

جداسازی صدای گریه نوزاد از محیط: استفاده از مقیاس خطی در ضرایب کپسترال

ادیب صدی^۱، محمدرضا یزدچی^۲ و مهدیس مصدق فر^۳

^۱دانشگاه اصفهان، adib.sadri@eng.ui.ac.ir

^۲دانشگاه اصفهان، yazdchi@eng.ui.ac.ir

^۳دانشگاه آزاد اسلامی واحد خمینی شهر، m_mosadeghfar@yahoo.com

چکیده - تشخیص گریه کردن نوزاد از صدای او می‌تواند سریع‌ترین راه ممکن برای اعلام نیاز به توجه اطرافیان نوزاد باشد. این مطالعه بر اساس داده‌های جمع‌آوری شده از صدای گریه نوزادان ایرانی و سایر اصوات محیطی صورت گرفته است. سه ویژگی فرکانسی شامل فرکانس پایه صوت، ضرایب کپسترال فرکانس مل و ضرایب کپسترال فرکانس خطی از سیگنال‌ها استخراج شده و توسط طبقه بند ماشین بردار پشتیبان جداسازی شده‌اند. استفاده از فرکانس با مقیاس خطی در جداسازی سیگنال‌های حاوی اطلاعات در فرکانس‌های بالا مفید باشد. نرخ صحت $99/4\%$ با استفاده از ضرایب کپسترال فرکانس خطی به دست آمد که در مقایسه با نرخ صحت $99/1\%$ با استفاده از ضرایب کپسترال فرکانس مل بهبود نشان می‌دهد. کلیدواژه- گریه نوزاد، فرکانس پایه، ضرایب کپسترال، فرکانس مل، ماشین بردار پشتیبان

۱- مقدمه

گریه کردن نوزاد به‌عنوان تنها سلاح او در برابر ناخوشایندی‌هایی مانند گرسنگی، درد و عفونت که برایش اتفاق می‌افتد به‌منزله هشدار است تا اطرافیان خود را برای کمک به خود فراخواند، این‌رو عدم پاسخگویی اطرافیان به این علامت هشدار می‌تواند موجب آسیب رسیدن به نوزاد و والدین او شود. سیگنال صوت گریه حاوی اطلاعات زیادی است که در صورت تحلیل درست می‌توان به پیام‌های صادر شده از مغز نوزاد دست‌یافت.

تشخیص و تمایز سیگنال گریه اولین مأموریت سیستم تشخیص و تحلیل خودکار گریه نوزاد است که دیاگرام آن در شکل ۱ به نمایش درآمده است.

در این مقاله، مرحله تشخیص و جداسازی سیگنال گریه نوزاد از سایر اصوات محیطی مورد توجه قرار گرفته است. مطالعات مشابهی در سال‌های اخیر روی این موضوع انجام شده است اما عمده مطالعات بر

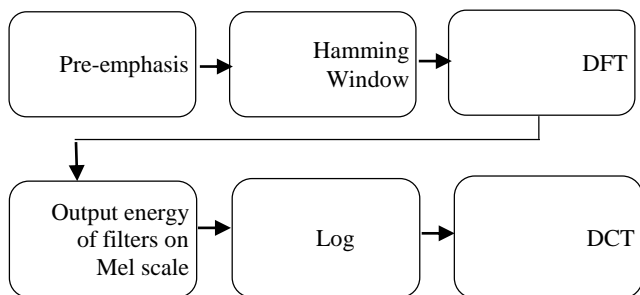
روی بازشناسی علت گریه تمرکز داشته است [۱]. در سال ۲۰۰۲ میکلسون و همکارانش ۱۷۲ نوزاد سالم را مورد بررسی قرار دادند و فرکانس پایه^۱ گریه نوزاد را بین ۴۰۰ تا ۶۰۰ هرتز اعلام کردند [۲]. کوهن و همکارش در سال ۲۰۱۲ مطالعه‌ای با موضوع تشخیص و تحلیل گریه نوزاد ارائه دادند و با استخراج فرکانس پایه، ضرایب کپسترال فرکانس مل^۲ (MFCC) و انرژی در هر قطعه یک‌ثانیه‌ای و استفاده از الگوریتم نزدیک‌ترین همسایه^۳ (KNN) نرخ 100% را برای جداسازی سیگنال‌های گریه با SNR بالا و اصوات محیط گزارش دادند [۳]. هم‌چنین در سال ۲۰۱۲ کیا و همکارانش سیستم تشخیص و هشدار گریه نوزاد را با استفاده از طبقه بند شبکه عصبی معرفی کردند که پس از تشخیص سیگنال گریه هشدار را از طریق شماره‌گیری تلفن اعلام می‌کرد [۴]. در این مقاله، مطالعه مشابهی برای ارزیابی کارایی ضرایب کپسترال فرکانس خطی^۴ برای جداسازی سیگنال‌های گریه و اصوات محیطی در مقایسه با ضرایب MFCC و فرکانس پایه ارائه می‌شود. در بخش دوم توضیح اجمالی پیرامون

^۳ K-Nearest Neighbor

^۴ Linear Frequency Cepstral Coefficients

^۱ F_0 ; Fundamental Frequency

^۲ Mel Frequency Cepstral Coefficients



شکل ۲: مراحل محاسبه ضرایب MFCC

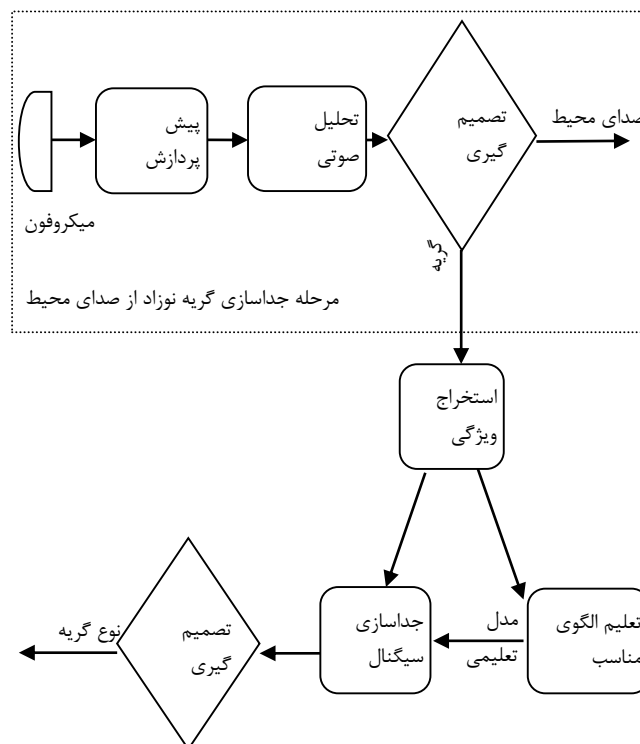
فرکانسی مفیدی است که در طیف فرکانسی وجود دارد که بقیه مؤلفه‌های سیگنال ضرایب صحیحی از این فرکانس هستند [۵]. یکی از متداول‌ترین تکنیک‌های به دست آوردن فرکانس پایه استفاده از تابع خودهمبستگی است. مقدار حداقل شیفتم انجام شده که منجر به ایجاد بیشترین شباهت با سیگنال اصلی شده است، زمانی است که با معکوس کردن آن فرکانس پایه به دست می‌آید. بیان ریاضی این روش استفاده از تابع خودهمبستگی است [۶].

$$R(\tau) = \sum_{\langle N \rangle} s(n)s(n+\tau), \tau = 0, 1, 2, \dots, N \quad (1)$$

که در آن $s(n)$ سیگنال اصلی، N طول سیگنال و τ میزان شیفتم سیگنال است. با معکوس کردن مقداری که باعث به دست آمدن بیشترین مقدار R شده است، فرکانس پایه F_0 به دست می‌آید. بسیاری از گروه‌های تحقیقاتی فرکانس پایه را برای گریه‌ی نوزاد بین ۳۰۰ تا ۷۵۰ هرتز گزارش کرده‌اند [۷-۱۰].

۲-۲ ضرایب کپسترال فرکانس مل (MFCC)

محاسبه‌ی ضریب کپستروم فرکانس مل یک روش رایج و معروف است که در زمینه‌ی استخراج ویژگی سیگنال گریه‌ی نوزاد به‌صورت مکرر استفاده شده است [۱۱-۱۳]. کپستروم فرکانس مل، طیف توان یک صوت را با استفاده از تبدیل کسینوسی خطی لگاریتم طیف توان در مقیاس مل نشان می‌دهد. مقیاس مل به‌صورت رابطه‌ی (۲) تعریف می‌شود [۱۳].



شکل ۱: سیستم تشخیص و تحلیل خودکار گریه نوزاد

استخراج ویژگی‌های فرکانسی به‌ویژه ضرایب LFCC و کاربرد آن ذکر می‌شود. در بخش سوم بلوک تصمیم‌گیری یا همان طبقه‌بندی گریه و صدای محیطی که با استفاده از الگوریتم ماشین بردار پشتیبان^۵ (SVM) صورت گرفته است توضیح داده می‌شود. در بخش چهارم نتایج پیاده‌سازی روی دادگان مورد استفاده شرح داده خواهد شد و در بخش پایانی به تفسیر نتایج پرداخته خواهد شد.

۲-۱ ویژگی‌های فرکانسی

۱-۲ فرکانس پایه (F_0)

فرکانسی که در آن تارهای صوتی در هنگام تولید صدا به ارتعاش درمی‌آیند، فرکانس پایه نامیده می‌شوند. فرکانس پایه کم‌ترین مؤلفه‌ی

^۵ Support Vector Machine

می‌دهد q کوفرانس یا فرکانس در حوزه کپسترال، n شیفت زمانی، $E_{l,t}(b)$ انرژی فیلتر b ام در فریم t ام از سگمنت l ام سیگنال است و B تعداد فیلترهای فیلتربانک و T تعداد فریم‌ها در هر سگمنت است.

۳- طبقه‌بندی

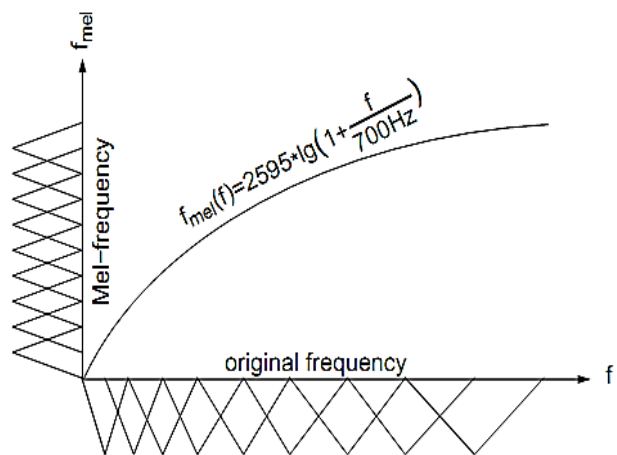
پس از استخراج بردارهای ویژگی مربوط به هر قطعه از سیگنال-های گریه و غیر آن نوبت به طبقه‌بندی آن‌ها و تعلیم سیستم جداسازی که بتواند دو نوع سیگنال گریه و صدای محیط را تشخیص دهد می‌رسد. در این مقاله از الگوریتم ماشین بردار پشتیبان (SVM) استفاده شده است که توسط واپنیک در سال ۱۹۹۵ ارائه شد و امروزه کاربرد زیادی در جداسازی داده‌هایی که درهم‌آمیختگی غیرخطی دارند دارد [۱۴]. این طبقه بند بر اساس کمینه کردن ریسک خطای مرزی بین دوطبقه عمل می‌کند. شیب این مرز فقط تابع تعدادی از بردارهای ورودی است که روی حاشیه‌ی مرز دوطبقه قرار می‌گیرند و بردارهای پشتیبان مرز نامیده می‌شوند. این طبقه بند باینری است ولیکن هر دو نوع خطی و غیرخطی آن موجود است. نوع غیرخطی

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right) \quad (2)$$

که در آن f مقدار فرکانس و M مقدار مل متناظر آن است. ضرایب MFCC را می‌توان تبدیل کسینوسی لگاریتم انرژی حاصل از اعمال فیلتربانک مل بر طیف سیگنال پنجره گذاری شده تعریف کرد. مراحل محاسبه ضرایب MFCC در شکل ۲ نشان داده شده است. مزیت مهم MFCC، مقاومت در برابر نویز خطاهای تخمین طیفی تحت شرایط مختلف است [۱۲].

۳-۲ ضرایب کپسترال فرکانس خطی (LFCC)

ضرایب کپسترال فرکانس مل به‌طور وسیعی در مباحث پردازش صوت و گفتار به‌خصوص در بازشناسی گفتار مورد استفاده قرار می‌گیرد، با این وجود بر اساس نظریه‌های موجود مبنی بر تأثیر ساختار مجرای صوتی در تولید مؤلفه‌های فرکانس بالای صوت، برخی ویژگی‌های پر کاربرد در تشخیص و تمایز اصوات در فرکانس‌های بالا پدیدار می‌شوند. از این رو استفاده از یک مقیاس خطی فرکانسی برای تشکیل فیلترهای پردازشی به‌جای استفاده از مقیاس مل می‌تواند نقش این مؤلفه‌های فرکانس بالا را بهتر نمایان سازد [۱۱، ۱۲]. شکل ۳ مقایسه بین فیلتربانک با مقیاس‌های مل و خطی را همراه با رابطه نگاشت فرکانس خطی به مل نشان می‌دهد.



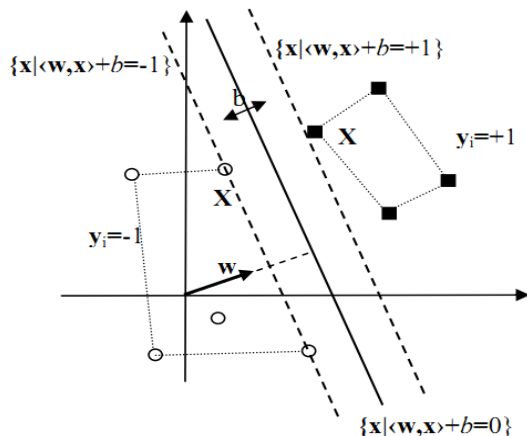
شکل ۳: مقایسه فیلتربانک با مقیاس‌های مل و خطی و نگاشت بین آن‌ها [۱۲]

نحوه محاسبه ضرایب LFCC نیز از مراحل محاسبه ضرایب MFCC پیروی می‌کند؛ بنابراین رابطه (۳) مبین نحوه محاسبه هر دو نوع ضرایب است.

$$C_l(q, n) = \sum_{t=0}^{T-1} \sum_{b=0}^{B-1} \log(E_{l,t}(b)) \times \cos\left(\frac{(2b+1)q\pi}{2B}\right) \times \cos\left(\frac{(2t+1)n\pi}{2T}\right), 0 \leq q < B, 0 \leq n < T \quad (3)$$

در این رابطه که یک مختصات دوبعدی فرکانس-زمان را تشکیل

رابطه (۶) یک شرط برای رابطه (۷) است زیرا برای محاسبه‌ی فاصله‌ی نقاط تا صفحه‌ی جداساز، سمت چپ نامساوی (۷) باید به نرم بردار متعامد که از رابطه‌ی (۶) محاسبه می‌شود تقسیم گردد؛ بنابراین با کمینه کردن نرم اقلیدسی بردار عمود بر صفحه‌ی جداساز رابطه‌ی (۷) بیشینه می‌گردد و بهترین جداسازی صورت می‌گیرد. توضیحات بیش‌تر از الگوریتم را در [۱۵] می‌توان یافت.



شکل ۴: صفحه جداساز بهینه با بردار عمود w و مقدار ثابت b . [۱۵]

۴- پیاده‌سازی

۴-۱- دادگان مورد استفاده

دادگان استفاده‌شده در این مطالعه شامل سیگنال‌های گریه ۴۹ نوزاد ۱ تا ۵ روزه که در سه بیمارستان شهر اصفهان ثبت شده است و هم‌چنین سیگنال‌های ثبت شده از صداهای محیطی شامل رفت‌وآمد وسایط نقلیه، صداهای ممکن الوقوع در منازل مانند صدای شیر آب و غیره است. تعداد سیگنال‌های گریه و اصوات محیطی یکسان و برابر ۵۷۵ ثانیه بود. سیگنال‌ها با کیفیت عالی MP3 و نرخ ۱۹۲ kbps Sony-PX333 به وسیله یک دستگاه ضبط خبرنگاری حرفه‌ای ضبط شده‌اند.

SVM از هسته‌های غیرخطی مخصوصاً توابع RBF^۶ و چندجمله‌ای^۷ استفاده می‌کند. اگر به‌طور ساده به مسئله‌ی جداسازی دو بعد پردازیم، وظیفه الگوریتم پیدا کردن پارامترهای مربوط به یک خط جداساز است که به بهترین شکل داده‌های مربوط به کلاس مثبت را از منفی جدا می‌سازد. این خط یا صفحه را با یک تابع تصمیم‌گیری مانند رابطه (۴) مشخص می‌سازیم که در آن w بردار وزن‌هایی است که بر صفحه‌ی جداساز عمود است و b یک کمیت اسکالر است که حاشیه‌ی جداسازی را مشخص می‌کند. تابع کرنل است که داده‌ها را به ابعاد بالاتری منتقل می‌کند تا جداسازی خطی ممکن شود و عملگر $\langle \cdot, \cdot \rangle$ نشان‌دهنده‌ی ضرب داخلی دو بردار است. تابع $\text{sgn}(x)$ به ازای محتوای مثبت خروجی $+1$ و برای محتوای منفی خروجی -1 را حاصل می‌کند. در صورتی که بردار w نرمال باشد عبارت $\langle w, \varphi(x) \rangle$ نشان‌دهنده‌ی طول در جهت بردار w است، از این رو بردار w را ابتدا با تقسیم بر $\|w\|$ نرمالیزه می‌کنیم.

$$f(x) = \text{sgn}(\langle w, \varphi(x) \rangle + b) \quad (۴)$$

شکل ۴ مسئله‌ی ساده جداسازی بین دو کلاس که با دایره‌ها و مربع‌ها نشان داده شده‌اند به نمایش درمی‌آورد. همان‌طور که در شکل ۴ قابل مشاهده است ممکن است تعداد صفحات زیادی قابلیت جداسازی بین دوطبقه را داشته باشند اما الگوریتم به دنبال یافتن جداساز بهینه است به طوری که بیش‌ترین حاشیه را میان دوطبقه ایجاد کند. این شرط به‌عنوان یک پایه و حقیقت در مسئله‌ی جداسازی، در تضاد با حقیقتی دیگر قرار می‌گیرد که با افزایش حاشیه‌ی جداسازی در طبقه‌بندی میان کلاس‌ها کاهش می‌یابد. این دو ویژگی متضاد، مسئله‌ی طبقه‌بندی را به یک مصالحه تبدیل می‌کند که با رابطه‌ی (۵) بیان می‌شود [۱۵].

$$\max_{w \in H, b \in R} \min \{ \|x - x_i\| \mid x \in H, \langle w, \varphi(x) \rangle + b = 0, i = 1, \dots \} \quad (۵)$$

صفحه‌ی جداساز بهینه را می‌بایست با حل دو تابع هزینه پیدا کرد:

$$\min \tau(w) = \frac{1}{2} \|w\|^2 \quad (۶)$$

$$y_i (\langle w, \varphi(x) \rangle + b) \geq 1 \quad (۷)$$

^۷ Polynomial Functions

^۶ Radial Basis Function

۲-۴ پیش پردازش

استاندارد فرکانس پایه را بین دو گروه سیگنال گریه و غیر آن مقایسه می کند.

جدول ۱: میانگین و پراکندگی فرکانس پایه در دو گروه گریه و غیر آن

پراکندگی استاندارد (Hz)	میانگین F0 (Hz)	تعداد نمونه	نوع سیگنال
۲۴/۰۶	۴۶۴/۶۶	۵۷۵	گریه
۷۵/۱۲	۳۰۲/۸۷	۵۷۵	غیر گریه

جدول ۲: میزان صحت طبقه بندی را با استفاده از ویژگی فرکانس پایه و طبقه بند SVM نشان می دهد.

جدول ۲: میزان صحت جداسازی بین دو گروه بر اساس فرکانس پایه

درصد صحت	تشخیص اشتباه	تشخیص صحیح	نوع سیگنال
۹۸/۲۶	۱۰	۵۶۵	گریه
۹۸/۷۸	۷	۵۶۸	غیر گریه
۹۸/۵۲	۱۷	۱۱۳۳	مجموع

استخراج ویژگی های MFCC و LFCC نیازمند پنجره گذاری برای هر قطعه یک ثانیه ای است. پنجره همینگ^{۱۱} بهترین انتخاب مطابق تجربه در این مطالعه و مطالعات مشابه است [۴, ۱۲, ۱۶]. طول هر پنجره برابر ۱۰۰ میلی ثانیه و بدون همپوشانی با یکدیگر در نظر گرفته شد. تعداد فیلترهای فیلتربانک اعمال شده بر هر پنجره برابر ۲۰ و طول بردار ویژگی های MFCC و LFCC برابر ۱۶ تنظیم شد. جدول ۳ و ۴ به ترتیب میزان صحت طبقه بندی را با استفاده از ویژگی های MFCC و LFCC نشان می دهند.

مرحله ی پیش پردازش جهت آماده سازی سیگنال برای ورود به کدهای پردازشی است. در این مرحله سیگنال های گریه توسط نرم افزار تخصصی پردازش صوت DAE^۸ از فرآیند حذف سکوت عبور داده می شوند که طی آن قطعاتی که صدای گریه به گوش نمی رسد حذف می گردد. این کار بر اساس الگوریتم ASD^۹ که حذف سکوت تطبیقی را با استفاده از فیلترهای وافی انجام می دهد انجام شده است. پس از این مرحله سیگنال به قطعات یک ثانیه ای تقسیم بندی می شوند. انتخاب طول یک ثانیه برای هر قطعه بر اساس تجربه و نتایج مطالعات مرتبط صورت گرفته است [۱۱, ۱۶, ۱۷]. هر کدام از این قطعات از فیلتر پیش تأکید $H(z) = 1 - 0.95z^{-1}$ عبور داده می شوند.

۳-۴ نتایج

تمامی مراحل استخراج هر سه نوع ویژگی مورد نظر در این مطالعه و طبقه بندی با روش SVM با استفاده از نرم افزار MATLAB انجام شده است [۱۸]. لازم به ذکر است در طبقه بندی با الگوریتم SVM بهترین نتایج طبقه بندی با استفاده از کرنل چند جمله ای به دست آمد، بنابراین تمامی نتایج در این گزارش بر اساس انتخاب این کرنل ارائه می شود. همچنین از روش ارزیابی مقطعی چند پوشه ای^{۱۰} با تعداد ۱۰ پوشه برای اعتبارسنجی صحت نتایج دادگان اختصاص یافته برای تعلیم و تست استفاده شد. به این ترتیب، در هر ارزیابی دادگان به ۱۰ قسمت تصادفی تقسیم شده و هر بار یک قسمت به عنوان دادگان تست و مابقی برای تعلیم استفاده می گردند. این عمل ۱۰ بار تکرار شده و هر بار یک قسمت برای تست تخصیص داده می شود.

برای استخراج فرکانس پایه، سیگنال ها فریم بندی شده و سپس از هر فریم ۱۰۰ میلی ثانیه ای یک فرکانس پیچ استخراج می شود و در یک نمونه از گریه ۱۰ عدد فرکانس پیچ حاصل می شود و در نهایت برای اینکه به هر نمونه ی یک ثانیه ای یک فرکانس پیچ نسبت دهیم، با توجه به تغییرات کم در ۱۰ عدد فرکانس پیچ روی این ۱۰ عدد عمل متوسط گیری انجام می شود. جدول ۱ میانگین و پراکندگی

^{۱۰} K-fold Cross Validation

^{۱۱} Hamming Window

^۸ Digital Audio Editor

^۹ Adaptive Silence Detector

for Recognition Based on Biometrics, ed: Springer, 2010, pp. 3-18.

- [۲] K. Michelsson, K. Eklund, P. Leppänen, and H. Lyytinen, "Cry characteristics of 172 healthy 1-to 7-day-old infants," *Folia phoniatrica et logopaedica*, vol. 54, pp. 190-200, 2002.
- [۳] R. Cohen and Y. Lavner, "Infant cry analysis and detection," in *Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Conventon of*, 2012, pp. 1-5.
- [۴] M. Kia, S. Kia, N. Davoudi, and R. Biniastian, "A detection system of infant cry using fuzzy classification including dialing alarm calls function," in *Innovative Computing Technology (INTECH), 2012 Second International Conference on*, 2012, pp. 224-229.
- [۵] I. McLoughlin, *Applied speech and audio processing: with Matlab examples*: Cambridge University Press, 20۰۹.
- [۶] D. Lederman, "Estimation of Infants' Cry Fundamental Frequency using a Modified SIFT algorithm," *arXiv preprint arXiv:1009.2796*, 2010.
- [۷] R. P. Daga and A. M. Panditrao, "Acoustical analysis of pain cries in neonates: fundamental frequency," *IJCA Spec Issue Electron, Inf Commun, Eng ICEICE (3)*, pp. 18-21, 2011.
- [۸] D. Bard, P. Runefors, and E. Arnbjornsson, "Newborn pain cry analysis based on pitch frequency tracking," *Journal of the Acoustical Society of America*, vol. 123, p. 3072, 2008.
- [۹] L. L. LaGasse, A. R. Neal, and B. M. Lester, "Assessment of infant cry: acoustic cry analysis and parental perception," *Mental retardation and developmental disabilities research reviews*, vol. 11, pp. 83-93, 2005.
- [۱۰] H. R. Gilbert and M. P. Robb, "Vocal fundamental frequency characteristics of infant hunger cries: birth to 12 months," *International journal of pediatric otorhinolaryngology*, vol. 34, pp. 237-243, 1996.
- [۱۱] M. J. Kim, Y. Kim, S. Hong, and H. Kim, "ROBUST detection of infant crying in adverse environments using weighted segmental two-dimensional linear frequency cepstral coefficients," in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, 2013, pp. 1-4.
- [۱۲] M. Hariharan, L. S. Chee, and S. Yaacob, "Analysis of infant cry through weighted linear prediction cepstral coefficients and probabilistic neural network," *Journal of medical systems*, vol. 36, pp. 1309-1315, 2012.
- [۱۳] C. A. R. García and C. Reyes Garcia, "Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neural networks," in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, 2003, pp. 3140-3145.
- [۱۴] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*: John Wiley & Sons, 2012.
- [۱۵] C. M. Bishop, *Pattern recognition and machine learning*: springer, 2006.
- [۱۶] M. Hariharan, R. Sindhu, and S. Yaacob, "Normal and hypoacoustic infant cry signal classification using time-frequency analysis and general regression neural network," *Computer methods and programs in biomedicine*, vol. 108, pp. 559-569, 2012.
- [۱۷] J. Saraswathy, M. Hariharan, T. Nadarajaw, W. Khairunizam, and S. Yaacob, "Optimal selection of mother wavelet for accurate infant cry classification," *Australasian Physical & Engineering Sciences in Medicine*, vol. 37, pp. 439-456, 2014.
- [۱۸] MATLAB, 8.0 ed: The MathWorks Inc., 2012.

جدول ۳: میزان صحت جداسازی بین دو گروه بر اساس ضرایب MFCC

نوع سیگنال	تشخیص صحیح	تشخیص اشتباه	درصد صحت
گریه	۵۷۱	۴	۹۹/۳۰
غیر گریه	۵۶۹	۶	۹۸/۹۵
مجموع	۱۱۴۰	۱۰	۹۹/۱۳

جدول ۴: میزان صحت جداسازی بین دو گروه بر اساس ضرایب LFCC

نوع سیگنال	تشخیص صحیح	تشخیص اشتباه	درصد صحت
گریه	۵۷۳	۲	۹۹/۶۵
غیر گریه	۵۷۰	۵	۹۹/۱۳
مجموع	۱۱۴۳	۷	۹۹/۳۹

۵- نتیجه گیری

در این پژوهش به مبحث تشخیص و تمایز سیگنال گریه نوزاد از سایر اصوات محیط پیرامون او پرداخته شد که مرحله‌ی ابتدایی طراحی و ساخت سیستم خودکار تشخیص و هشدار علت گریه نوزاد است. با توجه به چندوجهی و پویا بودن گریه نوزاد لزوم دقت و صحت هرچه بیشتر در تمامی جزئیات این سیستم انکارناپذیر است. در این مقاله استفاده از ضرایب LFCC به عنوان ایده‌ای نو برای تمایز سیگنال - های گریه و غیر آن مورد بررسی قرار گرفت که در مقایسه با ویژگی‌های فرکانس پایه و ضرایب MFCC که بیشترین کاربرد را نه تنها در مبحث جداسازی گریه و غیر آن بلکه در موضوع طبقه‌بندی نوع گریه نیز کاربردهای فراوانی دارند کارایی و نرخ صحت ۹۹/۴٪ را در مقایسه با ۹۹/۱٪ و ۹۸/۵٪ به ترتیب مربوط به ضرایب MFCC و فرکانس پایه حاصل کرد. از جمله مشکلات استفاده از ضرایب MFCC هزینه محاسباتی بالا در تشکیل و اعمال فیلتربانک روی سیگنال است که با بالا رفتن حجم داده‌ها مورد استفاده به شدت خودنمایی می‌کند. این محاسبات در استفاده از ضرایب LFCC به علت یکنواخت بودن مقیاس فیلترها به طور چشمگیری کاهش پیدا می‌کند. با این حال نمی‌توان قدرت ضرایب MFCC را در مباحث بازشناسی صوت و گفتار نادیده گرفت. در مطالعات آتی می‌توان از این ویژگی یا ترکیبی از آن و ضرایب MFCC استفاده نمود و مبحث جداسازی وفقی گریه نوزاد را در پیش گرفت.

مراجع

- [۱] C. A. Reyes-Garcia, O. F. Reyes-Galaviz, S. D. Cano-Ortiz, D. I. Escobedo-Becerro, R. Zatarain, and L. Barrón-Estrada, "Soft Computing Approaches to the Problem of Infant Cry Classification with Diagnostic Purposes," in *Soft Computing*