

غنی سازی سیگنالهای صوتی با استفاده از آنالیز LPC در روشهای مبتنی بر آستانه گذاری

مصطفی حیدری ، محمد رضا کریمی ملاتی ، احسان نادر نژاد

مجتمع آموزش عالی نوشیروانی بابل - دانشکده مهندسی برق و کامپیوتر

E-mail: baradaran5@yahoo.com, mkarami@nit.ac.ir, ehsan_nader@yahoo.com

چکیده

در این مقاله، دو روش جدید برای غنی سازی سیگنالهای صوتی ارائه شده است. این روشها که مبتنی بر روش آستانه گذاری داناها و روش شیخ زاده و ابوطالبی می باشند، از ضرایب پیشگویی خطی (LPC) برای تخمین و استخراج نویز استفاده شده است. سپس به مقایسه روشهای پیشنهادی با روش موجود پرداخته ایم و مشاهده شده است که روشهای پیشنهادی به میزان چشمگیری، نسبت سیگنال به نویز در سیگنالهای صوتی آغشته به نویز را در مقایسه با روشهای موجود، بهبود بخشیده اند. ضمن اینکه از نظر تست شنوایی نیز وضع به همین گونه بوده است##

کلمات کلیدی : سیگنال صوتی، نویز، غنی سازی سیگنال صوتی، تفریق طیفی، آنالیز LPC

۱- مقدمه

از میان روشهای موجود برخی همانند روش تفریق طیفی [۵،۶،۷]، تفریق طیفی چند بانده [۵]، تفریق طیفی معکوس فوریه و ... از تبدیل فوریه برای غنی سازی سیگنالهای صوتی استفاده نموده اند. در حالی که برخی از روشها همانند روش تفریق طیفی موجک [۱۷]، روش شیخ زاده و ابوطالبی [۱۳] و ... از تبدیل موجک برای غنی سازی سیگنالهای صوتی بهره جسته اند.

در استفاده از ضرایب موجک به چند طریق می توان عمل نمود، به عنوان مثال در روش تفریق طیفی موجک [۱۷]، عمل تفریق به جای اینکه بر روی تابع اندازه تبدیل فوریه سیگنال صوت آغشته به نویز و سیگنال نویز انجام شود، بر روی ضرایب موجک صورت می گیرد و یا اینکه در مابقی روشهای ذکر شده در فوق به جای استفاده مستقیم از نویز تخمینی، تنها برخی از پارامترهای آن استفاده می گردد و عمل غنی سازی با استفاده از آستانه گذاری [۹،۱۰،۱۲،۱۳،۱۴،۱۵] روی ضرایب موجک تحقق می یابد.

در روش داناها [۱۰،۱۴] از یک مقدار آستانه ثابت برای کلیه فریمها و فرکانسها برای کاهش نویز استفاده شده است. در روش شیخ زاده و ابوطالبی [۱۳]، از یک تابع آستانه گذاری جدید که در حقیقت شکل بهبود یافته تابع آستانه گذاری سخت می باشد، استفاده گردیده است. ضمن اینکه در این روش مقدار

کاهش و یا حذف نویز، یکی از مباحث مهم در پردازش سیگنالهای صوتی (به عنوان مثال در سیستمهای ارتباطی، کدینگ سیگنالهای صوتی، تشخیص صوت و ...) می باشد و به همین منظور روشهای زیادی برای کاهش میزان نویز در سیگنالهای صوتی ارائه شده است. از این میان می توان به روش تفریق طیفی [۵،۶،۷]، تفریق طیفی چند بانده [۵]، تفریق طیفی معکوس فوریه، فیلتر وقتی [۱۱]، فیلتر واینر [۸]، تبدیل موجک [۱۰،۱۲،۱۳،۱۴،۱۵،۱۶،۱۷] و ... اشاره نمود.

در روش تفریق طیفی سه فرض بایستی برقرار باشد:

۱- نویز و سیگنال جمع شونده باشند.

۲- نویز و سیگنال ناهمبسته می باشند.

۳- یک کانال در دسترس می باشد.

روش تفریق طیفی، اگرچه بسیار ساده و کارآمد می باشد اما سبب ایجاد نویز جدیدی به نام نویز موزیکال می گردد که جهت کاهش این نویز، از روش تفریق طیفی با استفاده از کف طیفی و تفریق بیش از حد که توسط بروتی [۷] ارائه شده است، استفاده می گردد.

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

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

$$R_{s,n}(\tau; m) = 0 \quad (۸)$$

که $x(n; m)$ سیگنال پنجره شده سیگنال صوتی $x(n)$ می باشد. با گرفتن تابع چگالی احتمال از طرفین رابطه (۶) داریم:

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

از طرفی نیز می دانیم [۱]:

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

با توجه به اینکه مقدار $|X(\omega; m)|^2$ ، به مراتب بزرگتر از مقدار N^2 می باشد، می توان به راحتی از ظریب $\frac{1}{N^2}$ چشم پوشی نموده و با تقریب خوبی، به رابطه ذیل رسید:

$$\Gamma_x(\omega; m) \approx |x(\omega; m)|^2 \quad (۱۱)$$

با استفاده از روابط (۱۱) و (۹) رابطه زیر حاصل می گردد:

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

که در روابط فوق $|X(\omega; m)|$ اندازه تبدیل فوریه سیگنال پنجره شده $x(n)$ و $|S(\omega; m)|$ و $|N(\omega; m)|$ به ترتیب اندازه تبدیل فوریه سیگنال صوتی تمیز پنجره شده و سیگنال نویز پنجره شده می باشد.

همانطور که از رابطه (۱۲) پیداست، برای بدست آوردن اندازه تبدیل فوریه سیگنال تمیز بایستی، اندازه تبدیل فوریه سیگنال تصادفی نویز را نیز در دست داشت، بدین منظور سیگنال تصادفی نویز را از قسمت سکوت سیگنال صوتی آغشته به نویز، تخمین می زنند. دلیل این روش تخمین نویز، بر این ایده استوار است که چون در قسمت سکوت هیچ سیگنال صوتی وجود ندارد لذا هر آنچه که در این قسمت در سیگنال صوتی آغشته به نویز وجود دارد ناشی از نویز است و عموماً اولین فریم صوتی را به عنوان نویز در نظر می گیرند.

اینک برای دستیابی به سیگنال صوتی تمیز در حوزه زمان لازم است که علاوه بر اندازه تبدیل فوریه، فاز آنرا نیز به دست آورده و با معکوس فوریه زمان کوتاه (St FFT) به سیگنال صوتی در حوزه زمان رسید.

در تمامی کاربردهای عملی می توان فاز سیگنال صوتی تمیز را همان فاز سیگنال صوتی آغشته به نویز در نظر گرفت [۶]:

$$\varphi_{\hat{s}(\omega; m)} = \varphi_{X(\omega; m)} \quad (۱۳)$$

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

با توجه به معادلات (۱۲) و (۱۳) سیگنال صوتی تمیز به صورت زیر تخمین زده می شود:

آستانه برای حروف صدا دار و بی صدا و نیز باندهای مختلف فرکانسی متفاوت می باشد.

ما در این مقاله ابتدا روشهای تفریق طیفی، آستانه گذاری داناها و شیخ زاده و ابوطالبی [۱۳] را شرح داده، سپس با استفاده از آنالیز LPC [۱،۲،۳،۴] به تخمین نویز از روی سیگنال صوتی پرداخته ایم. در نهایت نویز تخمینی را به روش آستانه گذاری داناها و روش شیخ زاده و ابوطالبی اعمال نموده ایم که در نتیجه به بهبود بسیار زیاد این روشها انجامید.

۲- تفریق طیفی توان (PSS):

فرض نموده ایم که سیگنال و نویز جمع شونده هستند لذا می توان یک سیگنال صوتی آغشته به نویز را بصورت زیر بیان داشت:

$$x(n) = s(n) + n(n) \quad (۱)$$

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

طبق فرض دوم، نویز و سیگنال ناهمبسته اند از طرفی نیز نویز نیز ناهمبسته است لذا می توان نوشت [۱]:

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

$$R_{s,n}(\tau) = 0 \quad (۳)$$

که $R_n(\tau)$ تابع خود همبستگی سیگنال تصادفی نویز و $R_{s,n}(\tau)$ نیز تابع همبستگی متقابل سیگنالهای s, n می باشد. با توجه به روابط فوق و با فرض ایستا بودن سیگنالهای s, n می توان نوشت:

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

که Γ_x و Γ_s و Γ_n به ترتیب چگالی طیف قدرت x, s, n می باشد. بنا به رابطه (۴) چنانچه چگالی طیف قدرت سیگنال تصادفی نویز در دسترس باشد می توان چگالی طیف قدرت سیگنال صوتی تمیز را بدست آورد:

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

رابطه (۴) و به دنبال آن (۵) با فرض ایستا بودن سیگنال صوتی تمیز و نیز نویز برقرار می باشند، اما در طبیعت چنین فرضی همواره برقرار نمی باشد.

از آنجایی که سیگنال صوتی تمیز در بازه های زمانی کوتاه ایستایی محلی دارند و نیز فرض ایستا بودن نویز در بازه های کوتاه منطقی تر می باشد لذا، ابتدا روی سیگنال صوتی آغشته به نویز عمل پنجره گذاری انجام می شود و بدین طریق سیگنال گفتاری به فریم های کوتاه تقسیم می گردد و سپس عمل تفریق طیفی را بر روی هر فریم انجام می دهند.

۳- غنی سازی سیگنال صوتی با استفاده از روش داناها :

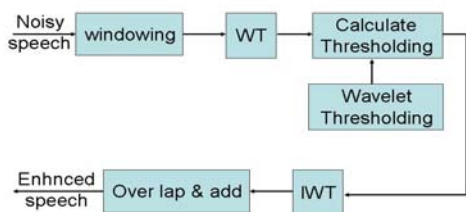
داناها در سال ۱۹۹۵ برای اولین بار ایده کاهش نویز سیگنالهای صوتی را با استفاده از آستانه گذاری بر روی تبدیل موجک ارائه نموده است. وی با فرض اینکه نویز سیگنال صوتی، نویز سفیدگوسی باشد رابطه زیر را برای آستانه گذاری بر روی ضرایب موجک ارائه نموده است [۱۰] :

$$T = \hat{\sigma} \sqrt{2 \ln(N)} \quad (18)$$

که در آن N طول سیگنال صوتی آغشته به نویز و $\hat{\sigma}$ انحراف معیار استاندارد سیگنال نویز سفید گوسی تخمین زده شده می باشد که بصورت زیر محاسبه می گردد [۱۴] :

$$\hat{\sigma} = \frac{\text{median}(|C|)}{0.6745} \quad (19)$$

که در آن C ، دنباله ضرایب موجک سیگنال نویز تخمین زده شده می باشد که همانند روش تفریق طیفی موجک از قسمت سکوت سیگنال و عموماً فریم اول آن تشخیص داده می شود.



شکل ۲: بلوک دیاگرام روش داناها

۴- غنی سازی سیگنال با استفاده از روش شیخ زاده و ابوطالبی

در برخی از روشهای غنی سازی سیگنالهای صوتی همانند روش داناها [۱۰، ۱۴] تنها از یک مقدار آستانه برای تمامی باندهای سیگنال صوتی استفاده شده است و به این نکته که در حروف صدادار غالب انرژی در باندهای پایین فرکانسی و در حروف بی صدار غالب انرژی در باندهای بالای فرکانسی متمرکز گردیده است، توجهی نشده است.

در روش پیشنهادی شیخ زاده و ابوطالبی برای باندهای مختلف، مقادیر آستانه مختلفی در نظر گرفته شده است و ضمناً به منظور مقابله با نویزهای رنگی و نالیستا، از یک واحد تشخیص گر سکوت استفاده شده است [۱۳].

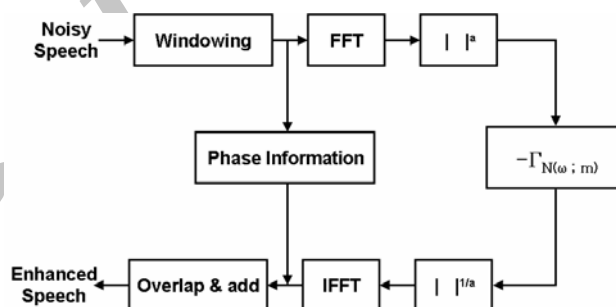
از سوی دیگر، مقدار آستانه در فریمهای متعلق به حروف صدادار به گونه ای است که مقدار آن در باندهای بالای فرکانسی نسبت

$$\begin{aligned} \hat{S}(\omega; m) &= \left| \hat{S}(\omega; m) \right| \exp\left\{i\phi_{\hat{S}(\omega; m)}\right\} \\ &= \left[\left| X(\omega; m) \right|^2 - \left| \hat{N}(\omega; m) \right|^2 \right]^{1/2} \exp\left\{i\phi_{\hat{S}(\omega; m)}\right\} \end{aligned} \quad (14)$$

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

به روش فوق، روش تفریق طیفی توان گفته می شود چرا که از توان دوم اندازه تبدیل فوریه که مبین قدرت و توان سیگنال می باشد استفاده کرده است، عموماً از یک ضریب توانی دیگری غیر از ۲، در روش تفریق طیفی استفاده می شود که مقدار آن را با بهینه سازی محاسبه می نمایند. در این حالت به روش فوق، روش تفریق طیفی تعمیم یافته می گویند [۱] :

$$\hat{S}(\omega; m) = \left[\left| X(\omega; m) \right|^a - \left| \hat{N}(\omega; m) \right|^a \right]^{1/a} \exp\left\{i\phi_{\hat{S}(\omega; m)}\right\} \quad (15)$$



شکل (۱) بلوک دیاگرام روش تفریق طیفی تعمیم یافته

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

الف) اصلاح نیم موج :

$$\left| \hat{S}(\omega; m) \right| = \begin{cases} \left| \hat{S}(\omega; m) \right| & \text{if } \left| \hat{S}(\omega; m) \right| > 0 \\ 0 & \text{elsewhere} \end{cases} \quad (16)$$

ب) اصلاح تمام موج:

$$\left| \hat{S}(\omega; m) \right| = \text{abs}\left\{ \left| \hat{S}(\omega; m) \right| \right\} \quad (17)$$

$$C_{ij(k)} = \left[\frac{\lambda}{e^{-\gamma} - 1} (e^{\frac{\gamma |C_{ij}(k)|}{\lambda}} - 1) \right] \times \quad (23)$$

$$\text{sign}(C_{ij}(k)) \text{ if } |C_{ij}(k)| \leq \lambda_{ij}$$

$$\lambda = \begin{cases} \lambda_{1j} & \text{else} \\ \lambda_{2j} & \text{last band} \end{cases} \quad (24)$$

که $C_{ij}(k)$ ، k امین ضریب تبدیل موجک باز در j امین باند فرکانسی از فریم i ام می باشد. و λ ضریبی است جهت کنترل سرعت اوج تابع نمای آستانه گذاری. برای فریمهای بی صدا نیز تابع آستانه گذاری بصورت زیر می باشد:

$$C_{ij(k)} = \left[\frac{\lambda}{e^{-\gamma} - 1} (e^{\frac{\gamma |C_{ij}(k)|}{\lambda}} - 1) \right] \times \quad (25)$$

$$\text{sign}(C_{ij}(k)) \text{ if } |C_{ij}(k)| \leq \lambda_{ij}$$

$$\lambda = \begin{cases} \lambda_{1j} & \text{last band} \\ \lambda_{2j} & \text{else} \end{cases} \quad (26)$$

و در مورد فریمهای سکوت داریم:

$$C_{ij}(k) = \begin{cases} 0 & |C_{ij}(k)| \leq \frac{\lambda_{ij}}{2} \\ \left(|C_{ij}(k)| - \frac{\lambda_{ij}}{2} \right) \times \text{sign}(C_{ij}(k)) & |C_{ij}(k)| > \frac{\lambda_{ij}}{2} \end{cases} \quad (27)$$

۵- ضرائب پیشگویی خطی (LPC)

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

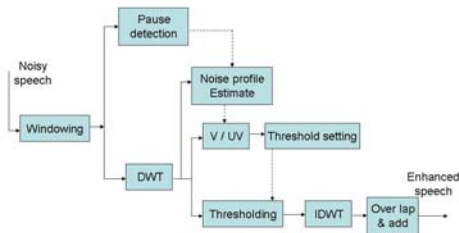
ضرایب پیشگویی خطی یکی از قدرتمندترین ابزار در پردازش صوت می باشد [۲]. ایده کلی این آنالیز این است که هر نمونه از سیگنال صوتی را می توان بصورت معادله ای خطی برحسب خروجی ها و ورودی های گذشته نوشت:

$$s(n) = \sum_{k=1}^p a_k s(n-k) + \sum_{l=1}^q b_l u(n-l) \quad (28)$$

که a_k و b_l به ترتیب ضرایب مخرج و صورت فیلتر و $u(n)$ سیگنال ورودی می باشد. که برای حروف صدادر یک قطار ضربه و برای حروف بی صدا یک رشته نویز تصادفی می باشد [۱،۲،۳،۴]. با تبدیل Z گرفتن از دو طرف معادله (۲۸) می توان به رابطه زیر رسید.

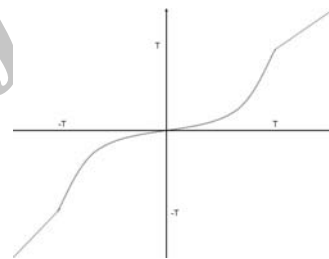
$$H(z) = \frac{S(z)}{U(z)} = \frac{1 + \sum_{l=1}^q b_l z^{-l}}{\sum_{k=1}^p a_k z^{-k}} \quad (29)$$

به باندهای پایین فرکانسی بیشتر باشد تا در فرکانسهای بالا که نویز بیشتر از سیگنال است نویز بیشتری حذف گردد و در حروف بی صدا بالعکس می باشد [۱۳]. میزان افزایش حد آستانه، به میزان ۱۰٪ نسبت به مقدار آستانه کلاسیک می باشد. در این روش تعداد سطوح تجزیه، γ سطح می باشد و چون از تبدیل موجک باز استفاده شده است، شاهد ۸ باند فرکانسی خواهیم بود.



شکل ۳: بلوک دیاگرام روش شیخ زاده و ابوطالبی

در این روش از یک تابع آستانه گذاری جدید که شکل بهبود یافته تابع آستانه گذاری سخت می باشد به جای توابع سخت و نرم استفاده شده است که در شکل (۴) آمده است.



شکل ۴: تابع آستانه گذاری پیشنهاد شده توسط

شیخ زاده و ابوطالبی

همانطور که گفته شد، در این روش مقدار آستانه با توجه به باندهای فرکانسی و فریم های بی صدا و صدادر متغیر می باشد که بصورت زیر محاسبه می گردد [۱۳]:

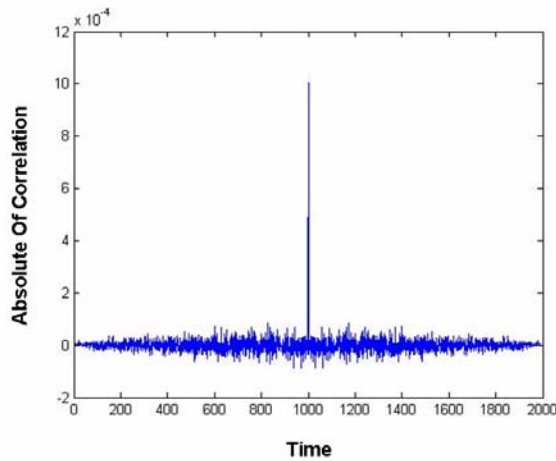
$$\hat{\sigma}_j = \frac{\text{median}(|C|)}{0.6745} \quad (20)$$

$$\lambda_{1j} = \hat{\sigma}_j \sqrt{2 \ln(N)} \quad (21)$$

$$\lambda_{2j} = \alpha \lambda_{1j} \quad (22)$$

که در آن $\hat{\sigma}_j$ ، انحراف معیار استاندارد نویز تخمینی در باند j ام فرکانسی می باشد و N طول باند j ام و α فاکتور افزایش مقدار آستانه می باشد که مبین میزان افزایش مقدار آستانه، نسبت به مقدار آستانه کلاسیک می باشد که پیش از این در مورد آن بحث شده است.

تابع آستانه گذاری برای فریم های صدادر بصورت زیر می باشد:



شکل ۵: تابع خود همبستگی سیگنال خطا

البته سیگنال $e(n)$ اگرچه همان سیگنال نویز اضافه شده به سیگنال صوتی تمیز نمی باشد اما اکثر مشخصات آنرا دربر دارد. از طرفی نیز در خروجی فیلتر $A(z)$ ، مقداری از سیگنال uncorrelate مربوط به سیگنال گفتاری نیز وجود دارد که در مقایسه با نویز قابل چشم پوشی است.

۶- اصلاح روش داناها با استفاده از آنالیز LPC :

در تخمین نویز از فریم اول سیگنال صوتی برای روش داناها، فرض بر این است که این فریم، فریم سکوت باشد و در صورت عدم برقراری چنین شرطی SNR خروجی به شدت کاهش می یابد.

جهت اصلاح و بهبود الگوریتم داناها، پیشنهاد می نمایم که از آنالیز LPC استفاده شود، بدین صورت که به جای اینکه اولین فریم را مستقیماً به روش داناها اعمال نمایم، ابتدا از آن ضرایب LPC را بدست آورده و خروجی فیلتر $A(z)$ را به عنوان تخمینی از نویز به روش داناها اعمال نمایم.

از آنجایی که در روش پیشنهادی، نویز تخمینی به نویز واقعی نزدیکتر می باشد و تا حد زیادی پارامترهای نویز واقعی را دربردارد، لذا نیازی به استفاده از تابع میانه نمی باشد و می توان رابطه (۱۹) را بصورت زیر اصلاح نمود:

$$\hat{\sigma} = \frac{\text{mean}(|C|)}{\alpha} \quad (34)$$

در رابطه فوق، برای محاسبه α ، مقدار آنرا از 0.2 تا 0.7 تغییر داده و در هر بار ده سیگنال صوتی از دیتا بیس $Timit$ را به روش پیشنهادی اعمال نموده و در هر بار میانگین SNR سیگنالهای صوتی غنی شده محاسبه گردیده است.

مشاهده شده است که به ازای $\alpha = 0.4$ بهترین SNR در خروجی ظاهر گردیده است.

در عمل برای سیگنالهای صوتی یک مدل تمام قطب، تقریب بسیار خوبی برای تابع تبدیل $H(z)$ می باشد [۱] و می توان آنرا بصورت زیر بیان داشت:

$$H(z) = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} = \frac{1}{A(z)} \quad (30)$$

در عمل و به خصوص برای حنجره انسان، p یک عدد صحیح در محدوده $10 \leq p \leq 14$ انتخاب می شود.

نکته اساسی در محاسبه ضرایب پیشگویی خطی این است که این ضرایب مستقیماً از سیگنال صحبت بدست آید به همین منظور و به علت ماهیت تغییر پذیری سیگنال گفتار (با زمان) ابتدا عمل پنجره گذاری بر روی سیگنال صوتی انجام می شود و سپس ضرایب LP در فریم های کوتاه محاسبه می شوند [۲].

با توجه به توضیحاتی که ارائه شده است، با تقریب خوبی می توان هر نمونه از سیگنال صوتی را تنها با P نمونه قبلی از همان سیگنال صوتی (بدون استفاده از P نمونه ماقبل ورودی) محاسبه نمود [۱]:

$$\hat{S}(n) = \sum_{k=1}^p a_k s(n-k) \quad (31)$$

سیگنال خطا در حقیقت، اختلاف بین سیگنال صوتی اصلی و سیگنال صوتی تخمین زده شده از روی P نمونه ماقبل می باشد:

$$e(n) = S(n) - \hat{S}(n) = S(n) - \sum_{k=1}^p a_k S(n-k) \quad (32)$$

که اگر از طرفین رابطه فوق تبدیل Z گرفته شود داریم:

$$E(z) = S(z) \left[1 - \sum_{k=1}^p a_k z^{-k} \right] = A(z) S(z) \quad (33)$$

تابع $E(z)$ ، تبدیل Z سیگنال خطا می باشد که دارای ماهیت نویزی است. برای اثبات این ادعا کافی است تابع خود همبستگی سیگنال $e(n)$ را محاسبه نمایم.

در شکل (۵) تابع خود همبستگی سیگنال خطا که متعلق به یک سیگنال صوتی از دیتا بیس $Timit$ می باشد رسم گردیده است. همانطور که مشاهده می شود، سیگنال خود همبستگی آن همانند تابع خود همبستگی سیگنال تصادفی نویز است.

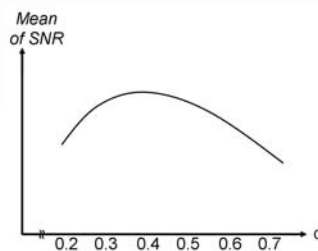
ما در الگوریتم پیشنهادی خود، از این سیگنال برای تخمین نویز استفاده نموده ایم که منجر به بهبود بسیار زیاد SNR سیگنالهای صوتی آغشته به نویز (در مقایسه با روشهای موجود) گردیده است.

در این مقاله روشی ارائه نموده ایم تا به اصلاح الگوریتم مزبور از دیدگاه تخمین واقعی تر نویز و پارامترهای آن بپردازیم.

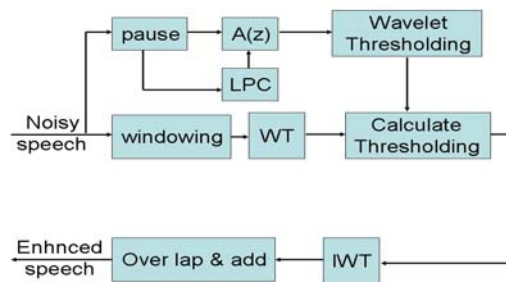
برای اینکه نویز تخمینی به نویز واقعی نزدیکتر باشد ابتدا از فریم سکوت ضرایب LP را محاسبه نموده و سپس، فریم مورد نظر را از یک فیلتر تمام صفر که از ضرایب LP همان فریم بدست آمده است عبور می دهیم. از آنجایی که ضرایب LP هر فریم حاوی مشخصات آن فریم می باشد لذا خروجی فیلتر $A(z)$ که در روابط (۳۰) و (۳۳) شرح داده شده است و مبین اختلاف سیگنال اصلی و سیگنال تخمینی می باشد، دارای ماهیت نویزی بیشتری نسبت به سیگنال اصلی می باشد. به همین دلیل از خروجی فیلتر $A(z)$ به عنوان نویز تخمینی استفاده می کنیم، چون این نویز تخمینی به نویز واقعی نزدیک می باشد لذا در محاسبه انحراف معیار استاندارد نیازی به اعمال تابع ریاضی میانه نمی باشد بلکه می توان از نویز تخمینی میانگین گرفت:

$$\hat{\sigma} = \frac{\text{mean}(|C|)}{0.4} \quad (35)$$

مقدار ۰,۴ همانند روش داناها به دست آمده است.



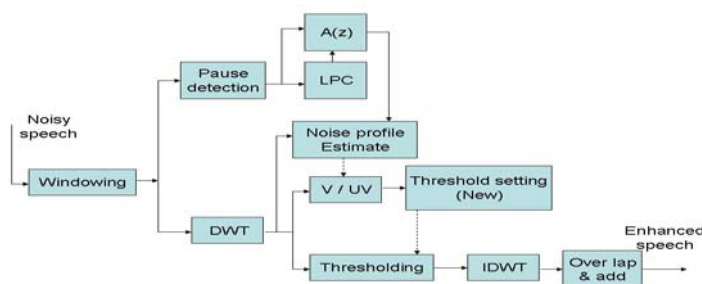
شکل ۶: میزان SNR سیگنال غنی شده بر حسب مقدار α



شکل ۷: روش پیشنهادی بر اساس الگوریتم LPC و روش داناها

۷- اصلاح روش ابوطالبی با استفاده از آنالیز LPC.

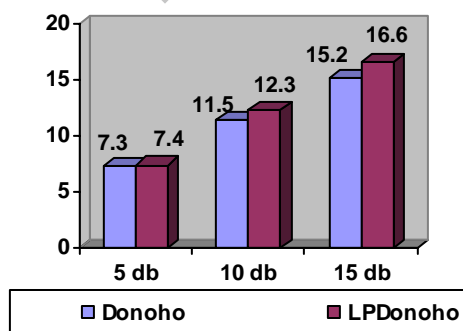
روش پیشنهادی شیخ زاده و ابوطالبی، از دیدگاههای متفاوتی مورد بررسی قرار گرفته است و از هر دیدگاه برخی از اصلاحات بر آن اعمال شده است. نظیر استفاده از تبدیل موجک بسته، اصلاح تابع آستانه گذاری و ...



شکل ۸: بلوک دیاگرام روش پیشنهادی

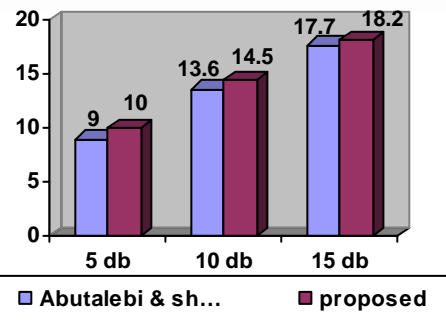
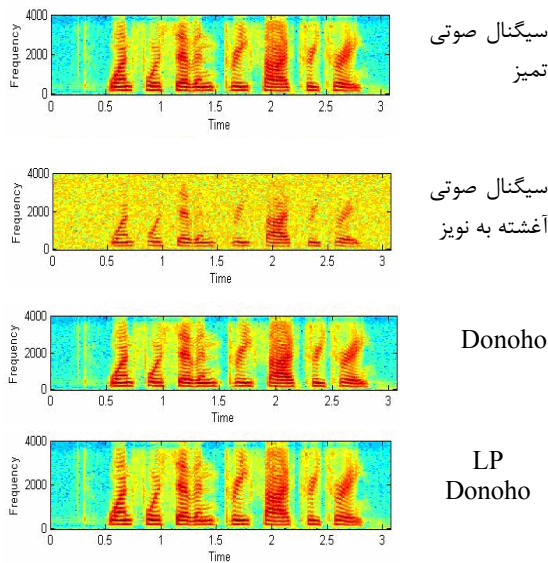
۸- نتایج

الگوریتمهای پیشنهادی خود را به همراه روشهای موجود بر روی پنجاه سیگنال صوتی از دیتابیس Timit با نسبت سیگنال به نویز اولیه 5, 10, 15 (نویز از نوع سفید گوسی) اعمال شده است. میانگین SNR های خروجی در نمودارهای (۱) و (۲) آمده است. همانطور که مشاهده می شود الگوریتمهای پیشنهادی، در مقایسه با روشهای موجود به میزان قابل توجهی، نسبت سیگنال به نویز را در سیگنال غنی شده، ارتقا داده است.



نمودار ۱: مقایسه روشهای Donoho و LPDonoho به ازای SNR= 5, 10, 15 dB (نویز از نوع سفید گوسی)

شکل ۹: سیگنال صوتی تمییز و سیگنال آغشته به نویز به همراه بهبود یافته آن

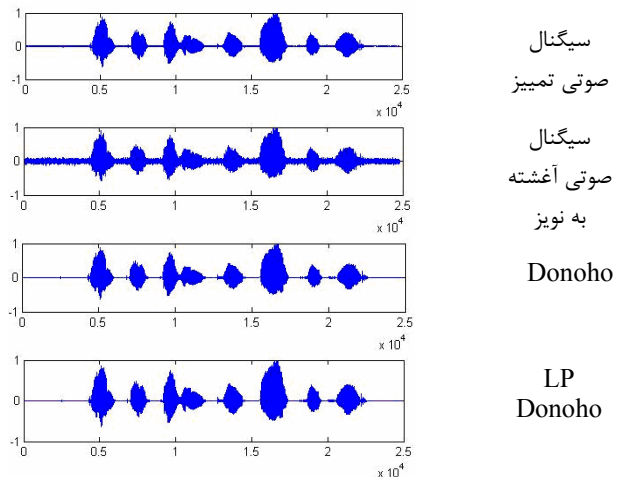


نمودار ۲: مقایسه روشهای Aboutalebi و LPAboutalebi به ازای SNR= 5, 10, 15 dB (نویز از نوع سفید گوسی)

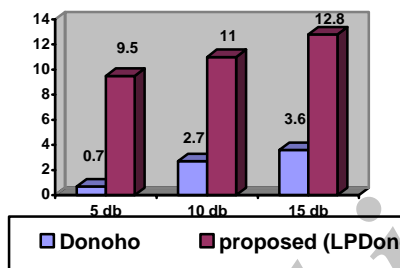
همانطور که پیش از این نیز بیان شده است در روشهای موجود، عموماً از فریم اول به عنوان قسمت سکوت استفاده می گردد. ضعف شدید روشهای موجود هنگامی آشکار می گردد که فریم اول، یک فریم سکوت نباشد، نمودار (۳) به مقایسه روش Donoho با LPDonoho پرداخته است و همانطور که مشاهده می شود، الگوریتم پیشنهادی از یک برتری کاملاً مشهود در مورد این گونه سیگنالها برخوردار است. در مورد روش ابوطالبی نیز برای این گونه سیگنالها وضع به همین گونه می باشد.

جهت بررسی بیشتر روشهای پیشنهادی و مقایسه آنها با روشهای موجود، به عنوان نمونه، در شکل (۹) سیگنال صوتی تمییز و سیگنال صوتی آغشته به نویز با SNR=10db به همراه سیگنالهای بهبود یافته توسط روشهای Donoho و LPDonoho و در شکل (۱۰)، نمودار اسپکتوگرام آنها رسم گردیده است.

همانطور که از این دو شکل برمی آید، در روشهای پیشنهادی سیگنال صوتی به میزان قابل توجهی نسبت به روشهای موجود بهبود یافته است.



شکل ۱۰: نمودار اسپکتوگرام سیگنال صوتی و آغشته به نویز به همراه بهبود یافته آن



نمودار ۳: مقایسه روشهای Donoho و LPDonoho وقتی که فریم اول فریم سکوت نباشد

۸-۱: تست شنوایی (Mean Opinion Score):

تا به حال ملاک مقایسه روشهای موجود و روشهای پیشنهادی، نسبت سیگنال به نویز سیگنال غنی شده صوتی بوده است. اینک می خواهیم مقایسه کیفی بین روشهای موجود و روشهای پیشنهادی انجام دهیم بدین منظور می خواهیم از تست شنوایی [۱۸] برای مقایسه روشها استفاده نماییم.

پنجاه سیگنال صوتی از دیتابیس Timit با SNR اولیه 10db با فرض اینکه نویز از نوع سفیدگوسی می باشد را به الگوریتمهای موجود و پیشنهادی اعمال نموده ایم سپس از شش نفر (سه نفر زن و سه نفر مرد و از طیف سنی جوان تا پیر) برای نمره دهی به سیگنالهای غنی شده براساس جدول (۱) استفاده نموده ایم [۱۸]، میانگین نمرات این شش نفر

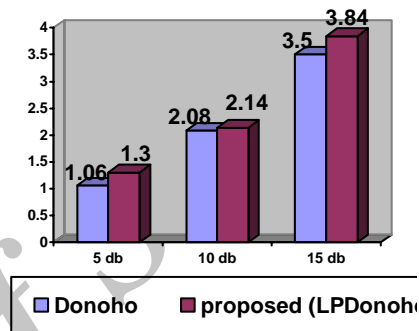
- [7] signal processing, vol. ASSP-27, April 1979, pp:113-120. M. Berouti, R. Schwartz, and J. Makhoul, enhancement of speech corrupted by acoustic noise, *proc. IEEE ICASSP*, Washington DC, April 1979, 208-211.
- [8] P. S. Whitehead, D.V. Andeson, M. A. Clements, Adaptive acoustic noise suppression for speech enhancement, *IEEE International Conference on Multimedia & Expo.*, July 2003.
- [9] H. Sameti, H. Sheichzadeh, Li Deng, R. L. Brennan, HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise, *IEEE Transactions on Speech and Audio processing*, vol. 6, No. 5, September 1998.
- [10] D. L. Donoho, De-noising by soft-thresholding, *IEEE Transactions on Information Theory*, vol. 41, No. 3, May 1995, pp:613-627.
- [11] K. Y. Lee, B. G. Lee, S. Ann, Adaptive filtering for speech enhancement in colored noise, *IEEE Trans. On Signal Processing Letters*, Vol. 4, October 1997, pp:277-279.
- [12] Ing Yann Soon, Soo Ngee Koh, Chai Liat Yeo, Wavelet for Speech Denoising, *TENCON 97*, Brisbane, Australia, 1997, pp: 479-482.
- [13] H.Sheikhzadeh, 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] D. L. Donoho, I.M. Johnston, "Ideal Spatial adaptive via wavelet shrinkage", *Biometrika*, vol. 81, pp:425-455, 1994.
- [15] I. M. Johnston, B. W. Silverman, "Wavelet threshold estimators for data with correlated noise", *J. Roy. Statist. Soc. B*, vol. 59, pp:319-351, 1997.
- [16] J. Seok, K. Bae, Speech enhancement with reduction of noise components in wavelet domain, *Proceeding of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'97)*, vol 2, April 21-24, 1997, pp:1323-1326.
- [17] Y. Ghanbari, M.R. Karami, "Spectral subtraction in the wavelet domain for speech Enhancement", *International Journal of software and Information Technologies (IJSIT)*, vol.1 no.1, pp:26-30,2004.

[۱۸] ی. قنبری، کاهش نویز در سیگنال گفتار با استفاده از تبدیل موجک، پایان نامه کارشناسی ارشد، دانشگاه مازندران .

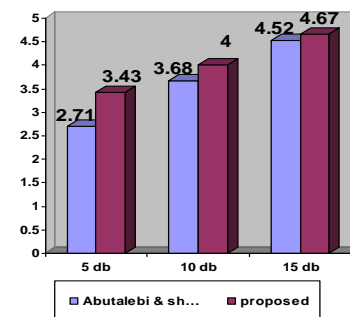
به این ده سیگنال صوتی (۳۰۰ بار آزمایش به ازای هر SNR اولیه) در نمودارهای (۴) و (۵) آمده است.

جدول ۱: معیار تست شنوایی

نمبره	کیفیت شنیدن نویز
۵	نویز غیر قابل درک است
۴	نویز قابل درک است اما اذیت کننده نیست
۳	نویز قابل درک است و کمی اذیت کننده است
۲	نویز اذیت کننده است اما نمی توان کلا مخالفت کرد
۱	نویز بسیار اذیت کننده است



نمودار ۴: مقایسه روشهای Donoho و LPDonoho براساس معیار تست شنوایی



نمودار ۵: مقایسه روشهای Abutalebi و LPAbutalebi براساس معیار تست شنوایی

۹- منابع

- [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] Tierney J., A study of LPC analysis of speech in additive noise, *IEEE trans. Acoust. Speech and signal process.* ASSP-28,4, pp:389-379 (Aug.1980).
- [4] Sambur M.r. and Jayant N.s., LPC analysis/synthesis from speech inputs containing quantizing 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] S. F. Boll, Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. On Acoust. Speech &*