ارتقاء روش تفریق طیفی موجک با استفاده از آنالیز LPC جهت غنی‌سازی سیگنال‌های صوتی

مصطفی حیدری، احسان نادرنژاد و محمدرضا کرمی ملایی

چکیده: در این مقاله، روشی جدید برای غنی‌سازی سیگنال‌های صوتی ارائه شده است. در این روش که مبتنی بر روش تفریق طیفی موجک می‌باشند، از ضرایب پیش‌گویی خطی (LPC) برای تخمین و استخراج نویز استفاده شده است. سپس به مقایسه روش پیشنهادی با روش تفریق طیفی موجک پرداخته‌ایم و مشاهده شده است که روش پیشنهادی به میزان چشم‌گیری، نسبت سیگنال به نویز در سیگنال‌های صوتی آغشته به نویز را در مقایسه با روش تفریق طیفی موجک، بهبود بخشیده است. ضمن اینکه از نظر تست شنوایی نیز وضع به‌همین گونه بوده است.

کلید واژه: **آنالیز LPC، تبدیل موجک، تفریق طیفی، غنی‌سازی سیگنال صوتی، نویز.**

۱- مقدمه

کاهش و یا حذف نویز، یکی از مباحث مهم در پردازش سیگنال‌های صوتی (به‌عنوان مثال در سیستم‌های ارتباطی، کدینگ سیگنال‌های صوتی، تشخیص صوت و ...) می‌باشد و به‌همین منظور روش‌های زیادی برای کاهش میزان نویز در سیگنال‌های صوتی ارائه شده است. از این میان می‌توان به روش تفریق طیفی [۱] تا [۳]، تفریق طیفی چندبانده [۱]، تفریق طیفی معکوس فوریه، فیلتر وفقی [۴]، فیلتر واینر [۵]، تبدیل موجک [۴] و [۶] تا [۱۱] و ... اشاره نمود.

در روش تفریق طیفی سه فرض بایستی برقرار باشد:

۱) نویز از نوع جمع‌شونده است.

۲) نویز و سیگنال ناهمبسته می‌باشند.

۳) یک کانال در دسترس می‌باشد.

روش تفریق طیفی، اگرچه بسیار ساده و کارآمد می‌باشد اما سبب ایجاد نویز جدیدی به‌نام 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 [۳] ارائه شده است، استفاده می‌گردد.

در روش تفریق طیفی از اندازه تبدیل فوریه برای کاهش نویز استفاده می‌گردد در حالی که علاوه بر تبدیل فوریه از تبدیل موجک نیز می‌توان برای کاهش نویز استفاده نمود.

از میان روش‌های مبتنی بر تبدیل موجک می‌توان به روش آستانه‌گذاری روی ضرایب موجک (مانند روش داناهو [۷] و [۱۰]، روش سون و همکاران [۸]، روش آقایان ابوطالبی و شیخ‌زاده [۹] و ...) و یا روش تفریق طیفی موجک [۶] و [۱۲] اشاره نمود. در استفاده از ضرایب موجک به چند طریق می‌توان عمل نمود. به‌عنوان مثال در روش تفریق طیفی موجک [۶] و [۱۲]، عمل تفریق به‌جای اینکه بر روی تابع اندازه تبدیل فوریه سیگنال صوت آغشته به نویز و سیگنال نویز انجام شود، بر روی ضرایب موجک صورت می‌گیرد و یا اینکه در مابقی روش‌های ذکرشده در فوق به‌جای استفاده مستقیم از نویز تخمینی، تنها برخی از پارامترهای آن استفاده می‌گردد و عمل غنی‌سازی با استفاده از آستانه‌گذاری روی ضرایب موجک تحقق می‌یابد.

ما در این مقاله ابتدا روش‌های تفریق طیفی و تفریق طیفی موجک را شرح داده، سپس با استفاده از آنالیز LPC [۱۳] تا [۱۶] به تخمین نویز از روی سیگنال صوتی پرداخته‌ایم. در نهایت نویز تخمینی را به روش تفریق طیفی موجک اعمال نموده‌ایم که در نتیجه به بهبود بسیار زیاد این روش انجامید.

۲- روش تفریق طیفی توان (PSS[1])

فرض نموده‌ایم که سیگنال و نویز جمع‌شونده هستند. لذا 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.

طبق فرض دوم، نویز و سیگنال ناهمبسته‌اند. از طرفی نویز نیز ناهمبسته است لذا می‌توان نوشت [۱۳]

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

$$R_{s,n}(\tau) = ۰ \qquad (۳)$$

که $R_n(\tau)$ تابع خودهمبستگی سیگنال تصادفی نویز و $R_{s,n}(\tau)$ نیز تابع همبستگی متقابل سیگنال‌های n و s می‌باشد. با توجه به روابط فوق و با فرض ایستابودن 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 می‌باشد. بنا به (۴) چنانچه چگالی طیف قدرت سیگنال تصادفی نویز در دسترس باشد، می‌توان چگالی طیف قدرت سیگنال صوتی تمیز را به‌دست آورد

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

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

این مقاله در تاریخ ۱۸ اردیبهشت ماه ۱۳۸۶ دریافت و در تاریخ ۲۹ مرداد ماه ۱۳۸۶ بازنگری شد.

مصطفی حیدری، دانشکده مهندسی برق و کامپیوتر، دانشگاه مازندران، بابل – کدپستی ۴۷۱۴۴ (email: baradaran5@yahoo.com).

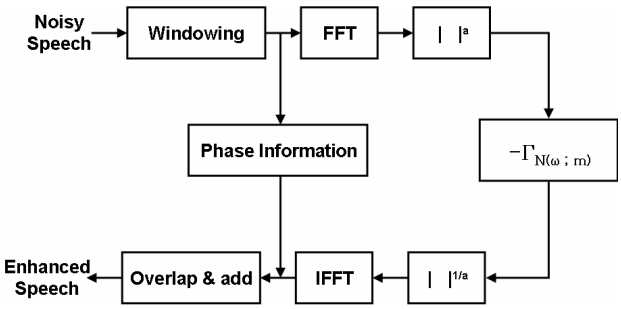
احسان نادرنژاد، دانشکده مهندسی برق و کامپیوتر، دانشگاه مازندران، بابل – کدپستی ۴۷۱۴۴ (email: ehsan_nader @yahoo.com).

محمدرضا کرمی ملایی، دانشکده مهندسی برق و کامپیوتر، دانشگاه مازندران، بابل – کدپستی ۴۷۱۴۴ (email: mkarami@nit.ac.ir).

1. Power Spectral Subtraction

Archive of SID

www.SID.ir



شکل ۱: بلوک دیاگرام روش تفریق طیفی تعمیم‌یافته.

از آنجایی که سیگنال صوتی تمیز در بازه‌های زمانی کوتاه ایستایی محلی دارند و نیز فرض ایستابودن نویز در بازه‌های کوتاه منطقی‌تر می‌باشد، لذا ابتدا روی سیگنال صوتی آغشته به نویز عمل پنجره‌گذاری انجام می‌شود و بدین طریق سیگنال گفتاری به فریم‌های کوتاه تقسیم می‌گردد و سپس عمل تفریق طیفی را بر روی هر فریم انجام می‌دهند

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

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

$$R_{s,n}(\tau;m) = ۰ \quad (۸)$$

که $x(n;m)$ سیگنال پنجره‌شده سیگنال صوتی $x(n)$ می‌باشد. با گرفتن تابع چگالی احتمال از طرفین (۶) داریم

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

از طرفی نیز می‌دانیم [۱۳]

$$\Gamma_x(\omega;m) = \frac{X(\omega;m)X^*(\omega;m)}{N^۲} = \frac{|X(\omega;m)|^۲}{N^۲} \quad (۱۰)$$

با توجه به اینکه مقدار $|X(\omega;m)|^۲$ به‌مراتب بزرگ‌تر از مقدار $N^۲$ می‌باشد، می‌توان به‌راحتی از ضریب $۱/N^۲$ چشم‌پوشی نموده و با تقریب خوبی به رابطه زیر رسید

$$\Gamma_x(\omega;m) \approx |x(\omega;m)|^۲ \quad (۱۱)$$

با استفاده از (۹) و (۱۱) رابطه زیر حاصل می‌گردد

$$|S(\omega;m)|^۲ = |X(\omega;m)|^۲ - |N(\omega;m)|^۲ \quad (۱۲)$$

که در روابط فوق $|X(\omega;m)|$ اندازه تبدیل فوریه سیگنال پنجره‌شده $x(n)$، $|S(\omega;m)|$ و $|N(\omega;m)|$ به‌ترتیب اندازه تبدیل فوریه سیگنال صوتی تمیز پنجره‌شده و سیگنال نویز پنجره‌شده می‌باشد.

همان‌طور که از (۱۲) پیداست، برای به‌دست 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 می‌زنند. 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 به‌عنوان 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 (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 [۱]

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

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

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

$$\begin{aligned}\hat{S}(\omega;m) &= \left|\hat{S}(\omega;m)\right| \exp\{i\varphi_{\hat{S}(\omega;m)}\} \\ &= [|X(\omega;m)|^۲ - |\hat{N}(\omega;m)|^۲]^{\frac{۱}{۲}} \exp\{i\varphi_{\hat{S}(\omega;m)}\}\end{aligned} \quad (۱۴)$$

که در روابط فوق $\hat{S}(\omega;m)$، $N(\omega;m)$ و $\hat{s}(\omega;m)$ به‌ترتیب تبدیل فوریه سیگنال تمیز تخمینی، تبدیل فوریه سیگنال نویز تخمینی و سیگنال تمیز تخمینی در حوزه زمان می‌باشند.

به روش فوق، روش تفریق طیفی توان گفته می‌شود، چرا که از توان دوم اندازه تبدیل فوریه که مبین قدرت و توان سیگنال می‌باشد استفاده کرده 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 تعمیم‌یافته می‌گویند که در شکل ۱ نشان داده شده است [۱۳]

$$\hat{S}(\omega;m) = [|X(\omega;m)|^a - |\hat{N}(\omega;m)|^a]^{\frac{۱}{a}} \exp\{i\varphi_{\hat{S}(\omega;m)}\} \quad (۱۵)$$

اما مسئله مهمی که در روش تفریق طیفی پیش می‌آید، مقادیر منفی برای اندازه تبدیل فوریه سیگنال تمیز می‌باشد. به‌عبارت دیگر هیچ تضمینی برای مثبت‌بودن اندازه تبدیل فوریه محاسبه‌شده برای سیگنال صوتی تمیز در هر یک از (۱۴) و (۱۵) وجود ندارد. دو روش برای اصلاح این مقادیر منفی وجود دارد [۱۳]

الف) اصلاح نیم‌موج

$$\left|\hat{S}(\omega;m)\right| = \begin{cases}\left|\hat{S}(\omega;m)\right| & \text{if} \quad \left|\hat{S}(\omega;m)\right| > ۰ \\ ۰ & \text{elsewhere}\end{cases} \quad (۱۶)$$

ب) اصلاح تمام‌موج

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

۳- تفریق طیفی روی ضرایب موجک

ایده کلی روش تفریق طیفی روی ضرایب موجک[1] (WSS) همانند روش تفریق طیفی می‌باشد. با این تفاوت که در این روش چون ضرایب تقریب و تفصیل تبدیل موجک دارای ماهیت زمانی هستند، عمل تفریق بر روی تبدیل فوریه ضرایب تقریب و تفصیل سیگنال صوتی آغشته به نویز و سیگنال نویز انجام می‌گیرد [۶] و [۱۲]

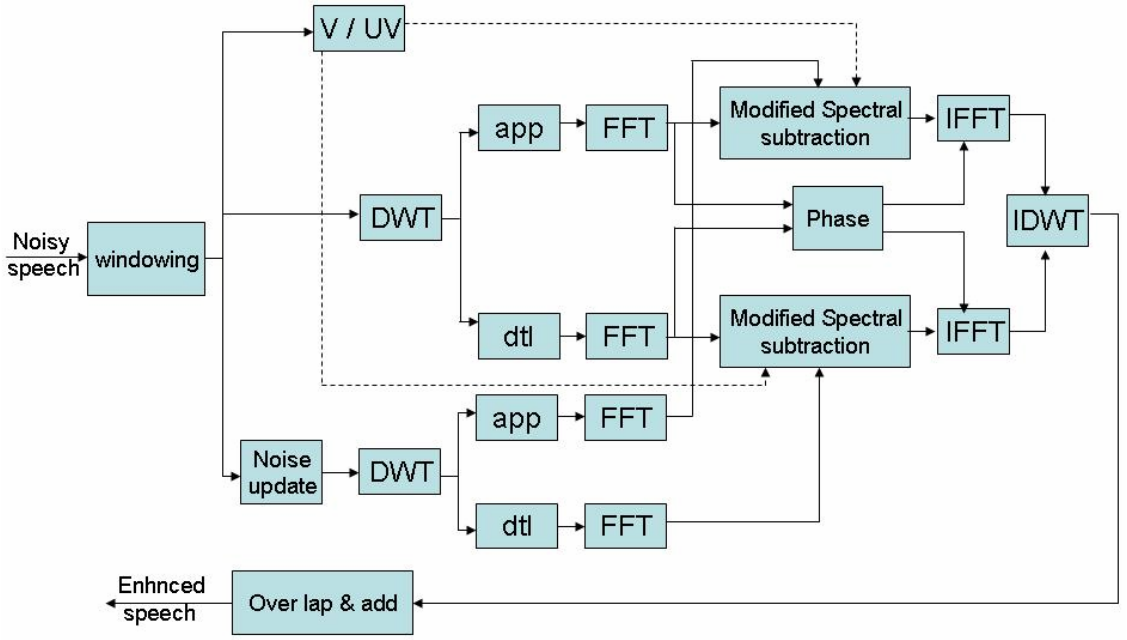
$$X_{app/dtl}(\omega;m) = S_{app/dtl}(\omega;m) + N_{app/dtl}(\omega;m) \quad (۱۸)$$

$$\left|\hat{S}_{app/dtl}(\omega;m)\right|^a = |X_{app/dtl}(\omega;m)|^a - \left|\hat{N}_{app/dtl}(\omega;m)\right|^a \quad (۱۹)$$

که در روابط فوق $X_{app/dtl}(\omega;m)$، $S_{app/dtl}(\omega;m)$ و $\hat{N}_{app/dtl}(\omega;m)$ به‌ترتیب تبدیل فوریه ضرایب تقریب و تفضیل سیگنال صوتی آغشته به

1. Wavelet Spectral Subtraction

Archive of SID
www.SID.ir



شکل ۲: بلوک دیاگرام روش تفریق طیفی روی ضرایب موجک.

نویز، سیگنال صوتی تمیز و سیگنال نویز تخمینی پنجره‌شده می‌باشد که نویز از قسمت سکوت سیگنال صوتی تخمین زده می‌شود.

همان‌طور که پیش از این گفته شد، روش تفریق طیفی سبب ایجاد نویز جدیدی به نام نویز موزیکال خواهد شد. بروتی [۲] 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. اگر پیشنهاد بروتی را در مورد تفریق طیفی موجک اعمال نماییم، (۱۹) به‌صورت زیر تغییر می‌یابد [۶]

$$\left|\hat{S}_{app/dtl}(\omega;m)\right|^a = \left|X_{app/dtl}(\omega;m)\right|^a - \alpha\left|\hat{N}_{app/dtl}(\omega;m)\right|^a \quad (۲۰)$$

که α فاکتور Over subtraction بوده و تابعی از SNR سیگنال صوتی می‌باشد. مقدار SNR در هر فریم نیز به‌صورت زیر محاسبه می‌گردد [۶]

$$SNR = ۱۰\log\frac{\sum_{\omega=b_i}^{e_i}\left|X_{app/dtl}(\omega;m)\right|^۲}{\sum_{\omega=b_i}^{e_i}\left|\hat{N}_{app/dtl}(\omega;m)\right|^۲} \quad (۲۱)$$

که b_i و e_i به‌ترتیب فرکانس‌های شروع و پایان تبدیل فوریه ضرایب تقریب و تفضیل در هر فریم می‌باشد.

بروتی برای اصلاح مقادیر اندازه تبدیل فوریه، فاکتور کف طیفی (β: Spectral floor) را معرفی نموده است که مطابق (۲۲) مقدار اندازه تبدیل فوریه را اصلاح می‌نماید [۲]

$$\left|\hat{S}_{app/dtl}(\omega;m)\right|^a = \begin{cases} \left|\hat{S}_{app/dtl}(\omega;m)\right|^a \\ \text{if } \left|\hat{S}_{app/dtl}(\omega;m)\right|^a > \beta\left|\hat{N}_{app/dtl}(\omega;m)\right| \\ \beta\left|\hat{N}_{app/dtl}(\omega;m)\right| \quad \text{else} \end{cases} \quad (۲۲)$$

٤- اصلاح روش تفریق طیفی روی ضرایب موجک (IWSS)

از آنجایی که سیگنال تحریک حروف بی‌صدا، نویز می‌باشد [۱۳] تا [۱۶]، از این رو به‌دلیل ماهیت نویزی حروف بی‌صدا در فریم‌های متعلق به حروف بی‌صدا فاکتور Over subtraction باعث می‌گردد تا انرژی تخمینی در (۲۰) برای فریم‌های بی‌صدا به میزان زیادی کاهش یابد ($\alpha > ۱$) [۶]، چرا که در این فریم‌ها دو انرژی از یک جنس و با اندازه تقریباً برابر از یکدیگر کم می‌شود. لذا بایستی به‌نحوی از مقدار α کاسته گردد تا مقدار انرژی تخمینی برای حروف بی‌صدا که از تفریق دو انرژی با ماهیت نویزی به‌دست می‌آید بیش از اندازه تضعیف نگردد. به‌همین منظور فاکتور جدیدی معرفی شده است که به‌صورت زیر تعریف می‌گردد [۶]

$$\eta = \begin{cases} ۰/۵ & \text{unvoiced frame} \\ ۱ & \text{else} \end{cases} \quad (۲۳)$$

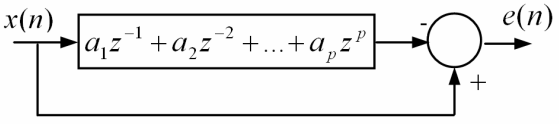
از این رو، (۲۰) به‌صورت زیر تغییر می‌یابد

$$\left|\hat{S}_{app/dtl}(\omega;m)\right|^a = \left|X_{app/dtl}(\omega;m)\right|^a - \eta\alpha\left|\hat{N}_{app/dtl}(\omega;m)\right|^a \quad (۲۴)$$

نحوه اصلاح روش تفریق طیفی موجک با استفاده از ضریب η در شکل ۲ آمده است. در این بلوک دیاگرام، ابتدا سیگنال صوتی به فریم‌های کوچک‌تر تقسیم می‌گردد، سپس ضمن محاسبه ضرایب تبدیل موجک برای هر یک از فریم‌ها، نوع فریم‌ها (صدادار و بی‌صدا) نیز تعیین می‌شود. در مرحله بعد اندازه تبدیل فوریه ضرایب موجک هر یک از فریم‌ها از ضرایب موجک متناظر در نویز تخمینی (که عموماً از فریم اول تخمین زده می‌شود) کم می‌گردد. میزان این کاسته‌شدن به نوع فریم بستگی دارد، به‌طوری که در فریم‌های بی‌صدا نویز تخمینی با یک ضریب کاهنده مواجه می‌شود تا از اندازه تبدیل فوریه ضرایب موجک در این فریم‌ها کمتر کاسته شود. در نهایت کافی است این اندازه جدید با فاز تبدیل فوریه هر یک از فریم‌ها ترکیب گردد تا با تبدیل معکوس فوریه و متعاقب آن تبدیل معکوس موجک، سیگنال بهبود یافته به‌دست آید.

Archive of SID

www.SID.ir



شکل ۳: نمودار بلوکی محاسبه سیگنال خطای $e(n)$.

۵- ضرایب پیشگویی خطی

از آنجایی که در الگوریتم پیشنهادی، از آنالیز پیشگویی خطی[1] (LPC) [۱۳] تا [۱۶] برای تخمین نویز استفاده شده است، لذا در این بخش به توضیح این آنالیز می‌پردازیم.

ضرایب پیشگویی خطی یکی از قدرتمندترین ابزار در پردازش صوت می‌باشد [۱۴]. ایده کلی این آنالیز این است که هر نمونه از سیگنال صوتی را می‌توان به‌صورت معادله‌ای خطی برحسب خروجی‌ها و ورودی‌های گذشته نوشت

$$s(n)=\sum_{k=1}^{p}a_k s(n-k)+\sum_{l=1}^{q}b_l u(n-l) \qquad (۲۵)$$

که a_k و b_l به‌ترتیب ضرایب مخرج و صورت فیلتر و $u(n)$ سیگنال ورودی می‌باشد که برای حروف صدادار یک قطار ضربه و برای حروف بی‌صدا یک رشته نویز تصادفی می‌باشد [۱۳] تا [۱۶]. با تبدیل z گرفتن از دو طرف (۲۵) می‌توان به رابطه زیر رسید

$$H(z)=\frac{S(z)}{U(z)}=\frac{1+\sum_{l=1}^{q}b_l z^{-l}}{\sum_{k=1}^{p}a_k z^{-k}} \qquad (۲۶)$$

در عمل برای سیگنال‌های صوتی یک مدل تمام‌قطب تقریب بسیار خوبی برای تابع تبدیل $H(z)$ می‌باشد [۱۳] و می‌توان آن را به‌صورت زیر بیان داشت

$$H(z)=\frac{1}{1-\sum_{k=1}^{p}a_k z^{-k}}=\frac{1}{A(z)} \qquad (۲۷)$$

در عمل و به‌خصوص برای حنجره انسان، p یک عدد صحیح در محدوده [۱۰] و [۱۳] انتخاب می‌شود.

نکته اساسی در محاسبه ضرایب پیشگویی خطی این است که این ضرایب مستقیماً از سیگنال صحبت به‌دست می‌آید. به‌همین منظور و به‌علت ماهیت تغییرپذیری سیگنال گفتار (با زمان) ابتدا عمل پنجره‌گذاری بر روی سیگنال صوتی انجام می‌شود و سپس ضرایب LP در فریم‌های کوتاه محاسبه می‌شوند [۱۴].

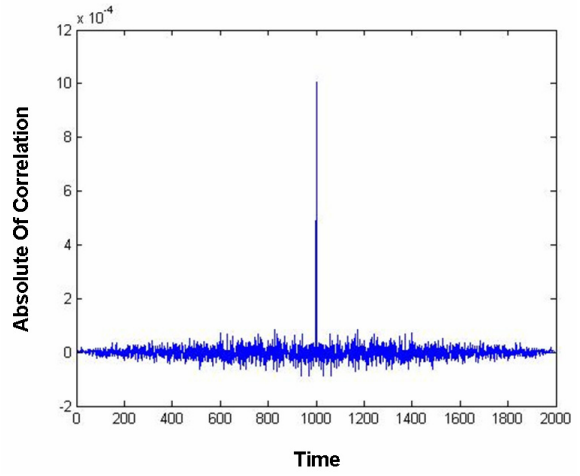
با توجه به توضیحاتی که ارائه شده است، با تقریب خوبی می‌توان هر نمونه از سیگنال صوتی را تنها با P نمونه قبلی از همان سیگنال صوتی (بدون استفاده از P نمونه ماقبل ورودی) محاسبه نمود [۱۳]

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

سیگنال خطا در حقیقت اختلاف بین سیگنال صوتی اصلی و سیگنال صوتی تخمین زده شده از روی P نمونه ماقبل می‌باشد

$$e(n)=S(n)-\hat{S}(n)=S(n)-\sum_{k=1}^{p}a_k S(n-k) \qquad (۲۹)$$

1. Linear Prediction Coding



شکل ۴: تابع خودهمبستگی سیگنال خطا.

که اگر از طرفین رابطه فوق تبدیل z گرفته شود داریم

$$E(z)=S(z)[1-\sum_{k=1}^{p}a_k z^{-k}]=A(z)S(z) \qquad (۳۰)$$

به عبارت دیگر، سیگنال $e(n)$، اختلاف بین سیگنال کدشده با ضرایب LPC و سیگنال اصلی می‌باشد (شکل ۳).

ضرایب LPC تنها قادر خواهند بود که سیگنال‌های همبسته را کد نمایند [۱۳] تا [۱۶]. بدین دلیل اگر ناهبستگی‌هایی در یک سیگنال وجود داشته باشد، ضرایب LPC تنها اجزای همبسته این سیگنال را کد می‌نمایند. بنابراین اگر این اجزای کدشده از سیگنال اصلی کم شود، اجزای ناهمبسته در خروجی ظاهر خواهند شد.

در روش‌های مبتنی بر تفریق طیفی، فرض بر این است که نویز با سیگنال ناهمبسته می‌باشد. بنابراین اگر سیگنال صوتی آغشته به نویز از فیلتری که اختلاف بین سیگنال کدشده و سیگنال اصلی را محاسبه می‌کند (نظیر فیلتر $A(z)$) عبور داده شود، در خروجی فیلتر نویز ناهمبسته اضافه شده به‌همراه سایر بخش‌های ناهمبسته سیگنال صوتی نویزی وجود خواهد داشت.

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.

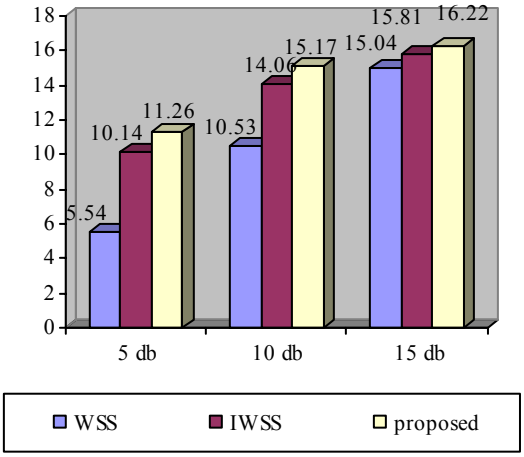
در شکل ۴ تابع خودهمبستگی سیگنال خطا ($e(n)$) که متعلق به یک سیگنال صوتی از دادگان TIMIT می‌باشد رسم گردیده است. همان‌طور که مشاهده می‌شود، سیگنال خودهمبستگی سیگنال خطا بسیار شبیه تابع خودهمبستگی یک سیگنال تصادفی نویز می‌باشد.

البته سیگنال $e(n)$ اگرچه همان سیگنال نویز اضافه‌شده به سیگنال صوتی تمیز نمی‌باشد، اما اکثر مشخصات آنرا در بر 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 چشم‌پوشی است.

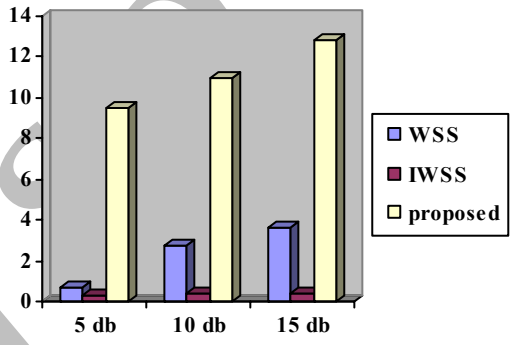
۶- اصلاح روش IWSS با استفاده ازآنالیز LPC (LPIWSS)

همان‌طور که پیش از این نیز گفته شد، در روش تفریق طیفی موجک از قسمت سکوت سیگنال برای تخمین نویز استفاده شده [۶] و [۱۲] که عموماً از فریم اول سیگنال صوتی آغشته به نویز به‌عنوان قسمت سکوت سیگنال استفاده می‌گردد. بدیهی است که در صورت عدم برقراری شرط فوق، روش تفریق طیفی موجک به‌شدت تضعیف می‌گردد. برای حل این

Archive of SID
www.SID.ir



شکل ۵: مقایسه SNR سیگنال غنی‌شده در روش‌های WSS، IWSS و LPIWSS (روش پیشنهادی).



شکل ۶: مقایسه SNR سیگنال غنی‌شده در روش‌های موجود و روش پیشنهادی (LPIWSS).

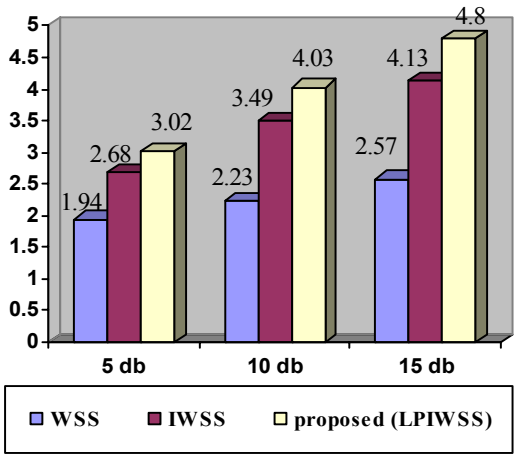
مشکل پیشنهاد می‌کنیم از آنالیز LPC استفاده گردد. بدین صورت که از فریم اول سیگنال آنالیز LPC گرفته و خروجی فیلتر $A(z)$ را نویز در نظر گرفت. از طرف دیگر برای اینکه اندازه تبدیل فوریه ضرایب موجک نویز تخمینی به نویز واقعی نزدیک‌تر باشد پیشنهاد می‌کنیم که از میانگین اندازه تبدیل فوریه ضرایب موجک سیگنال نویز تخمینی در (۲۰) استفاده گردد

$$\left|\hat{S}_{app/dtl}(\omega;m)\right|^{a} = \left|X_{app/dtl}(\omega;m)\right|^{a} - \eta\alpha\left|\overline{N}_{app/dtl}(\omega;m)\right|^{a} \qquad (۳۱)$$

۷- پیاده‌سازی و مقایسه الگوریتم‌ها

الگوریتم‌های تفریق طیفی موجک و تفریق طیفی موجک اصلاح‌شده را به‌همراه روش پیشنهادی بر روی پنجاه سیگنال صوتی از دادگان TIMIT با نسبت سیگنال به نویز اولیه ۵، ۱۰، ۱۵ (نویز از نوع سفید گوسی) اعمال شده است و همان‌طور که از شکل ۵ نیز پیداست، روش پیشنهادی به میزان زیادی سبب بهبود SNR سیگنال‌های صوتی گردیده است. در این نمودار مقدار $\beta = ۰٫۰۱$ در نظر گرفته شده است.

همان‌طور که پیش از این نیز بیان شده است، در روش‌های موجود عموماً از فریم اول به‌عنوان قسمت سکوت استفاده می‌گردد. ضعف شدید تفریق طیفی موجک هنگامی آشکار می‌گردد که فریم اول یک فریم سکوت نباشد. شکل ۶ به مقایسه این سه الگوریتم پرداخته است و همان‌طور که مشاهده می شود، الگوریتم پیشنهادی از یک برتری کاملاً مشهود در مورد این‌گونه سیگنال‌ها برخوردار است.



شکل ۷: مقایسه تست MOS سیگنال غنی‌شده در روش‌های WSS، IWSS و LPIWSS (روش پیشنهادی).

جدول ۱: نحوه امتیازگذاری پنج نمره‌ای در تست MOS.

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

تا به حال ملاک مقایسه روش موجود و روش پیشنهادی، نسبت سیگنال به نویز سیگنال غنی‌شده صوتی بوده است. اینک می‌خواهیم مقایسه کیفی بین روش موجود و روش پیشنهادی انجام دهیم. بدین منظور می‌خواهیم از تست شنوایی MOS[1] [۱۷] برای مقایسه روش‌ها استفاده نماییم.

پنجاه سیگنال صوتی از دادگان TIMIT با SNR اولیه ۵ db، ۱۰، ۱۵ 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 غنی‌شده بر اساس جدول ۱ استفاده نموده‌ایم [۱۷]. میانگین نمرات این شش نفر به این پنجاه سیگنال صوتی (۳۰۰ بار آزمایش به ازای هر SNR اولیه) در شکل ۷ آمده است.

۸- نتیجه‌گیری

در این مقاله روشی جدید برای غنی‌سازی سیگنال‌های صوتی بر مبنای تفریق طیفی موجک ارائه گردیده است. در این روش از آنالیز LPC برای تخمین میزان نویز از روی سیگنال صوتی استفاده گردیده است.

با توجه به منابع متعدد نویز و نیز تصادفی‌بودن آن، عموماً میزان قدرت نویز در تمام طول سیگنال به یک اندازه نمی‌باشد. لذا در روش پیشنهادی از هر فریم و با توجه به مشخصات آن فریم برای تخمین نویز در آن استفاده گردیده است.

نتایج آزمایشات نشان داده است که کیفیت سیگنال خروجی در روش پیشنهادی به میزان زیادی بهبود یافته است.

1. Mean Opinion Score

Archive of SID

www.SID.ir

مراجع

[1] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proc. of ICASSP'2002,* vol. 4, pp. 4160-4164, Orlando, FL, US, May 2002.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. on Acoust. Speech & Signal Processing*, vol. 27, no. 2-3, pp. 113-120, Apr. 1979.

[3] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE ICASSP,* vol. 4, pp. 208-211, Washington DC, US, Apr. 1979.

[4] 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, no. 10, pp. 277-279, Oct. 1997.

[5] P. S. Whitehead, D. V. Andeson, and M. A. Clements, "Adaptive acoustic noise suppression for speech enhancement," in *Proc. IEEE Int. Conf. on Multimedia & Expo.*, vol. 2, pp. 565-568, Jul. 2003.

[6] Y. Ghanbari and M. R. Karami, "Spectral subtraction in the wavelet domain for speech enhancement," *Int. J. of Software and Information Technologies (IJSIT)*, vol. 1, no. 1, pp. 26-30, Aug. 2004.

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

[8] I. Y. Soon, S. N. Koh, and C. L. Yeo, "Wavelet for Speech denoising," in *Proc. TENCON 97,* vol. 2, pp. 479-482, Brisbane, Australia, 2-4 Dec. 1997.

[9] H. Sheikhzadeh and H. R. Abutalebi, "An improved wavelet-based speech enhancement system," in *Proc. 7th European Conf. on Speech Communication and Technology, EuroSpeech'01,* pp. 1855-1858, Aalborg, Denmark, 3-7 Sep. 2001.

[10] D. L. Donoho and I. M. Johnston, "Ideal spatial adaptive via wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425-455, Sep. 1994.

[11] I. M. Johnston and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Static. Soc. B*, vol. 59, no. 2, pp. 319-351, 1997.

[12] Y. Ghanbari and M. R. Karami-Mollaei, "A new approach for speech enhancement based on adaptive thresholding of wavelet packets," *Speech Communication*, vol. 48, no. 8, pp. 927-940, Aug. 2006.

[13] J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*, 2nd edition, IEEE press, 2000.

[14] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.

[15] J. Tierney, "A study of LPC analysis of speech in additive noise," *IEEE Trans. Acoust. Speech and Signal Process.,* vol. 28, no. 4, pp. 379-389, Aug. 1980.

[16] 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.*, vol. 24, no. 6, pp. 488-494, Dec. 1976.

[۱۷] ی. قنبری، *کاهش نویز در سیگنال گفتار با استفاده از موجک*، پایان‌نامه کارشناسی ارشد دانشگاه مازندران، ۱۳۸۳.

مصطفی حیدری تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد اقتصاد به‌ترتیب در سال‌های ۱۳۸۲ و ۱۳۸۴ از دانشگاه مازندران به پایان رسانده است و هم‌اکنون عضو هیئت علمی مؤسسه آموزش عالی شمال می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: غنی‌سازی سیگنال گفتار، نهان‌سازی سیگنال‌های صوتی، قطعه‌بندی تصاویر و میدان‌های تصادفی مارکوف.

احسان نادرنژاد تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد اقتصاد به‌ترتیب در سال‌های ۱۳۸۱ و ۱۳۸۵ از دانشگاه مازندران به پایان رسانده است و هم‌اکنون عضو هیئت علمی مؤسسه آموزش عالی صنعتی مازندران می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش سیگنال، قطعه‌بندی تصاویر، معادلات با مشتقات جزیی و میدان‌های تصادفی مارکوف.

محمدرضا کرمی ملایی در سال ۱۳۷۰ مدرک کارشناسی مهندسی برق خود را از دانشگاه فردوسی مشهد و در سال ۱۳۷۳ مدرک کارشناسی ارشد مهندسی برق خود را در شاخه پردازش سیگنال از دانشگاه INPG فرانسه و دکترای خود را در سال ۱۳۷۷ در رشته مهندسی پزشکی از دانشگاه INPL فرانسه دریافت نمود. دکتر کرمی از سال ۱۳۷۷ عضو هیئت علمی دانشگاه مازندران می‌باشد و هم‌اکنون با عنوان دانشیار دانشکده مهندسی برق و کامپیوتر دانشگاه صنعتی بابل به کار مشغول است. زمینه‌های علمی مورد علاقه نام‌برده متنوع بوده و شامل موضوعاتی مانند ایده‌های نو در پردازش سیگنال صوت و تصویر، شناسایی آماری الگوها و مدارهای مخابراتی می‌باشد.

Archive of SID

www.SID.ir