیل شرایط مختلف آزمایش‌ها گویای شباهت‌ها و تفاوت‌ها به طور کامل نیستند، می‌توان از آنها برای یک مقایسه کلی استفاده نمود. در [۲۵] نرخ تشخیص ۸۶/۹٪ با ارزیابی ۱۰ تکه برابر به دست آمده است. مرجع [۲۶] نرخ تشخیص ۸۸/۸٪ را با یک طبقه‌بند سه طبقه برای جداسازی ۶ احساس گزارش نموده است. همچنین با استفاده از یک طبقه‌بند مبتنی بر

۶- نتیجه‌گیری

هدف این تحقیق ارزیابی یک طبقه‌بند دو طبقه است که در آن ابتدا نمونه‌ها با استفاده از ویژگی‌های عروزی و طیفی بر اساس سطح برانگیختگی طبقه‌بندی می‌شوند، سپس ویژگی‌های با سطح برانگیختگی یکسان با استفاده از ویژگی‌های دینامیکی غیر خطی طبقه‌بندی می‌شوند. بر اساس نتایج آزمایش‌های ما ویژگی‌های دینامیکی غیر خطی حاوی اطلاعات مهمی از سطح جاذبه گفتار می‌باشند. نتایج زیر از این مقاله قابل استحصال است:

اول این که شکل هندسی فضای فاز بازسازی شده ارتباط مهمی با احساس گفتار دارد، لذا به نظر می‌رسد منحنی فضای فاز بازسازی شده منبع خوبی برای استخراج ویژگی‌های مرتبط با احساس باشد.

دوم این که ویژگی‌های متداول عروزی و طیفی برای جداسازی احساس‌ها بر اساس سطح برانگیختگی مناسب می‌باشند، حال آن که این ویژگی‌ها برای جداسازی احساس‌های با سطح برانگیختگی یکسان از کارایی لازم برخوردار نیستند.

- [13] P. Prajith, Investigation on the Applications of Dynamical Instabilities and Deterministic Chaos for Speech Signal Processing, Ph.D Thesis, University of Calicut, 2008.
- [14] J. Sun, N. Zheng, and X. Wang, "Enhancement of chinese speech based on nonlinear dynamics," *Signal Processing*, vol. 87, no. 10, pp. 2431-2445, Oct. 2007.
- [15] J. Krajewski, S. Schnieder, D. Sommer, A. Batliner, and B. Schuller, "Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech," *Neurocomputing*, vol. 84, no. 1, pp. 65-75, May 2012.
- [16] T. Drugman, B. Bozkurt, and T. Dutoit, "Causal-anticausal decomposition of speech using complex cepstrum for glottal source estimation," *Speech Communication*, vol. 53, no. 6, pp. 855-866, Jul. 2011.
- [17] S. Gonzalez and M. Brookes, "A pitch estimation filter robust to high levels of noise (PEFAC)," in *Proc. European Signal Processing Conf., EUSIPCO'11*, pp. 451-455, Barcelona, Spain, 29 Aug.-2 Sept. 2011.
- [18] J. Kaiser, "On a simple algorithm to calculate the 'energy' of a signal," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP'90*, vol. 1, pp. 381-384, 3-6 Apr. 1990.
- [19] G. Zhou, J. Hansen, and J. Kaiser, "Nonlinear feature based classification of speech under stress," *IEEE Trans. Audio Speech Language Process.*, vol. 9, no. 3, pp. 201-216, Mar. 2001.
- [20] T. Polzehl, A. Schmitt, F. Metze, and M. Wagner, "Anger recognition in speech using acoustic and linguistic cues," *Speech Communication*, vol. 53, no. 5, pp. 1198-1209, Nov. 2011.
- [21] C. C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, "Emotion recognition using a hierarchical binary decision tree approach," *Speech Communication*, vol. 53, no. 9, pp. 1162-1171, Nov. 2011.
- [22] L. He, M. Lech, N. C. Maddage, and N. B. Allen, "Study of empirical mode decomposition and spectral analysis for stress and emotion classification in natural speech," *Biomedical Signal Processing and Control*, vol. 6, no. 1, pp. 139-146, May 2011.
- [23] P. Laukka, D. Neiberg, M. Forsell, I. Karlsson, and K. Elenius, "Expression of affect in spontaneous speech: acoustic correlates and automatic detection of irritation and resignation," *Computer Speech and Language*, vol. 25, no. 4, pp. 84-104, Sep. 2011.
- [24] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A database of German emotional speech," in *Proc. Interspeech*, vol. 5, pp. 1517-1520, Sept. 2005.
- [25] B. Schuller, D. Seppi, A. Batliner, A. Maier, and S. Steidl, "Emotion recognition in the noise applying large acoustic feature sets," in *Proc. Speech Prosody*, vol. 6, pp. 1802-1805, May. 2006.
- [26] M. Lugger and B. Yang, "Cascaded emotion classification via psychological emotion dimensions using a large set of voice quality parameters," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, vol. 4, pp. 4945-4948, May 2008.
- [27] E. M. Alborno, D. H. Milone, and H. L. Rufiner, "Spoken emotion recognition using hierarchical classifiers," *Computer Speech and Language*, vol. 25, no. 3, pp. 556-570, Aug. 2011.
- [28] N. Kamaruddin, A. Wahab, and C. Quek, "Cultural dependency analysis for understanding speech emotion," *Expert Systems with Applications*, vol. 39, no. 5, pp. 5115-5133, Apr. 2012.
- [29] M. Kotti and C. Kotropoulos, "Gender classification in two emotional speech databases," in *Proc. 19th Int. Conf. on Pattern Recognition, ICPR'08*, vol. 3, pp. 380-386, Aug. 2008.
- [30] C. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer, 2006.
- [31] J. R. Raudays and A. K. Jain, "Small sample size effects in statistical pattern recognition: recommendations for practitioners," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 3, pp. 252-264, Mar. 1991.
- [32] B. Yang and M. Lugger, "Emotion recognition from speech signals using new harmony features," *Signal Processing*, vol. 90, no. 5, pp. 1415-1423, May 2010.
- [33] S. Whittle, M. Yucel, M. B. H. Yap, and N. B. Allen, "Sex differences in the neural correlates of emotion: evidence from neuroimaging," *Biological Psychology*, vol. 87, no. 3, pp. 319-333, May 2011.
- [34] H. Altun and G. Polat, "Boosting selection of speech related features to improve performance of multi-class SVMs in emotion detection," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8197-8203, May 2009.

علی حریمی در سال ۱۳۸۴ مدرک کارشناسی خود را در رشته برق- الکترونیک از دانشگاه سجاد مشهد دریافت نمود. وی مقطع کارشناسی ارشد خود را در سال ۱۳۸۶ از دانشگاه صنعتی شاهرود دریافت کرد و در سال ۱۳۹۳ در رشته برق- مخابرات سیستم

سوم، ویژگی‌های مستخرج از منحنی‌های توصیفگر به دست آمده از فضای فاز بازسازی شده برای طبقه‌بندی احساس‌های عصبانیت و خوشحالی و همچنین خستگی و عادی کارآمد هستند.

همچنین نتایج آزمایش‌های ما نشان دادند احساس زنان با دقت بیشتری از احساس مردان قابل بازشناسی است. این امر به ویژه هنگامی که از ویژگی‌های عروسی و طیفی متداول برای جداسازی احساس‌ها بر اساس سطح برانگیختگی استفاده می‌کنیم نمود بیشتری دارد.

به علاوه، طبقه‌بند دو طبقه پیشنهادی نیز تأثیر به‌سزایی در کاهش خطای طبقه‌بندی دارد. با توجه به این که تداخل احساس‌های عصبانیت و خوشحالی و همچنین خستگی و عادی بخش مهمی از خطای طبقه‌بندی را به خود اختصاص می‌دهد، استفاده از یک طبقه‌بند متوالی برای جداسازی این احساس‌ها می‌تواند کمک شایانی به کاهش نرخ خطا نماید. در سال‌های اخیر، فضای فاز بازسازی شده توانسته است جایگاه خود را در کاربردهایی نظیر تشخیص گفتار [۱۲] و [۱۳]، ارتقای گفتار [۱۴] و تشخیص خواب‌آلودگی از روی گفتار [۱۵] پیدا کند. به عنوان پیشنهادی برای کارهای آینده، بهینه‌سازی ساختار طبقه‌بند و بردار ویژگی مورد استفاده در دستور کار قرار دارد. همچنین با توجه به این که هدف نهایی سیستم‌های تشخیص احساس از روی گفتار کارکردن در محیط واقعی است، ارزیابی سیستم پیشنهادی در شرایط حضور نویز نیز می‌تواند مفید باشد.

مراجع

- [1] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion recognition in speech using neural networks," *Neural Comput. Appl.*, vol. 9, no. 4, pp. 290-296, 2000.
- [2] B. Schuller, G. Rigoll, and M. Lang, "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture," in *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP'04*, vol. 1, pp. 577-580, 7-21 May 2004.
- [3] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, and M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE Trans. Biomedical Eng.*, vol. 47, no. 7, pp. 829-837, Jul. 2007.
- [4] J. Hansen and D. C. Icarus, "Source generator based real-time recognition of speech in noisy stressful and Lombard effect environments," *Speech Commun.*, vol. 16, no. 4, pp. 391-422, Jun. 1995.
- [5] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572-587, May 2011.
- [6] S. Wu, T. H. Falk, and W. Y. Chan, "Automatic speech emotion recognition using modulation spectral features," *Speech Communication*, vol. 53, no. 5, pp. 768-785, May 2011.
- [7] E. Kim, K. Hyun, S. Kim, and Y. Kwak, "Speech emotion recognition using eigen-fft in clean and noisy environments," in *Proc. 16th IEEE Int. Symp. on Robot and Human Interactive Communication, RO-MAN'07*, pp. 689-694, 26-29 Aug. 2007.
- [8] R. Cowie, et al., "Emotion recognition in human-computer interaction," *IEEE Signal Process. Mag.*, vol. 18, no. 1, pp. 32-80, Jan. 2001.
- [9] F. Takens, "Detecting strange attractors in turbulence," *Dynamical Systems and Turbulence, Warwick*, vol. 898, pp. 366-381, 1980.
- [10] H. M. Teager and S. M. Teager, "Evidence for nonlinear sound production mechanisms in the vocal tract," in *Speech Production and Speech Modelling*, W. J. Hardcastle and A. Marchal, Eds., NATO Advanced Study Institute Series D, vol. 55, Bonas, France, 1989.
- [11] X. Huang, A. Acero, and H. W. Hon, *Spoken Language Processing*, Upper Saddle River, NJ and Prentice Hall, 2001.
- [12] K. M. Indrebo, R. J. Povinelli, and M. T. Johnson, "Sub-banded reconstructed phase spaces for speech recognition," *Speech Communication*, vol. 48, no. 7, pp. 760-774, Dec. 2006.

علی شهزادی مدرک دکتری خود را در سال ۱۳۸۸ از دانشگاه علم و صنعت ایران دریافت نمود و در حال حاضر به‌عنوان عضو هیأت علمی دانشگاه سمنان مشغول به فعالیت است. زمینه‌های علمی مورد علاقه ایشان عبارتند از: مخابرات امن و سیستم‌های رمزنگاری و کدینگ.

خشایار یغمائی در سال ۱۹۹۷ م. مدرک دکتری خود را در رشته مهندسی برق- مخابرات سیستم از دانشگاه Surrey انگلستان اخذ نمود. وی از سال‌ها پیش از آن به‌عنوان عضو هیأت علمی دانشگاه سمنان مشغول بود و جزء پایه‌گذاران رشته مهندسی برق در این دانشگاه به شمار می‌رود. ایشان در سال ۱۳۹۰ در سن ۵۴ سالگی در اثر سکته قلبی مرحوم شدند.

مقطع دکتری از دانشگاه سمنان فارغ التحصیل شد. ایشان هم‌اکنون به‌عنوان عضو هیأت علمی دانشگاه آزاد اسلامی واحد شاهرود به فعالیت اشتغال دارد. زمینه‌های علمی مورد علاقه ایشان عبارتند از: بینایی ماشین و پیاده‌سازی سیستم‌های خیره.

علیرضا احمدی فرد موفق به اخذ درجه کارشناسی از دانشگاه صنعتی اصفهان در سال ۱۳۶۸ در رشته مهندسی برق- الکترونیک شد. وی در سال ۱۳۷۳ مدرک کارشناسی ارشد خود را از دانشگاه امیر کبیر ایران دریافت نمود. در سال ۲۰۰۳ م. مدرک دکتری خود را از دانشگاه Surrey انگلستان دریافت نمود. وی هم‌اکنون به‌عنوان عضو هیأت علمی و رئیس دانشکده برق دانشگاه صنعتی شاهرود به فعالیت اشتغال دارد. زمینه‌های علمی مورد علاقه ایشان عبارتند از: بینایی ماشین و پردازش سیگنال به‌ویژه پردازش تصویر.

Surf and download all data from SID.ir: www.SID.ir

Translate via STRS.ir: www.STRS.ir

Follow our scientific posts via our Blog: www.sid.ir/blog

Use our educational service (Courses, Workshops, Videos and etc.) via Workshop: www.sid.ir/workshop