

ارزیابی سه ویژگی مبتنی بر بعد فرکتال کوتاه مدت و مشخصات فرمت‌ها برای طبقه‌بندی خودکار گفتار طبیعی و پاتولوژیک

بهرام وزیرنژاد^۱، محمدحسن مرادی^۲
دانشگاه صنعتی امیرکبیر، دانشکده مهندسی پزشکی
{bvazirnezhad, mhmoradi}@cic.aut.ac.ir

چکیده

در این مقاله به منظور تشخیص نارسایی‌های گفتاری و با استفاده از طبقه‌بندی خودکار صداهای طبیعی و پاتولوژیک، به استخراج ویژگی‌هایی مرتبط با هدف از سیگنال صوتی واکه کشیده شده^۳ پرداخته‌ایم. دادگان مورد استفاده در این تحقیق شامل ۳۲۹ سیگنال صدای بیمار و ۵۸ سیگنال صدای طبیعی بود. ویژگی‌های استخراج شده شامل بعد فرکتال^۴ سیگنال، فرکانس فرمنت اول واکه و نسبت دامنه فرمنت اول به دوم واکه بوده‌است. بعد فرکتال سیگنال بیانگر میزان کمی نوسانات کوتاه مدت در داخل سیگنال است و مشخصات فرمنت‌ها حاوی اطلاعات تولیدی واکه می‌باشند. به منظور ارزیابی و مقایسه کارایی ویژگی‌ها برای طبقه‌بندی، نسبت پراکندگی بین کلاسی به مجموع پراکندگی‌های داخل کلاسها برای هر یک از سه ویژگی و برای ترکیب آنها محاسبه گردید. نتایج بیانگر قابلیت بالای ویژگی بعد فرکتال سیگنال برای تمایز سیگنال‌های طبیعی از پاتولوژیک می‌باشد. در ادامه با استفاده از یک شبکه عصبی جلوسوی پرسپترون به طبقه‌بندی سیگنال‌ها با توجه به این سه ویژگی پرداختیم و به درصد صحت ۸۵/۰۹٪ برای داده غیرتعلیم رسیدیم.

واژه‌های کلیدی: نارسایی‌های صوتی-پردازش صداهای پاتولوژیک-بعد فرکتال سیگنال-فرمنت واکه-شبکه عصبی

۱- مقدمه

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 استروبوکوپیی ویدئویی^۵، کیموگرافی ویدئویی^۶ و سیستم‌های نمایشگر حرکات تارهای صوتی در اختیار گفتاردرمانگرها قرار گرفته است. لاکن روش‌های ذکر شده تهاجمی، وقت گیر 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

^۱ دانشجوی دکتری مهندسی پزشکی، گرایش بیوالکتریک
^۲ دانشیار دانشکده مهندسی پزشکی، گرایش بیوالکتریک

^۳ Sustained Vowel
^۴ Fractal Dimension
^۵ Videostroboscopy
^۶ Videokymography

ریه‌ها یا نارسایی‌های حنجره یا دیگر جهاز صوتی و یا حتی مشکلات عصبی هستند [۱، ۲ و ۳]. در این مقاله ضمن معرفی و مقایسه سه ویژگی تمایزگر برای طبقه‌بندی صداهای طبیعی و پاتولوژیک با استفاده از ابزار شبکه عصبی به طبقه‌بندی سیگنالها توسط این سه ویژگی خواهیم پرداخت. مزایای این روش طبقه‌بندی در مقایسه با روش‌های دیگر در موارد زیر عنوان می‌گردد.

الف- این روش 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 روش امکان ساخت یک سیستم تشخیص خودکار رایانه‌ای وجود دارد.

ج- تشخیص شیء گرا^۷ 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 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 فرمنت‌ها 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 باشند یعنی بی‌نظمی‌هایی در نظر گرفته شوند که از یک سیستم غیرخطی ناشی شده باشد [۶]. در این مقاله به معرفی یک ویژگی مبتنی بر بعد هندسه فرکتال به منظور تمایز دادن سیگنال نرمال از پاتولوژیک پرداخته شده است. این ویژگی با دو ویژگی فرمنت اول و نسبت دامنه فرمنت اول به دوم بر مبنای نسبت پراکندگی بین کلاسی به مجموع پراکندگی‌های داخل کلاسی نمونه‌ها مقایسه شده است. 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 سه ویژگی با یکدیگر می‌پردازد، در بخش ۵ ضمن ارائه یک طبقه‌بندی‌کننده مبتنی بر شبکه عصبی پرسپترون به بررسی نتایج طبقه‌بندی توسط این سه ویژگی، می‌پردازیم. و در انتها به بحث و نتیجه‌گیری خواهیم پرداخت.

۲- دادگان

در این تحقیق از دادگان برجسب‌خورده 'Disordered Voice Database Model 4337' تهیه شده در مرکز تحقیقات بیمارستان تخصصی گوش و چشم ماساچوست، آزمایشگاه گفتار و صدا، جهت تعلیم طبقه‌بندی‌کننده‌ها استفاده گردید. برای ضبط داده از بیمار خواسته می‌شود 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.

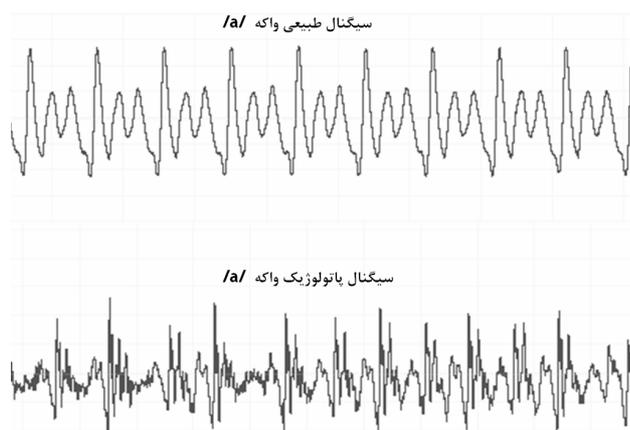
⁷ Objective

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

۳- فرآیند استخراج ویژگی‌ها

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 کوتاه‌مدت دامنه و یا انتقال نواحی پارانرژی در طیف فرکانسی سیگنال نمایان می‌شوند. در این بخش به توصیف سه ویژگی تمایزگر بعد فرکتال سیگنال، فرکانس فرمنت اول و نسبت دامنه فرمنت اول به دامنه فرمنت دوم خواهیم پرداخت و نحوه محاسبه آنها را بیان می‌نمائیم. در شکل ۱ 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. در این تحقیق سعی نمودیم با توجه به این نکته، با استخراج بعد فرکتال سیگنال به یک ویژگی تمایزگر بین سیگنال‌های طبیعی و پاتولوژیک دست یابیم. بعد فرکتال معیاری کمی از میزان این نوسانات و تغییرجهت‌ها در اختیار قرار می‌دهد.



شکل ۱- 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.

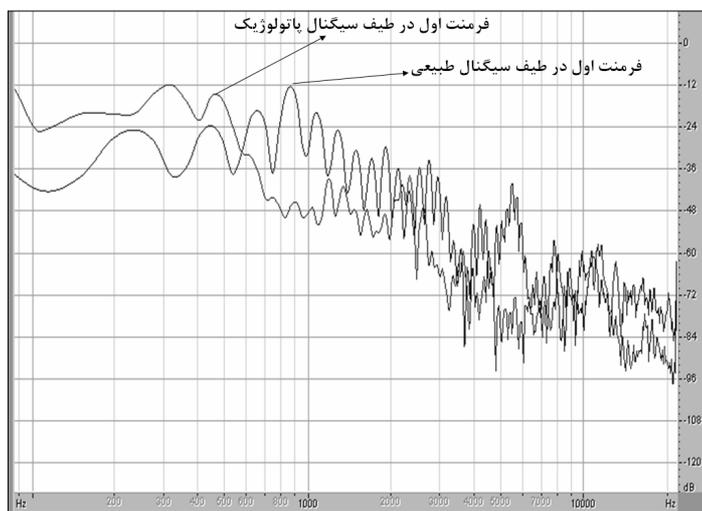
بسیاری از الگوهای موجود در طبیعت دارای خصوصیت خودهمانندی می‌باشند. به این مفهوم که اگر در مقیاس‌های مختلف آن الگو را رویت نمایید به ساختار خودهمانندی خواهید رسید. چنین الگوهایی با هندسه اقلیدسی قابل تحلیل نیستند و دارای ابعاد صحیح می‌باشند. ولی بسیاری از الگوها طبیعت دارای ابعاد اعشاری هستند. هندسه فرکتال 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.

$$D = \frac{\log_{10} n}{\log_{10} n + \log_{10} \left(\frac{n}{n + 0.4 \times N_{\Delta}} \right)} \quad (1)$$

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 فرکانس سیگنال در اختیار قرار می‌دهد. میانگین بعد فرکانس برای سیگنال‌های نرمال در حدود $1/0.4$ با انحراف معیار 0.2 و میانگین برای سیگنال‌های پاتولوژیک $1/0.9$ با انحراف معیار 0.25 محاسبه گردید.

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 هریک 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 فرمنت نامیده می‌شوند. مقدار فرکانس فرمنت‌ها توسط نحوه قرارگیری زبان در حفره دهان تعیین می‌شوند. مقادیر فرمنت‌ها 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 750 هرتز و میانگین فرمنت دوم 1750 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 ۲ طیف فرکانسی دو سیگنال طبیعی و پاتولوژیک واکه کشیده شده ملاحظه می‌شود. فرکانس فرمنت اول برای سیگنال طبیعی 850 هرتز است که به 490 هرتز در سیگنال پاتولوژیک منتقل شده است. مقدار میانگین فرکانس فرمنت اول برای سیگنال‌های طبیعی 760 هرتز با انحراف معیار 90 هرتز و برای سیگنال‌های پاتولوژیک 650 هرتز با انحراف معیار 160 هرتز محاسبه گردید. همانطور که در شکل دیده می‌شود نواحی انباشت انرژی نیز در این دو طیف تا حد زیادی متفاوت است. نسبت دامنه فرمنت اول به دامنه فرمنت دوم برای سیگنال طبیعی کمتر از سیگنال پاتولوژیک می‌باشد. بر این اساس نسبت دامنه فرمنت اول به دامنه فرمنت دوم را بعنوان ویژگی سوم تمایزگر انتخاب نمودیم. مقدار میانگین ویژگی سوم برای سیگنال‌های طبیعی $1/10$ با انحراف معیار 0.05 و برای سیگنال‌های پاتولوژیک $1/13$ با انحراف معیار 0.06 هرتز محاسبه شد.



شکل ۲- طیف فرکانسی دو سیگنال طبیعی و پاتولوژیک در یک شکل دیده می‌شود. همانطور که دیده می‌شود فرمنت اول در سیگنال طبیعی در 850 هرتز است ولی در سیگنال پاتولوژیک به 490 هرتز منتقل شده است. به عبارتی نواحی ذاتی مربوط به فرمنت‌های واکه در جای اصلی خود قرار ندارد.

به منظور محاسبه فرمنت اول ابتدا مقادیر فرکانسی کمتر از 480 هرتز و بالاتر از 850 هرتز را توسط دو فیلتر سری باترورث حذف نمودیم. سپس فرکانس غالب را از سری زمانی سیگنال فیلتر شده توسط رابطه زیر پیدا نمودیم.

$$F_1 = \frac{0.5 \times F_s \times ZC}{N} \quad (2)$$

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 فرمنت دوم توسط یک فیلتر بالاگذر باترورث مقادیر فرکانسی بالای 1000 هرتز را از

سیگنال اصلی حذف نمودیم و سپس با استفاده از فرمول (۲) فرکانس فرمنت دوم از سری زمانی سیگنال فیلتر شده استخراج گردید.

۴- مقایسه قدرت تمایز ویژگی‌ها

به منظور ارزیابی و مقایسه میزان کارایی ویژگی‌ها با یکدیگر، پراکندگی هر یک از ویژگی‌ها را در داخل کلاس‌های طبیعی و پاتولوژیک با توجه به فرمول‌های (۳ و ۴) و هم‌منطور پراکندگی بین کلاسی را با استفاده از فرمول‌های (۵ و ۶) محاسبه نمودیم.

$$\hat{\mu}_i = \frac{1}{N_i} \sum_{\beta \in w_i} \beta \quad i=1,2 \quad (3)$$

$$w_i = \sum_{\beta \in w_i} (\beta - \hat{\mu}_i)(\beta - \hat{\mu}_i)' \quad i=1,2 \quad (4)$$

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^2 N_i \hat{\mu}_i \quad (5)$$

$$N = N_1 + N_2$$

$$B = \sum_{i=1}^2 N_i (\hat{\mu}_i - \hat{\mu})(\hat{\mu}_i - \hat{\mu})' \quad (6)$$

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 (۷)، نسبت پراکندگی بین کلاسی را به مجموع پراکندگی ویژگی‌ها در داخل کلاس‌ها محاسبه نمودیم [۹]. در جدول (۲) مقادیر این معیار برای هر یک از سه ویژگی، و برای هر سه ویژگی با هم، ذکر گردیده اند.

$$J = \frac{B}{w_1 + w_2} \quad (7)$$

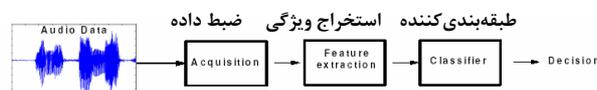
جدول ۲- قدرت تفکیک‌پذیری هر یک از ویژگی‌ها بر اساس معیار نسبت پراکندگی بین کلاسی به پراکندگی داخل کلاسی

عنوان ویژگی	معیار قدرت تفکیک‌کنندگی کلاس‌ها J
بعد فرکتال سیگنال	۰/۲۱
فرکانس فرمنت اول	۰/۰۹
نسبت دامنه فرمنت اول به فرمنت دوم	۰/۰۵
هر سه ویژگی	۰/۲۴

نتایج بیانگر کارایی بیشتر ویژگی بعد فرکتال سیگنال به‌منظور تفکیک کلاس‌هاست. پس از ویژگی بعد فرکتال، ویژگی فرکانس فرمنت اول که ویژگی بارز در ادای واکه‌هاست، بیشترین توانایی را در تفکیک کلاس طبیعی از پاتولوژیک داراست. به‌منظور ارزیابی قدرت تمایز مجموع هر سه ویژگی باهم، پارامتر J را برای ترکیب آنها محاسبه نمودیم. در این حالت β یک بردار 1×3 خواهد بود. در حالت ترکیبی مقدار J به ۰/۲۴ رسید. بنابراین بهتر است برای طبقه‌بندی از هر سه ویژگی استفاده نمود.

۵- طبقه‌بندی با استفاده از شبکه عصبی

برای طبقه‌بندی سیگنال صوتی واکه کشیده شده $a/$ به دو طبقه طبیعی و پاتولوژیک همانطور که در شکل ۳ 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 ۳- 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 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 (۸) 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.

$$\beta_n = \frac{\beta - \mu_\beta}{\sigma_\beta} \quad (8)$$

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 طبقه‌بندی کننده نسبتاً ساده به نتایج خوبی دست یافت.

۶- بحث و نتیجه‌گیری

سیگنال واکه کشیده شده منبع ارزشمندی برای تشخیص نارسایی‌های صوتی است. در این تحقیق سه ویژگی به منظور طبقه‌بندی سیگنال واکه /a/ به کلاس‌های طبیعی و پاتولوژیک، استخراج گردید. 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 دو ویژگی 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.

سپاسگذاری

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

مراجع

- [1] M.N. Vieira, F.R. McInnes, 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. Umopathy, 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 **مجموعه مقالات سیزدهمین کنفرانس مهندسی برق ایران**، 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.