علوم و مهندسی کامپیوتر
نشریه علمی پژوهشی انجمن کامپیوتر ایران
مجلد ۴ شماره ۳ (الف)، ۱۳۸۵
صفحات ۱-۸
مقاله عادی

ارایه یک روش مقاوم به نویز جهت تشخیص نواحی سکوت در سیگنال گفتار*

خدیجه آقاجانی[۱] محمد تقی منظوری شلمانی[۲،۱]

[۱]دانشکده مهندسی کامپیوتر, دانشگاه صنعتی شریف، تهران، ایران
[۲]پژوهشکده علوم کامپیوتر, پژوهشگاه دانشهای بنیادی، تهران، ایران

چکیده

تشخیص نواحی سکوت از نواحی غیر سکوت دارای کاربرد های فراوانی در شاخه های مختلف پردازش گفتار است که در این میان می توان به زمینه های تشخیص گفتار، فشرده سازی اطلاعات گفتار، تخمین و حذف نویز و غیره اشاره کرد. تاکنون روش های مختلفی برای انجام عمل جداسازی پیشنهاد شده است. در این مقاله پس از بررسی روش های موجود و رایج، روش جدیدی برای تشخیص نواحی سکوت از نواحی غیر سکوت با استفاده ازتبدیل موجک ارائه می شود. دلیل انتخاب ابزار موجک قابلیت بالای آن در تفکیک انرژی سیگنال در باند های مختلف و دلخواه است. نشان داده شده است که نتایج حاصل از این روش از بسیاری جهات بر روش های قبلی برتری دارد.

کلمات کلیدی: تشخیص نواحی سکوت، تبدیل موجک، فیلتر های OS

۱-مقدمه

سیگنال گفتار معمولا از دو ناحیه سکوت و غیر سکوت تشکیل می شود. در سیگنال گفتار حاوی نویز ناحیه سکوت، نویزی است. در کل، اصوات صدادار دارای انرژی بیشتری نسبت به اصوات بی صدا هستند. این در حالی است که اصوات بی صدا بیشتر شبیه نویزند و همین امر تشخیص نواحی غیر سکوت از نواحی سکوت, که اختصاراً VAD[1] نامیده می شود را در سیگنال گفتار آغشته به نویز مشکل تر می کند. تشخیص ناحیه سکوت در سیگنال گفتار از اهمیت بسزایی برخوردار است. سیستمهای VAD در کاربردهایی نظیر تشخیص گفتار، فشرده سازی اطلاعات گفتار، تخمین و حذف نویز و غیره مورد استفاده قرار می گیرند. در تشخیص گفتار باید بطور دقیق نقاط شروع و پایان بیان کلمات را مشخص نماییم. این کار با بهره گیری از یک VAD مناسب انجام پذیر می باشد. در سیستمهای انتقال گفتار, با اختصاص دادن بیت های کمتر به نواحی سکوت، قادر به فشرده سازی اطلاعات خواهیم بود و در الگوریتم های بهسازی گفتار با تشخیص نواحی سکوت در سیگنال گفتار نویزی می توانیم به تخمین صحیح تری از نویز در سیگنال برسیم[۱۳]. الگوریتم های تشخیص سکوت برای جداسازی و مشخص نمودن نواحی سکوت و صحبت, از برخی خصوصیات برای تمایز بین این دو ناحیه استفاده می کنند که در ادامه به شرح آنها پرداخته می شود.

این مقاله در۷ قسمت به شرح زیر ارائه شده است. روش های تشخیص گفتار از سکوت و کار های انجام شده در بخش ۲ مورد بررسی قرار گرفته اند.

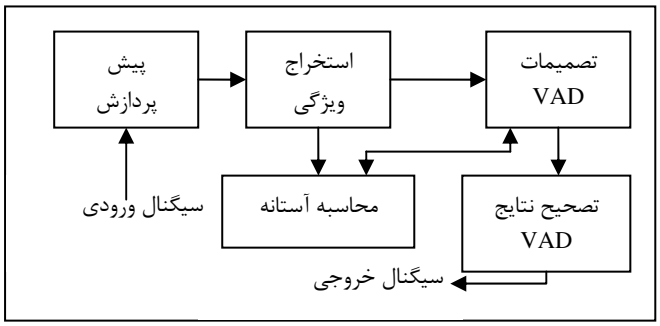
در بخش ۳ بطور خلاصه تبدیل موجک و در بخش ۴ فیلتر های OS مرور می شوند. راه حل پیشنهادی در بخش ۵ و ارزیابی کارایی آن در بخش ۶ آمده است. بخش ۷ این مقاله نیز به نتیجه گیری اختصاص یافته است.

۲- روش های تشخیص گفتار از سکوت

بطور معمول یک تشخیص دهنده نواحی سکوت در سیگنال گفتار دارای ساختار ارائه شده در شکل ۱ می باشد. همانگونه که در این شکل دیده می شود, تشخیص نواحی سکوت در سیگنال گفتار از مراحل اصلی زیر تشکیل شده است:

- استخراج ویژگی: در این مرحله پارامتر های مورد نیاز از فریم مربوطه استخراج می شود. بطور طبیعی پارامتر هایی انتخاب می شوند که فاکتور خوبی برای تمایز بین این دو ناحیه باشند.

- اعمال مقدار آستانه: در این مرحله مقدار آستانه ای بر اساس پارامترهای استخراج شده از مرحله قبل، برای تعیین هویت فریم ارائه می گردد. این مقدار می تواند ثابت و یا بطور وقفی باشد.



شکل ۱- دیاگرام یک VAD ساده

یکی از مشکلات عمده در پیاده سازی VAD ها تعیین مقدار آستانه می باشد. در بیشتر الگوریتم ها این مقدار از روی چند فریم اولیه که نویز فرض می شوند محاسبه می شود.

لازم به ذکراست که در برخی روش های پیاده سازی VAD، بعد از فاز استخراج ویژگی، برای تشخیص از روش های مبتنی بر منطق فازی [۲۳] و یا از روش های مبتنی بر مدل های آماری نظیر HMM[2] استفاده می شود [۱۷] و[۱۸]. در [۱۸] با کمک مدل مارکوف ۲ حالته (گفتار و سکوت) و در [۱۷] با کمک مدل مارکوف ۳ حالته (سکوت قبل از گفتار، گفتار و سکوت بعد از گفتار) و در نظر گرفتن مدل های گوسی مناسب برای هر حالت، عملیات تشخیص را انجام می دهند.

از پارامتر های رایج در پیاده سازی VADها می توان به سطح انرژی و نرخ عبور از صفر اشاره نمود [۵][۳][۱۹] و [۲۰].

در مرجع [۶] از ویژگی آنتروپی برای تشخیص استفاده شده است. ایده استفاده از این مشخصه در سیگنال گفتار برخاسته از منظم تر و به اصطلاح ساختاریتر[3] بودن طیف سیگنال گفتار نسبت به نویز می باشد. همانند محاسبه ضرایب آنتروپی در یک مجموعه از سمبل ها و محاسبه احتمال رخداد هر یک از آنها توسط قانون شانون، می توان این ضرایب را در مجموعه ای از نقاط فرکانسی در نمودار طیف سیگنال بدست آورد.

از پارامتر های رایج دیگر می توان به ضرایب کپسترال اشاره نمود. در این روش ضرایب کپسترال در فریم ها محاسبه می شود و سپس با محاسبه فاصله این ضرایب با مقادیری که به نویز اختصاص داده شده اند و در نهایت مقایسه این مقادیر با مقدار آستانه هویت فریم مشخص می شود [۱۰] و [۷].

ویژگی تواتر[4] در فریم مورد نظر نیز یکی دیگر از پارامتر هایی است که در جهت پیاده سازی VAD ها از آن استفاده می شود [۱۱]. اما این ویژگی نیز مانند ویژگی آنتروپی در صورت رخداد نویزهای موزیکال و نویز های همهمه به دلیل منظم بودن طیف نویز به خوبی عمل نمی کند.

از روش های دیگر در پیاده سازی VAD ها می توان به استفاده از ضرایب LPC[5] اشاره نمود [۱۲]. بطور معمول در الگوریتم های مبتنی بر LPC یک سری ضرایب میانگین برای حالت های باصدا، بی صدا و سکوت بطور تجربی در نظر می گیرند. سپس ضرایب LPC فریم مشکوک و فاصله آن با هر یک از میانگین ها را بدست می آورند و از روی این فاصله ها، ماهیت فریم را مشخص می کنند. بر فرض اگر فاصله LPC فریم با میانگین LPC فریم با صدا از همه کمتر باشد، احتمال تعلق فریم به گروه فریم های باصدا بالاتر خواهد بود.

از روش های دیگر پیاده سازی VAD می توان به روش VE[6] اشاره نمود [۸]. در این روش نویز به عنوان یک منبع مستقل از سیگنال گفتار استخراج شده و باعث بالا بردن کارایی VAD می شود.

در [۱۴] نیز تلفیقی از روش های انرژی، آنتروپی، نرخ عبور از صفر و کپسترال بصورت های مختلف ارائه و در نهایت در شرایط مختلف به ارزیابی آنها پرداخته شده است.

روش دیگر پیاده سازی VAD استفاده از LTSE[7] فریم می باشد[۴]. در این روش N(k) ضرایب فوریه نویز را به طور میانگین از روی فریم های اولیه به دست می آورند. سپس مولفه k ام LTSE فریم جاری را از روی ماکزیمم k امین ضریب فوریه بین M فریم مجاور به دست می آورند. پس از آن انحراف LTSE فریم جاری از N(k) را محاسبه می کنند و با مقایسه آن با یک مقدار آستانه وقفی، ماهیت فریم را مشخص می‌کنند. در این الگوریتم N(k) هم در صورت لزوم به روز می‌شود. از پارامتر های رایج دیگر در این الگوریتم ها می توان به ضرایب موجک [۱]، [۹]و SNR در زیر باندها[۲۲] و LSPE[8] [۱۹] وAMDF[10] [۲۴] اشاره نمود. در الگوریتم های مبتنی بر موجک با توجه به انرژی دربرخی از زیر باند های موجک، عملیات تشخیص انجام شده است.

۳- معرفی تبدیل موجک

این تبدیل اطلاعاتی راجع به اینکه چه مولفه فرکانسی در چه بازه زمانی رخ داده است به دست می دهد. تحلیل موجک اجازه می دهد که در بازه های زمانی بلند، به بررسی مولفه های فرکانسی پایین و در بازه های زمانی کوتاه، به بررسی مولفه های فرکانسی بالا بپردازیم. بطور خلاصه در مورد توابع موجک می توان گفت که موجک یک شکل موج محدود با مقدار متوسط صفر در آن ناحیه می باشد.

تبدیل موجک در کاربرد هایی از قبیل حذف نویز، فشرده سازی اطلاعات، بررسی پدیده های لحظه ای و غیره مورد استفاده قرار می گیرد. موجک های مختلف فراوانی وجود دارند و یک مساله مهم در الگوریتم های مبتنی بر موجک، انتخاب موجک مناسب، با توجه به کاربرد، جهت پیاده سازی می باشد [۲].

برای پیاده سازی الگوریتم تبدیل موجک از فرم ماتریسی استفاده نمی شود زیرا در این صورت برای بدست آوردن ضرایب نیاز به $O(N^2)$ عملیات خواهد بود. لذا برای سهولت در انجام کار از روش خطی برای پیاده سازی استفاده می شودکه عملیات مورد نیاز در هر مرحله از درجه $O(N)$ می باشد. برای محاسبه ضرایب موجک در هر مرحله یک سری ضرایب و یک سری میانگین محاسبه می شود. مثلا اگر داده ها Nتایی باشند در مرحله اول N/2 ضریب و N/2 میانگین محاسبه می شود. بر فرض، میانگین ها طبق الگوریتم در نیمه پایینی یک آرایه و ضرایب در نیمه بالایی آن ذخیره گردد. نیمه های تولید شده، در مرحله بعدی به عنوان ورودی دوباره تحت عملیات بالا قرار می گیرند. با انجام این عمل در نهایت یک درخت تجزیه موجک در چند سطح خواهیم داشت. به عنوان مثال برای موجک Db2 می توان روابط (۱) و (۲) را برای محاسبه میانگین ها و ضرایب در نظر گرفت.

$$a_i = h_0 s_{2i} + h_1 s_{2i+1} + h_2 s_{2i+2} + h_3 s_{2i+3} \qquad (۱)$$

$$c_i = g_0 s_{2i} + g_1 s_{2i+1} + g_2 s_{2i+2} + g_3 s_{2i+3} \qquad (۲)$$

که در آن h_i ها توابع scaling، g_i ها توابع wavelet، a_i ها میانگین و c_i ها ضرایب موجک می باشند.

بطور کلی می توان طرز نمایش ماتریسی شکل ۲ را برای این تبدیل در نظر گرفت. در پیاده سازی فرض شده است که تعداد داده ها در هر فریم توانی از ۲

باشد. در هر مرحله از تبدیل, سیگنال داده به دو قسمت میانگین[۹] و تفاوت[۱۰] تقسیم می شود, که در آن تفاوت ها اشاره به ضرایب موجک دارند. برای مثال اگر داده ۲۵۶ تایی باشد، در مرحله اول این سیگنال به یک دستهٔ ۱۲۸تایی میانگین و یک دسته ۱۲۸تایی تفاوت تقسیم می شود. سپس هر دو نیمه برای بدست آوردن سریهای دیگر میانگین و تفاوت در مرحلهٔ بعد به عنوان ورودی مورد استفاده قرار می گیرند. این روند ادامه می یابد تا به تعداد خاصی از میانگین و تفاوت برسیم.

در این کار از روش جابجایی پنجره[۱۱] استفاده می شود، به این ترتیب که یک پنجره با اندازهٔ مشخص (به عنوان مثال در Db2 اندازه پنجره برابر۴ است)، با قدم های ۲ تایی تا انتهای سیگنال حرکت می کند. لازم به ذکراست که در روش های مختلف، می توان اندازهٔ گام ها را متفاوت در نظر گرفت و در نهایت ضرایب موجک را محاسبه کرد.

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & g_0 & g_1 & g_2 & g_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_0 & g_1 \end{bmatrix} * \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{bmatrix}$$

شکل ۲- نمایش ماتریسی محاسبه ضرایب موجک برای یک آرایه ۸ تایی

از نکات مهم دیگر در پیاده سازی این الگوریتم, سیاست این الگوریتم در برابر نقاط انتهایی می باشد. فرضاً برای یک نقطه میانی با شاخص i، در Db2، ۴ نقطه بعد از $2i$ در محاسبه c_i و a_i ها دخیل می باشند، ولی برای نقاط انتهایی، سیگنال به طور آیینه ای در هر دو انتها تکرار می شود. در این مقاله از موجک Db4 برای پیاده سازی استفاده شده است.

۴- فیلتر های OS[۱۲]

نشان داده شده است که فیلتر های غیر خطی در برخی کاربرد ها قابلیت کاربردی بیشتری نسبت به فیلتر های خطی دارند. به عنوان مثال از این دسته فیلتر ها می توان به فیلتر های OS و از این دسته می توان به فیلتر میانه[۱۳] اشاره نمود که در کاربرد هایی نظیر پردازش تصویر برای بررسی نویز های ضربه ای مورد استفاده قرار می گیرند.

خروجی یک فیلتر OS بر روی یک مجموعه داده ای ورودی بصورت رابطه زیر می باشد[۲۱]:

$$y_s(k) = \sum_{i=1}^{L} \beta_i . x_i \qquad (۳)$$

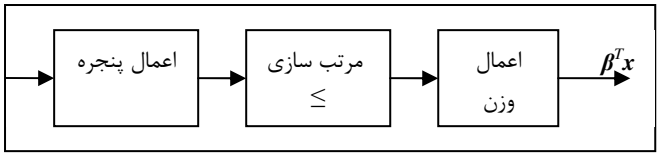
که در ان $L=2N+1$ و x_i نیز نمایانگر داده های مرتب شده صعودی ورودی و s درجه فیلتر می باشد بطوری که داریم: $x_1 \leq x_2 \leq \ldots \leq x_L$

وزن های β_i نمایانگر ضرایب فیلتر می باشند. برای مثال در فیلتر میانه $(s=0.5)$ مقدار β_i ها به ازاء $i=N+1$ برابر ۱ و برای بقیه مقادیر برابر صفر می باشد.

چگونگی اعمال فیلتر OS بر روی اطلاعات ورودی در دیاگرام شکل ۳ نمایش داده شده است.

۵- روش پیشنهادی

یکی از روش های پیاده سازی VAD ها استفاده از محاسبه SNR[14] در زیر باند های فرکانسی مختلف و در نهایت محاسبه میانگین آنها و مقایسه آن با یک مقدار آستانه می باشد [۱۵] و [۱۶]. در این الگوریتم پیشنهادی نیز به جای محاسبه ضرایب فوریه و محاسبه انرژی در زیر باند ها و سپس محاسبه SNR، در ابتدا انرژی ضرایب موجک فریم جاری در سطح چهارم در هر زیر باند محاسبه می شود. لازم به تذکر است که در این پژوهش تعداد زیر باند ها برابر با ۱۶ در نظر گرفته شده است.در گام بعدی, فاصله اقلیدسی این بردار ۱۶تایی و بردار مربوط به نویز بدست آمده و با یک مقدار آستانه مقایسه می گردد .



شکل ۳- نمایش طرز کار یک فیلتر OS

در این الگوریتم برای بالا بردن دقت و کارایی از بهنگام سازی طیف موجک نویز، توسط مکانیزم hangover، و به کار گیری یک فیلتر OS استفاده شده است. برای توضیح بیشتر, مراحل مختلف این الگوریتم به شرح زیر بیان می شود.

۱- محاسبه ضرایب موجک: در این قسمت, ضرایب موجک هر فریم توسط یک درخت تجزیه موجک با یک الگوریتم خطی مناسب که در قسمت قبل توضیح داده شده است، محاسبه می گردد. لازم به ذکراست که در صورت توانی از دو نبودن طول فریم به تعداد لازم صفر به آن اضافه می شود تا توانی از دو گردد[۱۵]. بعد از محاسبه ضرایب موجک، مقدار انرژی در زیر باند های مختلف محاسبه می شود. خروجی این قسمت از الگوریتم برای هر فریم, با توجه به سطح تجزیه, ۱۶ مقدار عددی می باشد. در اینجا موجک استفاده شده از نوع Db4 است. با توجه به پیاده سازی خطی این قسمت، الگوریتم پیاده سازی شده از سرعت بالایی بر خوردار می باشد.

۲- بدست آوردن تخمینی از طیف نویز: در اینجا فرض می شود که ۱۰ فریم اولیه نویز باشند. لذا طیف موجک نویز از روی میانگین گیری طیف موجک ۱۰ فریم اولیه بدست می آید. لازم به ذکر است که برای افزایش کارائی الگوریتم, طیف نویز در طول پردازش فایل های صوتی در زمان رخداد سکوت با ضریب $\gamma=0.97$، بهنگام می شود.

۳- محاسبه فاصله اقلیدسی و اعمال فیلتر OS بر روی فاصله ها: بعد از تخمین طیف موجک نویز و محاسبه طیف موجک در هر فریم، فاصله اقلیدسی بین این دو در هر فریم محاسبه می گردد. یعنی در حقیقت فاصله اقلیدسی دو بردار ۱۶ تایی در این مرحله محاسبه می شود. بعد از محاسبه این فاصله در هر فریم, نوبت به اعمال آستانه می رسد. می توان آستانه را بر روی تک تک فریم ها اعمال نمود, لیکن در اینجا برای بالا بردن دقت و کارایی یک فیلتر OS بر روی فریم جاری اعمال می گردد. این ایده در برخی VADها در قسمت اعمال آستانه مورد استفاده قرار گرفته است [۱۵] و [۱۶]. با توجه به توضیحات ارائه شده در رابطه با این نوع فیلتر در بخش قبل، برای فریم mام با توجه به $2N$ فریم مجاور(N فریم قبل و N فریم بعد) مجموعه زیر در نظر گرفته می شود:

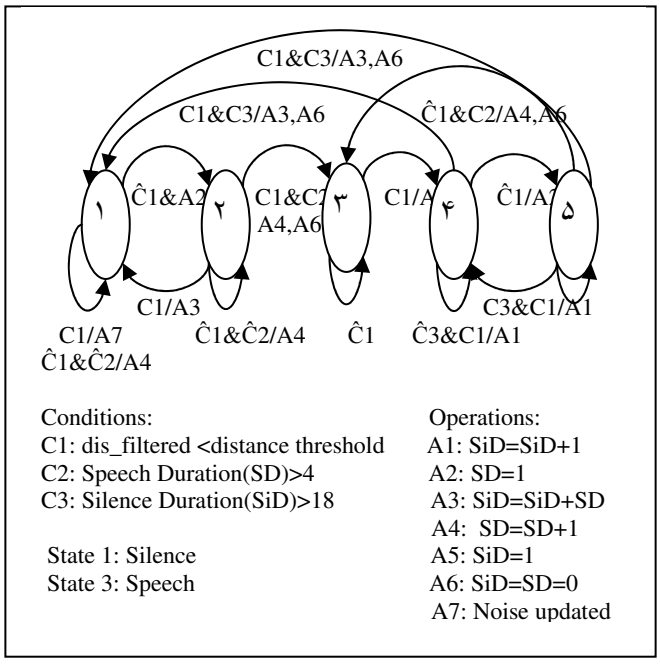
$$\{dis_{(m-N)}, dis_{(m-N+1)}, \ldots, dis_{(m)}, \ldots, dis_{(m+N-1)}, dis_{(m+N)}\}$$

ومقدار $dis\_filtered$ به ازای هر فریم از رابطه (۴) محاسبه می گردد:

$$dis\_filtered(m)=(1-f).dis_s(m)+f.dis_{(s+1)}(m) \qquad (۴)$$

که در اینجا s درجه فیلتر است که از رابطه $s = \lfloor 2pN \rfloor$ به دست می آید و $f = 2pN - s$. ضمناً فرض می شود که N=8، p=0.9 و $dis(m)$ فاصله اقلیدسی طیف موجک فریم mام از طیف نویز است. بررسی ها نشان می دهد که خصوصاً استفاده از این روش در صورت رخداد نویز ضربه موجب افزایش دقت نتایج می گردد.

۴- استفاده از ماشین حالت محدود Hangover : یکی از مشکلات VADها در کاربرد هایی نظیر تشخیص گفتار، حذف قسمت سکوت در واجهای انفجاری به عنوان غیر گفتار می باشد. برای جلوگیری از این مشکل و در ضمن جلوگیری از تشخیص نویز های ضربه ای به عنوان گفتار از ماشین حالت محدود[۱۶] Hangover استفاده شده، که در شکل ۴ آورده شده است[۲۵].



شکل ۴- ماشین حالت محدود Hangover

۵- بهنگام سازی طیف نویز : با توجه به روش Hangover مورد استفاده در این الگوریتم, بهنگام سازی طیف نویز را می توان در هر مرحله از کار که فریم سکوت تشخیص داده شد انجام داد (حالت ۱ در شکل ۴).

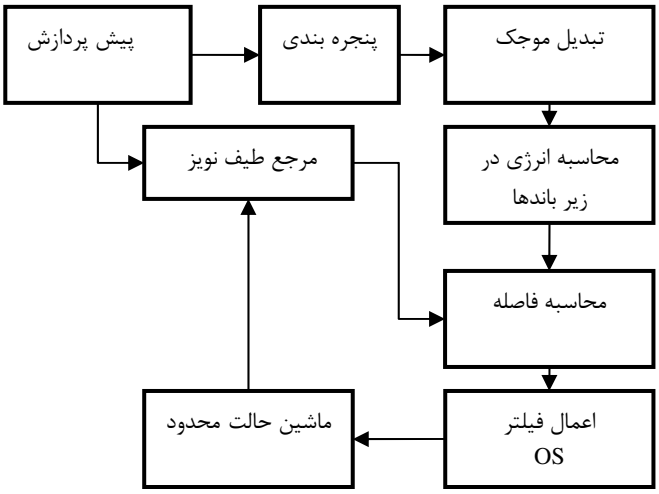
برای بالا بردن دقت, می توان این عمل را در صورت گذر از حالتهای۵ و ۴ به حالت ۱ نیز انجام داد. ولی این کار در گذر از حالت ۲ به حالت ۱(بدلیل رخداد احتمالی نویز ضربه) پیشنهاد داده نمی شود.

۶- تعیین مقدار آستانه : یکی ازمسائل اصلی در الگوریتم های VAD، تعیین مقدار آستانه می باشد. طبق رابطه (۵) این مقدار مضربی از میانگین فاصله در ده فریم اول سیگنال گفتار است که نویز فرض می شوند. بطور تجربی این مقدار وابسته به مقدار SNR در فایل صوتی مورد نظر می باشد. در SNRهای بالا این مقدار کم و در SNR های پایین این مقدار زیاد است.

$$threshold = \frac{\alpha}{10}\sum_{i=1}^{10} Dis(i) \qquad (۵)$$

می توان این آستانه را در SNR مختلف ثابت در نظر گرفت ولی در اینجا در ابتدا از یک متد ساده به عنوان پیش پردازش، SNR کل بطور تقریبی محاسبه می شود. سپس پارامتر α با توجه به SNR تقریب زده شده، بطور خطی در یک رنج خاص تغییر داده می شود. تخمین SNR از روش های مختلف امکان پذیر می باشد ولی در اینجا برای سادگی با فرض نویز بودن چند فریم اول میانگین انرژی فریم نویزی (از روی ۱۰ فریم اولیه) و میانگین انرژی فریم (از روی کل فایل صوتی) محاسبه و SNR کل از روی این دو مقدار تقریب زده می شود.

بطور خلاصه شمای کلی الگوریتم را میتوان به صورت شکل ۵ در نظر گرفت. در قسمت پیش پردازش, SNR به طور تقریبی محاسبه شده و از روی آن مقدار آستانه حساب می شود. همانگونه که در شکل ۵ دیده می شود طیف نویز توسط ماشین بهنگام می شود. این بهنگام سازی در تمام زیر باندها انجام می گردد.



شکل ۵- چگونگی بهنگام سازی طیف نویز

۵-۱ کاهش وابستگی الگوریتم به SNR محاسبه شده

در این قسمت می خواهیم الگوریتم را به صورتی تغییر دهیم تا از وابستگی آن به مقدار SNR محاسبه شده در قسمت پیش پردازش کاسته شود. برای انجام این کار بجای استفاده از یک مقدار آستانه از یک نوار آستانه و بجای استفاده از ماشین حالت محدود شکل ۴, از ماشین تلفیقی دیگری استفاده میکنیم. در این جا یک حد بالای آستانه Threshold_h و یک حد پایین آستانه Threshold_l در نظر می گیریم. برای سادگی فهم مطلب برای یک دنباله فاصله محاسبه شده متناظر با یک دنباله فریم، خروجی output را برای هر فریم توسط روابط زیر تعیین می شود:

IF (smoothed_distance(i) < Threshold_l) output(i)=0;
IF (smoothed_distance(i)>Threshold_h) output(i)=1;
IF (Threshold_l≤ smoothed_distance(i)≤Threshold_h) output(i)=0.5

حال می توان خروجی VAD را با توجه به سیگنال output و بر اساس توابع غیر خطی نشان داده شده در شکل۶ بدست آورد. عملیاتی که در این شکل نمایش داده شده, به ترتیب توسط ۴ تابع transition(00)، transition(11)، transition(10)و transition(01) بیان می شود. شکل ۷ نشان دهنده ماشین حالت مناسب برای انجام این عملیات می باشد.

همانگونه که قبلا ذکر شده است بررسی حالات ناپایدار, همچون نویز ضربه و سکوت واجهای انفجاری در VADها بسیار مهم می باشد. بنابر این در این شرایط، ماشین حالت شکل ۷ به تنهایی کارایی بالایی نخواهد داشت. لذا در اینجا سعی شده تا بطور منطقی, ماشین شکل ۴ با ماشین شکل ۷ تلفیق شود. نتیجه ادغام این دو در شکل ۸ آمده است. لازم به ذکر است که توابع ()transition بر روی فاصله های زمانی L_amb اعمال می شوند.

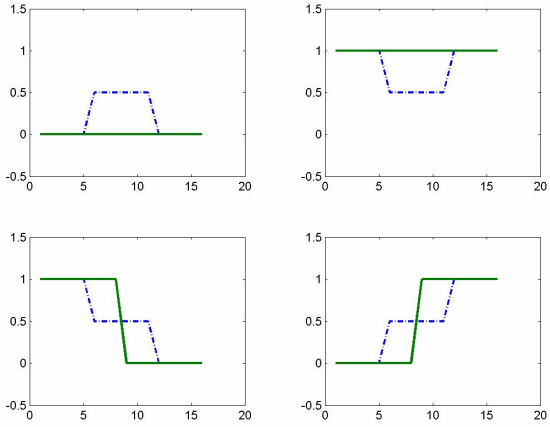
۶- آزمایشات انجام شده

برای بررسی کارایی الگوریتمهای الگوریتم ارائه شده و مقایسه عملکرد آنها، آنها را بر روی دادگان فارسی فارس دات شنوا, بـه حجـم ۱۴۰ فایـل بـا فرکـانس نمونـه برداری ۲۲۰۵۰ هرتز، همراه با نویز هایی از قبیل ماشـین، نمایـشگاه و همهمـه در SNR های ۲۰ و ۱۵ و۱۰ و ۵ و۰ و ۵- دسی بل اعمال نمودیم. این فایلها بیـانگر گفتار پنج گوینده مرد و دو گوینده زن است, که هر کدام بیست جمله متفـاوت را بیان کرده اند. برای نمونه, شکل ۹ نتیجه اعمال الگوریتم ارایه شـده بـر روی یـک فایل صوتی در SNR های۲۰، ۱۰و ۰ دسی بل می باشد.

نواحی زیر پنجره هـا بعنـوان گفتـار و بقیـه نـواحی بعنـوان سـکوت شناسـائی شده اند.

پارامتر های ارزیابی کارایی عبارتند از FA که بیانگر تعداد فریم هـای سـکوتی که به اشتباه گفتار تـشخیص داده شـده انـد و Miss نـشانگر تعـداد فـریم هـای گفتاری که به اشتباه سکوت تشخیص داده شده اند می باشـد. همچنـین Pe(%) بیانگر احتمال رخداد خطا, ویا به عبارت دیگـر P(FA+Miss) اسـت. همچنـین, Corr نماینده تعداد فریمهائی است که بصورت صحیح تشخیص داده شده اند.

به منظور بررسی کارایی و استخراج پارامترهای ذکر شده, برای هر فایل صـوتی در سطح واج, اطلاعات مربوط به سکوت بودن و یا نبودن هـر واحـد آن اسـتخراج گردید. برای اینکه بتوان نتایج این آزمایش ها را با معیـار هـای اسـتاندارد موجـود مقایسه کنیم از VAD اسـتاندارد ETSI بـه نـام AMR2 اسـتفاده شـده اسـت [۲۰].

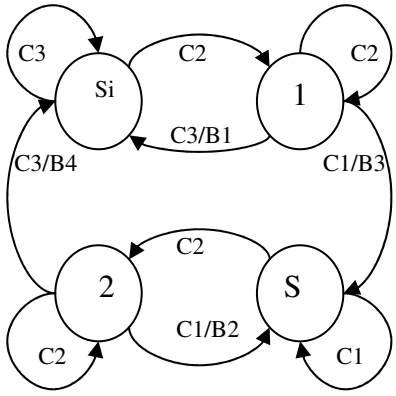


شکل ۶- بدست آوردن خروجی VAD از روی سیگنال output (خـط چـین هـا بیانگر سیگنال output و خطوط متدد خروجی VAD می باشند)

این VAD تصمیم گیری را بر اساس تحلیل در بانـد هـای فرکانـسی مختلـف انجام می دهد. در ابتدا این الگوریتم بر روی فایـل هـای دادگـان اعمـال گردیـد و مقادیر جدول ۱ بدست آمد. دلیل خطای بالای این الگوریتم, وابسته بـودن آن بـه SNR تخمین زده شده در قسمت پیش پردازش است, لذا در این الگـوریتم نمـی توان خطای بلوک تخمین SNR را نادیده گرفت. بنابر این سعی شد در الگـوریتم ارائه شده وابستگی به واحد پیش پردازش را کاهش دهیم (شکل ۷, ماشین ۱).

نتیجه حاصل از اعمال این روش بر روی یک نمونه از فایلهای صوتی در شکل ۹ آورده شده است که در آن قسمتهای زیر پنجره نشاندهنده نواحی غیر سکوت مـی باشد.

برای افزایش هر چه بیشتر کارائی الگوریتم, سعی شد که نیازی به تخمین دقیق SNR نباشد.



Conditions:

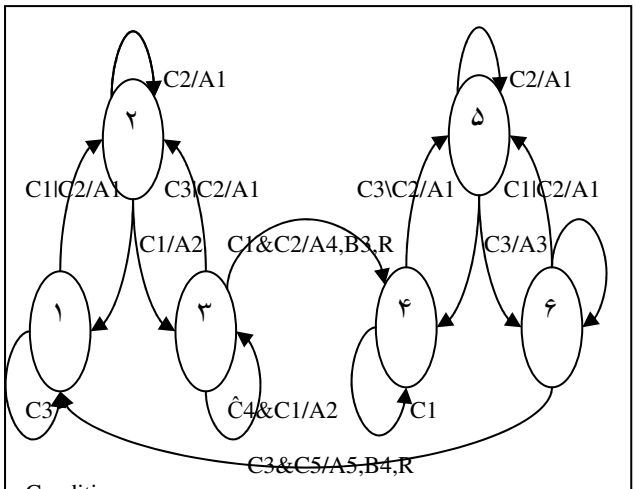
C1: output (i) = =1
C2: output (i) = = 0.5
C3: output (i) = = 0
S: Speech
Si: Silence

Operations:

B1: transition (00)
B2: transition (11)
B3: transition (01)
B4: transition (10)

شکل ۷ - ماشین حالت محدود پیشنهادی اول (ماشین ۱)

برای این منظور از ماشین حالت پیشنهادی ۲ استفاده شد (شکل ۸). نتایج حاصله از روش پیشنهادی دوم در جدول ۳ آورده شده است. در مقایسه جدول ۲ و ۳ مشاهده می شود که الگوریتم پیشنهادی دوم توانسته است در تمامی حالات کارائی سیستم را افزایش داده و نرخ خطا را به صورتی قابل قبول کاهش دهد.



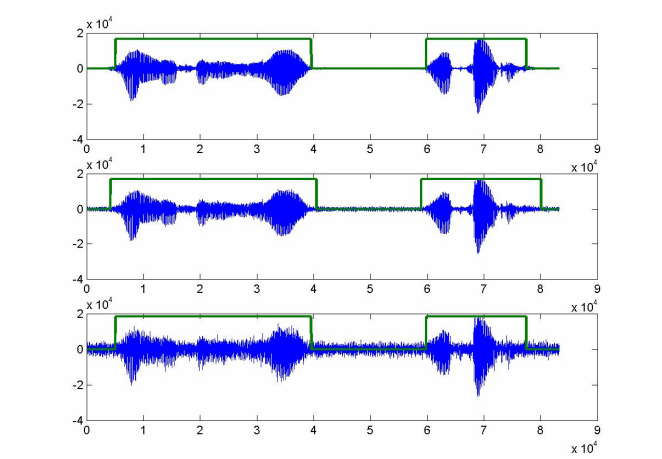
Conditions:
C1:smoothed_dis() > Threshold_h;
C2:Threshold_l<smoothed_dis() < Threshold_h;
C3:smoothed_dis() < Threshold_l
C4: sp_d>speech_duration
C5:si_d>silence_duratio

Operations:

B1: Transition (00)
B2: Transition (11)
B3: Transition (01)
B4: Transition (10)
R: sp_d = si_d = L_amb =0

A1: L_amb = L_amb + 1
A2:sp_d = sp_d + 1
A3:si_d = si_d + 1
A4:L_amb = L_amb + sp_d
A5:L_amb = L_amb + si_d

شکل ۸- ماشین حالت محدود تلفیقی جدید (ماشین ۲)



شکل ۹- اعمال الگوریتم پیشنهادی ۱ بر روی یک فایل در SNR های ۲۰، ۱۰و ۰ دسی بل آغشته به نویز ماشین

برای مقایسه هر چه بهتر، روش پیشنهادی خود را با الگوریتم معروفOSF مقایسه کردیم [۱۶]. این الگوریتم از کارائی بسیار بالائی در زمینه تشخیص نواحی سکوت برخوردار است و معمولاً در پژوهشهای مرتبط با VAD به عنوان مرجع مورد استفاده قرار می گیرد. نتایج حاصل از این مقایسه برای سه نوع مختلف از نویزهای استاندارد در شکل ۱۰ آورده شده است.

همانطور که از نمودارهای شکل ۱۰ آشکار است روش پیشنهادی در تمامی حالات عملکردی بمراتب بهتر را نسبت به روش ESTI-ARM2 نشان می دهد. در مقایسه با روش OSF الگوریتم ارائه شده عملکردهای متفاوتی را در نویز مختلف از خود بروز می هد. در رابطه با نویز ماشین، عملکرد روش پیشنهادی خصوصاً در محیطهای پر نویز بهتر است. برای نویز نمایشگاه عملکرد سیستم ارائه شده در تمامی موارد، بجز مواردی که سیگنال به نویز کمتر از صفر دسی بل است، بهتر از روش OSF است. نهایتاً اینکه، برای نویز همهمه نیز روش پیشنهادی رویهمرفته خطای کمتری را ایجاد می کند.

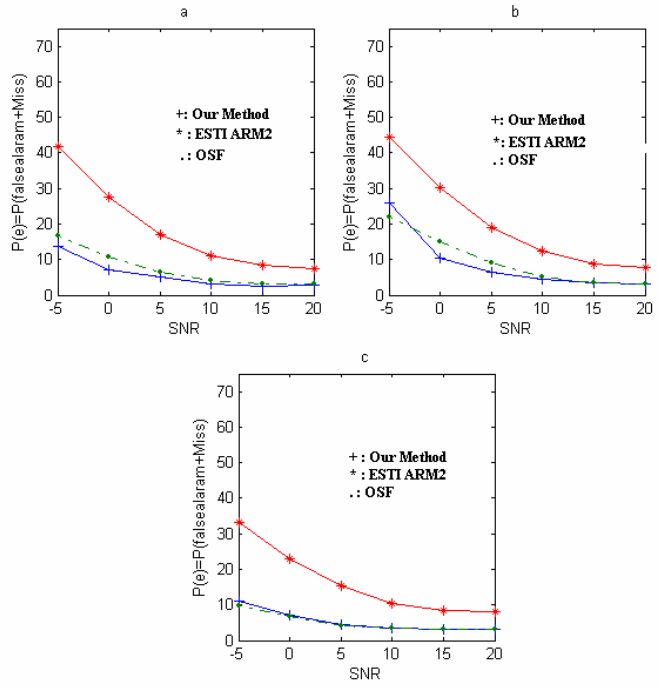
۷- نتیجه گیری

در این مقاله یک الگوریتم VAD به کمک تبدیل موجک ارائه شد. در این روش، با محاسبه فاصله طیف موجک فریم جاری و فریم های نویزی و اعمال آستانه بر روی فاصله مزبور، عملیات تشخیص انجام می شود. در این الگوریتم طیف موجک نویز با پارامترα=0.97 در طول فایل، در صورت رخداد سکوت بهنگام می شود. در ابتدا سعی شد تا با انتخاب یک مقدار آستانه، عملیات تشخیص توسط یک ماشین حالت محدود انجام شود. سپس به دلیل وابسته بودن مقدار آستانه به مقدار SNR تخمینی از مرحله ابتدایی (پیش پردازش)، سعی شد تا با کمک یک ماشین حالت محدود دیگر که تلفیقی از ماشین اول، و یک ماشین حالت خاص است که بجای استفاده از یک مقدار آستانه بخصوص از یک نوار آستانه بهره می برد، عملیات تشخیص را انجام داد.

همانگونه که در شکل(۱۰) نشان داده شد کارایی الگوریتم پیشنهادی بمراتب نسبت به کارائی الگوریتم AMR2 بهتر بود و نسبت به الگوریتم OSF نیز در قالب موارد عملکرد بهتری را نشان می داد.

جدول۱- نتایج حاصل از پیاده سازی الگوریتم ETSI-AMR2

نوع نویز	SNR	FA	Miss	Corr	P(e)%
ماشین	۲۰	۱۰۷۸	۲۵۴۳	۴۵۰۶۴	۷,۴۳
	۱۵	۹۸۸	۳۱۲۳	۴۴۵۷۵	۸,۴۴
	۱۰	۸۵۰	۴۶۱۳	۴۳۲۲۳	۱۱,۲۱
	۵	۷۸۵	۷۵۷۵	۴۰۳۲۶	۱۷,۱۶
	۰	۷۵۸	۱۲۷۱۴	۳۵۲۱۴	۲۷,۶۶
	-۵	۷۴۲	۱۹۶۶۳	۲۸۲۸۱	۴۱,۹۱
نمایشگاه	۲۰	۱۰۹۴	۲۶۵۴	۴۴۹۳۷	۷,۶۹
	۱۵	۹۵۸	۳۴۱۱	۴۴۳۱۶	۸,۹۱
	۱۰	۸۸۰	۵۱۰۲	۴۲۷۰۳	۱۲,۲۶
	۵	۷۸۹	۸۴۳۹	۳۹۴۵۸	۱۸,۹۵
	۰	۸۰۷	۱۳۹۱۵	۳۳۹۶۳	۳۰,۲۳
همهمه	۲۰	۱۱۲۴	۲۸۵۶	۴۴۷۰۶	۸,۱۷
	۱۵	۱۱۰۲	۲۹۹۲	۴۴۵۹۲	۸,۴۰
	۱۰	۱۰۰۶	۴۱۱۹	۴۳۵۶۰	۱۰,۵۲
	۵	۱۰۱۳	۶۵۳۸	۴۱۱۳۴	۱۵,۵۰
	۰	۱۰۸۴	۱۰۰۷۶	۳۷۵۲۵	۲۲,۹۲



شکل ۱۰- مقایسه خطای الگوریتم ها در SNR های مختلف برای سه نویز مختلف a: نویز ماشین b: نویز نمایشگاه و c: نویز همهمه

جدول۲- نتایج حاصل از پیاده سازی الگوریتم پیشنهادی اول (ماشین حالت ۱)

نوع نویز	SNR	FA	Miss	Corr	P(e)%
ماشین	۲۰	۷۲۴	۱۰۷۰	۴۶۸۹۰	۳,۶۸
	۱۵	۵۸۳	۱۴۵۳	۴۶۶۴۸	۴,۱۸
	۱۰	۴۷۰	۱۸۰۱	۴۶۴۱۳	۴,۶۶
	۵	۳۸۵	۲۴۱۵	۴۵۸۸۴	۵,۷۵
	۰	۴۳۹	۳۶۰۵	۴۴۶۴۰	۸,۳۰
	-۵	۶۱۳	۷۹۹۰	۴۰۰۸۱	۱۷,۶۷
نمایشگاه	۲۰	۷۷۱	۹۶۸	۴۶۹۴۵	۳,۵۷
	۱۵	۶۴۱	۱۱۷۶	۴۶۸۶۷	۳,۷۳
	۱۰	۵۰۳	۱۷۱۳	۴۶۴۶۸	۴,۵۵
	۵	۵۱۱	۲۲۹۹	۴۵۸۷۴	۵,۷۷
	۰	۱۰۴۵	۳۴۱۸	۴۴۲۲۱	۹,۱۶
همهمه	۲۰	۵۴۹	۲۱۲۸	۴۶۰۰۷	۵,۴۹
	۱۵	۴۳۷	۲۳۷۵	۴۵۸۷۲	۵,۷۷
	۱۰	۳۸۱	۲۳۷۳	۴۵۹۳۰	۵,۶۵
	۵	۴۳۳	۲۴۰۳	۴۵۸۴۸	۵,۸۲
	۰	۱۰۱۳	۲۶۰۵	۴۵۰۶۶	۷,۴۳

جدول۳- نتایج حاصل از پیاده سازی الگوریتم پیشنهادی دوم (ماشین حالت ۲)

نوع نویز	SNR	FA	Miss	Corr	P(e)%
ماشین	۲۰	۹۹۹	۴۰۱	۴۷۲۸۴	۲,۸۹
	۱۵	۶۶۶	۵۹۳	۴۷۴۲۵	۲,۵۹
	۱۰	۴۱۶	۱۱۵۲	۴۷۱۱۶	۳,۲۲
	۵	۲۸۵	۲۱۷۸	۴۶۲۴۸	۵,۰۱
	۰	۴۲۵	۳۰۲۴	۴۵۲۳۵	۷,۰۹
	-۵	۴۸۴	۶۲۳۰	۴۱۹۷۰	۱۳,۸
نمایشگاه	۲۰	۹۴۱	۵۱۰	۴۷۲۳۳	۲,۹۸
	۱۵	۷۳۶	۸۸۴	۴۷۰۶۴	۳,۳۳
	۱۰	۴۷۸	۱۶۹۳	۴۶۵۱۳	۴,۴۶
	۵	۴۳۷	۲۷۳۱	۴۵۵۱۶	۶,۵۱
	۰	۱۰۸۲	۴۰۶۳	۴۳۵۳۹	۱۰,۵۷
همهمه	۲۰	۱۰۳۵	۴۵۸	۴۷۱۹۱	۳,۰۶
	۱۵	۹۱۴	۵۷۹	۴۷۱۹۱	۳,۰۶
	۱۰	۸۳۸	۸۳۴	۴۷۰۱۲	۳,۴۳
	۵	۷۲۷	۱۴۷۹	۴۶۴۷۸	۴,۵۳
	۰	۶۴۴	۲۸۳۹	۴۵۲۰۱	۷,۱۵

مراجع

[1]. J. Shaojun, G. Hitato and Y. Fuliang, "A New Algorithm For Voice Activity Detection Based on Wavelet Transform," *Proc. of Int. symposium of intelligent multimedia, video and speech processing,* pp. 222-225, Hong Kong, Oct. 2004.

[2]. Y. long, L. Gang and G. Jun, "Selection of the Best Wavelet Based for Speech Signal," *Proc. of Int. symposium of intelligent multimedia, video and speech processing,* pp.218-221, Hong Kong, Oct. 2004.

[3]. B. Harsha, "A Noise Robust Activity Detection Algorithm," *Proc. of Int. symposium of intelligent multimedia, video and speech processing,* pp. 322-325, Hong Kong, Oct. 2004.

[4]. J. Ramırez, J. C. Segura and C. Benıtez, "A New Adaptive Long-Term Spectral Estimation Voice Activity Detector," *EuroSpeech,* pp. 3041-3044, Geneva, 2003.

[5]. J. Faneuff, "Spatial, Spectral, and Perceptual Nonlinear Noise Reduction for Hands-free Microphones in a Car," *Master Thesis, Electrical and Computer Engineering Dept.,* July 2002.

[6]. P. Renevey and A. Drygajlo, "Entropy Based Voice Activity Detection in Very Noisy Conditions," *EuroSpeech'01,* pp.1883-1886, 2001.

[7]. S. Skorik and F. Berthommier, "On a Cepstrum-Based Speech Detector Robust to White Noise," *Proc. of Int. Conference, Specom,* pp. 201-206, St. Petersburg, 2002.

[8]. D. R. Paoletti and G. Erten, "Enhanced Silence Detection in Variable Rate Coding Systems Using Voice Extraction," *Proc. 43rd IEEE Midwest Symp. on Circuits and Systems,* vol.2, pp.592-594, Lansing MI, 2000.

[9]. J. Stegmann and G. Schroeder, "Robust Voice Activity Detection Based on the Wavelet Transform," *Proc. IEEE Workshop on Speech Coding,* pp. 99-100, Pocono Manor, Pennsylvania, USA, Sep. 1997.

[10]. A. Haigh and J.S. Mason, "Robust Voice Activity Detection Using Cepstral Features," *Proc. of IEEE TENCON'93,* vol. 3, pp. 321-324, Beijing, 1993.

[11]. R. Tucker, "Voice Activity Detection Using a Periodicity Measure," *IEEE Proceeding,* vol.139, no. 4, pp. 377-380, August 1992.

[12]. L. R. Rabiner and M. R. Sambur, "Application of an LPC Distance Measure to the Voiced-Unvoiced-Silence Detection Problem," *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol. ASSP-25, no. 4, pp. 338-343, August 1977.

[۱۳].م. منظوری، س. آیت، ج. کبودیان، م. همایونپور، " ارائه روشی برای کاهش نویز موزیکال در روش تفریق طیفی با افزایش دقت تخمین نویز، "کنفرانس بین المللی فن آوری اطلاعات، دی ماه ۱۳۸۲.

[۱۴].م. همایونپور، ا. شریف نبوی، "مقایسه و ارزیابی روش های تشخیص گفتار از سکوت "، کنفرانس بین المللی فن آوری اطلاعات، دی ماه ۱۳۸۲.

[15]. J. Ramfrez, J.C. Segura, C. Benitez, A. de la Tora and A. Rubio, "A New voice Activity Detector Using Subband Order-Statistics Filters for Robust Speech Recognition," *IEEE Trans. on Acoustics, Speech, and Signal Processing,* vol. ASSP-52, no. 3, pp. 276-303, 2004.

[16]. J. Ramfrez, J.C. Segura, C. Benitez, A. de la Tora and A. Rubio, "An Effective Subband OSF-Based VAD with Noise Reduction for Robust Speech Recognition," *IEEE 2005*

[17]. W. H. Abdulla, "HMM Based Techniques for Speech Segments Extraction," *10th Int. symposium of Science Programming,* pp.221-239, 2002.

[18]. H.Othman, T.Abdulnasr, "A semi-continuous state transition probability HMM-based voice activity detection," *IEEE Proceeding-I.,* vol. 139, no. 4, pp.821-824, 2004.

[19]. G.S. Tanyer and H. Ozer, "Voice Activity Detection in Nonstationary Gaussian Noise," *proceeding of ICSP'98*, pp.1620-1623, 1998.

[20]. Sangwan, A., Chiranth, M. C., Jamadagni, H. S., Sah, R.,Prasad, R. V., Gaurav, V., "VAD Techniques for Real-Time Speech Transmission on the Internet," *5th IEEE International Conference on High-Speed Networks and Multimedia Communications,* pp. 46–50, 2002.

[21]. H. G. longbothom, A. C. Bovik and A.Restrepo, "Generalized Order Statistic Filters," *IEEE 1998,* pp.1610-1613.

[22]. A. Vahatalo and I. Johansson, "Voice Activity Detection for GSM Adaptive Multi-Rate Codec," *IEEE 1999,* pp. 55-57.

[23]. F. Beritelli S. Casale and A. Cavallaro, "A Robust Voice Activity Detector for Wireless Comminucation Using Soft Computing," *IEEE 1998,* pp.1818-1828.

[24]. M.Orlandi, a. santarelli and D. Falavigna, "Maximum Likelihood Endpoint Detection with Time-Domain Features," *Eurospeech03,* pp.1757-1760, Italy, 2003.

[25]. A. Martin, G. Damnati and L. Mauuary, "Robust Speech/Non-speech Detection Using LDA for Continuos Speech Recognition," *Eurospeech01*, pp. 675-678 Scandinavia.

خدیجه آقاجانی کارشناسی خود را در مهندسی کامپیوتر از دانشگاه شهید بهشتی در سال ۱۳۸۲ و کارشناسی ارشد خود را از دانشگاه صنعتی شریف در رشته مهندسی کامپیوتر, گرایش معماری کامپیوتر در سال ۱۳۸۵ دریافت نموده است. زمینه های پژوهشی مورد علاقه ایشان شامل پردازش سیگنال و گفتار و بازشناسی الگو می باشد.

محمد تقی منظوری شلمانی تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد مهندسی برق (الکترونیک) به ترتیب در سالهای ۱۳۶۴ و ۱۳۶۸ در دانشگاه صنعتی شریف به پایان رساند و سپس دکتری خود در مهندسی برق و کامپیوتر را در سال ۱۳۷۴ از دانشگاه صنعتی وین اتریش دریافت نمود. وی هم اکنون بعنوان استادیار در دانشکده مهندسی کامپیوتر دانشگاه صنعتی شریف مشغول به تدریس و تحقیق می باشد. ایشان قبل از پیوستن به دانشگاه صنعتی شریف, در طی سالهای ۱۳۷۴ الی ۱۳۷۵ در دانشگاه صنعتی وین بعنوان عضو هیئت علمی و محقق مشغول به کار بوده است. زمینه های تحقیقاتی مورد علاقه وی عبارتند از: پردازش سیگنال, پردازش گفتار و طراحی مدارهای مجتمع دیجیتال.
آدرس پست الکترونیکی ایشان عبارت است از:

manzuri@sharif.edu

[1] Voice Activity Detection
[2] Hidden Markov Model
[3] Organized
[4] Periodicity
[5] Linear Predictive Coding
[6] Voice Extraction
[7] Long Term Spectral Estimation
[8] Least Square Periodicity Estimator
[9] Average
[10] Difference
[11] Shifting Window
[12] Order Statistic
[13] Median
[14] Signal to Noise Ratio
[15] Zero Padding
[16] Finite State Automata

* این پژوهش تحت قرارداد شماره CS1385-4-03 توسط پژوهشگاه دانشهای بنیادی (IPM) حمایت شده است.