

بررسی سه ویژگی صحبت در دو حالت خوشحالی و ناراحتی و تفکیک این دو حالت

سعید راحتی قوچانی

استادیار گروه مهندسی پزشکی،

دانشگاه آزاد اسلامی واحد مشهد

rahati@mshdiau.ac.ir

سارا منعم خراسانی

دانشجوی کارشناسی ارشد مهندسی پزشکی،

دانشگاه آزاد اسلامی واحد مشهد

Sa_monem@yahoo.com

چکیده:

یکی از تفاوت‌های عمده بین بشر و سایر موجودات زنده، داشتن احساسات و ابراز آنها به صورتهای گوناگون است. به عنوان مثال، گاهی اوقات می‌توان احساسات فرد را از تغییر در حالت چهره او فهمید. در اکثر موارد به وجود آمدن یک حالت روحی در افراد، به طور ناخودآگاه روی نحوه ادای واژه‌ها و آهنگ گفتار آنها تأثیر می‌گذارد. این تغییر در گفتار می‌تواند ما را متوجه برخی خصوصیات فیزیکی موجود در سیگنال صحبت کند. چند خصوصیت فیزیکی سیگنال صحبت که در این تحقیق جهت تشخیص حالات روحی گوینده (شادی و غم) مورد استفاده قرار گرفته‌اند عبارتند از: میانگین فرکانس زیر و بمی^۱ پیوستگیهای منحنی تغییرات^۲ زیر و بمی، جیترا^۳ صدا، و نسبت فریمهای بی‌واک^۴ به کل فریمهای موجود در سیگنال. در اینجا از بانک داده صحبت با احساس (شادی و غم) به زبان آلمانی استفاده شده است. نشان داده شد که می‌توان به کمک این ویژگیها این دو حس موجود در صحبت را به طور کامل از هم جدا کرد زیرا، محدوده هر یک از ویژگیهای ذکر شده در مورد حس شادی و غم، به میزان قابل توجهی با یکدیگر متفاوتند.

در این تحقیق، ابتدا برای استخراج دقیق ویژگیها از نرم‌افزار پرت^۵ استفاده شده است سپس، از نرم‌افزار مطلب^۶ برای طراحی یک دسته‌بندی کننده (از نوع شبکه‌های عصبی) برای جداسازی حس شادی و غم استفاده شد.

۱- مقدمه:

هدف اصلی این تحقیق پرداختن به نحوه تأثیر حالات روحی فرد در طرز صحبت او و تشخیص ویژگیهای خاص در کلام است که در صحبت عادی و بدون احساس به چشم نمی‌خورند. همچنین دانستن این موضوع که این ویژگیها در احساسات مختلف با هم تفاوت دارند و تفاوتشان در حدی است که می‌توان از روی این خصوصیات به تفکیک آنها پرداخت، امر مهمی است. جدا از مسائل مربوط به

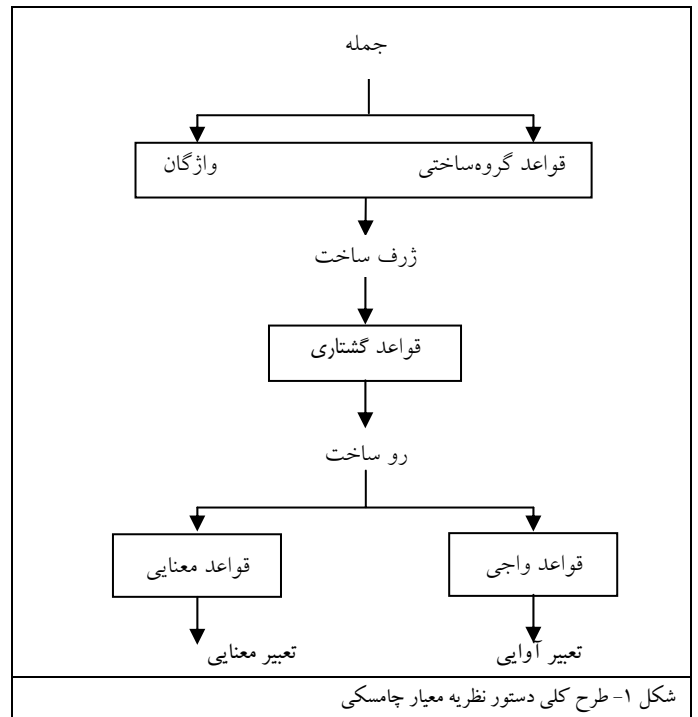
¹ pitch
² contour
³ jitter
⁴ unvoiced
⁵ PRAAT
⁶ Matlab

فیزیک صوت و پردازش سیگنال و روشهای محاسباتی استخراج این ویژگیها، باید ابتدا از نظر روان‌شناسی و زبان‌شناسی و جامعه‌شناسی به موضوع احساسات و صحبت آمیخته به آن نگاه کرد. چنین دیدگاهی می‌تواند در حل مسائل پیش‌رو و چگونگی تأثیر حالات روحی روی سیگنالهای حیاتی که یکی از پرمفهومترین آنها سیگنالهای صحبت است، مفید باشد. به منظور تحقیق درباره صحبت و نحوه دخالت احساسات در آن، ابتدا باید بدانیم در علم زبان‌شناسی چه رابطه‌ای بین واژگان جمله و خصوصیات احساسی که به الگوی آوایی تعبیر می‌شود وجود دارد. پس از آن، باید موضع خود را در مورد احساسات مشخص کرده و تعریف روشنی از آنها ارائه دهیم. همچنین باید بدانیم چرا احساس روی صحبت ما تأثیر می‌گذارد و آیا این تأثیر در شرایط مختلف یکسان است یا نه.

در سال ۱۹۶۵، چامسکی در کتابش تحت عنوان "جنبه‌هایی از نظریه نحو"^۷، نظریه دستوری را ارائه کرد که به نظریه معیار مشهور شد. دستور نظریه معیار به طور اساسی دربرگیرنده مجموعه‌ای از قواعد مختلف نحوی، معنایی و آوایی است. هر گروه از این قواعد به صورتی نظام‌مند برای ارائه یک توصیف یا نمود زبانی از تمام جملات در چهار سطح متفاوت به کار گرفته می‌شود. در سطح آوایی (تعبیر آوایی)، الگوی آوایی جمله نمود پیدا می‌کند، در سطح معنایی (تعبیر معنایی)، معانی و روابط منطقی جمله منعکس می‌شوند و در دو سطح نحوی (ژرف ساخت و روساخت)، جنبه‌های مختلف نحوی جمله متجلی می‌شوند. ژرف ساخت صورت نحوی زیرین جمله را نشان می‌دهد و روساخت دربرگیرنده صورت آشکارتر آن است. شمای کلی دستور نظریه معیار در شکل ۱ نشان داده شده است.

⁷ Aspects of the Theory of Syntax

همان طور که در شکل ۱ دیده می‌شود این دستور از مؤلفه‌های نحوی، معنایی، آوایی و یک واژگان یا مخزن لغات که دربرگیرنده تکواژها^۸ و اصطلاحات^۹ است، تشکیل می‌شود. بخش نحوی دو نوع قاعده نحوی را شامل می‌شود، قواعد گروه‌ساختی^{۱۰} (که قواعد پایه نیز نامیده می‌شوند) و قواعد گشتاری^{۱۱}، هر گروه از این قواعد مسئولیت شکل‌گیری یک بخش از جمله را به عهده دارند. [۱]



باید توجه داشت که پس از تولید روساخت هر جمله، تعبیر آن در قالب توالی نمادهای صوتی به عهده قواعد بخش آوایی است. این توالیهای صوتی تعبیر آوایی جمله نامیده می‌شوند. به عنوان مثال در یک جمله، این قواعد آوایی هستند که مشخص می‌کنند حروف انتهایی کلمات به حروف ابتدای کلمات بعدی به صورت چسبیده بیان شوند و یا واضح و جداگانه. همچنین در این مرحله است که براساس ویژگیهای نحوی جمله، یک الگوی تکیه نیز برای آن در نظر گرفته می‌شود. الگوی تکیه در جملات پرسشی و خبری و عاطفی متفاوت است. مورد دیگری که در الگوی آهنگ جمله به گوش می‌خورد، زیر و بمی ادای کلمات است که باز هم در جملات خبری و پرسشی تفاوت می‌کند.

⁸ morphemes
⁹ idioms
¹⁰ Phrase Structure rules
¹¹ Transformational rules

بنابراین می‌بینیم که خصوصیات صوتی ادای کلمات و جمله می‌توانند در تعبیری که ما از جمله داریم تأثیر بسیاری بگذارند. همچنین این خصوصیات صوتی در علم زبان‌شناسی تحت عنوان تعبیر آوایی مطرح می‌شوند و در مرحله خاصی از ساخت جمله شکل می‌گیرند. در شکل ۲ تغییراتی که در اثر الگوی آوایی در جمله به وجود می‌آید را به صورت نموداری از فرکانس بر حسب زمان (طیف‌نگاری) نمایش داده‌ایم.

حال باید بدانیم که احساسات یا همان عاملی که سبب ایجاد بخش عمده الگوی آوایی در جمله می‌شود از کجا نشأت می‌گیرد و چگونه می‌تواند در صحبت کردن ما تغییر ایجاد کند. در علم روان‌شناسی چهار دیدگاه یا چهار نظریه در تعریف احساسات وجود دارد، که هر نظریه طرفدارانی نیز دارد و طبق آن مسائل زیادی را توجیه می‌کنند. برای ورود به موضوع، این چهار نظریه را به طور خلاصه توضیح می‌دهیم. [۲]

۱-۱) نظریه داروین:

تفکر اصلی در این نظریه این است که احساسات، نتایج توابع سازگاری هستند که در طول دوران تکامل بشر به عنوان یک گونه، در برابر مشکلات پیش رو انتخاب شده و باقی مانده‌اند [۳]. بنابراین باید در تمام انسانها احساسات کم و بیش مشابهی انتظار داشته باشیم. همچنین چون انسان نسبت به سایر پستانداران تکامل یافته‌تر است پس باید چیزی شبیه این احساسات را در پستانداران دیگر نیز ببینیم. مثلاً ممکن است ترس در انسان و خرگوش متفاوت باشد، اما در هر دو گونه بخشهایی از مغز که درگیر ایجاد این حس می‌شوند مشابهند و عملکرد یکسانی دارند. بر اساس این نظریه "احساسات اولیه یا اصلی" مطرح شدند، یعنی احساساتی که نقش تطبیقی خاصی در برخورد با اتفاقاتی که در سیر تکاملی بشر رخ داده، داشته‌اند و نیز سایر احساسات از این احساسات اولیه مشتق می‌شوند.

۲-۱) نظریه جیمز:

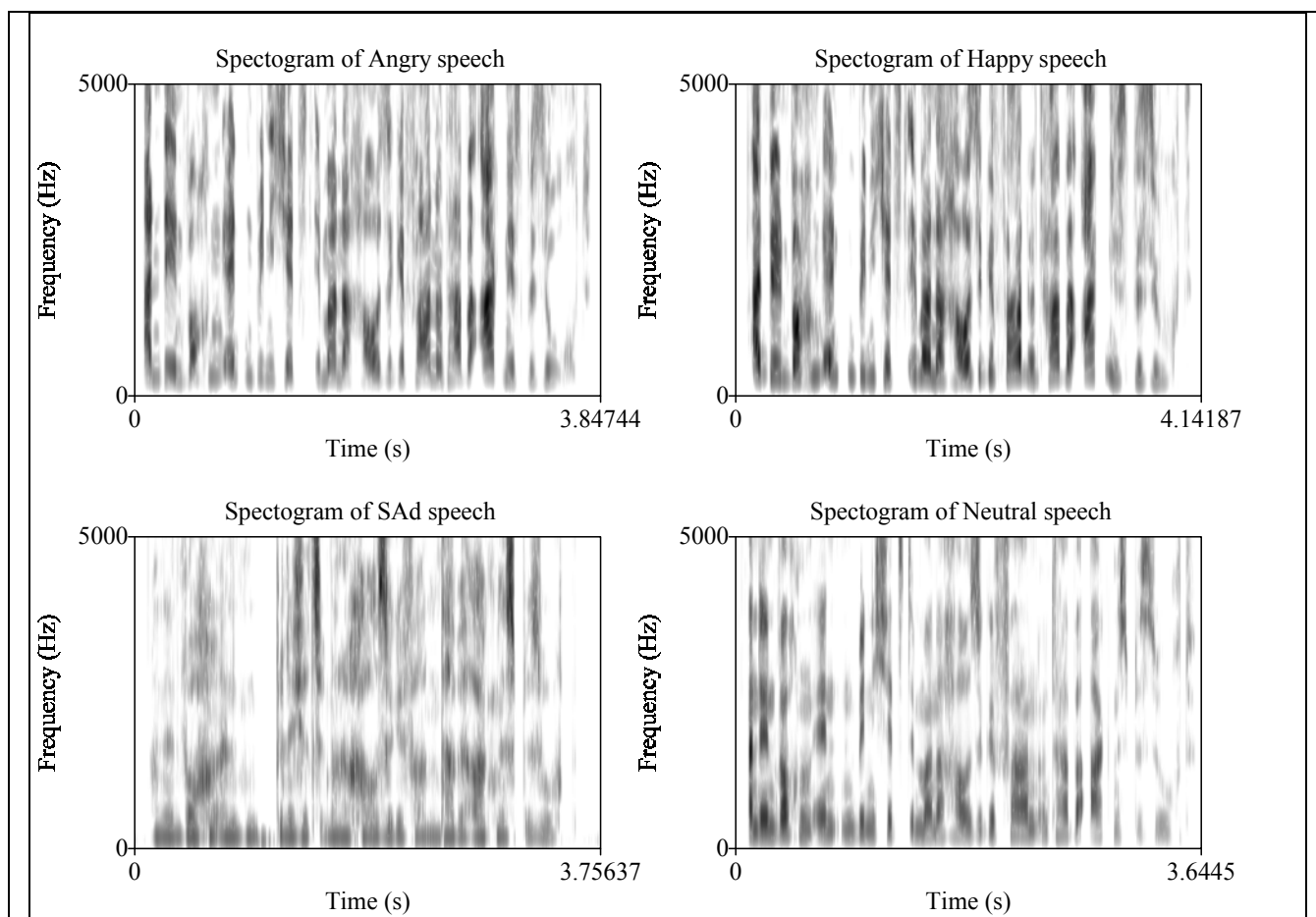
ویلیام جیمز در سال ۱۸۸۴ مقاله‌ای تحت عنوان "احساسات چیست؟" منتشر کرد که در آن احساسات را به نوعی با تغییرات بدنی معادل دانسته بود. به این صورت که "تغییرات جسمی مستقیماً از درک محرکها تبعیت می‌کنند و نحوه درک ما از این تغییرات، احساسات است." جیمز عقیده دارد که احساسات را نمی‌توانیم از تغییرات جسمی جدا فرض کنیم و همیشه این تغییرات بدنی هستند که پیش از احساسات رخ می‌دهند. او می‌گوید: "ما احساس شرم می‌کنیم چون مثلاً گریه کرده‌ایم."

می توان نتیجه گرفت که داروین و جیمز، احساسات را نوعی تطبیق با محیط و پاسخهایی خودکار در برابر اتفاقاتی می دانند که از محیط به ارگانسیم زنده تحمیل می شود.

شخصیت و حالت روانی او با حالت و موقعیت خاصی که در آن قرار گرفته ارتباط برقرار می کنند.

۱-۴) نظریه ساختار اجتماعی:

این دیدگاه جدیدترین و بحث برانگیزترین نظریه در بین چهار نظریه



شکل ۲- نمونه ای از طیف فرکانسی ۳ حالت خوشی و ناراحتی و عصبانیت و حالت بدون احساس در بیان یک جمله ثابت

۱-۳) نظریه شناختی^{۱۲}

آغاز نظریه معرفتی جدید بر اساس مطالعات آرنولد در دهه ۶۰ است که البته اصل نظریه بسیار قدیمتر است. فرض اصلی این دیدگاه این است که تفکر و احساسات از هم مستقل نیستند. تمام احساسات فرآیندی هستند که طی آن اتفاقات محیطی برای ما خوب یا بد جلوه می کنند. در نظریه آرنولد و موافقانش چیزی که جیمز به نام ادراک^{۱۳} از محرک قلمداد کرده بود به نام ارزیابی^{۱۴} می شناسند. همان طور که جیمز نمی توانست هیچ احساسی را مستقل و بدون جسم موجود زنده فرض کند، آرنولد هم به احساسات بدون ارزیابی اعتقادی ندارد. به عقیده او هر احساسی مربوط به الگوی متفاوت و خاصی از ارزیابی می شود. این الگوها، بین خصوصیات یک فرد یا ارگانسیم، فطرت و

مطرح شده است و بر خلاف سه مورد قبلی که احساسات را به زیست شناسی مربوط می دانستند، این نظریه عقیده دارد که احساسات، نتایج فرهنگی هستند که مفهوم و ارتباط آنها به دانستن قوانین اجتماعی مربوط می شود. طبق نظریه جیمز آوریل که در سال ۱۹۸۰ مطرح شد، احساسات نه تنها حاصل دوران تکامل زندگی ما هستند و نه تنها می توان آنها را توسط مفاهیم روان شناسی توضیح داد، بلکه مفاهیمی اجتماعی هستند و فقط وقتی می توان آنها را کامل فهمید که در سطح اجتماعی و گروهی تحلیل کرد. به عقیده او، عصبانیت از عدم رضایت ما از فردی دیگر ناشی می شود. مثلاً اگر دست شما به طور ناخودآگاه و تصادفی به صورت فرد دیگری برخورد از شما عصبانی نمی شود. اما اگر این کار را عمدی انجام دهید او از کار شما عصبانی خواهد شد. همچنین در این نظریه، فرهنگ هر جامعه نقشی اصلی و بنیادی در شکل گیری احساسات در سطوح مختلف دارد.

¹² cognitive

¹³ perception

¹⁴ appraisal

با توجه به این دیدگاهها و مطالعات جدید می‌توانیم به تعاریف دقیقی از هر یک از احساسات و انواع آنها و تعامل آنها با بدن انسان برسیم. باید بدانیم که تقریباً تمام نظراتی که در مورد منشأ و ماهیت احساسات مختلف داده می‌شود به نوعی به همین چهار نظریه مربوط می‌شوند و همچنین در قسمت‌های مختلف این چهار نظریه می‌توان ویژگی‌های مشترک و همپوشانی‌هایی پیدا کرد. حال جهت ادامه بحث و کار روی صحبت همراه با احساسات لازم است بدانیم که به نظر ما احساسات طبق کدام نظریه تعریف می‌شوند و یا این که در هر بخش از تحقیق کدام دیدگاه را مد نظر قرار دهیم. به عنوان مثال اگر در مورد وابستگی‌های صوتی صحبت با احساس در ۵ حس عمده عصبانیت، ناراحتی، شادی، تنفر و هیجان مطالعه می‌کنیم باید بدانیم این احساسات را چطور تعبیر کنیم. آیا باید طبق نظر داروین فرض کنیم که حالات قابل تشخیص در چهره همراه با این احساسات وجود دارند. یا باید بنا بر نظر جیمز بتوانیم فرض کنیم که این حالات چهره و تغییرات به وجود آمده در صدا و نحوه بیان فرد ممکن است با الگوهای منحصربه‌فردی از فعالیت سیستم عصبی خودکار همراه باشند. همچنین باید بدانیم که طبق نظریه شناختی هر یک از احساسات، الگوی منحصربه‌فردی در ارزیابی دارند. و نیز این که آیا احساسات مختلف در جوامع گوناگون بشری با فرهنگها و آداب و سنن متفاوت، یکسانند یا نحوه بروز آنها تفاوت خواهد کرد.

هدف از این تحقیق بررسی ویژگی‌های صوتی قابل تفکیک در صحبت با احساس است، لذا مبنا را بر وجود الگوهای خاص فعالیت سیستم عصبی خودکار، یعنی نظریه جیمز قرار می‌دهیم. البته فقط در این حد که حالات روحی مختلف تأثیری بر فعالیت‌های عصبی ما می‌گذارند و این تأثیر باعث ترشح هورمون‌ها و فعالیت برخی عضلات خواهد شد. بنابراین مثلاً در هنگام ناراحتی آب دهان خشک می‌شود، عضلات تارهای صوتی ضعیفتر از معمول عمل می‌کنند و صدای فرد، آرام به گوش می‌رسد. و یا هنگام هیجان و خوشحالی ضربان قلب سریعتر می‌شود و خون بیشتری به عضلات درگیر صحبت می‌رسد و در نتیجه فرد با انرژی بیشتری می‌تواند کلام خود را بیان کند. این تغییرات فیزیولوژیک، تأثیراتی روی صحبت ما می‌گذارند که می‌توان ویژگی‌های خاصی را در هر حالت استخراج کرد و از روی این خصوصیات به درک تفاوت بین احساسات مختلف رسید. همچنین به دلیل استفاده از بانک داده به زبان آلمانی در این تحقیق، می‌توانیم از نظر اجتماعی یعنی نظریه چهارم هم مسئله را بررسی کنیم. ممکن است فرهنگها و زبانهای مختلف، مفاهیم احساسی متفاوتی داشته باشند و یا از کلمات متفاوتی در بیان احساسات خود استفاده کنند اما حالات

عمده‌ای مثل خوشحالی و ناراحتی و عصبانیت و هیجان، در تمام ملیتها تعریف شده است. بنابراین اگر به جملات با احساسی که افراد با زبانهای خارجی بیان کرده‌اند توجه کنیم، فقط از روی تغییرات موجود در صدای گوینده می‌توانیم تا حد قابل توجهی پی به حالت روحی او ببریم. در زبان فارسی نیز می‌توان مشابه سایر زبانها تفاوت‌های زیادی بین صحبت افراد با حالات روحی مختلف یا در هنگام بیان جملات پرسشی یا تعجبی مشاهده کرد. این تفاوتها ممکن است در فرکانس زیر و بمی آنها و یا آهنگ کلی جمله باشد [۱۰].

در ادامه این مقاله در بخش دوم بانک داده را به صورت مختصر توضیح می‌دهیم. در بخش سوم ویژگی‌های صوتی را که در این تحقیق با استفاده از نرم‌افزار پرت محاسبه و مورد استفاده قرار گرفته‌اند را معرفی می‌کنیم. در بخش چهارم به بیان روش تفکیک دو حالت روحی خوشحالی و ناراحتی بر اساس داده‌های استخراج شده و توسط شبکه‌های عصبی می‌پردازیم. در پایان در بخش پنجم به بحث و نتیجه‌گیری پرداخته می‌شود.

۲- بانک داده:

بانک داده‌ای که در اینجا مورد استفاده قرار می‌گیرد توسط انجمن علوم ارتباطات دانشگاه فنی برلین^{۱۵} جهت آزمایش وابستگی‌های صوتی صحبت با احساس جمع‌آوری شده است. در این مجموعه، جملات با شش احساس عصبانیت، خوشحالی، ناراحتی، ترس، تنفر، و کسالت به همراه حالت بدون احساس و خنثی ادا شده‌اند. این جملات توسط ۱۰ هنرپیشه حرفه‌ای آلمانی (۵ زن و ۵ مرد) ادا شده‌اند. هر یک از این افراد، ۱۰ جمله را با شبیه‌سازی احساس مورد نظر بیان کرده‌اند. این جملات عبارتند از ۵ جمله‌ی کوتاه و ۵ جمله‌ی بلندی که در محاورات روزانه کاربرد زیادی دارند و قابل اجرا در تمام احساسات ذکر شده می‌باشند. همراه ضبط صدا، الکتروگلو توگرام افراد توسط یک لارنژوگراف پرتابل ثبت شده است، که ما در این تحقیق از این بخش از داده‌ها استفاده‌ای نکرده‌ایم. این بانک داده از ۸۰۰ جمله تشکیل شده است. جهت تست قابل تشخیص بودن و کیفیت ادای هر احساس، جملات توسط ۲۰ تا ۳۰ نفر مورد قضاوت قرار گرفته‌اند. جملاتی که حدود ۸۰٪ شنونده‌ها آنها را با همان احساس ثبت شده تشخیص دادند، برای تحلیل‌های بعدی مورد استفاده قرار گرفتند. نشانه‌گذاری جملات به این صورت است که هر نشانه شامل ۷ حرف است و هر حرف بیان‌کننده اطلاعاتی خاص در مورد گوینده و احساس و جمله مورد نظر است. در اینجا جهت اطلاع بیشتر به دو جمله از جملات بانک داده اشاره می‌کنیم [۴]:

¹⁵ TU-Berlin (Technical University of Berlin)

a07- این اتفاق ظرف هفت روز رخ خواهد داد.

b03- معمولاً آخر هفته‌ها برای دیدن آگنس به خانه می‌رفتم.

در این تحقیق چون فقط دو احساس خوشحالی و ناراحتی مد نظر بودند، لذا تمام جملات موجود در بانک داده که توسط تمام ۱۰ هنرپیشه در دو حالت خوشحال و ناراحت ادا شده بودند مورد استفاده قرار گرفتند.

انرژی صدا	بالا	بالا	پایین	معمولی
کیفیت صدا	از قفسه سینه	با نفس و فریاد	طین‌دار	حالت نامنظم
تغییرات زیر و بمی	تغییر ناگهانی در هجاهای تکیه‌دار	انحناهای نرم رو به بالا	انحنا رو به پایین	معمولی
طرز تلفظ	شدید	معمولی	شل و سست	حالت انسدادی

جدول ۱- وابستگی صوتی احساسات مختلف به برخی ویژگی‌های سیگنال صحبت

۳-۱) میانگین فرکانس زیر و بمی در بخشهای

پیوسته منحنی تغییرات زیر و بمی:

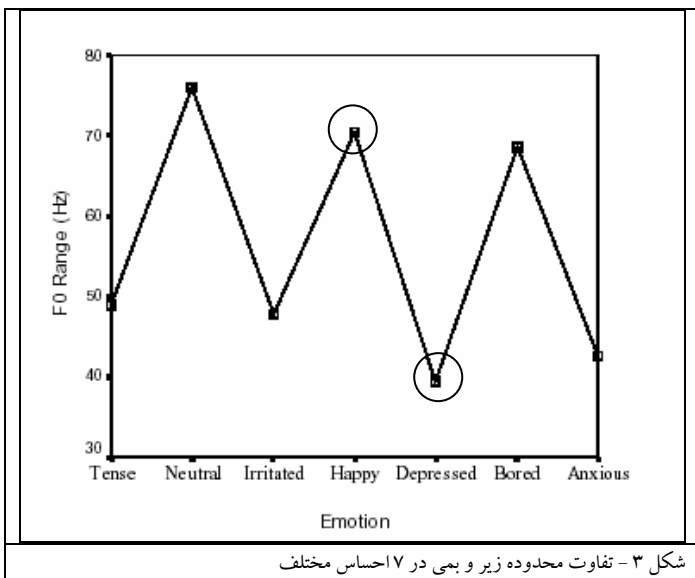
فرکانس زیر و بمی و تغییرات و مشخصات آماری آن از جمله حداکثر، حداقل، انحراف معیار، میانگین، بازه تغییرات و شیب تغییرات آن تقریباً در تمام تحقیقات مربوط به تشخیص احساسات از روی ویژگی‌های صحبت مورد استفاده قرار گرفته است [۵]. در نرم‌افزار پرت هم برای هر صدا و یا بخش انتخابی از آن ۵ پارامتر مربوط به فرکانس زیر و بمی محاسبه می‌شوند که عبارتند از: میانگین، میانه، حداکثر، حداقل، و انحراف معیار. در این تحقیق به جای استفاده از این پارامترها، سعی شد پارامتر جدیدی که البته چیزی جز میانگین فرکانس زیر و بمی در هر قطعه پیوسته از منحنی تغییرات زیر و بمی نیست، استفاده شود.

در تمام مطالعات انجام شده این نتیجه کاملاً مشخص است که مشخصه فرکانس زیر و بمی در مورد حس خوشحالی در سطح بالاتری نسبت به حس ناراحتی قرار دارد. این موضوع را در جدول ۱ هم می‌توان مشاهده کرد، همان طور که می‌بینید در ستون دوم و سوم که خوشحالی و ناراحتی با هم مقایسه شده‌اند، مشخص شده است که مقدار میانگین زیر و بمی در حالت خوشحالی بیشتر از حالت ناراحتی است. این مطلب را در شکل ۳ هم می‌توان مشاهده کرد [۶].

۳- ویژگی‌های صوتی سیگنالهای صحبت:

هر سیگنال صحبت، به دلیل داشتن ماهیت صوتی دارای ویژگی‌هایی است که این ویژگی‌ها در هنگام شنیدن سیگنال تا حدی حس می‌شوند. ویژگی‌هایی چون کم و زیاد بودن انرژی یا دامنه سیگنال، تغییرات فرکانسی و زیر و بم شدن صدا، تغییرات در ریتم و سرعت صحبت، نسبت بخش‌های سکوت موجود در جمله به بخش‌های حاوی اطلاعات صوتی و همچنین گنگ و مبهم یا شفاف بودن صحبت، خصوصیات هستند که گوش انسان قادر به تشخیص آنها می‌باشد. درک همین خصوصیات و تفاوتهاست که می‌تواند ما را متوجه حالات روحی گوینده هنگام صحبت و حتی برخی پارامترهای مربوط به دستور زبان مثل سؤالی یا خبری بودن جملات سازد. در جدول ۱ می‌توانید تأثیر احساسات را روی برخی از این ویژگی‌ها ببینید [۵].

برای استفاده از این ویژگی‌ها در تحقیقات مختلف باید بتوانیم آنها را به گونه‌ای کمی محاسبه کنیم. روشهای مختلفی جهت محاسبه این پارامترها وجود دارد که می‌توان به کمک نرم‌افزارهای متفاوت اقدام به محاسبه‌ی آنها نمود. در این تحقیق از محیط نرم‌افزار پرت استفاده شده است. در این محیط می‌توانیم فایل‌های صوتی موجود را بار کرده و یا صدای جدید ضبط کنیم. همچنین به راحتی می‌توان هر قطعه‌ای از کل جمله را انتخاب کرد و پارامترهایی همچون فرکانس زیر و بمی و مشخصات آماری آن، انرژی صوت، اطلاعات مربوط به فورمانتها و چند مشخصه دیگر را به دست آورد یا حتی تغییر داد. در این تحقیق، ویژگی‌های مورد نیاز خود را به کمک نرم‌افزار پرت استخراج کرده‌ایم. در اینجا به توصیف و تعریف دقیقتر ویژگی‌های به کار رفته در این مطالعه می‌پردازیم.

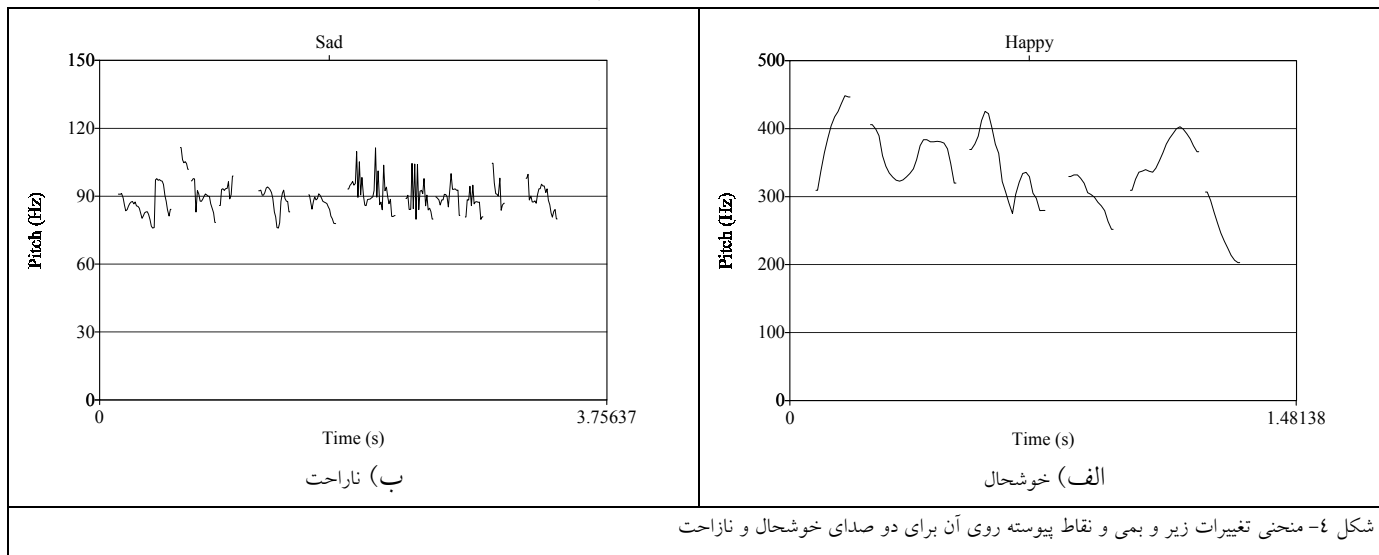


شکل ۳- تفاوت محدوده زیر و بمی در ۷ احساس مختلف

عصبانیت	خوشحالی	ناراحتی	ترس
کمی سریع	سریع یا آهسته	کمی آهسته	خیلی سریع
خیلی بالا	بالا	کمی پایین	خیلی بالا
بسیار وسیع	بسیار وسیع	کمی باریک	بسیار وسیع

سیگنالها، چه صحبت و چه سیگنالهای مخابراتی، این عدم همزمانی به صورت تکرار مجدد سیگنال با کمی تأخیر روی خودش اتفاق می افتد. به عنوان مثال اگر بخواهید یک صدای معمولی را به صورت مصنوعی جیترا دار کنید می توانید خود سیگنال را با تأخیرهای متفاوت (از کم به زیاد) ضبط کرده و دوباره روی صدای اصلی پخش کنید.

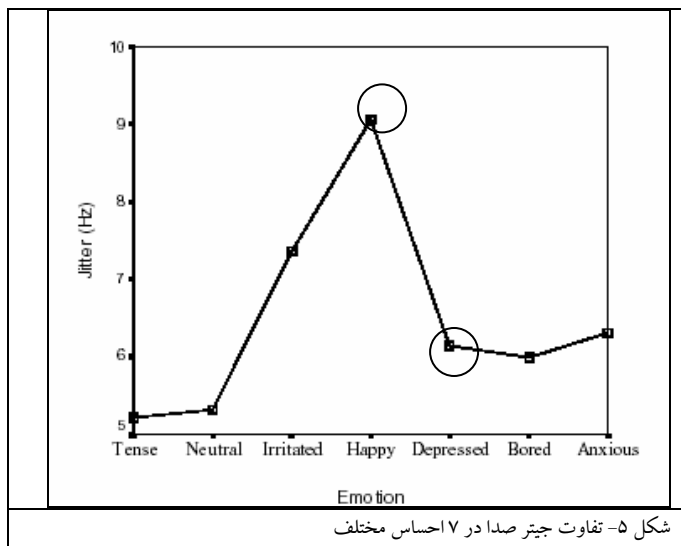
در این تحقیق هم پارامترهای زیر و بمی و به خصوص "میانگین زیر و بمی" در نقاط پیوستگی منحنی تغییرات زیر و بمی "در ۶۲ داده مربوط به حس خوشحالی و ۵۲ داده مربوط به ناراحتی، مورد بررسی قرار گرفت. برای نمونه، منحنی تغییرات زیر و بمی یک صدای خوشحال و یک صدای ناراحت در شکل ۴ آورده شده است.



با این کار هر چه تأخیر صدا بیشتر باشد درصد جیترا بیشتر خواهد شد [۸]. نتیجه این کار این است که صدا به صورت مبهمتر از حالت اصلی به گوش خواهد رسید. به عبارت دیگر نوعی نویز به آن اضافه می شود.

در این شکل مشاهده می شود که پیوستگیهای موجود در صدای خوشحال در محدوده ۲۰۰ تا ۴۰۰ هرتز قرار دارند، در حالی که همین نقاط در صدای ناراحت بین ۸۰ و ۱۱۰ قرار می گیرند. چنین الگویی تقریباً در تمام صداهای موجود دیده شد. لذا می توانیم از این ویژگی به عنوان یکی از معیارهای جداسازی دو حالت خوشحال و ناراحت استفاده کنیم.

به منظور رسیدن به ویژگی مورد نظر، در هر قسمت پیوستگی میانگین زیر و بمی را محاسبه کردیم. به طور متوسط تمام صداها ۶ بخش پیوسته داشتند. در صداهایی که این تعداد بیشتر بود، محللهای پیوستگی نزدیک به هم را یکی فرض کرده و میانگین فرکانس زیر و بمی را در کل آن به دست آوردیم.



03a01Fa	03a02Fc	03a04Fd	03a05Fc
156.88	228.09	303.2	202.42
197.48	273.53	237.9	193.32
168.98	266.79	221.3	224.69
180.42	288.38	221.57	325.11
244.26	360.2	171.98	201.37

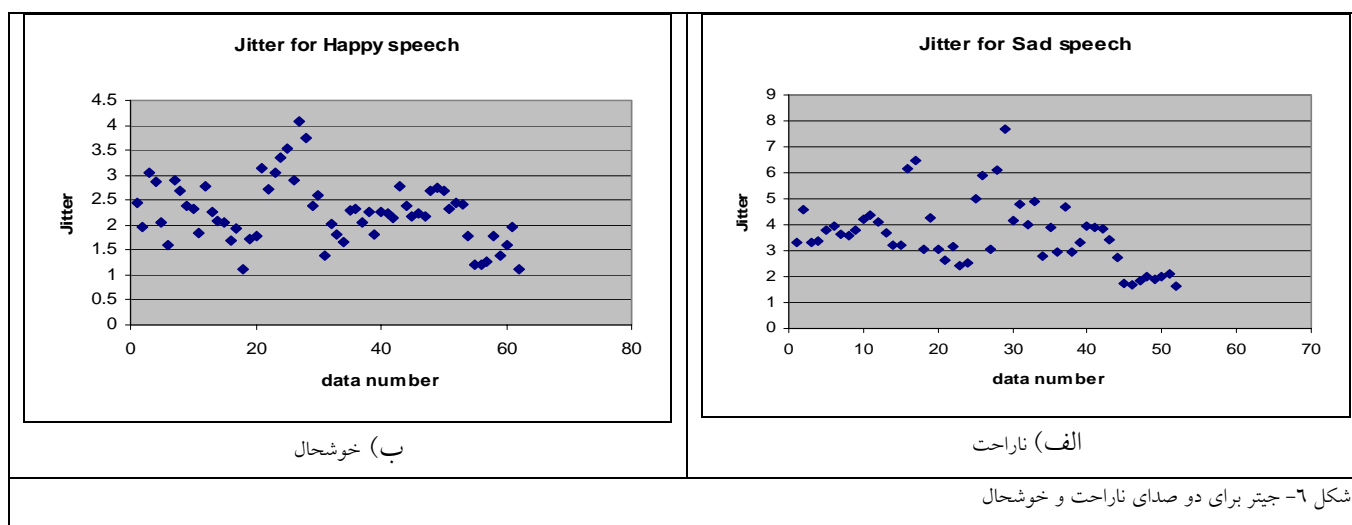
جدول ۲- میانگین زیر و بمی در پیوستگیهای منحنی تغییرات زیر و بمی برای ۴ صدای خوشحال

به هنگام صحبت با وجود برخی حالات روحی نیز، این تغییرات در صدا به وجود می آید و صدا را از شفافیت همیشگی اش خارج می سازد. این تغییرات در صدا (جیترا)، قابل محاسبه می باشند. یکی از روشهای محاسبه جیترا در نرم افزار پرت عبارت است از:

۲-۳) جیترا کل صدا:
جهت بررسی این ویژگی ابتدا باید آن را به طور ساده تعریف کنیم. به طور کلی به هر نوع عدم همزمانی ناخواسته، جیترا می گویند. در

تفاوت میانگین بین پریودهای متوالی تقسیم بر میانگین پریود کل صدا.

یکی دیگر از ویژگیهای مورد بررسی در این تحقیق، تعداد کل فریمهای بدون صدا یا بی‌واک به کل فریمهای داده مورد نظر



است. این کار را به کمک نرم‌افزار پرت انجام داده و به صورت درصد محاسبه می‌کنیم. حال به توضیح بیشتر این خصوصیت می‌پردازیم.

هر سیگنال صحبت را می‌توان به فریمهای کوچکتر تقسیم کرد. اکنون هر فریم یا واک‌دار یا بی‌واک یا سکوت است. بخشهای واک‌دار مربوط به حروف صدا دار در جمله و بخشهای بی‌واک مربوط به حروف بی صدا می‌باشند. حال می‌توان این محاسبات را با دقت بالا و با استفاده از روشهای محاسبه تعداد گذر از صفر^{۱۶} و محاسبه خود همبستگی^{۱۷} انجام داده‌ایم.

یکی از معیارهایی که در این تحقیق برای تفاوت دادن بین دو حالت روحی خوشحالی و ناراحتی مد نظر قرار گرفته است، جیتر می‌باشد. طبق اندازه‌گیریهایی که روی تمام ۶۲ صدای خوشحال و ۵۲ صدای ناراحت انجام گرفت، مشاهده شد که مقدار جیتر برای صداهای ناراحت بیشتر است. این موضوع را می‌توان به این صورت تفسیر کرد که در حالت ناراحتی، فرد کمی آهسته و ناواضح صحبت می‌کند، لذا صدای نفسها و هوایی که از مسیر هوایی خارج می‌شود می‌تواند حالت جیتردار به صدا بدهد.

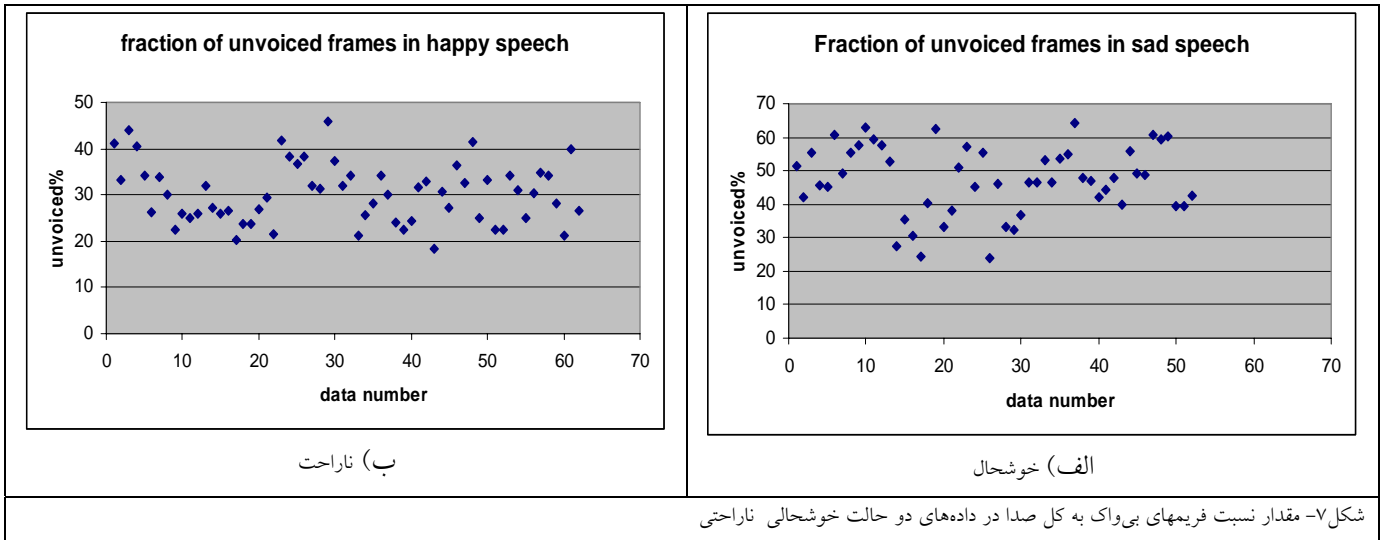
این مشاهدات با برخی مراجع هماهنگی نداشت [۷ و ۹]. در این مقالات آمده است که مقدار جیتر موجود در صحبت برای حالت روحی خوشحالی بیشتر از ناراحتی است (شکل ۵). همچنین گفته شده که در حالتی که فرد تحت استرس و فشار و هیجان است، مقدار جیتر صدا بیشتر است [۹].

در شکل ۶ نمودار جیتر به دست آمده برای دو حالت روحی خوشحال و ناراحت را در این تحقیق می‌آوریم.

همان طور که در شکل ۶-الف و ۶-ب دیده می‌شود مقدار جیتر برای صداهای ناراحت بیشتر بین ۲,۵ تا ۷ و برای صداهای خوشحال تقریباً بین ۱ و ۳ است. لذا چیزی که در این مشاهدات به دست آمد بر خلاف مراجع ذکر شده است. اما در هر حال به دلیل تفاوت کامل، می‌تواند معیار و ویژگی خوبی جهت تفکیک بین دو حالت روحی باشد.

۳-۳ نسبت فریمهای بی‌واک در کل صدا:

¹⁶ zero crossing
¹⁷ auto correlation



صداست. جدول ۳ یک نمونه از بردار ویژگی را برای یک صدای خوشحال و یک صدای ناراحت نشان می‌دهد.

	خوشحالی	ناراحتی
Mean pitch-1	367.29	160.12
Mean pitch-2	271.96	162.24
Mean pitch-3	438.26	153.82
Mean pitch-4	261.28	169.22
Mean pitch-5	285.45	443
Mean pitch-6	180.75	149.31
Jitter	2.09	3.8
Unvoiced frames	27.11	51.4

جدول ۳- یک نمونه از بردار ویژگی برای دو صدای خوشحال و ناراحت

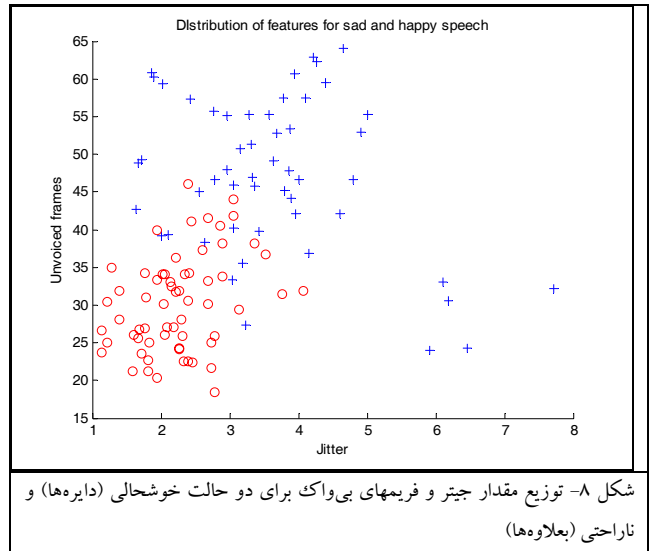
همچنین جهت مشاهده بهتر مقادیر ویژگی، دو سطر آخر بردار را برای تمام داده‌های هر دو حالت روحی در یک نمودار رسم کرده‌ایم که در شکل ۸ آمده است.

در نهایت جهت آزمودن این که این ویژگی‌ها می‌توانند به خوبی دو حالت مورد نظر را از هم جدا کنند یا خیر، از شبکه‌های عصبی استفاده کردیم. یک شبکه پرسپترون دو لایه که به تعداد ویژگی‌های مورد استفاده در هر مرحله، عصب ورودی دارد و یک عصب خروجی با دو حالت پاسخ طراحی شد.

شاید این سؤال پیش بیاید که واک‌دار یا بی‌واک بودن فقط به کلمات گفته شده و تعداد حروف صدادار و بی‌صدای موجود در آنها وابسته است، و در بیان جمله با حالات روحی مختلف تفاوتی نخواهند کرد. اما در این تحقیق پس از اندازه‌گیریها در مورد جملات یکسانی که با دو حالت خوشحالی و ناراحتی بیان شده بودند، مشاهده شد که درصد بخشهای بی‌واک به کل فریمهای صدا در دو حالت متفاوت است. این موضوع را می‌توانیم این‌گونه توضیح داد که چون در هنگام ناراحتی، فرد کلمات را با مکث بیشتری بیان می‌کند و در هنگام خوشحالی برعکس با سرعت و انرژی بیشتری جمله را ادا می‌کند، لذا کل زمان بیان حروف بی‌صدا در هنگام ناراحتی نسبت به خوشحالی بیشتر خواهد بود. این مشاهدات در شکل ۸ آورده شده است. همان طور که دیده می‌شود درصد بخشهای بی‌واک به کل صدا در حالت ناراحتی بیشتر بین ۴۰ و ۶۰ است. در حالی که همین مقدار در حالت خوشحالی چیزی بین ۲۰ و ۴۰ می‌باشد. لذا این ویژگی هم می‌تواند با دقت بالای دو حالت روحی ذکر شده را از هم تفکیک دهد.

۴- تفکیک دو حالت روحی با استفاده از ویژگی‌های ذکر شده:

همان طور که گفته شد و در شکل‌های ۴، ۶ و ۷ مشاهده می‌شود، سه ویژگی به دست آمده قابلیت تفکیک خوبی برای دو حالت روحی خوشحالی و ناراحتی خواهند داشت. اکنون از ترکیب تمامی ویژگی‌هایی که در بالا به آنها اشاره کردیم یک بردار ویژگی درست می‌کنیم. این بردار ویژگی یک بردار ۸ بعدی می‌باشد. در این بردار ویژگی، ۶ سطر اول مربوط به میانگین فرکانس زیر و بمی در پیوستگی‌های منحنی تغییرات زیر و بمی می‌باشند و سطر هفتم مقدار جیتر برای هر صدا و سطر هشتم بیانگر نسبت قابهای بی‌واک در کل



همان طور که در شکل ۹ مشاهده می‌کنید، شبکه برای تمام ۲۲ مورد داده‌های تست صدای خوشحال، خروجی مثبت به دست آورده است. همین‌طور خروجی شبکه برای ۱۲ مورد داده‌های تست صدای ناراحت، عددی منفی و بسیار نزدیک به -۱ شد که برای شبکه از پیش تعیین شده بود. یعنی اگر از یک آشکارساز آستانه استفاده کنیم و خروجی مثبت را مشخص کننده صدای خوشحال و خروجی منفی را نشان دهنده صدای ناراحت قرار دهیم، این دو ویژگی (جیتر و فریمهای بی‌واک) با دقت ۱۰۰٪ صداهای خوشحال و ناراحت را از هم تفکیک خواهند کرد.

لازم به ذکر است که جهت پیاده‌سازی شبکه عصبی و کار با آن از nntool GUI در محیط مطلب استفاده شد.

۵- نتیجه و بحث:

در این تحقیق سه ویژگی سیگنالهای صحبت که تاکنون زیاد مورد بررسی قرار نگرفته بودند، مورد مطالعه قرار گرفتند. هدف از این کار به دست آوردن ویژگیهایی از سیگنالهای صحبت است که قابلیت تفکیک بیشتری برای شناسایی احساسات مختلف به وجود آورند، و در احساسات بیان شده توسط افراد در هنگام صحبت تفاوت داشته باشند. در تمام کارهایی که در این زمینه انجام می‌شود بیشتر روی فرکانس زیر و بمی و خصوصیات آماری آن متمرکز می‌شوند. در این تحقیق، یکی از ویژگیها مربوط به فرکانس زیر و بمی است اما با کمی تفاوت نسبت به قبل. در اینجا برای هر صدا، ۶ فرکانس زیر و بمی در طول منحنی تغییرات زیر و بمی به دست آمد که در مجموع تفاوت بیشتری بین دو حالت روحی مورد نظر ایجاد کنند. ویژگی دیگری که محاسبه شد، مربوط به مفهوم یا نامفهوم بودن صحبت بود. می‌توان این

خصوصیت را به عنوان جیتر تعریف کرد. مشاهده شد که جیتر در صداهای خوشحال کمتر از صداهای ناراحت است. آخرین ویژگی مورد استفاده، میزان نسبت فریمهای بی‌واک به کل فریمهای موجود در صحبت بود. این پارامتر هم به دلیل تفاوت سرعت بیان کلمات در احساسات مختلف متفاوت است. در پایان بر اساس این ویژگیها به خوبی توانستیم دو حالت خوشحالی و ناراحتی را از هم تشخیص دهیم.

مبنای اصلی کارهای آینده این است که این ویژگیها را در احساسات دیگر نیز بررسی کنیم و ببینیم در آنها هم به همین اندازه قابلیت تفکیک وجود خواهد داشت یا خیر. همچنین در جستجوی ویژگیهای قابل محاسبه دیگری (مانند فرکانسهای فورمانت و یا استرس) باشیم که تفاوت حالات روحی موجود در صحبت را بارزتر نشان دهند.

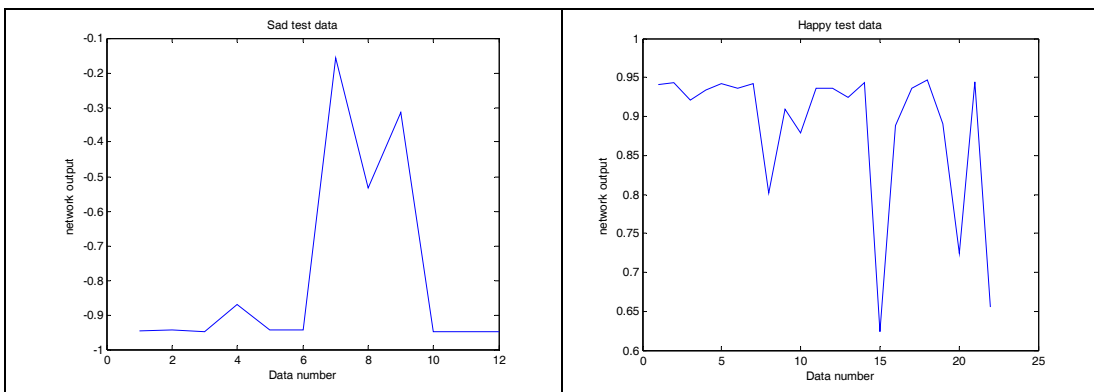
یافتن ویژگیهای مفیدتر و راههای تشخیص بهتر، می‌تواند در کاربردهایی همچون بهبود کیفیت صدایی که افراد کاشت حلزون شده می‌شنوند، و یا شبیه‌سازی احساس در رباتها استفاده کرد.

مراجع:

۱- دنی استاینبرگ، در آمدی بر روان‌شناسی زبان، ترجمه دکتر ارسلان گلغام،

۱۳۸۱

- [2] Randolf R, Cornelius, Theoretical approaches to emotion, Vassar college. NY USA. (1999).
- [3] Darwin, the Expression of Emotion in Man and Animals, 1872.
- [4] Prof. W. Sendlmeier, Berlin database of emotional speech, <http://pascal.kgw.tu-berlin.de/emodb/>
- [5] Sofia Gustafson-Capková, Emotions in Speech: Tagset and Acoustic Correlates, Department of Linguistics, Stockholm University, autumn -2001.
- [6] Valery A. Petrushin, Emotion recognition in speech signals: experimental study, development and application, Center for Strategic Technology Research (CSTaR). ICSLP 2000.
- [7] Tom Johnstone and Klaus R. Scherer, The effects of emotions on voice quality, University of Geneva. 1995.
- [8] M. Kathleen Pichora-Fuller, Temporal Jitter Disrupts Speech Intelligibility: Simulations of Auditory Aging. School of Audiology and Speech Sciences, University of British Columbia, March 1999.
- [9] Smith, G. A. Voice analysis for the measurement of anxiety. British Journal of Medical Psychology, 50, 367-373. . 1977
- [10] Mahjani, Behzad, An instrumental study of prosodic features and intonation in Modern Farsi (Persian). MS thesis, University of Edinburgh. 2003.
- http://www.ling.ed.ac.uk/teaching/postgrad/mcsclp/archive/dissertations/2002-3/behzad_mahjani.pdf



ب

الف

شکل ۹- خروجی شبکه برای داده‌های تست الف) صدای خوشحال و ب) صدای ناراحت