

# روش تفریق طیفی چندباندی اصلاح شده برای بهبود صحبت

مهدی بکرانی، هادی صدوقی‌یزدی و مجتبی لطفی‌زاد

یکسان نمی‌باشد؛ بنابراین برخی فرکانسها بیش از سایر فرکانسها به نوبز آلوده می‌شوند؛ ثانیاً نوبز اثر غیریکسانی روی مصوتها<sup>۱</sup> و صامت‌های<sup>۲</sup> مختلف دارد [۱۵]. اندازه‌گیری‌های SNR قطعه‌ای<sup>۳</sup> برای صحبت آمیخته شده با نوبزهای محیطی نیز این مطلب را تأیید می‌کند [۱۵].

روش تفریق طیفی چندباندی [۱۶] نمونه‌ای از روشهای تفریق طیفی غیرخطی می‌باشد. این روش از لحاظ میزان محاسبات، جزء ساده‌ترین روشهای غیرخطی تفریق طیفی می‌باشد که در سالهای اخیر پیشنهاد شده است. شکل ۱ دیاگرام بلوکی این روش را نشان می‌دهد.

ابتدا سیگنال پنجره بندی شده و تبدیل فوریه آن گرفته می‌شود. آنگاه بعد از مرحله پیش‌پردازش که شامل میانگین‌گیری وزن‌دار از طیف سیگنال ورودی است، به چند باند تجزیه شده و در هر باند طیف نوبز تخمین زده می‌شود و سپس برای هر باند بر حسب SNR قطعه‌ای آن باند، فاکتور فوق‌تفریق ( $\alpha_i$ ) بدست می‌آید. از طرفی فاکتور وابسته به فرکانس ( $\delta_i$ ) نیز از روی طیف سیگنال نوبزی بدست آمده و سپس فرایند تفریق طیفی صورت می‌گیرد. نهایتاً از روشی به نام تعیین کف طیفی<sup>۴</sup> به عنوان مرحله پس‌پردازش استفاده شده تا نوبز مانده به حداقل برسد و نهایتاً از طیف کوتاه دوره حاصل شده، تبدیل معکوس گرفته می‌شود و قابهای زمانی حاصل شده، با تکنیک همپوشانی و افزودن کنار هم قرار می‌گیرند. نتیجه، صحبت بهبودیافته است.

بلوک‌های پر رنگ شکل ۱، قسمتهایی را نشان می‌دهند که در این مقاله اصلاحات لازم در آنها صورت گرفته است. کارهای انجام شده در هر بلوک بطور مختصر عبارتند از:

- در مرحله پیش‌پردازش که شامل یک مرحله میانگین‌گیری وزن‌دار از هر ۵ قاب سیگنال ورودی است، عمل میانگین‌گیری با فاکتوری وابسته به SNR قطعه‌ای ورودی تنظیم می‌شود، بطوری که در قسمتهایی که عمدتاً از نوبز تشکیل شده هموارسازی<sup>۵</sup> بیشتر و در قسمتهایی که عمدتاً شامل سیگنال هست، هموارسازی ضعیفتر صورت می‌گیرد. این ایده سبب می‌شود اعوجاج ناشی از میانگین‌گیری وزن‌دار به حداقل برسد.
- در قسمت تخمین‌گر نوبز از تکنیک جدیدی برای تخمین نوبز استفاده شده است که نیاز به آشکارساز صحبت/سکوت نداشته و علاوه بر نوبز ایستان<sup>۶</sup>، برای نوبزهای غیرایستان<sup>۷</sup> و رنگی<sup>۸</sup> نیز مفید می‌باشد. تخمین نوبز با کمک پارامتر هموارسازی که وابسته به SNR زیرباندی<sup>۹</sup> تخمینی است، بطور پیوسته و وقتی بروز می‌گردد.
- نهایتاً با استفاده از تخمین SNR، فاکتور فوق‌تفریقی پیشنهاد شده است که سبب به حداقل رسیدن نوبز مانده می‌گردد.

چکیده: در این مقاله یک روش اصلاح شده تفریق طیفی چندباندی برای حذف نوبز از سیگنال صحبت ارائه شده است. در این روش با تخمین آمارگان نوبز غیرایستان، تخمینی از SNR برای سیگنال صحبت بدست می‌آید که در میانگین‌گیری وقتی از قابهای سیگنال صحبت بکار گرفته می‌شود. همچنین با استفاده از آن یک فاکتور فوق‌تفریق بدست می‌آید که سبب کاهش قابل توجه نوبز موزیکال می‌شود. تنظیم پارامترهای الگوریتم نهایی بر اساس SNR بدست آمده انجام می‌شود؛ بطوری که الگوریتم نهایی دارای حداقل نوبز موزیکال و اعوجاج در سیگنال صحبت است.

نتایج حاصل از معیارهای ایتاکورا-سایتو، SNR سراسری و SNR قطعه‌ای نشان می‌دهد که الگوریتم پیشنهادی دارای سیگنال خروجی با کیفیت بهتری نسبت به دو الگوریتم تفریق طیفی چندباندی و نوع بهبودیافته آن می‌باشد.

کلید واژه: تفریق طیفی چندباندی، بهبود صحبت، نوبز غیرایستان، فاکتور فوق‌تفریق، نوبز موزیکال.

## ۱- مقدمه

نوبز صوتی ناشی از محیط، بطور آکوستیکی به سیگنال صحبت اضافه می‌شود. عمده‌ترین نوبزهایی که در مباحث بهبود صحبت مطرح هستند، نوبزهای باند وسیع (نوبز سفید و نوبز رنگی) می‌باشند [۱]. در میان روشهای حذف نوبز باند پهن، روشهای مبتنی بر تفریق طیفی بدلائل متعددی همچون سادگی پیاده‌سازی، توان بالا در حذف نوبز، سرعت بالای الگوریتم و قابلیت پیاده‌سازی تک‌کاناله از اهمیت زیادی برخوردار هستند، بطوری که امروزه این روش در بسیاری از سیستم‌های بازنمایی صحبت، کدکننده‌ها، گوشی‌های موبایل، سمعکها و غیره برای کاهش نوبز محیطی مورد استفاده قرار می‌گیرد و با آنکه حدود سه دهه از ارائه روش تفریق طیفی می‌گذرد، هنوز موضوع بسیاری از کارهای تحقیقاتی امروزی می‌باشد [۱] تا [۱۰].

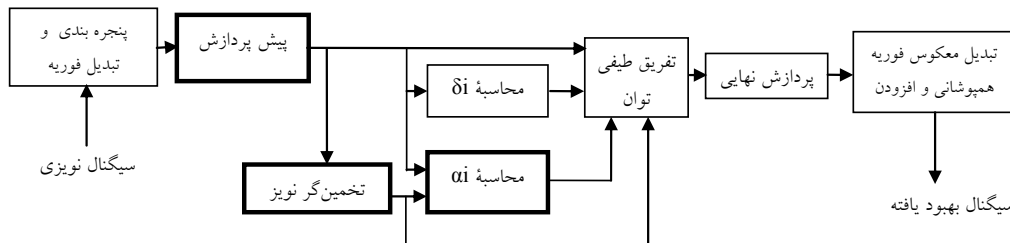
مهمترین چالشهای موجود در روش تفریق طیفی، ایجاد اعوجاج در سیگنال صحبت پردازش شده و ایجاد نوبز مانده‌ای که به نام "نوبز موزیکال" مشهور است، می‌باشد. تاکنون روشهای زیادی برای بهبود عملکرد تفریق طیفی پیشنهاد شده‌اند که اکثر آنها روشهایی غیرخطی هستند [۱۱] تا [۱۴]. این روشها برخلاف روش کلاسیک، حذف نوبز را در هر مؤلفه فرکانسی و یا در هر باند فرکانسی بطور مجزا انجام می‌دهند تا تنها مقادیر لازم از طیف نوبز در هر مؤلفه فرکانسی را تفریق کنند تا بدین وسیله از تفریق مخرب طیف جلوگیری کرده و در عین حال تا حد ممکن نوبز مانده حذف گردد. علت کارایی مؤثر روشهای غیرخطی تفریق طیفی اینست که اولاً نوبز موجود در محیط بر خلاف نوبز سفید، دارای طیف یکنواخت نبوده و نوبز در کل بازه طیف سیگنال صحبت دارای اثر

این مقاله در تاریخ ۳ دی ماه ۱۳۸۳ دریافت و در تاریخ ۱۳ آذر ماه ۱۳۸۴ بازنگری شد.

مهدی بکرانی و مجتبی لطفی‌زاد، دانشکده فنی، دانشگاه تربیت مدرس، تقاطع بزرگراه‌های شهید دکتر چمران و آل احمد، صندوق پستی ۱۱۱-۱۴۱۱۵ تهران، ایران (email: m\_bekrani@yahoo.com).

هادی صدوقی‌یزدی، دانشکده فنی، دانشگاه تربیت معلم سبزوار، سبزوار، ایران (email: sadoghi@sttu.ac.ir).

1. Vowel
2. Consonant
3. Segmental SNR
4. Spectral Flooring
5. Smoothing
6. Stationary
7. Non-Stationary
8. Colored Noise
9. Sub-Band



شکل ۱: بلوک دیاگرام تفریق طیفی چندباندی [۱۶].

که در آن  $|Y_i(k)|^2$  طیف سیگنال نویزی پس از مرحله پیش‌پردازش می‌باشد. در مرحله پیش‌پردازش، واریانس مؤلفه‌های فرکانسی سیگنال، با کمک میانگین‌گیری از قابها کاهش می‌یابد، این عمل برای کاهش میزان نویز مانده در صحبت و رسیدن به کیفیت خوب صحبت لازم است. حالت کلی عمل میانگین‌گیری به صورت رابطه زیر است

$$\bar{Y}_i(k) = \sum_{j=-M}^M W_j Y_{i-j}(k) \quad (4)$$

که در آن  $i$  اندیس قاب و  $W_j$  ضرایب مورد استفاده برای وزن‌دهی قابها است. وقتی  $W_j = 1$  (به ازای هر  $j$ ) است میانگین‌گیری از نوع ساده آن می‌شود و در غیر اینصورت میانگین‌گیری وزن‌دار خواهد بود. البته به دلیل اینکه سیگنال صحبت در بازه‌های کوتاهی ایستاد است، تعداد قابهای همسایه برای میانگین‌گیری محدود است. وزنه‌های استفاده شده در روش MBSS بطور تجربی برای میانگین‌گیری از هر ۵ قاب همسایه بدست آمده است و مقادیر ضرایب وزن‌دهی بصورت  $W_L = [0.09, 0.25, 0.32, 0.25, 0.09]$  می‌باشد.

$\alpha_i$  فاکتور فوق‌تفریق وابسته به SNR قطعه‌ای در باند  $i$  ام بوده و به‌صورت زیر تعریف می‌شود [۱۶]

$$\alpha_i = \begin{cases} 5, & SNR_i \leq -5 \text{ dB} \\ 4 - \frac{3}{20} SNR_i, & -5 \text{ dB} < SNR_i \leq 20 \text{ dB} \\ 1, & SNR_i > 20 \text{ dB} \end{cases} \quad (5)$$

در این رابطه SNR قطعه‌ای در باند  $i$  ام به‌صورت زیر محاسبه می‌شود

$$SNR_i = 10 \log \frac{\sum_{k=b_i}^{e_i} |Y_i(k)|^2}{\sum_{k=b_i}^{e_i} |\hat{N}_i(k)|^2} \quad (6)$$

$\delta_i$  فاکتوری وابسته به باند بوده و مقدار آن در هر باند از رابطه زیر بدست می‌آید

$$\delta_i = \begin{cases} 1, & f_i \leq 1 \text{ kHz} \\ 2/5, & 1 \text{ kHz} < f_i \leq \frac{Fs}{2} - 2 \text{ kHz} \\ 1/5, & f_i > \frac{Fs}{2} - 2 \text{ kHz} \end{cases} \quad (7)$$

در رابطه فوق  $f_i$  فرکانس بالایی باند  $i$  ام بوده و  $Fs$  فرکانس نمونه‌برداری است.  $b_i$  و  $e_i$  مؤلفه‌های ابتدا و انتهای باند فرکانسی  $i$  ام می‌باشد [۱۶]. تعداد باندهای فرکانسی در این روش ۴ باند می‌باشد که با ملاحظه میزان بهبود حاصل شده در ازای میزان افزایش پیچیدگی، انتخاب شده است.

پس از عمل تفریق یک مرحله پردازش نهایی روی طیف حاصل شده، انجام می‌شود. این مرحله برای اصلاح مقادیر منفی طیف مورد استفاده قرار می‌گیرد. رابطه زیر نحوه این عمل را نشان می‌دهد

بخش‌بندی مقاله به صورت زیر است: بخش دوم به معرفی اصول تفریق طیفی ساده و چندباندی می‌پردازد. بخش سوم شامل ایده‌های پیشنهادی برای اصلاح روش تفریق طیفی چندباندی است که شامل چهار زیربخش میانگین‌گیری وقتی، تخمین نویز، فاکتور فوق‌تفریق وابسته به SNR پسین و تنظیم پارامترهای الگوریتم بر اساس SNR پسین می‌باشد. بخش چهارم به پیاده‌سازی الگوریتم و نتایج اختصاص دارد و نهایتاً جمع‌بندی و نتیجه‌گیری در بخش پایانی می‌آید.

## ۲- اصول و روابط تفریق طیفی

فرض اصلی در روش تفریق طیفی این است که سیگنال صحبت و نویز بطور آکوستیکی با هم جمع شده‌اند و این فرض در عمل برای بسیاری از محیط‌های واقعی، فرض درستی است. با توجه به اینکه خاصیت جمع‌شوندگی در حوزه فرکانس (یعنی پس از تبدیل فوری گسسته) همچنان صحیح است، سعی می‌شود که با تخمین طیف فرکانسی نویز و کم کردن آن از طیف فرکانسی سیگنال نویزی، تخمینی از طیف سیگنال خالص بدست آید. بر مبنای این فرضیه معادله اساسی تفریق طیفی به صورت زیر است [۱]

$$|\hat{X}(f)|^b = |Y(f)|^b - \alpha |N(f)|^b \quad (1)$$

که در آن  $|\hat{X}(f)|^b$  تخمینی از طیف سیگنال  $|X(f)|^b$  بوده و  $|N(f)|^b$  طیف نویز متوسط‌گیری شده و یا هموارسازی شده در زمان می‌باشد. برای تفریق طیفی توان  $b=2$  است. مقدار  $\alpha$  در ساده‌ترین حالت یک است، لیکن بسته به کاربرد الگوریتم و شرایط محیط می‌تواند مقادیر دیگری داشته باشد یا بطور وقتی تعیین شود. برای بازیافت سیگنال زمانی، اندازه طیف تخمینی با فاز سیگنال نویزی ترکیب می‌شود و سپس با تبدیل معکوس DFT به حوزه زمان برگشت داده می‌شود [۱].

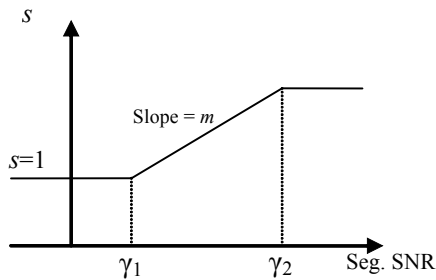
تفریق طیفی غیرخطی با رابطه زیر داده می‌شود [۱]

$$|\hat{X}(f)| = |Y(f)| - \alpha (SNR(f)) |N(f)|_{NL} \quad (2)$$

که در آن  $\alpha(SNR(f))$  فاکتور فوق‌تفریق وابسته به SNR بوده و  $|N(f)|_{NL}$  تخمین غیرخطی طیف نویز می‌باشد. در این رابطه  $b=1$  در نظر گرفته شده است و در حالت کلی می‌تواند مخالف یک باشد. تفریق طیفی غیرخطی کارایی و انعطاف‌پذیری بیشتری نسبت به تفریق طیفی کلاسیک دارد؛ زیرا دارای سازگاری با SNR سیگنال ورودی در بازه‌های فرکانسی مورد نظر می‌باشد [۱].

روش تفریق طیفی چندباندی<sup>۱</sup> (MBSS) یکی از روشهای غیرخطی تفریق طیفی می‌باشد [۱۶]. رابطه تفریق طیفی چندباندی به‌صورت زیر است

$$|\hat{X}_i(k)| = |Y_i(k)| - \alpha_i \delta_i |\hat{N}_i(k)| \quad b_i \leq k \leq e_i \quad (3)$$



شکل ۳: نمودار تغییرات پارامتر وزن دهی  $s$  بر حسب SNR قطعه‌ای (Seg. SNR).

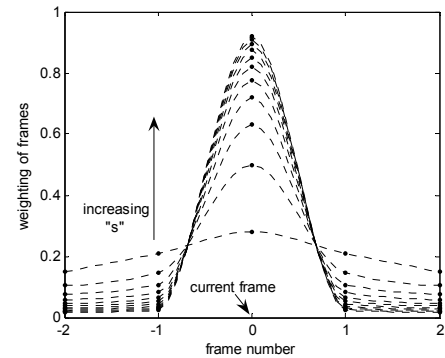
منظور ما ضرایب وزن دهی را به صورت  $W = (1/s^2)[\omega_1 \ \omega_2 \ s^2 - 2(\omega_1 + \omega_2) \ \omega_3 \ \omega_4]$  در نظر گرفتیم که در آن  $s$  فاکتوری وابسته به SNR قطعه‌ای می‌باشد. وزنهای این رابطه بطور تجربی با تأکید بر کاهش نویز مانده در  $s = 1$  انتخاب شده‌اند. از آنجا که مجموع وزن‌ها باید یک باشد، مقدار وزن وسطی  $(\omega_1 + \omega_2) - 2s^2$  در نظر گرفته شده است. شکل ۲ نمودار ضرایب وزن دهی قاب‌ها را بر حسب  $s$  های مختلف نشان می‌دهد.

مقدار  $s$  بستگی به میزان SNR قطعه‌ای دارد. هر جا که SNR قطعه‌ای زیاد شود،  $s$  باید افزایش یابد تا درصد کمتری از قاب‌های مجاور در میانگین‌گیری شرکت کند و بالعکس. بنابراین یک رابطه خطی به صورت شکل ۳ بین  $s$  و SNR قطعه‌ای پیشنهاد می‌شود.

حداقل مقدار  $s$  برابر ۱ است که این حالت برای قسمت‌های سیگنال با حداقل SNR های قطعه‌ای اعمال می‌شود. مقادیر  $\gamma_1$ ،  $\gamma_2$  و  $m$  را می‌توان بسته به محدوده SNR ورودی تنظیم کرد. علت محدود کردن  $s$  در SNR های قطعه‌ای زیاد، اطمینان از میانگین‌گیری با درصد کم در تمام قسمت‌های سیگنال صحبت است. این باعث می‌شود نویز مانده حاصل از خطای تخمین SNR قطعه‌ای، کاهش یابد. با توجه به اینکه تا این مرحله، تخمین گر و فاکتور فوق تفریق در الگوریتم پیشنهادی نظیر روش تفریق طیفی چندباندی هستند، ضرایب مناسب برای وزن دهی بصورت  $\omega_1 = 0.15$ ،  $\omega_2 = 0.21$  و ضرایب نمودار شکل ۳ بصورت  $\gamma_1 = 0$  dB،  $\gamma_2 = 20$  dB،  $m = 0.5$  در محدوده SNR سراسری ورودی برابر با صفر تا ۱۰ دسی‌بل بدست آمده است. این مقادیر بطور تجربی در جهت بهبود سیگنال صحبت و کاهش نویز مانده انتخاب شده‌اند. بدین ترتیب علاوه بر اینکه نویز موزیکال مثل حالت میانگین‌گیری وزن دار ساده که در روش [۱۶] بکار رفته حذف می‌شود، اعوجاج صحبت نیز به حداقل می‌رسد.

در ادامه کار، روش تفریق طیفی چندباندی بهبودیافته (IMBSS) [۱۷] که برای نویزهای گوسی سفید پیشنهاد شده است، با روش پیشنهادی MMBSS و همچنین روش MBSS مقایسه می‌شود. در روش IMBSS با توجه به اینکه نویز موجود در سیگنال صحبت از نوع نویز سفید است، پس از اینکه تخمینی از نویز در هر باند بدست می‌آید، از نویز تخمینی روی کل باند فرکانسی میانگین‌گیری کرده و حاصل آنرا جایگزین نویز تخمینی در تمام باندها می‌کند. سایر مراحل این الگوریتم نظیر الگوریتم MBSS است. البته الگوریتم IMBSS با  $b = 1/5$  پیاده‌سازی شده است (رابطه (۱) ملاحظه شود) و ما برای مقایسه صحیح، آنرا با  $b = 2$  پیاده کرده‌ایم.

مقایسه الگوریتم‌ها با سه نوع نویز یعنی نویز سفید، نویز کارخانه<sup>۳</sup> و نویز همهمه<sup>۴</sup> که از پایگاه داده NoiseX گرفته شده بود [۱۸]، انجام شده



شکل ۲: نمودار مقدار وزن  $s$  قاب همسایه با توجه به مقدار پارامتر  $s$ .

$$|\hat{X}(f)|^2 = \begin{cases} |\hat{X}(f)|^2 & \text{if } |\hat{X}(f)|^2 > \beta |\hat{N}(f)|^2 \\ \beta |\hat{N}(f)|^2 & \text{else} \end{cases} \quad (8)$$

که در آن پارامتر کف طیفی  $\beta$  برابر  $0.02$  می‌باشد [۱۶].

در این روش تخمین گر پیشنهادی از نوع ابتدایی آن؛ یعنی استفاده از آشکارساز صحبت/سکوت و میانگین‌گیری از طیف سیگنال در بازه‌های سکوت برای تخمین طیف نویز در تمام بازه سیگنال ورودی می‌باشد. عیب این روش، نیاز آن به آشکارساز دقیق صحبت/سکوت است تا خطای تخمین قابل قبول باشد. مشکل دیگر این است که صرفاً استفاده از آشکارساز برای تخمین نویزهای غیرایستاد ایده مناسبی نیست؛ زیرا در صورتی که پارامترهای نویز در بازه‌های صحبت تغییر کنند، خطای تخمین نویز زیاد می‌گردد. در بخش بعد روش دیگری برای تخمین نویز شرح داده می‌شود که نیاز به آشکارساز صحبت/سکوت نداشته و می‌تواند نویزهای غیرایستاد را با دقت خوبی دنبال کند.

### ۳- اصلاح تفریق طیفی چندباندی

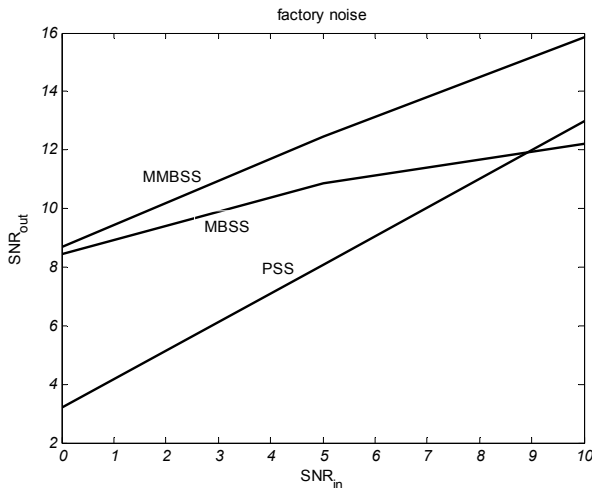
در این قسمت اصلاحاتی که به روش تفریق طیفی چندباندی و روش بهبودیافته آن اعمال شده است، شرح داده می‌شود. این اصلاحات به ترتیب شامل اصلاح مرحله پیش پردازش، افزودن تخمین گر نویز غیرایستاد و اصلاح فاکتور فوق تفریق می‌باشد. ما روش پیشنهادی را در مقایسه با روش MBSS (۳)، روش تفریق طیفی چندباندی اصلاح شده<sup>۲</sup> (MMBSS) می‌نامیم.

#### ۳-۱- اصلاح مرحله پیش پردازش

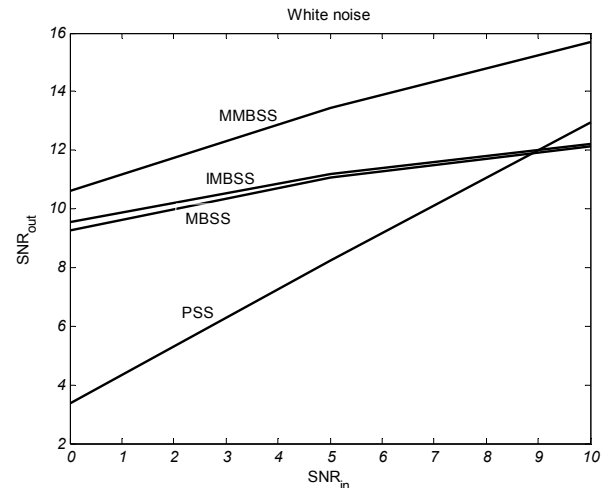
همانطور که اشاره شد میانگین‌گیری از قاب‌ها سبب کاهش واریانس سیگنال نویزی و در نتیجه کاهش واریانس نویز روی سیگنال صحبت نویزی می‌گردد. علاوه بر آن محتویات طیفی صحبت را تقویت می‌کند، بنابراین از تفریق مخرب جلوگیری می‌کند [۱۵]. اگر چه میانگین‌گیری سبب کاهش نویز مانده می‌شود، ولی بدلیل ایستاد نبودن صحبت باعث بروز اعوجاج در صحبت پردازش شده، گشته و منجر به کاهش فهم صحبت می‌گردد. به این خاطر ما برای حداقل کردن اعوجاج صحبت، تکنیکی برای میانگین‌گیری پیشنهاد کرده‌ایم که در آن نحوه وزن دهی به قاب‌ها برای میانگین‌گیری متناسب با میزان SNR قطعه‌ای می‌باشد. در این روش عمل میانگین‌گیری در نواحی که نویز سیگنال زیاد است، کاملاً صورت گرفته و در نواحی که میزان نویز در آن کم بوده و بیشتر شامل سیگنال‌های صحبت است، تقریباً میانگین‌گیری صورت نمی‌گیرد. به این

3. Factory Noise  
4. Babble Noise

1. Spectral Floor  
2. Modified Multi-Band Spectral Subtraction



شکل ۳: مقایسه میزان بهبود SNR روشهای تفریق طیفی در سیگنال صحبت آغشته به نویز رنگی (کارخانه).



شکل ۴: مقایسه میزان بهبود SNR روشهای تفریق طیفی در سیگنال صحبت آغشته به نویز سفید.

جدول ۱: مقایسه روشهای PSS و MBSS و IMBSS و روش پیشنهادی از نظر میزان SNR سیگنال بهبود یافته (نویز سفید گوسی) برای SNR<sub>in</sub> = ۵ dB

Seg. SNR <sub>in</sub> = -۶٫۱۹ dB	PSS	MBSS	IMBSS	MMBSS
Seg. SNR <sub>out</sub> (dB)	-۲٫۸۳	۵٫۳۳	۵٫۶۵	۶٫۳۷
Avg. IS	۷٫۸۸	۳٫۶۲	۳٫۳۶	۳٫۱۹

جدول ۲: مقایسه روشهای PSS و MBSS و روش پیشنهادی از نظر میزان SNR سیگنال بهبود یافته (نویز رنگی - همهمه) برای SNR<sub>in</sub> = ۵ dB

Seg. SNR <sub>in</sub> = -۵٫۸۷ dB	PSS	MBSS	MMBSS
Seg. SNR <sub>out</sub> (dB)	-۳٫۲۷	۲٫۳	۲٫۷۵
Avg. IS	۰٫۵۹	۰٫۵۵	۰٫۵۱

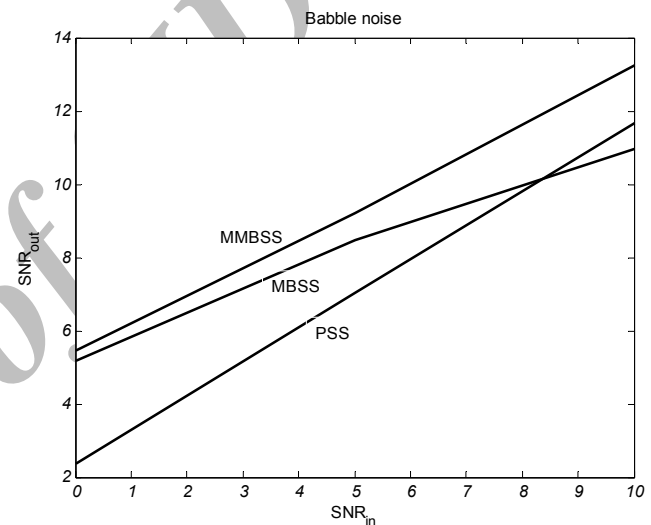
در روش پیشنهادی (به ازای هیچ کدام از نویزها) دیده نمی‌شود؛ علت این امر را می‌توان به این صورت بیان کرد که با افزایش SNR ورودی، SNR قطعه‌ای نیز به طور متوسط افزایش می‌یابد. بنابراین با توجه به نحوه وزن دهی درصد کمتری از قابها در عمل میانگین‌گیری، که عامل اعوجاج صحبت است، شرکت می‌کنند، در نتیجه بهبود بیشتری در سیگنال حاصل خواهد شد.

جدول ۱ و ۲ مقایسه این سه روش را با معیارهای ایتاکورا-سایتو و SNR قطعه‌ای برای سیگنال ورودی با SNR = ۵ dB نشان می‌دهد. در شبیه‌سازیها از متوسط معیار ایتاکورا-سایتو و همچنین متوسط SNR قطعه‌ای روی تمام قابهای صحبت استفاده شده است. رابطه متوسط معیار ایتاکورا-سایتو به صورت زیر است [۱۹]

$$d(\mathbf{c}_1, \mathbf{c}_2) = \frac{(\mathbf{c}_1 - \mathbf{c}_2)^T \mathbf{R} (\mathbf{c}_1 - \mathbf{c}_2)}{\mathbf{c}_1^T \mathbf{R} \mathbf{c}_1} \quad (9)$$

که در آن  $\mathbf{c}_1$  بردار ضرایب پیشگویی سیگنال خالص بوده و بردار  $\mathbf{R}$  ماتریس خودهمبستگی توپلیتز<sup>۲</sup> سیگنال خالص و  $\mathbf{c}_2$  بردار ضرایب پیشگویی سیگنال بهبود یافته می‌باشد.

با توجه به جداول بالا ملاحظه می‌شود معیار فاصله ایتاکورا-سایتو در روش پیشنهادی برای هر دو نوع نویز کمتر از سایر روشها است که این مبین کیفیت بهتر صوت خروجی نسبت به سایر روشها است. معیار SNR



شکل ۵: مقایسه میزان بهبود SNR روشهای تفریق طیفی در سیگنال صحبت آغشته به نویز رنگی (همهمه).

است. نتایج حاصل از شبیه‌سازیها در شکل‌های ۴ تا ۶ و جداول ۱ و ۲ ملاحظه می‌گردد. برای ارزیابی اثر میانگین‌گیری و فاکتور فوق تفریق روی تفریق طیفی، نتایج حاصل از تفریق طیفی ساده<sup>۱</sup> (PSS) (رابطه (۱) با پارامترهای  $\alpha = 1$  و  $b = 2$ ) نیز در اینجا ارائه شده است. نتایج حاصل شده، از میانگین‌گیری روی ۲۰ نمونه مختلف سیگنال صحبت مرد و زن، هر کدام با طول ۶ ثانیه بدست آمده است. لازم به ذکر است که نمونه‌های سیگنال آزمون متفاوت از نمونه‌های مورد استفاده برای تنظیم پارامترها می‌باشد.

همان طور که در شکل‌های ۴ تا ۶ ملاحظه می‌شود، بهبود حاصل از روش پیشنهادی (MMBSS) در SNRهای بین صفر تا ۱۰ دسی‌بل بیشتر از روشهای IMBSS (برای نویز سفید)، MBSS و PSS می‌باشد؛ نکته دیگر اینکه اگر چه روش IMBSS نسبت به MBSS کارایی بهتری از خود نشان می‌دهد، لیکن بهبود آن نسبت به MBSS چندان قابل توجه نیست. در SNRهای بالا (حدود ۱۰ دسی‌بل) بهبود SNR سراسری در روش PSS بیشتر از MBSS و IMBSS است؛ این امر نشان می‌دهد که در SNRهای زیاد که انرژی سیگنال صحبت در مقایسه با نویز زیاد است اثر افت SNR ناشی از میانگین‌گیری قابل ملاحظه می‌گردد. این مسئله

دیگر اگر صحبت موجود باشد، توان سیگنال جدید  $\hat{\sigma}_{x_i}^2(p)$  از تخمین نویز قدیم  $\hat{\sigma}_{w_i}^2(p-1)$  بسیار بزرگتر است، بنابراین از (۱۳)،  $\alpha_i(p) \approx 1$  نتیجه می‌شود. از این رو بهنگام‌سازی نویز در (۱۲) کندتر خواهد بود. در طول قابهای صحبت داریم  $\hat{\sigma}_{w_i}^2(p) \approx \hat{\sigma}_{w_i}^2(p-1)$ ، یعنی فرآیند بهنگام‌سازی نویز تقریباً متوقف می‌شود و تخمین نویز تقریباً برابر تخمین نویز در قاب قبلی است؛ زیرا  $\alpha_i(p)$  تقریباً برابر یک است. برای هموارسازی بیشتر عمل فیلترینگ زیر برای تخمین نهایی بکارگرفته می‌شود

$$\hat{\sigma}_{w_i, final}^2(p) = \alpha_{final} \hat{\sigma}_{w_i, final}^2(p-1) + (1 - \alpha_{final}) \hat{\sigma}_{x_i}^2(p) \quad (14)$$

در این رابطه مقدار  $\alpha_{final}$  در حدود ۰.۵ تا ۰.۸ می‌باشد [۲۰]. در نهایت تخمین واریانس سیگنال خالص با کمک رابطه زیر محاسبه می‌گردد

$$\hat{\sigma}_{s_i}^2(p) = \max\{\hat{\sigma}_{x_i}^2(p) - \hat{\sigma}_{w_i, final}^2(p), 0\} \quad (15)$$

در اینجا عملگر max برای اجتناب از تخمینهای منفی در رابطه وارد شده است [۲۱].

برای محاسبه ضرایب بهره، ابتدا سیگنال صحبت به کمک یک بانک فیلتری به چند باند فرکانسی تجزیه می‌شود. سپس ضریب بهره سیگنال برای هر باند از رابطه زیر بدست می‌آید که برای تخمین نویز بکار می‌رود [۲۱]

$$Ky_i = \frac{\sigma_{w_i}^2}{\sigma_{s_i}^2 + \sigma_{w_i}^2} \quad (16)$$

در این رابطه  $\sigma_{s_i}^2$  واریانس سیگنال خالص،  $\sigma_{w_i}^2$  واریانس نویز و  $Ky_i$  ضریب بهره سیگنال در باند  $i$  ام می‌باشد [۴] و [۲۰]. سپس طبق رابطه زیر تخمینی از نویز بدست می‌آید

$$\hat{N}_i(n) = Ky_i * \bar{y}_i(n) \quad (17)$$

که در آن  $\hat{N}_i(n)$  و  $\bar{y}_i(n)$  به ترتیب نویز تخمینی و سیگنال نویزی در باند  $i$  ام هستند. این روش در واقع اعمال فیلتر وینر در حوزه زمان به سیگنال نویزی می‌باشد. ذکر این نکته لازم است که از طریق فیلترسازی وینر نیز می‌توان به سیگنال بهبود یافته در حوزه زمان دست یافت؛ به این صورت که با توجه به فرض جمع‌شوندگی سیگنال صحبت و نویز در حوزه زمان، می‌توان به سادگی صحبت نویزی را از نویز در حوزه زمان تفریق نمود. لیکن در این حالت با توجه به وجود خطا در تخمین واریانس نویز و همچنین نگاشت غیرخطی ناشی از صفر کردن تفاضل واریانس نویز از واریانس صحبت نویزی، زمانی که حاصل آن منفی می‌شود (۱۵)، مقادیر قابل توجهی نویز مانده، در سیگنال صحبت بازمی‌ماندند. در حالی که در روش تفریق طیفی به کمک پردازش در حوزه فرکانس تا حد زیادی به این مشکل غلبه می‌شود (استفاده از فاکتور فوق تفریق و تعیین کف طیفی).

ذکر این نکته خالی از فایده نیست که در صورتی که بتوان یک آشکارساز دقیق صحبت/سکوت را با الگوریتم تخمین نویز پیشنهاد شده، تلفیق نمود خطای تخمین در بازه‌های سکوت کاهش می‌یابد؛ بنابراین به دقت کلی تخمین‌گر (به قیمت پیچیدگی بیشتر) افزوده خواهد شد.

### ۳-۲-۱ تنظیم پارامترهای تخمین‌گر

تخمین بهینه نویز غیرایستاد بطور کلی نیاز به اطلاعاتی پیشینه راجع به نوع نویز و میزان تغییرات پارامترهای آن در حوزه زمان (نسبت به تغییرات پارامترهای صحبت) دارد. با توجه به اینکه تخمین‌گر مورد

قطعه‌ای نیز این مسئله را تأیید می‌کند. در زیر-بخش بعد یک روش جدید تخمین وقتی آمارگان نویز پیشنهاد می‌شود و سپس از آن در تنظیم پارامترهای الگوریتم MMBSS و یک فاکتور فوق تفریق مناسب استفاده می‌شود.

### ۳-۲ تخمین نویز

همانطور که اشاره شد روش تخمین نویز پیشنهادی در تفریق طیفی چندباندی (MBSS و IMBSS) نیاز به آشکارساز دقیق صحبت/سکوت دارد که پیچیدگی زیادی دارد. بنابراین برای تخمین نویز ما روشی که در [۲۰] برای رفع این عیب پیشنهاد شده است را مورد استفاده قرار داده و پس از آن ضرایب فاکتور فوق تفریق را به نحوی تنظیم می‌کنیم که میزان بهبود صحبت بر اساس معیارهای در نظر گرفته شده به حداکثر برسد. به منظور تخمین نویز، ضرایب بهره‌ای برای هر باند سیگنال صحبت در حوزه زمان بدست می‌آید که برای تخمین نویز در آن باند مورد استفاده قرار می‌گیرد.

برای محاسبه ضرایب بهره سیگنال نیاز به داشتن واریانس نویز و واریانس سیگنال خالص است. از طرفی تنها سیگنال نویزی و واریانس آن در دسترس است. بنابراین لازم است که واریانس سیگنال خالص و نویز به‌طریقی از آن استخراج گردد. با فرض اینکه  $\sigma_{s_i}^2$  واریانس سیگنال خالص،  $\sigma_{w_i}^2$  واریانس نویز و  $\sigma_{x_i}^2$  واریانس سیگنال نویزی هستند و در نظر گرفتن ناهمبستگی بین سیگنال و نویز، داریم [۴]

$$\sigma_{x_i}^2 = \sigma_{s_i}^2 + \sigma_{w_i}^2 \quad (10)$$

معمولاً  $\sigma_{w_i}^2$  ناشناخته بوده و با زمان تغییر می‌کند. در اینجا فرض بر اینست که صحبت و نویز سیگنالهایی غیرایستاد بوده، ولی تغییرات توان نویز نسبتاً آرام صورت می‌گیرد. واریانس سیگنال نویزی از رابطه زیر بدست می‌آید

$$\hat{\sigma}_{x_i}^2(p) = \frac{1}{N} \sum_{n=p-N}^{p-1} x_i^2(pN+n) \quad (11)$$

واریانس تخمینی نویز نیز از رابطه زیر بدست می‌آید [۴] و [۲۰]

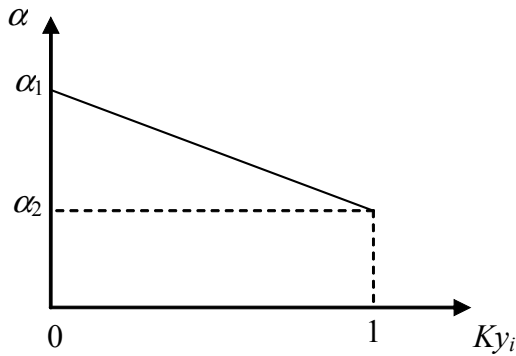
$$\hat{\sigma}_{w_i}^2(p) = \alpha_i(p) \hat{\sigma}_{w_i}^2(p-1) + (1 - \alpha_i(p)) \hat{\sigma}_{x_i}^2(p) \quad (12)$$

که در آن پارامتر هموارسازی  $\alpha_i(p)$  در قاب  $p$  از رابطه وقتی زیر بدست می‌آید [۲۰]

$$\alpha_i(p) = 1 - \min\left\{1, \left(\frac{\hat{\sigma}_{x_i}^2(p)}{\hat{\sigma}_{w_i}^2(p-1)}\right)^{-Q}\right\} \quad (13)$$

در این رابطه  $Q$  عددی صحیح بوده و  $\hat{\sigma}_{w_i}^2(p-1)$  میانگین یا میانه تخمینهای نویز در ۵ تا ۱۰ قاب قبلی می‌باشد. نسبت  $\hat{\sigma}_{x_i}^2(p) / \hat{\sigma}_{w_i}^2(p-1)$  تقریبی از SNR پسین، یعنی  $\sigma_{s_i}^2 / \sigma_{w_i}^2 = (\sigma_{s_i}^2 + \sigma_{w_i}^2) / \sigma_{w_i}^2$  می‌باشد. عملگر min در (۱۳) برای اجتناب از مقادیر منفی  $\alpha_i(p)$  و بهنگام‌سازی غلط استفاده می‌شود.

حال عمل این فاکتور را در تخمین نویز توضیح می‌دهیم. در ابتدا اگر صحبت در قاب  $p$  موجود نباشد، توان نویز جدید  $\hat{\sigma}_{x_i}^2(p)$  باید بسیار نزدیک به تخمین متوسط نویز قدیم  $\hat{\sigma}_{w_i}^2(p-1)$  باشد. بنابراین طبق (۱۳)  $\alpha_i(p) \approx 0$  است. همچنین از (۱۲) داریم  $\hat{\sigma}_{w_i}^2(p) \approx \hat{\sigma}_{x_i}^2(p)$  چرا که مقدار  $\alpha_i(p)$  کوچک است. بنابراین تخمین توان نویز در قاب  $p$  بسرعت توان سیگنال نویزی را در غیاب صحبت دنبال می‌کند. از طرف

شکل ۷: رابطه  $\alpha$  با  $Ky_i$ .

### ۴-۳- تنظیم پارامترهای مرحله پیش‌پردازش و فاکتور فوق تفریق

در این قسمت نحوه تنظیم پارامترهای  $\gamma_1$  و  $\gamma_2$  مربوط به مرحله پیش‌پردازش و  $\alpha_1$  و  $\alpha_2$  مربوط به فاکتور فوق تفریق در الگوریتم اصلاح شده تفریق طیفی چندباندی توصیف می‌شود. دقت شود که هدف اصلی، رسیدن به حداقل نویز مانده در تفریق طیفی چندباندی است؛ با حداقل اعوجاج در سیگنال صحبت. برای تنظیم این پارامترها بصورت زیر عمل می‌کنیم:

الف- ابتدا بدون مرحله پیش‌پردازش پارامترهای فاکتور فوق تفریق ( $\alpha_1$  و  $\alpha_2$ ) طوری تنظیم می‌شوند که حتی‌الامکان بدون افزایش قابل توجه نویز مانده، مؤلفه‌های ضعیف صحبت در اسپکتروگرام باقی بمانند. این کار را در SNR ورودی ۵ دسی‌بل که در آزمون‌ها متداولترین SNR ورودی است، انجام می‌دهیم.

ب- سپس میانگین‌گیری وزن‌دار با وزنهای ثابت اعمال می‌شود. وزنهای در این حالت طوری انتخاب می‌شوند که حتی‌الامکان نویز موزیکال باقی نمانده باشد.

ج- در نهایت رابطه‌ای خطی بین مقادیر وزن‌ها و میزان SNR هر قاب، شبیه روش [۲۲] در نظر می‌گیریم. البته فرق کوچکی بین رابطه مذکور و رابطه پیشنهادی در این بخش است و آن اینکه SNR مورد استفاده در [۲۲] SNR قطعه‌ای است؛ ولی SNR این قسمت، SNR پسین تقریبی  $\hat{\sigma}_{y_i}^2(p) / \hat{\sigma}_{y_i}^2(p-1)$  (بخش ۳-۱) می‌باشد. نتایج تجربی حاصل شده به قرار زیر است:

$$\alpha_1 = 8, \alpha_2 = 1, \gamma_1 = -0.4 \text{ dB}, \gamma_2 = 20 \text{ dB}, m = 0.75$$

$$W = \left(\frac{1}{S^2}\right) * [0.06 \quad 0.15 \quad 0.15 \quad 0.42 \quad 0.15 \quad 0.06]$$

### ۴- پیاده‌سازی الگوریتم و نتایج آن

شبیه‌سازی‌ها با استفاده از ۴۰ نمونه صحبت زن و مرد و دو نوع نویز سفید ایستادن (با میانگین صفر و واریانس ثابت) و غیرایستادن (با واریانس متغیر که ابتدا زیاد شده و سپس شروع به کاهش می‌کند) انجام شده است. سیگنال صحبت با کمک پنجره همینگ به قاب‌هایی تقسیم‌بندی گردید. طول هر پنجره ۲۵۶ نمونه، معادل ۲۳ میلی‌ثانیه و ۵۰٪ همپوشانی در نظر گرفته شد و پس از تبدیل فوریه گسسته، سیگنال به ۴ باند تجزیه گردید. سپس تخمینی از نویز با استفاده از تخمین‌گر بخش ۳-۲ (۱۷) بدست آمده و بدین ترتیب با کمک (۱۱) و (۱۲) میزان SNR پسین

استفاده، از نوع چندباندی است؛ بنابراین نوع نویز (سفید یا رنگی بودن آن) محدودیت مهمی برای تخمین‌گر بوجود نمی‌آورد؛ چرا که در هر زیرباند تغییرات فرکانسی نویز رنگی کمتر بوده و می‌توان آنرا با تقریب معادل با نویز سفید در آن باند در نظر گرفت. لیکن میزان تغییرات مشخصات آماری نویز در حوزه زمان تأثیر مستقیمی بر عملکرد تخمین‌گر دارد. هر چه میزان تغییرات واریانس نویز در حوزه زمان بیشتر باشد، لازم است بهنگام‌سازی تخمین نویز سریعتر صورت گیرد تا بتوان تغییرات نویز را بخوبی دنبال کرد. بنابراین مقدار  $Q$  در (۱۳) باید عدد کوچکی باشد. بالعکس چنانچه مشخصات آماری نویز با زمان تغییر نکند و یا تغییرات اندکی داشته باشد (که در محیط واقعی معمولاً اینگونه نیست)، مقدار بهینه برای  $Q$  عددی بزرگ خواهد بود. بنابراین  $Q$  بهینه باید با توجه به حداکثر تغییرات زمانی پارامترهای نویز تعیین شود تا تخمین‌گر بتواند در دنبال کردن نویزهای شدیداً غیرایستادن نیز عملکرد قابل قبولی داشته باشد. از طرفی حداکثر تغییرات واریانس نویز نباید بیشتر از تغییرات واریانس سیگنال صحبت باشد. بر این اساس آزمایشهای متعدد در SNR های مختلف نشان داد مقدار مناسب  $Q$  (بدون هموارسازی نهایی) حدود ۲ است و چون  $Q$  بمنظور کاهش حجم محاسبات، عددی صحیح فرض شده است [۱۷]، آنرا برابر ۲ در نظر گرفته‌ایم.

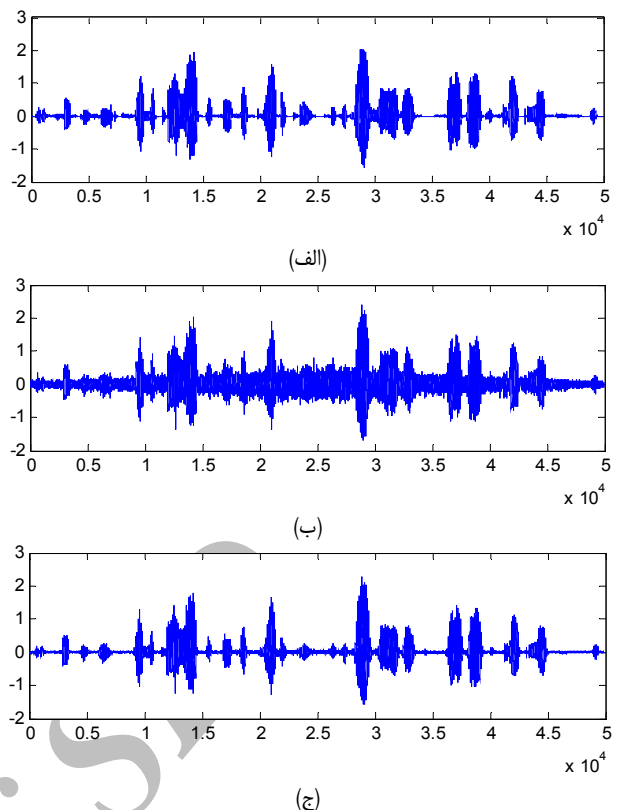
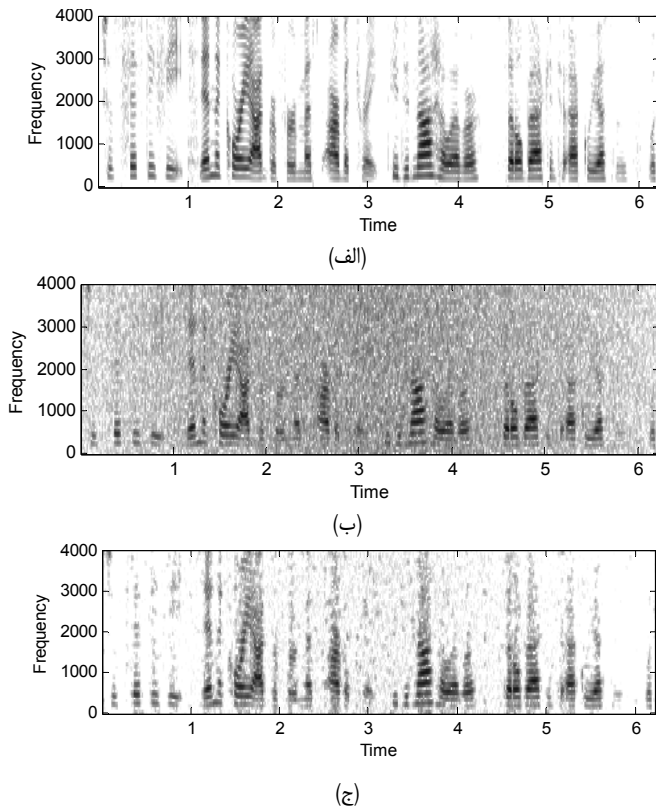
سپس تأثیر مقدار پارامتر هموارسازی نهایی  $\alpha_{final}$  در (۱۴) بر نرم خطای تخمین مورد بررسی قرار گرفت. همانطور که ذکر شد، مقدار  $\alpha_{final}$  پیشنهادی در [۱۷]، حدود ۰.۵ تا ۰.۸ می‌باشد. تأثیر  $\alpha_{final}$  در تخمین نویز بدین صورت است که هر چه مقدار این پارامتر بزرگتر در نظر گرفته شود خطای روی قاب‌های صحبت کمتر شده و در نتیجه اعوجاج صحبت کمتر می‌شود؛ اما این اقدام سبب افزایش خطای تخمین شده و بنابراین باید مصالحه‌ای بین میزان اعوجاج و مقدار نویز موزیکال برقرار شود. بر این اساس مقدار ۰.۵۵ برای  $\alpha_{final}$  انتخاب گردید.

### ۳-۳- ارائه فاکتور فوق تفریق وابسته به SNR پسین

ایده‌ای که در این قسمت پیشنهاد شده است، استفاده از ضریب  $Ky_i$  بدست آمده در (۱۶) برای تعیین فاکتور فوق تفریق می‌باشد. لازم به ذکر است که یافتن فاکتور فوق تفریق بهینه برای نویزهای غیرایستادن دشوارتر از نویزهای ایستادن است زیرا نویزهای غیرایستادن را معمولاً نمی‌توان به دقت تخمین زد و علاوه بر این به طور کلی خطای تخمین در قسمتهایی که نویز در حال افزایش است، ممکن است با خطای تخمین در قسمتهایی که نویز در حال کاهش است متفاوت باشد. در تفریق طیفی چندباندی فاکتور فوق تفریق، وابسته به SNR قطعه‌ای است و در اینجا فاکتور فوق تفریق را تابعی از SNR پسین بدست آورده‌ایم تا برای نویزهای غیرایستادن نیز مناسب باشد. به این منظور رابطه  $\alpha$  (فاکتور فوق تفریق) را با  $Ky_i$  به صورت رابطه خطی زیر

$$\alpha = (\alpha_2 - \alpha_1)Ky_i + \alpha_1 \quad (18)$$

که متناظر با شکل ۷ است، در نظر می‌گیریم. سپس این مقدار  $\alpha$  را در رابطه متعارف تفریق طیفی توان استفاده کردیم. نویز مورد استفاده مانند نویز قسمت دوم غیر ایستادن بوده و تخمین آن مانند همان قسمت صورت می‌گیرد. نتایج نشان می‌دهد که تغییر مقدار  $\alpha$  بدون اینکه روی اعوجاج صحبت اثر قابل ملاحظه‌ای داشته باشد، تأثیر زیادی روی افزایش و کاهش نویز مانده دارد. مقادیر بهینه  $\alpha_1$  و  $\alpha_2$  بسته به محدوده SNR ورودی دارند. نحوه تنظیم این پارامترها در بخش ۳-۴ آمده است.



شکل ۹: نمودار اسپکتروگرام سیگنال‌ها؛ (الف) سیگنال خالص، (ب) سیگنال نویزی و (ج) سیگنال حذف نویز شده.

شکل ۸: (الف) سیگنال خالص، (ب) سیگنال نویزی و (ج) سیگنال حذف نویز.

جدول ۳: نتایج ارزیابی محاسباتی الگوریتم

	Stationary Noise			Non-Stationary Noise		
	۵	۱۰	۱۰	۵	۱۰	۱۰
SNRin (dB)	-	۵	۱۰	-	۵	۱۰
Segmental SNRin (dB)	-۸,۳۴	-۳,۲۵	۱,۷۱	-۹,۷۴	-۱,۹۴	۳,۱۱
SNRo (dB)	۹,۳۴	۱۳,۰۶	۱۵,۲۵	۹,۳۲	۱۳,۱۸	۱۵,۴۴
Segmental SNRo (dB)	۴,۱۲	۷,۲۵	۹,۳۲	۴,۸۷	۷,۴۹	۹,۶۶
IS measure - in	۲۹,۳۶	۱۵,۳۹	۸,۴۳	۲۸,۹۰	۱۵,۳۳	۸,۳۲
IS measure - out	۴,۸۷	۴,۱۲	۳,۳۷	۴,۶۵	۳,۶۹	۳,۱۳

جدول ۴: نحوه ارزیابی ادراکی

سطح اغتشاش	کیفیت صحبت	امتیاز
غیر قابل درک	عالی	۵
قابل درک، بدون مزاحمت	خوب	۴
کمی آزاردهنده	متوسط	۳
آزاردهنده	ضعیف	۲
غیر قابل تحمل	بد	۱

برای انجام آزمایش سه نمونه سیگنال صحبت با طول ۶ ثانیه با نویزهای مختلف در SNR های مختلف آغشته شده و بوسیله ۴ شنونده امتیازدهی شد. جدول ۵ میانگین نتایج ارزیابی این سه نمونه صحبت را نشان می‌دهد.

$[\hat{\sigma}_{x_i}^2(p) / \bar{\sigma}_{w_i}^2(p-1)]$  محاسبه شد. با استفاده از این رابطه ما ضرایب بردار  $W$  را بدست آورده و در (۴) مورد استفاده قرار دادیم. سپس تفریق طیفی توان با فاکتور فوق تفریق (۱۸) بر اساس SNR پسین جدید، که از طیف میانگین‌گیری شده بدست آمده است، برای هر باند انجام شد. جدول ۳ نتایج حاصل شده را نشان می‌دهد. این نتایج از میانگین‌گیری از نتایج آزمایشهای ۴۰ نمونه صحبت مذکور بدست آمده است.

همانطور که از جدول ملاحظه می‌شود در الگوریتم پیشنهادی، SNR سراسری هم برای سیگنال آغشته به نویز ایستان و هم آغشته به نویز غیرایستان بیش از ۵ دسی‌بل بهبود پیدا کرده است و میزان SNR قطعه‌ای نیز بیش از ۶ دسی‌بل بهبود یافته است. شکل ۸ یک نمونه سیگنال صحبت خالص با طول ۶ ثانیه و سیگنال آغشته به نویز غیرایستان با SNR=۵ dB و سیگنال پردازش شده را نشان می‌دهد. در شکل ۹ نیز اسپکتروگرام این سیگنال‌ها آمده است. با توجه به شکلها ملاحظه می‌شود که الگوریتم پیشنهادی بخوبی نویز را کاهش می‌دهد و میزان نویز مانده نیز بسیار کم است.

برای ارزیابی ادراکی<sup>۱</sup> الگوریتم از آزمون MOS<sup>۲</sup> استفاده شد [۲۳]. در

1. Subjective  
2. Mean Opinion Score



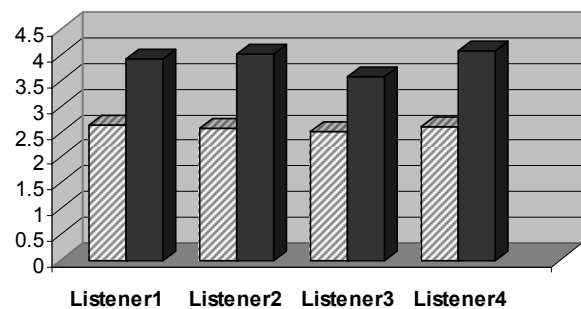
جدول ۵: نتایج ارزیابی ادراکی الگوریتم توسط ۴ شنونده

SNR	White-nonstationary noise		White-stationary noise		Babble noise		Factory noise		
	Noisy	Proposed	Noisy	Proposed	Noisy	Proposed	Noisy	Proposed	
شنونده اول	۰ dB	۱	۳٫۷	۳	۳٫۷	۱	۳	۴	۴٫۷
	۵ dB	۱٫۳	۴	۳٫۳	۴	۱٫۳	۳	۴	۵
	۱۰ dB	۳	۴٫۳	۴	۴٫۳	۱٫۳	۴٫۳	۴٫۳	۵
شنونده دوم	۰ dB	۱	۴	۲٫۳	۲٫۷	۱	۳	۲٫۳	۴
	۵ dB	۳٫۷	۴٫۳	۳٫۳	۳٫۷	۱٫۷	۳٫۳	۲٫۷	۵
	۱۰ dB	۳	۴٫۷	۴٫۳	۴٫۷	۲٫۳	۳٫۷	۳	۵
شنونده سوم	۰ dB	۱٫۳	۳٫۳	۲	۳٫۳	۱	۲٫۳	۳٫۷	۴
	۵ dB	۲	۳٫۳	۲٫۳	۳٫۷	۱٫۳	۲٫۷	۴	۴٫۷
	۱۰ dB	۲٫۷	۳٫۷	۳٫۷	۴	۱٫۷	۳	۴٫۳	۵
شنونده چهارم	۰ dB	۱	۳٫۳	۳	۴	۱٫۳	۳٫۳	۳	۴٫۷
	۵ dB	۱٫۳	۳٫۳	۳٫۷	۴٫۳	۲٫۳	۳٫۷	۳٫۳	۵
	۱۰ dB	۱٫۷	۳٫۷	۴٫۳	۴٫۷	۲٫۳	۴	۴	۵

## مراجع

- [1] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, John Wiley & Sons, Ltd., 2000.
- [2] H. Gustafsson, S. E. Nordholm, and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 8, pp. 799-807, Nov. 2001.
- [3] K. Yamashita, S. Ogata and T. Shimamura, "Spectral subtraction iterated with weighting factors," in *Proc. IEEE Speech Coding Workshop*, pp. 138-140, Oct. 2002.
- [4] L. Lin, E. Ambikairajah, and H. Holmes, "Speech enhancement for nonstationary noise environment," in *Proc. Asia-Pacific Conf. on Circuits and Systems*, vol. 1, pp. 177-180, Oct. 2002.
- [5] D. Almagor, *Enhancement of Noisy Speech Signals by Low-Cost, Adaptive Methods*, National Semiconductor Corporation, DSP Technology Group, pp. 25-28, 1986.
- [6] B. L. Sim, Y. C. Tong, J. S. Chang, and C. T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 4, pp. 328-337, Jul. 1998.
- [7] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, pp. 113-120, Apr. 1979.
- [8] J. Poruba, "Speech enhancement based on nonlinear spectral subtraction," in *Proc. Fourth IEEE Int. Caracas Conf. on Devices, Circuits, and Systems*, pp. T031-1-T031-4, 2002.
- [9] S. F. Boll, "A spectral subtraction algorithm for suppression of acoustic noise in speech," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 200-203, Apr. 1979.
- [10] A. Lagler, G. Gruhler, P. Sovka, V. Davidek, and P. Pollak, *Real-time Fixed-point DSP-implementation of Spectral Subtraction Algorithm for Speech Enhancement in Noisy Environment*, Fachhochschule Heilbronn, University for Applied Sciences, D-74081 Heilbronn, Germany, Czech Technical University Prague, Fac. of Electrical Engineering, Dpt. Of Circuit Theory Technicka2, 16627, Praha 6, Czech Republic, 1996.
- [11] P. Lockwood and J. Boudy, "Experiments with a non-linear spectral subtractor (NSS), hidden Markov models and the projections, for robust recognition in cars," *Speech Comm.*, vol. 11, no. 2-3, pp. 215-228, June 1992.
- [12] C. He and G. Zweig, "Adaptive two-band spectral subtraction with multi-window spectral estimation," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 793-796, Mar. 1999.
- [13] Md. Kamrul Hasan, S. Salahuddin, and M. Rezwana Khan, "A modified a priori SNR for speech enhancement using spectral subtraction rules," *IEEE Signal Processing Letters*, vol. 11, no. 4, pp. 450-453, Apr. 2004.
- [14] K. Wu and P. Chen, "Efficient speech enhancement using spectral subtraction for car hands-free application," in *Proc. Int. Conf. on Consumer Electronics*, vol. 2, pp. 220-221, 2001.
- [15] S. D. Kamath, *A Multi-Band Spectral Subtraction Method for Speech Enhancement*, M.Sc Thesis in Electrical Eng., University of Texas at Dallas, Dec. 2001.

## MOS Measure



شکل ۱۰: مقایسه ادراکی کیفیت سیگنال صحبت نویزی و سیگنال صحبت حذف نویز شده، ستونهای هاشور خورده: صحبت نویزی، ستونهای پررنگ: صحبت حذف نویز شده.

## ۵- نتیجه گیری

در این مقاله چند ایده برای بهبود کارایی روش تفریق طیفی پیشنهاد شد و پس از پیاده‌سازی، میزان بهبود حاصل شده، مقایسه و ارزیابی گردید. ابتدا تفریق طیفی چندباندی MBSS و تفریق طیفی چندباندی بهبود یافته IMBSS پیاده‌سازی شده و سپس به مرحله پیش‌پردازش آن شرطی اضافه شد که وزنهای مربوط به میانگین‌گیری، متناسب با SNR قطعه‌ای تخمینی تنظیم شوند. این عمل باعث شد در عین حال که نویز موزیکال تا حد امکان توسط مرحله پیش‌پردازش کم می‌شود، اعوجاج صحبت به حداقل برسد. نتایج تستهای محاسباتی و ادراکی نیز این موضوع را تأیید می‌کند. در مرحله دوم یک تخمین‌گر وقتی نویز پیاده‌سازی شد و پارامترهای آن طوری تنظیم گردید که بتواند به شکل مؤثرتری نویزهای غیرایستاد را دنبال کند. سپس این تخمین‌گر در تفریق طیفی مورد استفاده قرار گرفت. در این حالت، روش تفریق طیفی قادر به کاهش نویزهای غیرایستاد نیز می‌گردد. در مرحله سوم با کمک SNR پسین که از مرحله دوم بدست آورده بودیم، یک فاکتور فوق‌تفریق پیشنهاد شده و جایگزین فاکتور فوق‌تفریق MBSS گردید و در نهایت پارامترهای میانگین‌گیری وقتی و فاکتور فوق‌تفریق به‌گونه‌ای تنظیم شد که بتواند نویزهای غیرایستاد را با حداقل اعوجاج و نویز مانده حذف کند. نتایج حاصل شده نشان‌دهنده کارایی مؤثر الگوریتم پیشنهاد شده برای حذف نویز می‌باشد.



**مهدی بکرانی** تحصیلات خود را در مقطع کارشناسی مهندسی برق - الکترونیک از دانشگاه فردوسی مشهد و در مقطع کارشناسی ارشد مهندسی برق - الکترونیک از دانشگاه تربیت مدرس بترتیب در سالهای ۱۳۸۱ و ۱۳۸۳ به پایان رسانده است و هم اکنون دانشجوی دکترای مهندسی برق - الکترونیک دانشگاه تربیت مدرس می‌باشد. زمینه‌های تحقیقاتی مورد علاقه وی در حال حاضر، پردازش سیگنال‌های صوتی، تصویری، ویدیویی و پزشکی زیستی است.

**هادی صدوقی یزدی** تحصیلات کارشناسی الکترونیک خود را در سال ۱۳۷۳ در دانشگاه فردوسی مشهد و کارشناسی ارشد الکترونیک را در ۱۳۷۵ در دانشگاه تربیت مدرس به پایان رساند. او دکترای مهندسی الکترونیک را در زمینه پردازش صحنه‌های ترافیک در دانشگاه تربیت مدرس در سال ۱۳۸۴ اخذ کرد و هم اکنون استادیار دانشکده فنی و مهندسی دانشگاه تربیت معلم سبزوار می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان فیلترهای وقتی، پردازش تصاویر ویدیویی و شناسایی الگو است.

**مجتبی لطفی زاد** درجه کارشناسی مهندسی برق را در ۱۳۵۹ از دانشگاه صنعتی امیرکبیر دریافت کرد و کارشناسی ارشد و دکترای خود را بترتیب در ۱۳۶۴ و ۱۳۶۷ از دانشگاه ولز انگلستان دریافت نمود. هم اکنون عضو هیأت علمی بخش مهندسی برق دانشگاه تربیت مدرس است. زمینه‌های علمی مورد علاقه او پردازش سیگنال، فیلترهای وقتی و پردازش صوت است.

- [16] S. Kamath, and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 4, p. 4164, Orlando, US, May 2002.
- [17] Y. Ghanbari, M. R. Karami-Mollaei, and B. Amelifard, "Improved multi-band spectral subtraction method for speech enhancement," in *Proc. The Sixth IASTED Int. Conf. on Signal and Image Processing*, pp. 225-250, Honolulu, US, Aug. 2004.
- [18] <http://www.speech.cs.cmu.edu/comp.speech/section1/data/noisex.html>
- [19] J. R. Deller, Jr., J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*, IEEE Press, Piscataway, NJ, 2000.
- [20] L. Lin, W. H. Holmes, and E. Ambikairajah, "Subband noise estimation for speech enhancement using a perceptual wiener filter," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 80-83, Apr. 2003.
- [21] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, Jul. 2001.
- [22] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 208-211, Apr. 1979.
- [23] — "IEEE recommended practice for speech quality measurement," *IEEE Trans. on Audio and Electroacoustics*, vol. 17, no. 3, pp. 227-246, Sept. 1969.

Archive of SID