

در این تحقیق به منظور تشخیص خودکار نارسایی‌های گفتاری، به تعریف و استخراج ده ویژگی از سیگنال صوتی واکه کشیده شده^۱ /a/ پرداختیم. این ویژگی‌ها شامل میانگین فرکانس پایه، بالاترین و پایین‌ترین مقادیر فرکانس پایه در طول سیگنال، انحراف معیار مقادیر فرکانس پایه، میزان نوسانات کوتاه‌مدت فرکانس پایه، انحراف معیار ضرایب دامنه، میزان اغتشاشات کوتاه‌مدت دامنه، فرکانس فرمانت اول، نسبت دامنه فرمانت اول به فرمانت دوم، و بعد فرکتال کوتاه‌مدت سیگنال می‌باشند. به منظور ارزیابی قدرت هر یک از ویژگی‌ها در تمایز دادن سیگنال‌های طبیعی و پاتولوژیک، نسبت پراکندگی بین کلاسی به مجموع پراکندگی‌های داخل کلاسها برای هر یک از ده ویژگی محاسبه گردید. نتایج بیانگر قابلیت بالای ویژگی بعد فرکتال سیگنال برای تمایز سیگنال‌های طبیعی از پاتولوژیک می‌باشد. در ادامه با استفاده از شبکه عصبی جلوسوی پرسپترون به طبقه‌بندی سیگنال‌ها با استفاده از سه ویژگی برتر پرداختیم. فرآیند طبقه‌بندی با استفاده از شبکه عصبی مستقل از جنسیت گوینده صورت گرفت. و به درصد صحت ۸۵/۰۹٪ برای داده غیرتعلیم رسیدیم.

: نارسایی‌های صوتی-پردازش صداهای پاتولوژیک-فرکانس پایه-فرمانت-بعد فرکتال سیگنال-شبکه عصبی

یکی از اهداف بسیار مهم بررسی کلینیکی گفتار، مطالعه کمی گفتار طبیعی و پاتولوژیک می‌باشد. گفتار پاتولوژیک نتیجه نارسایی‌هایی در کنترل اندام‌های تولید است. نارسایی‌های صوتی شامل انواع گوناگونی با منشاءهای فیزیکی یا عصب-عضله‌ای و یا صرفاً عصبی می‌باشند، این عوامل بصورت مستقیم کیفیت صدای تولیدی توسط بیمار را تحت تاثیر قرار می‌دهند. در حال حاضر ابزارهای تشخیصی نظیر استروپوسکوپ و ویدئویی، کیموگرافی ویدئویی و سیستم‌های نمایشگر حرکات تارهای صوتی در اختیار گفتاردرمانگرها قرار گرفته است. لاکن روش‌های ذکر شده تهاجمی، وقت‌گیر و گران می‌باشند [۱]. واکه‌های کشیده شده که با طول گام و سطح بلندی دلخواه سوژه تولید می‌شوند، از نظر کلینیکی برای تشخیص بسیاری از نارسایی‌های گفتاری مورد توجه هستند. بسیاری از محققان عنوان نموده‌اند که واکه‌های کشیده شده منعکس کننده بسیاری از نارسایی‌ها و نکات غیرطبیعی در گفتار مانند اغتشاشات فرکانسی سیکل به سیکل در واکه، و اغتشاشات دامنه یا قطع غیرارادی صدا یا لرزش صدا می‌باشند. این موارد بیانگر ضعف در کنترل فشار هوای تولیدی توسط ریه‌ها یا نارسایی‌های حنجره یا دیگر جهاز صوتی و یا حتی مشکلات عصبی هستند [۱، ۲ و ۳]. در حین تولید واکه، دهان و اندام تولیدی در یک وضعیت ثابت و پایدار قرار می‌گیرند و افرادی که دارای مشکل کنترل اندام تولیدی خود هستند، اغلب توانایی این کار را ندارند و اندام تولیدی آنها در حین تولید واکه با لرزش و حرکات کنترل نشده همراه است [۴]. این لرزشها بصورت نوسانات فشار هوای تولیدی خود را نمایان می‌سازد و اثر آن نیز روی سیگنال مشهود است. اغلب تحلیل آکوستیکی گفتار پاتولوژیک از طریق اندازه‌گیری طول سگمنتهای گفتاری و مسیر فرمانت‌ها و تحلیل نحوه توزیع فرکانس پایه در کوتاه‌مدت یا بلندمدت می‌باشد [۵]. ویژگی‌های نامنظم حاضر در سیگنال گفتار بیانگر دینامیک‌های تولید گفتار می‌باشد، این بی‌نظمی‌ها می‌توانند ناشی از نویز، یا یک فرآیند تصادفی، در نظر گرفته شوند. اما می‌توانند ناشی از آشوب هم

¹ Sustained Vowel

باشند یعنی بی‌نظمی‌هایی در نظر گرفته شوند که از یک سیستم غیرخطی قطعی ناشی شده باشد [۶]. در این تحقیق ضمن معرفی و مقایسه ویژگی‌های تمایزگر برای طبقه‌بندی صداهای طبیعی و پاتولوژیک با استفاده از شبکه عصبی به طبقه‌بندی سیگنالها توسط برخی از این ویژگی‌ها خواهیم پرداخت. مزایای این روش طبقه‌بندی در مقایسه با روش‌های دیگر در موارد زیر عنوان می‌گردد.

الف- این روش‌ها غیرتهاجمی بوده و اجرای آن برای درمانگر و بیمار بسیار ساده می‌باشد، تنها لازم است بیمار صدای /a/ را برای مدت چند ثانیه تولید نماید.

ب- با استفاده از این روش‌ها امکان ساخت یک سیستم تشخیص خودکار رایانه‌ای وجود دارد.

ج- تشخیص شیءگرا است چراکه بر مبنای مقادیر ویژگی‌های صوتی انجام می‌گیرد.

د- این روش برای ارزیابی روند پیشرفت درمان مثل جراحی، درمان‌های دارویی یا فرآیندهای بازتوانی، قابل استفاده می‌باشد. و می‌توان با استفاده از مقادیر ویژگی‌های صوتی از میزان پیشرفت روند درمان ارزیابی کمی بدست آورد.

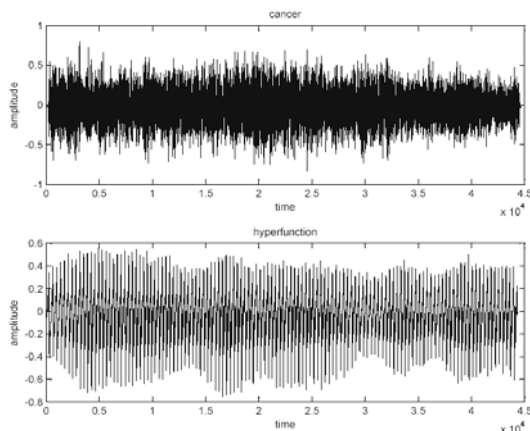
در این مقاله به معرفی ده ویژگی به منظور تمایز دادن سیگنال نرمال از پاتولوژیک پرداخته شده است. برخی از این ویژگی‌ها مبتنی بر اغتشاشات فرکانس پایه و دامنه سیگنال می‌باشند. مقادیر مطلق فرکانس پایه، فرمانت اول و نسبت دامنه فرمانت اول به دامنه فرمانت دوم نیز از ویژگی‌های مورد توجه بوده‌اند. این ویژگی‌ها بر مبنای نسبت پراکندگی بین کلاسی به مجموع پراکندگی‌های داخل کلاسی نمونه‌ها مقایسه شده‌اند. نتایج نشان از برتری ویژگی بعد فرکتال کوتاه‌مدت برای تفکیک نمونه‌ها دارد. در بخش ۲ به توضیح دادگان مورد استفاده برای انجام این تحقیق می‌پردازیم. در بخش ۳ ضمن معرفی ویژگی‌ها فرآیند استخراج آن‌ها را بررسی می‌نماییم. بخش ۴ به مقایسه ده ویژگی با یکدیگر می‌پردازد، در بخش ۵ ضمن ارائه یک طبقه‌بندی‌کننده مبتنی بر شبکه عصبی پرسپترون به بررسی نتایج طبقه‌بندی توسط سه ویژگی برتر، می‌پردازیم. و در انتها به بحث و نتیجه‌گیری خواهیم پرداخت.

در این تحقیق از دادگان برچسب‌خورده 'Disordered Voice Database Model 4337' تهیه شده در مرکز تحقیقات بیمارستان تخصصی گوش و چشم ماساچوست، آزمایشگاه گفتار و صدا، جهت تعلیم طبقه‌بندی‌کننده استفاده گردید. برای ضبط این داده از بیمار خواسته شده تا به مدت حدود ۳ الی ۵ ثانیه به تولید بدون وقفه واکه /a/ اقدام نماید. صدای تولید شده با فرکانس نمونه‌برداری ۴۴۱۰۰ هرتز و ۱۶ بیتی و بصورت تک‌کاناله ضبط می‌گردد و جهت پردازش در رایانه ذخیره می‌گردد. این دادگان شامل تعداد ۳۸۷ فایل صوتی واکه کشیده شده /a/ می‌باشد. بیمارانی که صدای آنها در این دادگان در اختیار است مبتلا به نارسایی‌های صوتی از نظر بیماری‌های جهاز صوتی و یا سیستم عصبی بوده‌اند. تعداد فایل‌های صوتی موجود در دادگان برای مردان و زنان در دو دسته طبیعی و پاتولوژیک در جدول ۱ ذکر شده است. برای طبقه‌بندی ۳/۴ دادگان جهت تعلیم و ۱/۴ آن برای آزمون در نظر گرفته شد. به این معنا که ۳/۴ از مردان پاتولوژیک، زنان پاتولوژیک، مردان طبیعی و زنان طبیعی برای تعلیم و ۱/۴ باقی‌مانده از هر یک از این چهار دسته برای آزمون در نظر گرفته شد.

جدول ۱- تعداد فایل‌های صوتی موجود در دادگان برای مردان و زنان در دو دسته طبیعی و پاتولوژیک

جنسیت	تعداد نمونه‌های پاتولوژیک	تعداد نمونه‌های طبیعی
مردان	۱۴۱	۲۲
زنان	۱۸۸	۳۶
مجموع	۳۲۹	۵۸

سیگنال صوتی واکه کشیده شده منبع بسیار ارزشمندی برای تشخیص انواع نارسایی‌های صوتی است. لرزش یا قطع غیرارادی صدا از نشانه‌های نارسایی‌های صوتی است. که ناشی از عملکرد نامناسب عضلات و جهاز صوتی دخیل در تولید سیگنال واکه می‌باشد. عوارض ناشی از نارسایی‌های صوتی بصورت نوسانات و اغتشاشات کوتاه مدت و بلندمدت فرکانس پایه و دامنه و یا انتقال نواحی پرنرزی در طیف فرکانسی سیگنال نمایان می‌شوند. برای نمونه شکل ۱ دو نمونه از سیگنال واکه $a/$ در دو بیمار را نشان می‌دهد. سیگنال بالایی مربوط به بیمار مبتلا به سرطان حنجره و سیگنال پایینی مربوط به بیمار مبتلا به Hyperfunction است.



شکل ۱- دو نمونه سیگنال پاتولوژیک واکه کشیده شده $a/$ سیگنال بالایی مربوط به بیمار مبتلا به سرطان حنجره و سیگنال پایینی مربوط به بیمار مبتلا به Hyperfunction است. اغتشاشات فرکانسی و نوسانات دامنه در هر دو سیگنال کاملاً مشهود است. چنین حالتی در سیگنال طبیعی وجود ندارد.

با توجه به آنچه ذکر شد، باید ویژگی‌ها را طوری تعریف نماییم که بتواند بین افراد طبیعی و بیمار تمایز ایجاد نماید. این ویژگی‌ها را طوری طراحی نمودیم که بتواند معیاری کمی از اغتشاشات فرکانسی و نوسانات دامنه‌ای در اختیار قرار دهد. ویژگی‌هایی که به عنوان مشخصه‌های تمایزگر از سیگنال استخراج گردید، ده ویژگی شامل میانگین فرکانس پایه، بالاترین مقدار فرکانس پایه در طول سیگنال، پایین‌ترین مقدار فرکانس پایه در طول سیگنال، انحراف معیار فرکانس پایه در طول سیگنال، پارامتر jitter یا ضریب اغتشاشات فرکانسی محلی و انحراف معیار ضریب دامنه در فریم‌ها و پارامتر shimmer ضریب اغتشاشات محلی دامنه، فرکانس فرمانت اول، نسبت دامنه فرمانت اول به دامنه فرمانت دوم و بعد فرکتال کوتاه مدت سیگنال بوده است.

برای استخراج فرکانس پایه می‌توان از روش‌های آنالیز گپستروم و یا روش آنالیز تابع خودهمبستگی استفاده نمود. طول فریم‌ها ۴۰ میلی ثانیه‌ای در نظر گرفته شد، و از این فریم‌ها فرکانس پایه با یکی از روش‌های ذکر شده به علاوه ضریب مربوط به دامنه استخراج گردید. مقادیر استخراج شده از تمام فریم‌ها در بردارهایی قرار می‌گیرند. سپس مقدار میانگین فرکانس پایه و انحراف معیار فرکانس‌های پایه و انحراف معیار ضریب دامنه، با توجه به فرمول‌های ۱، ۲ و ۳ محاسبه می‌شوند:

$$avgF_0 = \frac{1}{N} \sum_{i=1}^N F_0(i) \quad (1)$$

$$STD = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_0(i) - avgF_0)^2} \quad (2)$$

$$STDa = \sqrt{\frac{1}{N} \sum_{i=1}^N (A(i) - avgA)^2} \quad (3)$$

مقادیر اغتشاشات محلی فرکانس jitter و اغتشاشات محلی دامنه shimmer نیز بصورت زیر محاسبه می‌شوند [۱].

$$jitter = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{F_0(i+1) - F_0(i)}{\frac{1}{2}[F_0(i+1) + F_0(i)]} \times 100\% \quad (4)$$

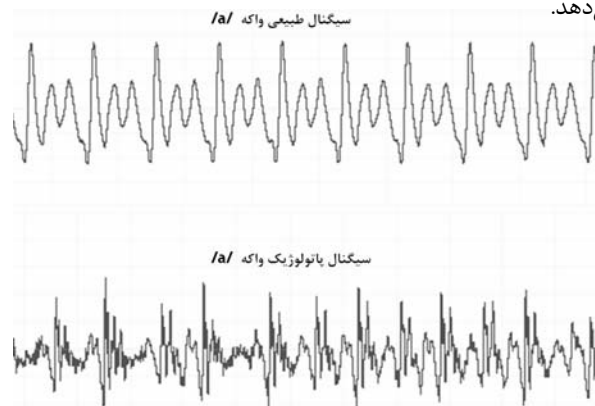
$$shimmer = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{A(i+1) - A(i)}{\frac{1}{2}[A(i+1) + A(i)]} \times 100\% \quad (5)$$

بسیاری از الگوهای موجود در طبیعت دارای خصوصیت خودهمبستگی می‌باشند. به این مفهوم که اگر در مقیاس‌های مختلف آن الگو را رویت نمایید به ساختار خودهمبستگی خواهید رسید. چنین الگوهایی با هندسه اقلیدسی قابل تحلیل نیستند و دارای ابعاد غیر صحیح می‌باشند. بسیاری از الگوها طبیعت دارای ابعاد اعشاری هستند. هندسه فرکتال ابزاری برای توصیف این الگوهاست. روشهای مختلفی برای محاسبه بعد فرکتال نظیر روش شمارش جعبه و روش مبتنی بر بعد همبستگی وجود دارد. یکی از روشهایی که برای اندازه‌گیری سریع بعد فرکتال از سیگنال‌های با ساختار فرکتال وجود دارد، روش پتروشن می‌باشد. فرمول زیر نحوه محاسبه را بر اساس این روش بیان می‌کند.

$$D = \frac{\log_{10} n}{\log_{10} n + \log_{10} \left(\frac{n}{n + 0.4 \times N_{\Delta}} \right)} \quad (6)$$

که n طول سیگنال و N_{Δ} تعداد تغییر علامت مشتق سیگنال است، که چون با سیگنال گسسته سر و کار داریم از تفاضل متوالی برای محاسبه مشتق استفاده می‌شود. این کمیت معیاری برای ارزیابی میزان نوسانات کوتاه‌مدت سیگنال مبتنی بر بعد فرکتال سیگنال در اختیار قرار می‌دهد. میانگین بعد فرکتال برای سیگنال‌های نرمال در حدود $1/0.4$ با انحراف معیار $0/0.2$ و میانگین برای سیگنال‌های پاتولوژیک $1/0.9$ با انحراف معیار $0/0.25$ محاسبه گردید.

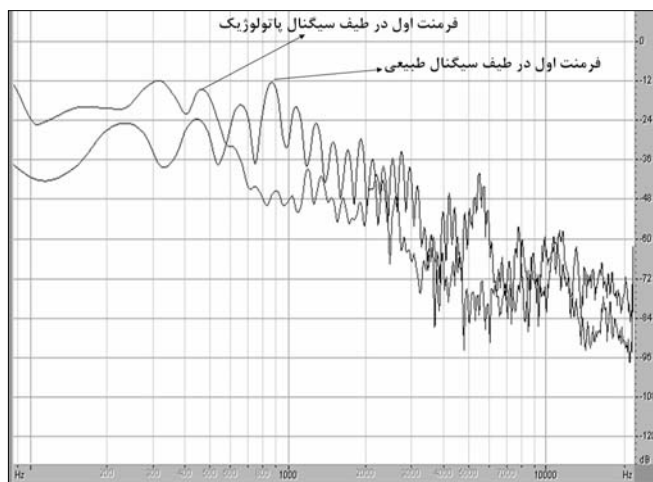
در شکل ۲ دو نمونه از سیگنال واکه $/a/$ در یک فرد طبیعی و یک فرد مبتلا به سرطان حنجره را می‌بینید. وجود نوسانات و تغییرجهت‌های متعدد در نمونه پاتولوژیک کاملاً مشهود است. بعد فرکتال معیاری کمی از میزان این نوسانات و تغییرجهت‌ها در اختیار قرار می‌دهد.



شکل ۲- نمونه سیگنال طبیعی و پاتولوژیک واکه کشیده شده $/a/$ همانطور که در سیگنال پاتولوژیک در قسمت پایین ملاحظه می‌شود. نوسانات کوتاه‌مدت دامنه در هر سیکل واکه ملاحظه می‌شود. این حالت لرزش و نوسان در سیگنال طبیعی وجود ندارد. سیگنال پاتولوژیک مربوط به بیمار مبتلا به سرطان حنجره می‌باشد.

نحوه قرارگیری زبان و جهاز صوتی در حین ادای واکه‌ها به شکلی است که ممانعت کمی برای عبور هوا ایجاد می‌شود. درعین حال وضعیت زبان در حین ادای هر یک از واکه‌ها خصوصیات خاص فرکانسی آنها را، با تغییر ویژگی‌های رزونانس حفره صوتی رقم می‌زند. هوایی که از حنجره وارد حلق می‌گردد با فرکانس پایه نوسان می‌نماید. فرکانس‌های تشدید اصلی حفره دهان در حین تولید واکه‌ها F_1 و F_2 به ترتیب اولین و دومین فرمانت نامیده می‌شوند. مقدار فرکانس فرمانت‌ها توسط نحوه قرارگیری زبان در حفره دهان تعیین می‌شوند. مقادیر فرمانت‌ها بیانگر کیفیت تولید واکه‌ها می‌باشند. F_1 بیشتر مربوط به فرکانس تشدید ناحیه خلفی حفره صوتی و F_2 وابسته به اندازه و شکل قسمت قدامی دهان می‌باشد. بر

اساس منابع معتبر، برای واکه $a/$ میانگین فرمنت اول ۷۵۰ هرتز و میانگین فرمنت دوم ۱۷۵۰ هرتز می‌باشد [۷]. چنانچه فرد قادر به کنترل جهاز صوتی خود به نحو مطلوبی نباشد، این مقادیر فرکانسی بدلیل وجود عواملی نظیر لرزش در حفره صوتی و زبان تغییر می‌کند. ضمن آنکه نواحی انباشت انرژی نیز در طیف فرکانسی تغییر خواهد نمود [۸]. در شکل ۳ طیف فرکانسی دو سیگنال طبیعی و پاتولوژیک واکه کشیده شده ملاحظه می‌شود. فرکانس فرمانت اول برای سیگنال طبیعی ۸۵۰ هرتز است که به ۴۹۰ هرتز در سیگنال پاتولوژیک منتقل شده است. مقدار میانگین فرکانس فرمانت اول برای سیگنال‌های طبیعی ۷۶۰ هرتز با انحراف معیار ۹۰ هرتز و برای سیگنال‌های پاتولوژیک ۶۵۰ هرتز با انحراف معیار ۱۶۰ هرتز محاسبه گردید. همانطور که در شکل دیده می‌شود نواحی انباشت انرژی نیز در این دو طیف تا حد زیادی متفاوت است. نسبت دامنه فرمانت اول به دامنه فرمانت دوم برای سیگنال طبیعی کمتر از سیگنال پاتولوژیک می‌باشد. بر این اساس نسبت دامنه فرمانت اول به دامنه فرمانت دوم را بعنوان ویژگی سوم تمایزگر انتخاب نمودیم. مقدار میانگین ویژگی سوم برای سیگنال‌های طبیعی ۱/۱۰ با انحراف معیار ۰/۰۵ و برای سیگنال‌های پاتولوژیک ۱/۱۳ با انحراف معیار ۰/۰۶ هرتز محاسبه شد.



شکل ۳- طیف فرکانسی دو سیگنال طبیعی و پاتولوژیک در یک شکل دیده می‌شود. همانطور که دیده می‌شود فرمانت اول در سیگنال طبیعی در ۸۵۰ هرتز است ولی در سیگنال پاتولوژیک به ۴۹۰ هرتز منتقل شده است. به عبارتی نواحی ذاتی مربوط به فرمانت‌های واکه در جای اصلی خود قرار ندارد.

به منظور محاسبه فرمانت اول ابتدا مقادیر فرکانسی کمتر از ۴۸۰ هرتز و بالاتر از ۸۵۰ هرتز را توسط دو فیلتر سری باترورث حذف نمودیم. سپس فرکانس غالب را از سری زمانی سیگنال فیلتر شده توسط رابطه زیر پیدا نمودیم.

$$F_1 = \frac{0.5 \times F_s \times ZC}{N} \quad (7)$$

که در آن F_s فرکانس نمونه برداری سیگنال اصلی، ZC تعداد عبور از صفر سیگنال فیلتر شده و N تعداد نمونه‌های سیگنال می‌باشد. برای محاسبه فرمانت دوم توسط یک فیلتر بالاگذر باترورث مقادیر فرکانسی بالای ۱۰۰۰ هرتز را از سیگنال اصلی حذف نمودیم و سپس با استفاده از فرمول (۷) فرکانس فرمنت دوم از سری زمانی سیگنال فیلتر شده استخراج گردید. در جدول ۲ ویژگی‌های مورد استفاده به همراه توصیف کوتاه آنها ملاحظه می‌شوند.

جدول ۲-توصیف ده مشخصه استخراج شده از سیگنال

شماره	تعریف مشخصه	علامت اختصاری	گروه مشخصه
۱	میانگین فرکانس پایه در طول سیگنال	$avg(F_0)$	فرکانس پایه
۲	انحراف معیار فرکانس پایه در طول سیگنال	$STD(F_0)$	

اغتشاشات محلی فرکانسی	$jitter$	ضریب اغتشاش فرکانسی بین فریم‌های ۴۰ میلی‌ثانیه‌ای	۳
	F_{0Low}	پایین‌ترین فرکانس پایه دیده شده در تمام فریم‌های ۴۰ میلی‌ثانیه‌ای	۴
	F_{0High}	بالا‌ترین فرکانس پایه دیده شده در تمام فریم‌های ۴۰ میلی‌ثانیه‌ای	۵
اغتشاش دامنه	$STD(a)$	انحراف معیار دامنه قله‌های موجود در فریم‌های ۴۰ میلی‌ثانیه‌ای	۶
اغتشاشات محلی دامنه	$shimmer$	ضریب اغتشاشات دامنه بین فریم‌های ۴۰ میلی‌ثانیه‌ای	۷
فرمانت اول	F_1	میانگین فرمانت اول در طول سیگنال	۸
توزیع انرژی در طیف سیگنال	$A(F_1)/A(F_2)$	نسبت دامنه فرمانت اول به دامنه فرمانت دوم در طول سیگنال	۹
بعد فرکتال	D	بعد فرکتال کوتاه مدت سیگنال	۱۰

همانطور که ملاحظه می‌شود، ویژگی‌های مشابه در گروه‌های یکسانی قرار داده شده‌اند. و ترتیب آنها در جدول با ترتیبی که در متن مقاله تشریح شده‌اند متفاوت است. شماره ویژگی‌ها در نمودار شکل ۴ مطابق با جدول ۲ می‌باشد.

به منظور ارزیابی و مقایسه میزان کارایی ویژگی‌ها در تفکیک بین کلاس‌ها، پراکندگی هر یک از ویژگی‌ها را در داخل کلاس‌های طبیعی و پاتولوژیک با توجه به فرمول‌های (۸) و (۹) و همینطور پراکندگی بین کلاسی را با استفاده از فرمول‌های (۱۰) و (۱۱) محاسبه نمودیم.

$$\hat{\mu}_i = \frac{1}{N_i} \sum_{\beta \in w_i} \beta \quad i=1,2 \quad (8)$$

$$w_i = \sum_{\beta \in w_i} (\beta - \hat{\mu}_i)(\beta - \hat{\mu}_i)' \quad i=1,2 \quad (9)$$

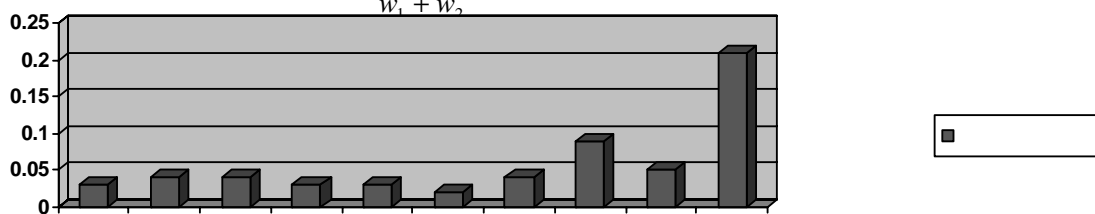
$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^2 N_i \hat{\mu}_i \quad (10)$$

$$N = N_1 + N_2$$

$$B = \sum_{i=1}^2 N_i (\hat{\mu}_i - \hat{\mu})(\hat{\mu}_i - \hat{\mu})' \quad (11)$$

این فرمول‌ها در حالت تعمیم یافته و برای بردارهای ویژگی چندبعدی نوشته شده است. β در حالت یک‌بعدی اسکالر و در حالت چندبعدی بردار است. $\hat{\mu}_i$ تخمینی از میانگین بردار ویژگی در کلاس i ام و w_i پراکندگی داخل کلاس i ام است. به عبارتی w_i تخمینی از کوواریانس کلاس i ام در فضای ویژگی‌ها است. $\hat{\mu}$ تخمین میانگین بردار ویژگی بین تمام کلاس‌هاست. و B ماتریس پراکندگی بین کلاسی در فضای ویژگی‌هاست. این ماتریس بیانگر میزان پراکندگی میانگین کلاس‌ها می‌باشد. بعنوان معیاری از قدرت تفکیک ویژگی‌ها بین دو کلاس، با توجه به فرمول (۱۲)، نسبت پراکندگی بین کلاسی را به مجموع پراکندگی ویژگی‌ها در داخل کلاس‌ها محاسبه نمودیم [۹]. در شکل (۴) مقادیر این معیار برای هر یک از ده ویژگی، ذکر گردیده‌اند.

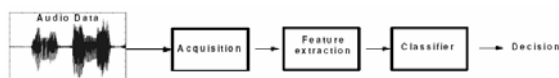
$$J = \frac{B}{w_1 + w_2} \quad (12)$$



شکل ۴- قدرت تفکیک‌پذیری هر یک از ویژگی‌ها بر اساس معیار نسبت پراکندگی بین کلاسی به پراکندگی داخل کلاسی

نتایج بیانگر کارایی بیشتر ویژگی بعد فرکتال سیگنال یا ویژگی ۱۰ام به منظور تفکیک کلاس‌هاست. پس از ویژگی ۱۰ام یا بعد فرکتال، ویژگی فرکانس فرمنت اول یا ویژگی ۸ام که ویژگی بارز در ادای واکه‌هاست، بیشترین توانایی را در تفکیک کلاس طبیعی از پاتولوژیک داراست. نسبت دامنه فرمنت اول به دامنه فرمنت دوم و یا ویژگی ۹ام سومین ویژگی برتر از نظر قدرت تفکیک‌کنندگی می‌باشد. برای طبقه‌بندی کلاسها از سه ویژگی برتر استفاده نمودیم. به‌منظور ارزیابی قدرت تمایز ترکیب سه ویژگی باهم، پارامتر J را برای ترکیب آنها محاسبه نمودیم. در این حالت β یک بردار 3×1 خواهد بود. در حالت ترکیبی مقدار J به $0/24$ رسید. استفاده از ویژگی‌های بیشتر بهبود قابل توجهی در مقدار پارامتر J نمی‌دهد. لذا به منظور کاهش پیچیدگی محاسبات و زمان محاسبات استفاده از سه ویژگی برتر بطور همزمان برای انجام طبقه‌بندی کافی بنظر می‌رسد.

برای طبقه‌بندی سیگنال صوتی واکه کشیده شده $a/$ به دو طبقه طبیعی و پاتولوژیک همانطور که در شکل ۵ آمده بعد از ضبط سیگنال و استخراج ویژگی‌های تمایزگر، به یک طبقه‌بندی‌کننده غیرخطی احتیاج داریم. چنانچه ویژگی‌های استخراج شده با یک طبقه‌بندی‌کننده مناسب همراه باشند، نتایج خوبی قابل انتظار خواهد بود.



شکل ۵- نمودار بلوکی فرآیند طبقه‌بندی صدای پاتولوژیک

برای طبقه‌بندی از یک شبکه عصبی جلوسوی پرسپترون با یک لایه پنهان استفاده گردید. لایه ورودی شامل سه نرون ورودی مربوط به سه ویژگی برتر استخراج شده از هر سیگنال صوتی بوده‌است و لایه میانی شامل ۱۰ نرون و لایه خروجی شامل دو نرون در نظر گرفته شد. در حین تعلیم برای سیگنال طبیعی نرون اول لایه خروجی معادل ۱ و دیگری معادل ۰ در نظر گرفته می‌شود. و در صورتی که سیگنال از طبقه پاتولوژیک باشد نرون دوم معادل ۱ و نرون اول ۰ در نظر گرفته می‌شود. در مرحله آزمون نرونی که سطح فعالیت بیشتری در خروجی دارد بعنوان نرون برنده انتخاب می‌شود. اگر نرون اول حائز این شرط باشد سیگنال از نوع طبیعی تشخیص داده می‌شود و در غیر اینصورت سیگنال از نوع پاتولوژیک در نظر گرفته می‌شود. با توجه به اینکه ورودی‌ها دارای رنج دینامیک وسیعی بودند، در حین تعلیم و آزمون باید ورودی‌ها را نرمالیزه نماییم [۱۰]. این کار برای تعلیم و همگرایی شبکه نیز ضروری می‌باشد. برای این منظور هر ویژگی را با توجه به مقدار میانگین و انحراف معیار آن ویژگی در داده تعلیم با استفاده از فرمول (۱۳) نرمالیزه نمودیم، اینکار هم در داده تعلیم و هم در داده آزمون انجام پذیرفت. بنابراین در مرحله آزمون بردار ویژگی‌ها با میانگین و انحراف معیار ویژگی‌ها در داده تعلیم نرمالیزه می‌گردد.

$$\beta_n = \frac{\beta - \mu_\beta}{\sigma_\beta} \quad (13)$$

β ویژگی، μ_β میانگین ویژگی در داده تعلیم و σ_β انحراف معیار ویژگی در داده تعلیم می‌باشد. نتایج طبقه‌بندی برای داده تعلیم و آزمون در جدول زیر ذکر شده است.

جدول ۳- درصد صحت طبقه‌بندی با استفاده از هر سه ویژگی توسط شبکه عصبی پرسپترون با یک لایه پنهان

درصد صحت طبقه‌بندی با استفاده از شبکه عصبی	
نتیجه طبقه‌بندی برای داده تعلیم	٪۹۱/۶۱
نتیجه طبقه‌بندی برای داده آزمون	٪۸۵/۰۹

دستیابی به این درصد صحت طبقه‌بندی، بیشتر از آنکه به عملکرد طبقه‌بندی‌کننده وابسته باشد، مدیون ویژگی‌های استفاده شده است. هریک از ویژگی‌های بعد فرکتال، فرکانس فرمنت اول و نسبت دامنه فرمنت اول به فرمنت دوم، توانایی ایجاد تمایز بین سیگنال‌های طبیعی و پاتولوژیک را دارند. بعد فرکتال سیگنال در نمونه‌های پاتولوژیک، به دلیل وجود نوسانات کوتاه‌مدت در دامنه سیگنال افزایش می‌یابد. این نوسانات ناشی از وجود لرزش در اندام‌های تولید حین ادای واکه در بیماران است. ویژگی بعدی یا فرکانس فرمنت اول در اغلب بیماران حین ادای واکه نسبت به میزان مناسب آن کمتر است، که نتیجه عدم قرارگیری صحیح زبان و جهاز صوتی می‌باشد. نسبت دامنه فرمنت‌ها نیز در بیماران به دلیل تغییر نواحی انباشت انرژی در طیف فرکانسی سیگنال، با افزایش نسبت به میزان طبیعی آن همراه است. همانگونه که در نتایج در جدول ۳ مشخص است، با در اختیار داشتن ویژگی‌های با قدرت تمایز مناسب می‌توان با یک طبقه‌بندی‌کننده نسبتاً ساده به نتایج خوبی دست یافت.

سیگنال واکه کشیده شده منبع ارزشمندی برای تشخیص نارسایی‌های صوتی است. در این تحقیق ده ویژگی به منظور طبقه‌بندی سیگنال واکه /a/ به کلاس‌های طبیعی و پاتولوژیک، استخراج گردید. که از این تعداد سه ویژگی برتر با توجه به معیار قدرت تفکیک‌کنندگی کلاسها برای طبقه‌بندی انتخاب شدند؛ و بصورت ترکیبی برای این منظور مورد استفاده قرار گرفتند. لرزش‌های نامنظم اندام تولیدی در بیماران گفتاری بصورت نوسانات کوتاه‌مدت دامنه که حالتی شبه‌نویز دارد، بروز می‌کند [۶]. بعد فرکتال سیگنال پاتولوژیک بدلیل همین نوسانات، بیشتر از بعد فرکتال سیگنال طبیعی است. سیگنال واکه حاوی فرکانس‌های شاخص در نواحی خاصی از باند فرکانسی است. که به اصطلاح فرمنت نام دارند. مشخصات فرکانس و دامنه فرمنت‌ها به نحوه قرارگیری زبان و اندام تولیدی وابسته می‌باشند. لذا در بیماران گفتاری که توانایی کنترل دقیق اندام تولیدی خود را ندارند، مشخصات فرمنت‌ها چه از نظر فرکانس و چه از نظر دامنه متفاوت خواهد بود و نواحی انباشت انرژی از آنچه در مورد سیگنال واکه /a/ باید باشد، تفاوت دارد. براین مبنا ویژگی‌های فرمنت اول و نسبت دامنه فرمنت‌ها می‌توانند در تمایز سیگنال‌های طبیعی و پاتولوژیک مفید باشند. با توجه به مقایسه‌های صورت گرفته قدرت تفکیک بعد فرکتال سیگنال واکه بیشتر از ویژگی‌های دیگر است. این نکته می‌تواند نشان از قابلیت بالای روش‌های غیرخطی مبتنی بر فرضیات آشوبگونه بودن سیگنال گفتار، که یک سیگنال تولیدشده توسط موجود زنده است، باشد. باید توجه کرد ویژگی بعد فرکتال از سری زمانی سیگنال به روشی غیرخطی استخراج شده است. تلفیق سه ویژگی قدرت تفکیک‌کنندگی بیشتری ایجاد می‌نماید، لذا با استفاده از هر سه ویژگی و بهره‌گیری از یک شبکه عصبی پرسپترون با یک لایه پنهان، به طبقه‌بندی سیگنال‌ها پراخته شد، و درصد صحت بالای ۸۵/۰۹٪ برای داده آزمون ایجاد گردید.

از مسؤولین محترم دانشکده ادبیات و علوم انسانی دانشگاه تهران، به ویژه دکتر محمود بیجن‌خان سرپرست آزمایشگاه زبانشناسی و مهندس وحید صادقی، بدلیل در اختیار گذاشتن دادگان لازم برای انجام این تحقیق، بسیار سپاسگزاریم.

[1] M.N. Vieira, F.R. McInnis, M.A. Jack, "On the influence of laryngeal pathologies on acoustic and electroglottographic jitter measures," *J. Acoustical Society of America*, Vol. 111, No. 2, pp. 1045-55, February, 2002.

[2] C. H. Espinosa, M. F. Redondo, "Diagnosis of Vocal and Voice Disorders by the Speech Signal," *Proc. Int. conf. on neural networks*. Como, Italy, Vol. 4, p. 4253, July 24-27, 2000.

- [3] A. Fourcin, "Patterns, Structure and Measurement in Voice Pathology," *IEEE Trans. On Biomedical Engineering*, Vol. 45, No. 3, pp. 300-312, March, 2000.
- [4] K. Umapathy, S. Krishnan, V. Parsa and D. Jamieson, "Time-Frequency Modeling and Classification of Pathological Voices," *Proc. 2nd Joint EMBS/BMES Conf.* Houston, Texas, U.S.A, pp. 116-117, October 23-26, 2002.

[۵] ب. وزیرنژاد، م.ح. مرادی، «تشخیص خودکار نارسایی‌های صوتی از طریق پردازش صدا»، در

۱۳۸۴.

- [6] A. P. Accardo, E. Mumolo, "An Algorithm for the Automatic Differentiation between the Speech of Normals and Patients with Friedreich's Ataxia based on the Short-time Fractal Dimension," *J. of Computers in Biology and Medicine.*, Vol. 28, pp. 75-89, September 22 1998.
- [7] X. Huang, A. Acero, H. W. Hon, *Spoken Language Processing*, Prentice-Hall, 2000.
- [8] B. R. Gerratt, J. Kreimann and N. A. Barroso, "Comparing Internal and External Standards in Voice Quality Judgments," *J. of Speech and Hearing Research*, Vol. 36, pp. 14-20, 1993.
- [9] A. Cohen, *Biomedical Signal Processing*, CRC Press, 2000.
- [10] A. Webb, *Statistical Pattern Recognition*. Arnold publishers, 2003.