روشی جدید برای غنی‌سازی سیگنال‌های صوتی با استفاده از آنالیز LPC در روش‌های تفریق طیفی پایه، چند بانده و معکوس فوریه

مصطفی حیدری[۱]، مربی، محمد رضا کرمی‌ملایی[۲]، دانشیار

۱- گروه مهندسی برق- دانشکده فنی و مهندسی- دانشگاه شمال- آمل- ایران- m.heydari@shomal.ac.ir
۲- دانشکده مهندسی برق و کامپیوتر- دانشگاه صنعتی نوشیروانی- بابل- ایران- mkarami@nit.ac.ir

چکیده: در این مقاله، سه روش جدید برای غنی‌سازی سیگنال‌های صوتی ارائه شده است. در این روش‌ها که مبتنی بر روش تفریق طیفی، روش تفریق معکوس تبدیل فوریه و تفریق طیفی چند بانده می‌باشند، از ضرایب پیشگویی خطی (LPC)[۱] و آنالیز VAD[۲] و نیز تشخیص‌گر حروف صدادار و بی‌صدا (V/UV detector[۳]) برای تخمین و استخراج نویز استفاده شده است. سپس به مقایسه روش‌های پیشنهادی و روش‌های موجود پرداخته شده و مشاهده شده است که روش‌های پیشنهادی به میزان چشمگیری، نسبت سیگنال به نویز در سیگنال‌های صوتی آغشته به نویز را بهبود بخشیده‌اند. نتیجهٔ آزمایش شنوایی نیز بیان‌گر این نتایج هستند.

واژه‌های کلیدی: غنی‌سازی سیگنال‌های صوتی، تفریق طیفی، آنالیز LPC ، VAD، V/UV detector.

A New Method for Speech Enhancement with LPC Analysis in Spectral Subtraction, Multi Band Spectral and Inverse Fourier Spectral Subtraction

M. Heidari, Faculty of Engineering-Shomal University, Amol-Iran
M. R. Karami Mollaei, Faculty of Electrical and Computer Engineering-Babol University of technology, Babol-Iran

Abstract: In this paper 3 new methods for the speech Signal enhancement were proposed, which are based on spectral subtraction, inverse Fourier transform subtraction and multi-band spectral subtraction methods. We use linear predicative coding (*LPC*), *VAD* analysis, and voice/unvoiced (V/UV) detector for noise estimation and extraction, then we compare the proposed methods with the previous ones. Our proposed method have improved signal to noise ratio and, also, good results have been achieved in the auditory tests.

Keywords: Speech enhancement, Spectral subtraction, LPC analysis, VAD, V/UV detector

۱- مقدمه

کاهش و یا حذف نویز، یکی از مباحث مهم درسیستم‌های پردازش سیگنال‌های صوتی مانند سیستم‌های ارتباطی، کدینگ سیگنال‌های صوتی و تشخیص صوت می‌باشد. به همین منظور روش‌های زیادی برای کاهش میزان نویز در سیگنال‌های صوتی ارائه شده است. از این میان می‌توان به روش‌های مبتنی بر تفریق طیفی (پایه و چند بانده) [۵، ۶، ۷ و ۸]، فیلتر وفقی [۱۱]، فیلتر وینر [۹]، تبدیل موجک [۱۰، ۱۲،۱۳،۱۴ و ۱۵] اشاره نمود.

در روش‌های مبتنی بر تفریق طیفی سه فرض می‌بایست برقرار باشد:

۱- نویز و سیگنال جمع شونده باشند.

۲- نویز و سیگنال ناهمبسته باشند.

۳- یک کانال در دسترس باشد.

روش تفریق طیفی پایه، اگرچه بسیار ساده و کارامد می‌باشد اما سبب ایجاد نویز جدیدی به نام نویز موزیکال می‌گردد که جهت کاهش این نویز، از روش تفریق طیفی با استفاده از کف طیفی و تفریق بیش از حد که توسط بروتی [۶] ارائه شده است، استفاده می‌گردد. بعدها کاماث و لویی‌زو [۵]، روش تفریق طیفی چند بانده را ارائه نموده‌اند. در این روش، ابتدا سیگنال گفتاری آغشته به نویز به چندین باند فرکانسی تقسیم می‌گردد و سپس روش تفریق طیفی در هر باند انجام می‌شود.

همان طور که پیش از این گفته شده است، فرض بر این است که سیگنال و نویز ناهمبسته‌اند اما در طبیعت این امر به ندرت اتفاق می‌افتد. لذا روش تفریق طیفی معکوس فوریه ارائه شده است که همانند روش تفریق طیفی می‌باشد با این تفاوت که در آن عمل تفریق بر روی معکوس تبدیل فوریه انجام می‌گیرد. در این روش تا حدودی مسئله وابستگی سیگنال و نویز حل می‌شود. روش‌های دیگری نظیر تفریق کپسترال و تبدیل موجک [۱۰، ۱۲،۱۳،۱۴ و ۱۵] نیز برای حذف و یا کاهش نویز وجود دارند.

در این مقاله ابتدا روش‌های تفریق طیفی پایه، تفریق طیفی چند بانده و تفریق طیفی معکوس فوریه را شرح داده و سپس با استفاده از آنالیز LPC [۱، ۲، ۳ و ۴] به تخمین نویز از روی سیگنال صوتی پرداخته می‌شود. مشاهده شده است که پارامترهای نویز تخمینی به میزان زیادی به پارامترهای نویز واقعی نزدیک است و به همین دلیل باعث بهبود چشمگیر روش‌های تفریق طیفی، تفریق طیفی چند بانده و تفریق طیفی معکوس فوریه گردیده است. اما در گام بعدی به دنبال بهبود هرچه بیشتر روش پیشنهادی پرداخته‌ایم و از الگوریتم‌های VAD [۱۶ و ۱۷] برای تشخیص قسمت‌های سکوت بهره جسته‌ایم که این امر به بهبود بیشتر نتایج منجر گردیده است.

۲- تفریق طیفی توان[۴] (PSS)

با فرض این که که سیگنال و نویز جمع‌شونده هستند می‌توان یک سیگنال صوتی آغشته به نویز را به‌صورت زیر بیان داشت:

$$x(n) = s(n) + n(n) \qquad (۱)$$

که $x(n)$ سیگنال صوتی آغشته به نویز، $s(n)$ سیگنال صوتی تمیز و $n(n)$ سیگنال تصادفی نویز می‌باشد.

با فرض سفید بودن نویز و ناهمبسته بودن نویز و سیگنال می‌توان نوشت [۱]:

$$R_n(\tau) = D_0\delta(\tau) \qquad (۲)$$

$$R_{s,n}(\tau) = 0 \qquad (۳)$$

D_0 عددی ثابت، $R_n(\tau)$ تابع خود همبستگی سیگنال تصادفی نویز و $R_{s,n}(\tau)$ نیز تابع همبستگی متقابل سیگنال‌های n و s می‌باشند. با توجه به روابط فوق و با فرض ایستا بودن سیگنال‌های n و s می‌توان نوشت:

$$\Gamma_x(\omega) = \Gamma_s(\omega) + \Gamma_n(\omega) \qquad (۴)$$

که Γ_x، Γ_s و Γ_n به‌ترتیب چگالی طیف قدرت x , s و n می‌باشند. بنابر رابطه (۴) چنانچه چگالی طیف قدرت سیگنال تصادفی نویز تخمین زده شود می‌توان چگالی طیف قدرت سیگنال صوتی تمیز را از روی رابطه (۵) تخمین زد.

$$\hat{\Gamma}_s(\omega) = \Gamma_x(\omega) - \hat{\Gamma}_n(\omega) \qquad (۵)$$

$\hat{\Gamma}_s(\omega)$ و $\hat{\Gamma}_n(\omega)$ ، به‌ترتیب، تخمینی از $\Gamma_s(\omega)$ و $\Gamma_n(\omega)$ می‌باشند.

رابطه (۴) و به دنبال آن رابطه (۵) با فرض ایستا بودن سیگنال صوتی تمیز و نویز برقرار می‌باشند، اما در طبیعت چنین فرضی همواره برقرار نمی‌باشد.

از آنجایی که سیگنال‌های صوتی تمیز در بازه‌های زمانی کوتاه ایستایی محلی دارند و نیز فرض ایستا بودن نویز در بازه‌های کوتاه منطقی‌تر می‌باشد، لذا، ابتدا روی سیگنال صوتی آغشته به نویز عمل

پنجره‌گذاری انجام می‌شود. بدین طریق سیگنال‌های گفتار به فریم‌های کوتاه تقسیم می‌گردد و سپس عمل تفریق طیفی بر روی هر فریم انجام می‌گیرد. با در نظر گرفتن m به‌عنوان شماره پنجره داریم:

$$x(n;m) = s(n;m) + n(n;m) \quad (۶)$$

$$R_n(\tau;m) = D_0\delta(\tau) \quad (۷)$$

$$R_{s,n}(\tau;m) = 0 \quad (۸)$$

که $x(n;m)$ سیگنال پنجره شده سیگنال صوتی $x(n)$ می‌باشد. با محاسبه تابع چگالی طیفی از طرفین رابطه (۶) داریم:

$$\Gamma_s(\omega;m) = \Gamma_x(\omega;m) - \Gamma_n(\omega;m) \quad (۹)$$

از طرفی نیز می‌دانیم [۱]:

$$\Gamma_x(\omega;m) = \frac{X(\omega;m)X^*(\omega;m)}{N^2} = \frac{|X(\omega;m)|^2}{N^2} \quad (۱۰)$$

که N طول پنجره و X سیگنال صوتی می‌باشد. به دلیل بزرگی مقدار $|X(\omega;m)|^2$ در مقابل مخرج، به راحتی می‌توان از ضریب $1/N^2$ چشم‌پوشی نموده و نوشت:

$$\Gamma_x(\omega;m) = |X(\omega;m)|^2 \quad (۱۱)$$

با توجه به روابط (۱۱) و (۹)، رابطه زیر حاصل می‌گردد:

$$|S(\omega;m)|^2 = |X(\omega;m)|^2 - |N(\omega;m)|^2 \quad (۱۲)$$

که در روابط فوق $|X(\omega;m)|$ اندازه تبدیل فوریه سیگنال پنجره شده $x(n)$ و $|S(\omega;m)|$ و $|N(\omega;m)|$، به‌ترتیب، اندازهٔ تبدیل فوریه سیگنال صوتی تمیز پنجره شده و سیگنال نویز پنجره شده می‌باشند. همان طور که از رابطه (۱۲) پیداست، برای به‌دست آوردن اندازه تبدیل فوریه سیگنال تمیز می‌بایست اندازه تبدیل فوریه سیگنال تصادفی نویز را نیز داشت. بدین منظور سیگنال تصادفی نویز را از قسمت سکوت سیگنال صوتی آغشته به نویز، تخمین می‌زنند. دلیل این روش تخمین نویز، بر این ایده استوار است که چون در قسمت سکوت هیچ سیگنال صوتی وجود ندارد، هر آنچه که در این قسمت از سیگنال صوتی آغشته به نویز وجود دارد ناشی از نویز است و عموماً اولین فریم صوتی را به عنوان نویز در نظر می‌گیرند.

اینک برای دستیابی به سیگنال صوتی تمیز در حوزه زمان لازم است که علاوه بر اندازه تبدیل فوریه، فاز آن را نیز به‌دست آورده و با معکوس فوریه زمان کوتاه[۵] (ST FFT) به سیگنال صوتی در حوزه زمان رسید.

در تمامی کاربردهای عملی می‌توان فاز سیگنال صوتی تمیز را همان فاز سیگنال صوتی آغشته به نویز در نظر گرفت [۸]:

$$\phi_{\hat{S}(\omega;m)} = \phi_{X(\omega;m)} \quad (۱۳)$$

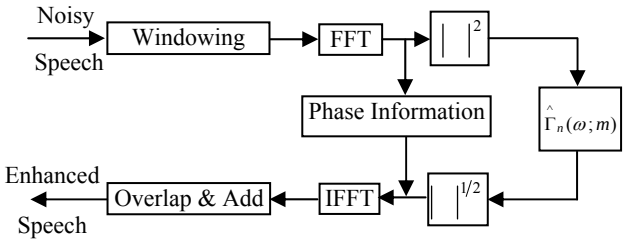
این بدان معناست که تأثیرگذاری نویز، بر روی فاز سیگنال‌های صوتی برای گوش انسان چندان محسوس نمی‌باشد.

با توجه به معادلات (۱۲) و (۱۳) سیگنال صوتی تمیز به‌صورت زیر تخمین زده می‌شود:

$$\begin{aligned}\hat{S}(\omega;m) &= \left|\hat{S}(\omega;m)\right| \exp\left\{i\phi_{\hat{S}(\omega;m)}\right\} \\ &= \left[|X(\omega;m)|^2 - \left|\hat{N}(\omega;m)\right|^2\right]^{1/2} \exp\left\{i\phi_{\hat{S}(\omega;m)}\right\}\end{aligned} \quad (۱۴)$$

که در روابط فوق $\hat{S}(\omega;m)$ و $\hat{N}(\omega;m)$ به ترتیب تبدیل فوریه سیگنال تمیز تخمینی و تبدیل فوریه سیگنال نویز تخمینی می‌باشند.

به روش فوق، روش تفریق طیفی توان (PSS) گفته می‌شود (شکل ۱). چرا که از توان دوم اندازه تبدیل فوریه که مبین قدرت و توان سیگنال می‌باشد استفاده شده است. عموماً از یک ضریب توانی دیگری غیر از ۲، در روش تفریق طیفی استفاده می‌شود که مقدار آن را با بهینه‌سازی محاسبه می‌نمایند. در این حالت به روش فوق، روش تفریق طیفی تعمیم یافته[۶] (GSS) [۱] می‌گویند.



شکل (۱): بلوک دیاگرام روش تفریق طیفی توان (PSS) [1]

$$\hat{S}(\omega;m)=\left[|X(\omega;m)|^{a}-\left|\hat{N}(\omega;m)\right|^{a}\right]^{1/a}\exp\left\{i\phi_{\hat{S}(\omega;m)}\right\} \quad (۱۵)$$

اما مسئله مهمی که در روش تفریق طیفی پیش می‌آید، مقادیر منفی برای اندازه تبدیل فوریه سیگنال تمیز می‌باشد. به عبارت دیگر هیچ تضمینی برای مثبت بودن اندازه تبدیل فوریه محاسبه شده برای سیگنال صوتی تمیز در هر یک از روابط (۱۵) و (۱۴) وجود ندارد. دو روش برای اصلاح این مقادیر منفی وجود دارد [۱]:

الف) اصلاح نیم موج:

$$\left|\hat{S}(\omega;m)\right|=\begin{cases}\left|\hat{S}(\omega;m)\right| & if\ \left|\hat{S}(\omega;m)\right|>0\\ 0 & elsewhere\end{cases} \quad (۱۶)$$

ب) اصلاح تمام موج:

$$\left|\hat{S}(\omega;m)\right|=abs\left\{\left|\hat{S}(\omega;m)\right|\right\} \quad (۱۷)$$

۲-۱- اصلاح روش تفریق طیفی

همان طور که پیش از این گفته شد، روش تفریق طیفی سبب ایجاد نویز جدیدی به نام نویز موزیکال خواهد شد. بروتی [۶] برای کاهش این نویز روشی را پیشنهاد کرده است که به روش تفریق طیفی با استفاده از کف طیفی و تفریق بیش از حد موسوم است. در این روش داریم:

$$\left|\hat{S}(\omega;m)\right|^{a}=|X(\omega;m)|^{a}-\alpha\left|\hat{N}(\omega;m)\right|^{a} \quad (۱۸)$$

$$\left|\hat{S}(\omega;m)\right|^{a}=\begin{cases}\left|\hat{S}(\omega;m)\right|^{a} & if\ \left|\hat{S}(\omega;m)\right|^{a}>\beta\left|\hat{N}(\omega;m)\right|^{a}\\ \beta\left|\hat{S}(\omega;m)\right|^{a} & elsewhere\end{cases} \quad (۱۹)$$

که α فاکتور *Over subtraction* بوده و تابعی از نسبت سیگنال به نویز($SNR^{۷}$) می‌باشد و به‌صورت زیر محاسبه می‌گردد:

$$\alpha=\begin{cases}\alpha_0+3/4 & SNR\le -5\,dB\\ \alpha_0-3/20 & -5\,dB<SNR<20\,dB\\ \alpha_0-3 & SNR\ge 20\,dB\end{cases} \quad (۲۰)$$

که α_0 مقدار مطلوب α در $SNR=0$ می‌باشد. بروتی [۶] مقدار آن را در محدودهٔ $3\le\alpha_0\le 6$ پیشنهاد کرده و در این مقاله مقدار $\alpha_0=4$ برای آن در نظر گرفته شده است.

اما همان طور که از رابطه (۲۰) مشخص است، قبل از استفاده از این روش، برای محاسبه ضریب α در هر پنجره؛ به مقدار SNR در آن پنجره احتیاج می‌باشد. نسبت سیگنال به نویز در هر فریم برحسب dB به‌صورت زیر محاسبه می‌شود:

$$SNR_i=10\log\frac{\sum_{\omega=b_i}^{e_i}\left|X(\omega;m_i)\right|^{2}}{\sum_{\omega=b_i}^{e_i}\left|\hat{N}(\omega;m_i)\right|^{2}} \quad (۲۱)$$

b_i و e_i به ترتیب فرکانس‌های ابتدایی و انتهایی در فریم i ام می‌باشند.

طبق پیشنهاد کاماث و لویی‌زو [۵]، $\beta<<1$ است (β فاکتور تفریق بیش از حد[۸] نامیده می‌شود).

۳- تفریق طیفی چند بانده[۹] (MBSS)

همان طور که پیش از این نیز بیان شده است، در روش تفریق طیفی پایه (PSS و GSS) فرض بر آن است که تأثیر نویز در سراسر سیگنال به یک میزان است اما وقوع چنین امری در طبیعت به ندرت اتفاق می‌افتد. چرا که علاوه بر وجود منابع مختلف نویز، حقیقت دیگری نیز وجود دارد و آن این که نویز در برخی از فرکانس‌ها، بیشتر از بقیهٔ فرکانس‌ها بر روی سیگنال گفتاری اثر می‌گذارد [۷]. این وابستگی اثر نویز به فرکانس، ما را به این حقیقت رهنمون می‌سازد که نویز بر روی حروف صدادار و بی‌صدا نیز به یک میزان اثر نمی‌گذارد [۷]. شکل (۲)، SNR را در هر فریم و در چهار باند فرکانسی برای عبارت *'one four seven three'* که توسط یک گوینده زن تلفظ شده است را نشان می‌دهد.

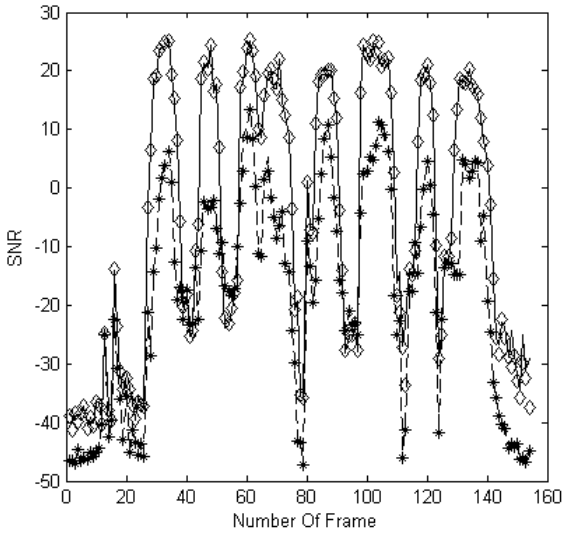
بنابر دلایل ذکر شده، کاماث و لویی‌زو [۵] پیشنهاد استفاده از روش تفریق طیفی بروتی [۶] را در چندین باند فرکانسی ارائه نموده‌اند. آن‌ها هر فریم صوتی را به چندین باند در حوزهٔ فرکانس تقسیم نموده و سپس روابط (۱۵) الی (۱۹) را در هر باند فرکانسی بر روی سیگنال پنجره شده اعمال نموده‌اند. در این روش میزان نویز موزیکال نیز کاهش می‌یابد.

اما آنچه که در این روش مهم می‌باشد نحوهٔ محاسبهٔ نسبت سیگنال به نویز در هر فریم و در هر باند فرکانسی از فریم مورد نظر

می‌باشد. مقدار SNR در هر فریم و باند فرکانسی به صورت زیر محاسبه می‌گردد:

$$SNR_{i,j} = 10\log\frac{\sum_{w=b_i}^{e_i}\left|X(\omega;m_i)\right|^2}{\sum_{w=b_i}^{e_i}\left|\hat{N}(\omega;m_i)\right|^2} \qquad (۲۲)$$

$SNR_{i,j}$، مقدار SNR در باند فرکانسی j ام از فریم iام، $\left|X(\omega;m_i)\right|^2$ مربع اندازه تبدیل فوریهٔ سیگنال آغشته به نویز در فریم iام، $\left|\hat{N}(\omega;m_i)\right|^2$ مربع اندازه تبدیل فوریه نویز تخمین زده شده در فریم iام و b_i و e_i به ترتیب فرکانس‌های ابتدایی و انتهایی در باند فرکانسی j ام می‌باشند.



شکل (۲): مقدار SNR (dB) مقطعی در۴ باند فرکانسی در هر فریم [۷]

۳-۱- اصلاح روش تفریق طیفی چند بانده (IMBSS)[۱۰]

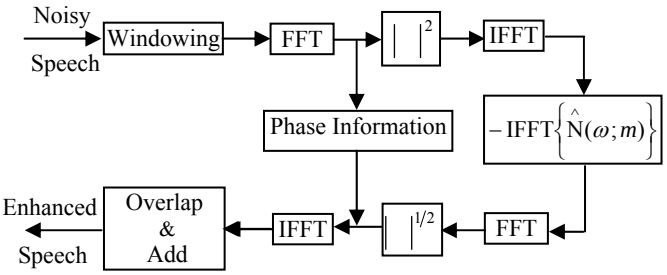
همانند روش GSS می‌توان برای بهبود نسبت سیگنال به نویزِ سیگنالِ غنی شده، به جای توان ۲ از توان α استفاده نمود. از طرف دیگر چون در این روش نیز از فریم سکوت برای تخمین نویز استفاده شده است، که عموماً همان فریم اول را به عنوان فریم سکوت در نظر می‌گیرند، برای افزایش دقت در تخمین نویز و نیز کاهش تفاوت طیف تخمینی نویز با طیف واقعی آن، از طیف تخمینی باندهای فرکانسی در فریم سکوت متوسط‌گیری به عمل آمده و این مقدار جایگزین تخمین اولیه می‌شود.

$$\left|\hat{N}(\omega;m)\right|^a = \left(mean\left(abs\left(\hat{N}(\omega;m)\right)\right)\right)^a \qquad (۲۳)$$

۴- تفریق طیفی معکوس فوریه (IFSS)[۱۱]

در روش تفریق طیفی [۵، ۶، ۷ و ۸] فرض بر این است که نویز و سیگنال ناهمبسته باشند. این شرط به واسطه اعمال تابع خود همبستگی بر دو طرف رابطه (۱)، به‌وجود می‌آید. حال اگر به نحوی از الزام برقراری رابطه (۴) و یا (۹) در روش تفریق طیفی کاسته شود، به همان میزان از الزام فرض ناهمبستگی سیگنال و نویز کاسته می‌شود. در روش تفریق طیفی معکوس فوریه، از معکوس فوریه اندازه تبدیل فوریه سیگنال آغشته به نویز و سیگنال نویز تخمینی برای عمل تفریق استفاده می‌گردد.

به طور شهودی می‌توان گفت که در روش تفریق معکوس فوریه عمل تفریق در حوزه زمان انجام شده است که در آن شرط ناهمبستگی نویز و سیگنال از الزام کمتری برخوردار می‌باشد. زیرا معمولا نویز با سیگنال در حوزه زمان جمع می‌شود که الزام مستقل بودن در آن وجود ندارد ولی جمع شدن در حوزه فرکانس الزام استقلال را دارد. در این روش سیگنال صوتی تمیز تخمینی طبق شکل (۳) به دست می‌آید.



شکل (۳): بلوک دیاگرام روش تفریق طیفی معکوس فوریه (IFSS) [۱]

۵- ضرایب پیشگویی خطی (LPC)

از آنجایی که در الگوریتم پیشنهادی ما، از آنالیز LPC [۱، ۲، ۳ و ۴] برای تخمین نویز استفاده شده است، لذا در این بخش به اختصار به توضیح این آنالیز می‌پردازیم.

ضرایب پیشگویی خطی یکی از قدرتمندترین ابزارها در پردازش گفتار می‌باشد [۲]. ایده کلی این آنالیز این است که هر نمونه از سیگنال صوتی را می‌توان به‌صورت معادله‌ای خطی برحسب خروجی‌ها و ورودی‌های قبلی نوشت:

$$s(n)=\sum_{k=1}^{p} a_k\, s(n-k)+\sum_{l=0}^{q} b_l\, u(n-l) \qquad (۲۴)$$

a_k و b_l به ترتیب ضرایب مخرج و صورت فیلتر می‌باشند و u(n) سیگنال ورودی است که برای حروف صدادار یک قطار ضربه و برای حروف بی‌صدا یک رشته نویز تصادفی می‌باشد [۱، ۲، ۳ و ۴]. تابع تبدیل سیستم با به کار بردن تبدیل z روی معادله (۲۴) قابل حصول می‌باشد.

$$H(z)=\frac{S(z)}{U(z)}=\frac{\sum_{l=0}^{q} b_l z^{-l}}{1-\sum_{k=1}^{p} a_k z^{-k}} \qquad (۲۵)$$

در عمل برای سیگنال‌های صوتی، یک مدل تمام قطب تقریب بسیار خوبی برای تابع تبدیل H(z) است [۱] و می‌توان آن را به‌صورت زیر بیان داشت:

$$H(z)=\frac{1}{1-\sum_{k=1}^{p} a_k z^{-k}}=\frac{1}{A(z)} \qquad (۲۶)$$

برای سیگنال صحبت انسان، p یک عدد صحیح در محدوده $10 \le p \le 14$ انتخاب می‌شود.

نکته اساسی در محاسبه ضرایب پیشگویی خطی این است که این ضرایب بایستی مستقیماً از سیگنال صحبت به‌دست آیند به همین منظور و به علت ماهیت تغییرپذیری سیگنال گفتار (با زمان)، ابتدا عمل پنجره‌گذاری بر روی سیگنال انجام می‌شود و سپس ضرایب LPC در فریم‌های کوتاه محاسبه می‌شوند [۲].

۵-۱- تخمین نویز

با توجه به توضیحات ارائه شده، با تقریب خوبی می‌توان هر نمونه از سیگنال صوتی را تنها با p نمونهٔ قبلی از همان سیگنال صوتی (بدون استفاده از p نمونهٔ ماقبل ورودی) محاسبه نمود [۱]:

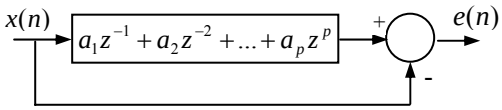
$$\hat{s}(n)=\sum_{k=1}^{p} a_k s(n-k) \qquad (۲۷)$$

سیگنال خطا در حقیقت، اختلاف بین سیگنال صوتی اصلی و سیگنال صوتی تخمین زده شده از روی p نمونهٔ ماقبل می‌باشد:

$$e(n)=s(n)-\hat{s}(n)=s(n)-\sum_{k=1}^{p} a_k s(n-k) \qquad (۲۸)$$

که اگر از طرفین رابطهٔ فوق تبدیل z گرفته شود داریم:

$$E(z)=S(z)\left[1-\sum_{k=1}^{p} a_k z^{-k}\right]=A(z)S(z) \qquad (۲۹)$$

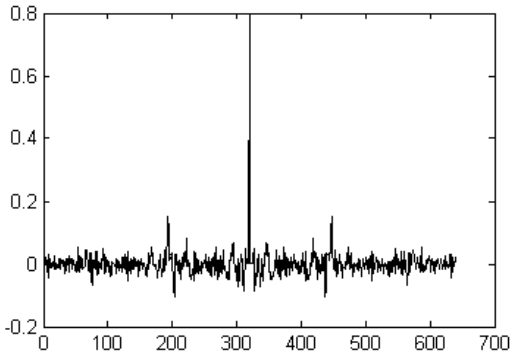


شکل (۴): بلوک دیاگرام نحوهٔ محاسبهٔ تابع خطا

تابع E(z)، تبدیل z سیگنال خطا می‌باشد که دارای ماهیت نویزی است. می‌توان گفت که فیلتر خطی، قسمت غیر وابسته سیگنال را جدا می‌کند که قسمت اعظم آن نویز می‌باشد. برای اثبات این ادعا کافی است تابع خود همبستگی سیگنال e(n) را محاسبه نماییم.

در شکل (۵) تابع خودهمبستگی سیگنال خطا که متعلق به یک سیگنال صوتی از پایگاه داده‌های Timit می‌باشد رسم گردیده است. همان‌طور که مشاهده می‌شود، سیگنال e(n) دارای ماهیت نویزی می‌باشد چرا که سیگنال خودهمبستگی آن همانند تابع خودهمبستگی سیگنال تصادفی نویز است.

ما در الگوریتم پیشنهادی خود، از این سیگنال برای تخمین نویز استفاده نموده‌ایم که منجر به بهبود بسیار زیاد SNR سیگنال‌های صوتی آغشته به نویز (در مقایسه با روش‌های موجود) گردیده است.



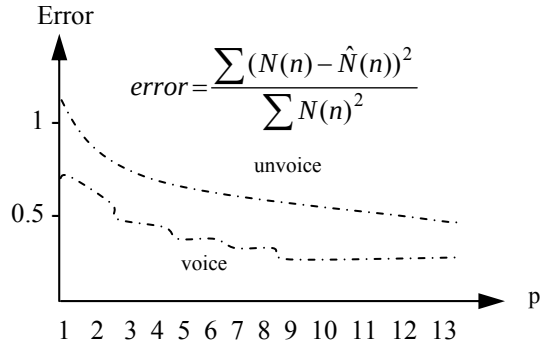
شکل (۵): تابع خودهمبستگی سیگنال خطا

البته سیگنال e(n) اگرچه همان سیگنال نویز اضافه شده به سیگنال صوتی تمیز نمی‌باشد اما اکثر مشخصات آن را در بر دارد. از طرفی نیز در خروجی فیلتر A(z)، مقداری از سیگنال ناهمبسته مربوط به سیگنال گفتار نیز وجود دارد که در مقایسه با نویز قابل چشم‌پوشی است.

۵-۲- اصلاح خروجی فیلتر A(z) با استفاده از الگوریتم‌های VAD و V/UV detector

همان طور که پیش از این بیان شده است، در روش‌های تفریق طیفی [۵، ۶، ۷ و ۸] و تفریق طیفی معکوس فرض بر آن است که تأثیر نویز در سراسر سیگنال به یک میزان است، در صورتی که وقوع چنین امری در طبیعت به ندرت اتفاق می‌افتد. چرا که علاوه بر وجود منابع مختلف نویز، حقیقت دیگری نیز وجود دارد و آن این که نویز در برخی از فرکانس‌ها، بیشتر از بقیهٔ فرکانس‌ها بر روی سیگنال گفتاری اثر می‌گذارد [۷]. این وابستگی نویز به فرکانس، ما را به این حقیقت رهنمون می‌سازد که نویز بر روی حروف صدادار و بی‌صدا نیز به یک میزان اثر نمی‌گذارد [۷]. به همین دلیل، به دنبال روشی هستیم تا با تفکیک فریم‌های صدادار و بی‌صدا [۱۶، ۱۷ و ۱۸]، با دقت بیشتری این تأثیر متفاوت نویز را در نظر داشته باشد.

علاوه بر مطالب بیان شده، چون ما در روش‌های پیشنهادی خود سعی داریم از آنالیز LPC برای تخمین نویز استفاده نماییم، در نظر گرفتن این نکته ضروری است که نویز تخمینی در فریم‌های متعلق به حروف صدادار، به نویز واقعی نزدیک‌تر است تا در فریم‌های متعلق به حروف بی‌صدا. شکل (۶) نمایشی از میزان این خطا را برای حروف صدادار و بی‌صدا نسبت به تعداد قطب‌های فیلتر $H(z)$، p، نشان می‌دهد.



شکل (۶): خطای خروجی فیلتر تمام قطب $H(z)$ برای حروف صدادار و بی‌صدا نسبت به تعداد قطب‌های فیلتر

حقیقت دیگر این است که یک فریم سکوت تنها حاوی نویز است. وقتی این فریم از فیلتر $A(z)$ ،گذر می‌کند و تا حدودی تضعیف می‌گردد. لذا می‌بایست در یک بهره تقویت‌کننده، ضرب شود و به منظور کاهش میزان خطا در حروف بی‌صدا، نویز تخمینی بایستی تضعیف گردد.

$$\hat{N}(\omega\ ;m) \leftarrow \gamma\ \hat{N}(\omega;m) \qquad (۳۰)$$

عبارت فوق بیان می‌دارد مقدار نویز تخمینی بایستی با یک مقدار تضعیف و یا تقویت شده جایگزین گردد که γ(ضریب تضیف و یا تقویت) با توجه به مطالب فوق؛ برای فریم‌های سکوت، صدادار و بی‌صدا متفاوت می‌باشد.

۶- بهبود الگوریتم‌ها با استفاده از روش پیشنهادی تخمین نویز

۶-۱- تفریق طیفی با استفاده از آنالیز LPC [۱۲](LPSS)

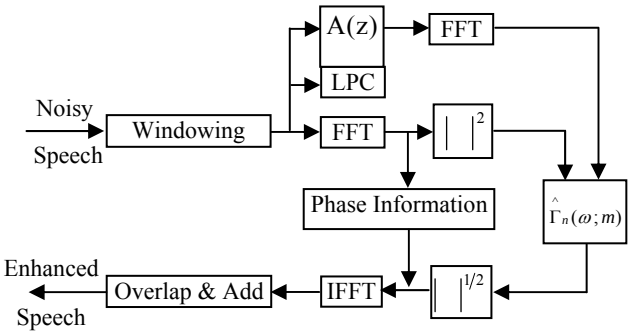
همان طور که در بخش (۲) بیان شده است، در روش‌های تفریق طیفی و تفریق طیفی معکوس، نویز از قسمت سکوت سیگنال تخمین زده می‌شود که عموماً از فریم اول سیگنال آغشته به نویز، به عنوان قسمت سکوت سیگنال استفاده می‌گردد. این روش تخمین نویز نیازمند دو فرض اساسی می‌باشد:

۱- فریم اول سیگنال صوتی واقعاً قسمت سکوت سیگنال باشد.

۲- نویز در سراسر سیگنال به یک اندازه تاثیر گذاشته باشد.

در عمل لزوماً دو شرط فوق برقرار نمی‌باشد. به خصوص در مورد فرض دوم که با توجه به منابع متعدد نویز و نیز تصادفی بودن آن عموماً میزان نویز در تمام طول سیگنال به یک اندازه نمی‌باشد. به همین منظور ما در روش پیشنهادی خود، از هر فریم و با توجه به مشخصات آن فریم برای تخمین نویز در آن استفاده نموده‌ایم. بدین طریق مسئله تغییر میزان نویز در طول سیگنال حل می‌شود.

بدین منظور ابتدا سیگنال صوتی به فریم‌های کوچک‌تر (10ms-20ms) تقسیم و سپس ضرایب LP هر یک از فریم‌ها محاسبه می‌گردد. در مرحله بعد سیگنال صوتی از فیلتر $A(z)$ عبور داده می‌شود. در نهایت از خروجی فیلتر $A(z)$ که مبین اختلاف سیگنال اصلی و سیگنال تخمینی و بیشتر از سیگنال اصلی دارای ماهیت نویزی می‌باشد، به عنوان نویز تخمینی استفاده می‌گردد. در شکل (۷) بلوک دیاگرام روش تفریق طیفی با استفاده از آنالیز LPC آمده است.



شکل (۷) بلوک دیاگرام روش پیشنهادی LPSS با استفاده از آنالیز LPC

۶-۱-۱- تفریق طیفی با استفاده از آنالیز LPC و الگوریتم‌های VAD و V/UV detector (Advanced LPSS)

همان طور که بیان شده است، نویز بر روی فریم‌های سکوت، بی‌صدا و صدادار به یک میزان اثر نمی‌گذارد [۷]. لذا روش LPSS بدین صورت اصلاح می‌گردد:

$$\hat{s}(n;m) = IFFT\{[|X(\omega;m)|^{a} - \gamma a \times |\hat{N}(\omega;m)|^{a}]^{1/a} \exp\{i\phi_{s(\omega;m_{kj})}\}\} \quad (۳۱)$$

$$\gamma = \begin{cases} 1.1 & Pause \\ 1 & voice \\ 0.9 & unvoice \end{cases} \quad (۳۲)$$

در روش پیشنهادی برای اصلاح مقادیر اندازه منفی، از روش اصلاح شده نیم‌موج استفاده نموده‌ایم.

۶-۲- تفریق طیفی چند بانده با استفاده از آنالیز LPC و الگوریتم‌های VAD و V/UV detector (LPMBSS و LPIBSS)

در روش تفریق طیفی چند بانده، گرچه به این مسئله که نویز در برخی از فرکانس‌ها، بیشتر از بقیهٔ فرکانس‌ها بر روی سیگنال گفتاری اثر می‌گذارد [۷] توجه شده است اما به تغییر میزان نویز اضافه شده در طول سیگنال و نیز تاثیر متفاوت آن بر روی حروف صدادار و بی‌صدا توجهی نشده است. بدین منظور، در روش پیشنهادی ابتدا سیگنال صوتی به فریم‌های زمانی کوچک‌تر تقسیم و ضرایب LP و خروجی فیلتر $A(z)$ در هر فریم محاسبه گردیده و به عنوان نویز تخمینی در هر فریم در نظر گرفته می‌شود. سپس هر فریم به چندین باند فرکانسی تقسیم و از این نویز تخمینی با توجه به نوع فریم (سکوت، صدادار و بی‌صدا) در غنی‌سازی سیگنال صوتی استفاده می‌گردد. بدین ترتیب روش‌های MBSS و IMBSS با استفاده از رابطه (۳۳) اصلاح می‌گردند:

$$\hat{s}(n;m_{kj}) = IFFT\{[|X(\omega;m_{kj})|^{a} - \gamma a \times |\hat{N}(\omega;m_{kj})|^{a}]^{1/a} \exp\{i\phi_{s(\omega;m_{kj})}\}\} \quad (۳۳)$$

$\hat{s}(n;m_{kj})$ سیگنال تمیز تخمینی در باند فرکانسی jام از فریم kام و $|X(\omega;m_{kj})|$ و $\hat{N}(\omega;m_{kj})$ به ترتیب اندازه تبدیل فوریه سیگنال صوتی آغشته به نویز و سیگنال نویز تصادفی در باند فرکانسی jام از فریم kام می‌باشد.

بقیه مراحل نظیر محاسبه α و اصلاح مقادیر منفی همانند روش چندبانده شرح داده شده در بخش (۲) می‌باشد.

۶-۳- تفریق معکوس فوریه با استفاده از آنالیز LPC و V/UV detector، VAD (LPIFFS)

همان طور که از شکل (۲) مشخص است، در روش تفریق طیفی معکوس فوریه، همانند روش تفریق طیفی [۵، ۶، ۷ و ۸]، به تخمینی از نویز احتیاج می‌باشد که برای تخمین سیگنال نویز، از قسمت سکوت سیگنال گفتاری آغشته به نویز استفاده می‌گردد. عموماً فریم اول سیگنال گفتاری به عنوان قسمت سکوت سیگنال در نظر گرفته می‌شود، در این روش فرض بر این است که:

۱- فریم اول سیگنال آغشته به نویز، جزء قسمت سیگنال می‌باشد.

۲- نویز در تمام طول سیگنال، به یک اندازه اثر گذاشته باشد.

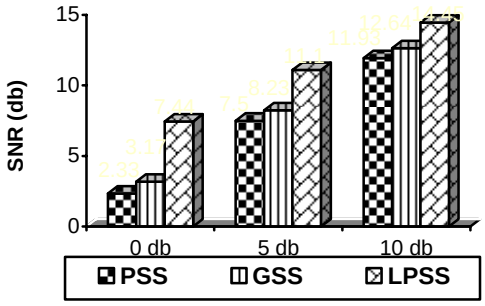
برای بهبود روش فوق پیشنهاد می‌شود که به جای تخمین نویز از قسمت سکوت سیگنال، ابتدا آن را از $A(z)$ عبور داده و سپس خروجی فیلتر را به عنوان نویز در نظر گرفته و به الگورتیم تفریق طیفی معکوس فوریه اعمال شود. چرا که خروجی این فیلتر به نویز نزدیکتر است تا تخمین اولیه نویز. از طرف دیگر برای رفع مسئله تغییرپذیری میزان تاثیر نویز در طول سیگنال صوتی آغشته به نویز، نویز تخمینی در هر فریم را از روی نزدیکترین فریم سکوت به‌دست می‌آوریم. به منظور نزدیکی هر چه بیشتر نویز تخمینی با نویز واقعی، پیشنهاد می‌شود که از تبدیل فوریه تخمینی متوسط گرفته شود.

۷- پیاده‌سازی و مقایسه روش‌ها

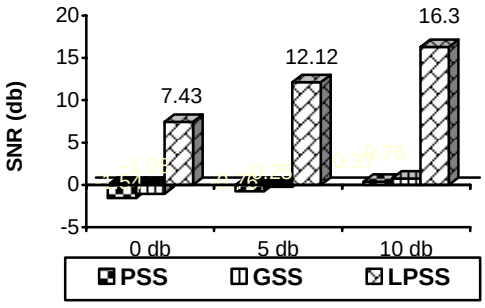
در این قسمت به مقایسه روش‌های پیشنهادی با روش‌های موجود می‌پردازیم. بدین منظور در نمودار (۱) ابتدا روش‌های PSS، GSS و LPSS را به ازای $SNR = 0, 5, 10\ dB$ (نویز از نوع سفیدگوسی) با یکدیگر مقایسه نموده‌ایم تا میزان توانایی روش تخمین نویز پیشنهادی در بهبود نسبت سیگنال به نویز سیگنال‌های صوتی در تمامی محدوده *SNR* های اولیه کم تا خوب، مشخص گردد.

شکل (۸)، برتری و قدرت روش LPSS را در کاهش نویز نشان می‌دهد.

همان طور که پیش از این نیز بیان شده است در روش‌های موجود، عموماً از فریم اول به عنوان قسمت سکوت استفاده می‌گردد. ضعف شدید روش‌های موجود هنگامی آشکار می‌گردد که فریم اول، یک فریم سکوت نباشد. شکل (۹) به مقایسه روش LPSS با روش‌های PSS و GSS به ازای چنین سیگنال‌های صوتی از پایگاه داده‌های *Timit* پرداخته و برتری کاملاً مشهود روش تخمین نویز پیشنهادی در مورد این گونه سیگنال‌ها نمایش داده شده است.



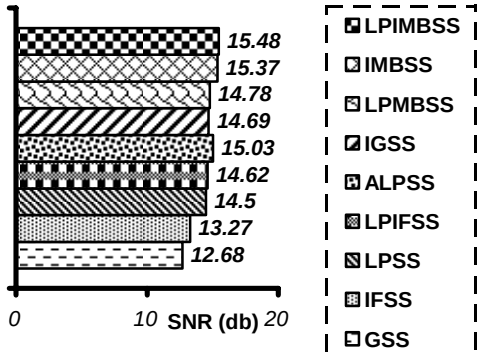
شکل (۸): مقایسه روش‌های PSS، GSS و LPSS به ازای SNR = 0,5,10 dB (نویز از نوع سفیدگوسی)



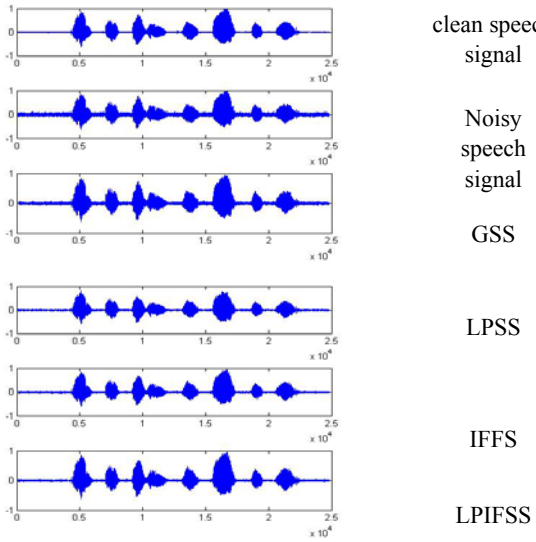
شکل (۹): مقایسه روش LPSS با روش‌های PSS و GSS هنگامی که فریم اول یک فریم سکوت نباشد

در مرحله بعد روش‌های پیشنهادی LPSS، LPSS با استفاده از VAD و V/UV detector (Advanced LPSSS)، LPMBSS، LPIMBSS و نیز روش‌های GSS، IGSS، MBSS و IMBSS را بر روی پنجاه سیگنال صوتی آغشته به نویز (نویز از نوع سفیدگوسی) از پایگاه داده‌های *Timit* با *SNR* اولیه *10 dB*، اعمال نموده‌ایم و میانگین *SNR* های خروجی در شکل (۱۰) آمده است. همان طور که ملاحظه می‌شود اعمال روش پیشنهادی تخمین نویز، بر روی هر یک از روش‌ها، سبب بهبود نسبت سیگنال به نویز سیگنال غنی شده خروجی می‌گردد.

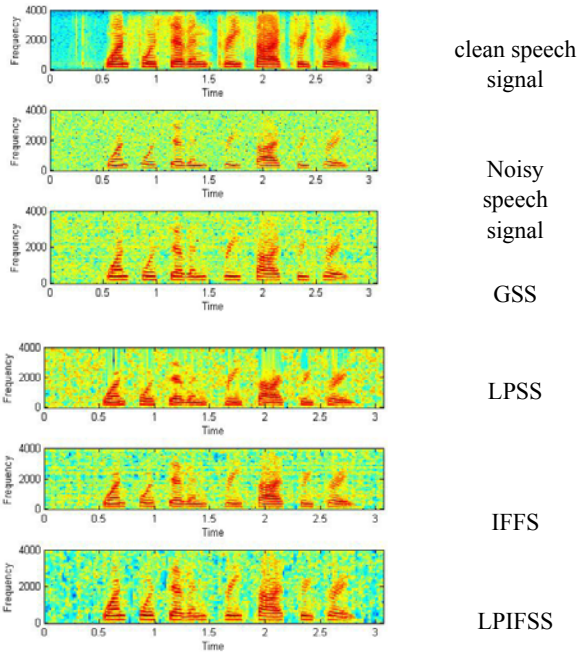
جهت بررسی بیشتر روش‌های پیشنهادی و مقایسه آن‌ها با روش‌های موجود، به عنوان نمونه، در شکل (۱۱) سیگنال صوتی تمیز و سیگنال صوتی آغشته به نویز با *SNR=10 dB*، به همراه سیگنال‌های بهبود یافته توسط روش‌های GSS، IFSS، LPSS و LPIFSS و در شکل (۱۲)، نمودار اسپکتوگرام آن‌ها رسم گردیده است. همان طور که از شکل (۱۱) و (۱۲) مشخص است، در روش‌های پیشنهادی، سیگنال صوتی خروجی به میزان قابل توجهی نسبت به روش‌های موجود بهبود یافته است.



شکل (۱۰): مقایسه SNR خروجی روش‌های پیشنهادی (LPIMBSS, LPMBSS, ALPSS, LPIFSS, LPSS) و روش‌های موجود



شکل (۱۱): شکل موج سیگنال خروجی در روش‌های پیشنهادی (LPSS,LPIFSS) و روش‌های موجود



شکل (۱۲): اسپکتوگرام سیگنال خروجی در روش‌های پیشنهادی (LPSS,LPIFSS) و روش‌های موجود

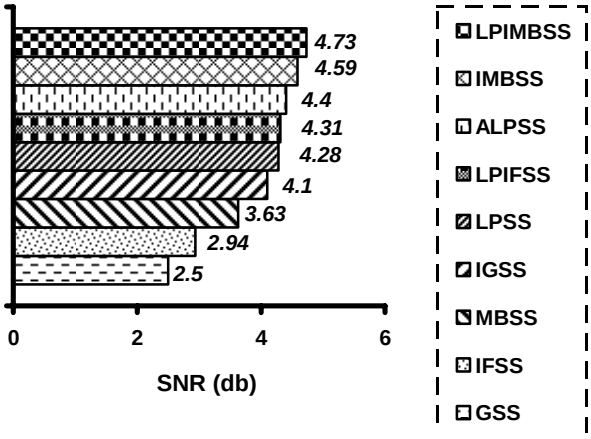
۸- آزمایش شنوایی(MOS[۱۳])

در بحث‌های قبلی مقایسه روش‌های موجود و روش‌های پیشنهادی، نسبت سیگنال به نویز سیگنال غنی شده صوتی بوده است. اینک می‌خواهیم مقایسه کیفی بین روش‌های موجود و روش‌های پیشنهادی انجام دهیم. بدین منظور می‌خواهیم از تست شنوایی [۱۴] برای مقایسه روش‌ها استفاده نمائیم.

ده سیگنال صوتی از پایگاه داده‌های *Timit* با *SNR* اولیه *10 dB* با فرض اینکه نویز از نوع سفیدگوسی می‌باشد را به الگوریتم‌های موجود و پیشنهادی اعمال نموده‌ایم، سپس از شش نفر (سه نفر زن و سه نفر مرد و از طیف سنی جوان تا پیر) برای نمره‌دهی به سیگنال‌های غنی شده براساس جدول (۱) استفاده نموده‌ایم [۱۴]. میانگین نمرات این شش نفر به این ده سیگنال صوتی(۶۰ بار آزمایش به ازای هر *SNR* اولیه) در شکل (۱۳) آمده است.

جدول (۱): معیار آزمایش شنوایی [۱۴]

کیفیت شنیدن نویز	نمره
نویز غیر قابل درک است	۵
نویز قابل درک است اما اذیت‌کننده نیست	۴
نویز قابل درک است و کمی اذیت‌کننده است	۳
نویز اذیت کننده است اما نمی‌توان کلا مخالفت کرد	۲
نویز بسیار اذیت‌کننده است	۱



شکل (۱۳): مقایسه خروجی روش‌های پیشنهادی (LPIMBSS, ALPSS, LPIFSS) و روش‌های موجود با استفاده از تست MOS

۹- نتیجه‌گیری

در این مقاله روش‌های مختلف بیان و تشریح شده‌اند. این روش‌ها عبارتند از: تفریق طیفی توان، تفریق طیفی توسعه‌یافته، تفریق طیفی با استفاده از کف طیفی و تفریق بیش از حد، تفریق طیفی چند بانده، تفریق طیفی معکوس تبدیل فوریه.

با مطالعه این روش‌های عمده به‌سازی گفتار، نقطه ضعف آن‌ها را در عدم تخمین صحیح نویز تشخیص دادیم. به منظور رفع این مشکل، ایده استفاده از آنالیز LPC برای تخمین نویز شکل گرفته شد. از بررسی روش‌ها به این نکته رسیدیم که در کلیه روش‌ها به تخمینی از نویز و یا برخی از پارامترهای آن نیاز می‌باشد. بدین منظور به دنبال روشی رفتیم تا بتواند تخمین بهتر و درست‌تری از نویز را در اختیار قرار دهد.

در آنالیز LPC به دنبال فیلتر و مدلی از حنجره هستیم تا بتواند کلیه مشخصه‌های حنجره را در خود داشته باشد و در صورت اعمال نویز در ورودی، در خروجی آن سیگنال گفتار را داشته باشیم. پس در صورت اعمال سیگنال گفتار به مدل (فیلتر) معکوس، بایستی سیگنال نویز را در خروجی آن داشته باشیم. قسمت ناهمبسته سیگنال گفتار نویزی در خروجی فیلتر ظاهر می‌شود که به علت خطی بودن فیلتر قسمت اعظم آن نویز می‌باشد. از این نویز در غنی‌سازی گفتار استفاده نموده و از آن در اصلاح روش‌های فوق کمک گرفته‌ایم.

در مرحله بعد به اصلاح این نویز تخمینی پرداختیم و برای اصلاح آن از الگوریتم‌های VAD و V/UV detector استفاده کردیم. پس از اصلاح روش‌های موجود، روشی را ارائه نمودیم که برای بهبود سیگنال گفتار، به تخمینی از نویز و یا پارامترهای آن نیاز نداشته باشد. با مقایسه روش‌های پیشنهادی و روش‌های موجود، مشاهده نمودیم که روش‌های پیشنهادی نه تنها *SNR* سیگنال‌های غنی‌سازی شده را

بهبود بخشیده‌اند بلکه در آزمایش شنوایی نیز بهتر از روش‌های موجود پاسخ داده‌اند.

مراجع

[1] J. R. Deller, J. H. L. Hansen and J. G. proakis, "Discrete-time processing of speech signals", 2nd edition, IEEE press, 2000.

[2] L. R. Rabiner and R. W. Schafer, "Digital processing of speech signals", Prentice Hall, 1978.

[3] J. Tierney, "A study of LPC analysis of speech in additive noise", IEEE Trans. Acoust. Speech and Signal processing, ASSP-28, 4, pp. 389-379, Aug. 1980.

[4] M. R. Sambur and N. S. Jayant, "LPC analysis/synthesis from speech inputs containing guantizing noise or additive white noise", IEEE Trans. Acoust. Speech and signal process. ASSP-24, 6, pp. 488-494, Dec. 1976.

[5] S. Kamath and P. Loizou, "A Multi-band spectral subtraction method for Enhancing speech corrupted by colored noise", Proceedings of ICASSP-2002, Orlando, FL, May 2002.

[6] M. Berouti, R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise", Proceeding IEEE ICASSP, Washington DC, pp. 208-211, Apr. 1979.

[7] Y. Ghanbari, M. R. Karami and B. Amelifard, "Improved multi-band spectral subtraction method for speech enhancement", Proceedings of the 6th ISTED International Conference SIGNAL AND IMAGE PROCESSING, pp. 225-230, Agu. 23-25, 2004, Honolulu, Hawaii, USA.

[8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. On Acoust. Speech and signal processing, vol. ASSP-27, pp. 113-120, Apr. 1979.

[9] P. S. Whitehead, D. V. Andeson and M. A. Clements, "Adaptive acoustic noise suppression for speech enhancement", IEEE International Conference on Multimedia & Expo., Jul. 2003.

[10] D. L. Donoho, "De-noising by soft-thresholding, IEEE Transactions on Information Theory", vol. 41, no. 3, pp. 613-627, May 1995.

[11] K. Y. Lee, B. G. Lee and S. Ann, "Adaptive filtering for speech enhancement in colored noise", IEEE Trans. On Signal Processing Letters, vol. 4, pp. 277-279, Oct. 1997.

[12] I. Y. Soon, S. N. Koh and C. L. Yeo, "Wavelet for Speech Denoising", TENCON 97, pp. 479-482, Brisbane, Australia, 1997.

[13] H. Sheikhzadeh and H. R. Abutalebi, "An Improved Wavelet-Based Speech Enhancement System", in proc. 7th European Conference on Speech Communication and Technology (EuroSpeech), Aalborg, Denmark, Sep. 2001.

[14] Y. Ghanbari and M. R. Karami, "Spectral Subtraction in the Wavelet Domain for Speech Enhancement", International Journal of Imaging Systems and Technology, vol. 1, no. 1, Aug. 2004.

[15] Y. Ghanbari and M. R. Karami-Mollaee, "A new approach for speech enhancement based on adaptive thresholding of wavelet packets", Speech communication 48 (2006) 927-940.

[16] F. Khakpoor, G. Ardeshir, 'Using PCA and SVD to improve wavelet-based method for detection voice and silence in speech', European Journal of Scientific Research, vol. 37, no. 4 (2009), pp. 641-648

[17] "M. Jeub, D. Kolassa, R. F. Astudillo, R. Orglmeister, 'Performance analysis of wavelet base voice activity detection', NGA/DAGA 2009, Rotterdam, pp: 407-410

[18] M. R. Karami and M. Eshaghi, "A new algorithm for voice activity detection based on wavelet packets", IJE Trans. A: vol. 22, no. 3, pp. 225-232, (2009).

زیرنویس‌ها

1- linear predictive coding
2- voice activity detector
3- voice /unvoiced
4- power spectral subtraction
5- short time fast fourier transform
6- general spectral subtraction
7- signal to noise ratio
8- over spectral
9- multi band spectral subtraction
10- improved multi band spectral subtraction
11- inverse fourier spectral subtraction
12- linear predictive spectral subtraction
13- mean opinion score