

# بهسازی گفتار با استفاده سیستم ترکیبی جداسازی کورمنابع و فیلتر کالمن

مجید پوراحمدی

حمید سلطانیان زاده

دانشجوی دکتری دانشکده برق

استاد گروه مهندسی برق و کامپیوتر

دانشگاه یزد

دانشگاه تهران

Email: [hszade@ut.ac.ir](mailto:hszade@ut.ac.ir), [m\\_pourahmadi@stu.yazduni.ac.ir](mailto:m_pourahmadi@stu.yazduni.ac.ir)

چکیده - در این مقاله، یک روش جدید بهسازی گفتار در محیط های شامل دو منبع صوتی و همراه با منابع نویز زمینه را معرفی خواهیم کرد. روش ما شامل ترکیب یک واحد جداسازی کور منابع (BSS) به همراه یک ساختار فیلتر کالمن بهینه می باشد. واحد اول منابع نقطه ای مثل سیگنالهای صوتی دو گوینده نزدیک به هم را جدا کرده درحالیکه واحد دوم نویز زمینه را از این سیگنالها حذف خواهد کرد. شبیه سازیهای مختلف، کارکرد مناسب این سیستم را در مواجهه با سیگنالهای گفتار آغشته به نویز سفید نشان می دهد.

کلیدواژه- بهسازی گفتار، جداسازی کور منابع، گفتارنشتی، نویز زمینه.

## ۱- مقدمه

در روش فیلتر کالمن، عموماً سیگنال گفتار به صورت یک فرایند AR شناخته می شود و عملیات حذف نویز در دو مرحله صورت می گیرد: در مرحله اول واریانس نویز و پارامترهای گفتار تخمین زده شده و در مرحله دوم با استفاده از فیلتر کالمن سیگنال گفتار تخمین زده خواهد شد. الگوریتم ارائه شده در این مقاله از نوعی فیلتر کالمن بهینه استفاده می کند که در آن نیازی به VAD نیست.

البته عملکرد فیلتر کالمن و سایر روشهای تک کاناله در مواقعی که انعکاس سیگنال در محیط نسبتاً زیاد باشد تا حد قابل توجهی کاهش خواهد یافت. در چنین مواردی استفاده از روش جداسازی کور منابع یا BSS (Blind Source Separation) به منظور تبدیل منابع مخلوط شده به مؤلفه های سازنده می تواند به عنوان جایگزینی برای این روشها مطرح باشد. در روشهای متداول BSS هیچ اطلاع

از آنجایی که عملکرد مناسب بسیاری از سیستم های پردازش گفتار مستلزم کارکرد آنها در شرایط نویزی است، از این رو نیاز به یک واحد حذف نویز در آنها احساس می شود. در میان روش های مختلف تک کاناله حذف نویز، روش فیلتر کالمن (Kalman Filtering) یکی از معروفترین آنها می باشد. در حالت کلی فیلتر کالمن با استفاده از یک واحد VAD (Voice Activity Detector) با شناسایی بخش های سکوت گفتار، نویز موجود در این قسمتها را استخراج نموده و به عنوان نویز کلی سیستم می شناسند. مسلماً سیستم در صورتی به طور مناسب عمل خواهد نمود که نویز مخلوط شده با گفتار، نویزی ایستاد باشد. به طور قطع در این حالت فرض ایستاد بودن نویز، فرضی محدود کننده خواهد بود.

(Multiple Adaptive Decorrelation) [۱] استفاده شده است. این الگوریتم برای جداسازی منابع غیر ایستان کانالو شده بصورت فریم به فریم مورد استفاده قرار می گیرد. ما این الگوریتم را به علت سرعت بالا و نتایج مطلوب آن در مورد سیگنالهای گفتار انتخاب کرده ایم. در روش MAD با  $m$  سیگنال  $x(t)$  حاصل از  $m$  میکروفون که شامل سیگنالهای  $m$  منبع نقطه ای  $s(t)$  و نویز زمینه جمع شونده  $n(t)$  می باشد داریم:

$$x(t) = \sum_{\tau=0}^p A(\tau) s(t-\tau) + n(t) \quad (1)$$

که  $P$  درجه کانولوشن و  $A(t)$  یک ماتریس ترکیب کننده  $m \times m$  می باشد. با این فرض که انعکاسهای محیط خیلی زیاد نباشد، منابع نقطه ای با دامنه و تأخیرهای مختلف به میکروفون ها می رسند. وظیفه سیستم جداکننده پیدا کردن این تأخیرها و اعمال آن روی سیگنالهای مخلوط به منظور جداسازی آنها است. اما نویز زمینه که از تعداد زیادی منبع ناشی می شود تقریباً به طور یکسان و بدون تأخیر به همه میکروفون ها می رسد [۲]. در اینصورت سیستم BSS قادر به جداسازی آن از سیگنالهای گفتار نبوده و در نتیجه نویز زمینه را در همه سیگنالهای جدا شده خواهیم داشت. وظیفه فیلتر کالمن حذف این نویز بوده که در بخشهای بعدی به طور مفصل توضیح داده خواهد شد.

## ۲-۱- روش MAD

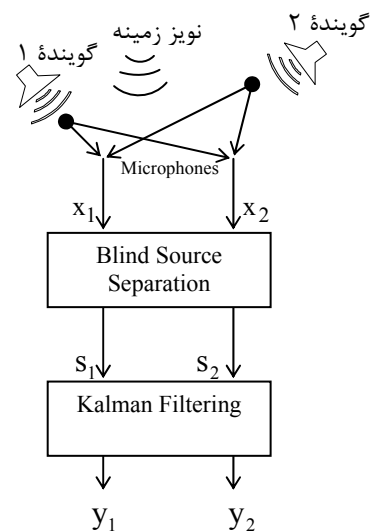
در سالهای اخیر روشهای مختلفی برای جداسازی کورمنابع با استفاده از همبستگی زدایی آمارگانههای درجه دو یا آمارگانههای با درجه بالاتر، معرفی شده اند [۲]. روش MAD که ما از آن استفاده کرده ایم به منظور همبستگی زدایی آمارگانههای درجه ۲ سیگنالهای غیر ایستان مورد استفاده قرار می گیرد. در این الگوریتم هدف یافتن یک دنباله از ماتریس های جداکننده  $W(t)$  از روی سیگنالهای ترکیب شده  $x(t)$  معادله (۱) بگونه ایست که:

$$\hat{s}(t) = \sum_{\tau=0}^Q W(\tau) x(t-\tau) \quad (2)$$

که در آن  $Q$  طول فیلتر جداکننده و  $\hat{s}(t) = [\hat{s}_1(t), \dots, \hat{s}_m(t)]^T$  بردار تخمین منابع در لحظه  $t$

قبلی در مورد منابع یا کانال ترکیب کننده در اختیار نیست اما فرض براین است که منابع نقطه ای بوده و هیچ نویز زمینه ای هم در محیط موجود نباشد. با این وجود در محیط های واقعی، منابع نویز زیادی موجود می باشد که در ترکیب با هم نویز زمینه پیوسته ای ایجاد کرده که با سیگنال منابع نقطه ای ترکیب می شود. در چنین شرایطی BSS قادر به حذف اینگونه نویزها نخواهد بود. در [۱] نشان داده شده که با افزایش تعداد میکروفون ها، از لحاظ تئوری روش BSS می تواند از نویز زمینه جلوگیری نماید؛ اما نتایج تجربی نشان می دهد که در چنین شرایطی هم سیستم عملکرد قابل توجهی نخواهد داشت. روش پیشنهادی ما که شامل ترکیب BSS با یک بلوک حذف نویز زمینه می باشد در شکل ۱ نشان داده شده است. در این بلوک از یک ساختار فیلتر کالمن اصلاح شده که در آن نیازی به VAD نمی باشد، استفاده کرده ایم.

فرض بر این است که محیط شامل دو منبع نقطه ای گفتار همراه با یک نویز زمینه (مثلاً نویز همهمه) باشد. اثر این نویز زمینه را در همه میکروفونونها یکسان در نظر می گیریم. در حالی که BSS منابع نقطه ای را از هم جدا می کند فیلتر کالمن نویز زمینه را از خروجیهای جداسازی شده حذف خواهد نمود.



سیگنالهای بهسازی شده خروجی  
شکل ۱: سیستم پیشنهادی بهسازی گفتار

## ۲- جداسازی کورمنابع

در این تحقیق برای جداسازی منابع از روش MAD

می باشد.

مؤلفه هایی از منابع دیگر را به همراه خواهد داشت.

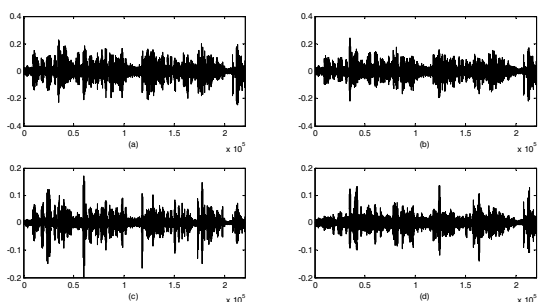
شرط اول معادله (۴) تحمیل می کند که طول فیلتر جداکننده Q خیلی کوچکتر از T انتخاب شود تا مؤلفه های زمانی این فیلتر هموارتر شده و مسأله جایگشت فرکانسی مرتفع گردد [۱]. البته باید توجه داشت که این شرط همیشه نمی تواند مانع از جایگشت شود [۳]. شرط دوم معادله (۴) نیز برای جلوگیری از تغییر مقیاس سیگنالهای بازسازی شده نسبت به منابع اصلی اعمال می شود.

در این صورت قانون یادگیری نهایی الگوریتم MAD به صورت زیر نوشته می شود:

$$\Delta W^*(\omega) \sim \sum_k (W \hat{R}_x(\omega, t_k) W^H - \hat{\Lambda}_s(\omega, t_k)) W(\omega) \hat{R}_x(\omega, t_k) \quad (5)$$

آزمایشهای مختلف عملکرد بسیار مؤثر این الگوریتم در شرایط مختلف مثلاً در یک مهمانی را نشان می دهد. شکل ۲ سیگنالهای ورودی و خروجی حاصل از اجرای الگوریتم MAD در یک محیط با انعکاس نسبتاً زیاد را نشان می دهد. سیگنالهای ورودی شامل گفتار دو گوینده بوده که همراه با نویز زمینه سفید ضبط شده است.

مشاهده می شود که در سیگنالهای بازسازی شده نویز زمینه اصلی هنوز هم وجود دارد. علاوه بر این به علت انعکاس های بالای سیگنالهای گفتار در محیط، هر سیگنال خروجی هنوز هم شامل مؤلفه هایی از سیگنال دیگر می باشد که الگوریتم BSS نتوانسته آنها را جدا کند.



شکل ۲: جداسازی کورمنابع نقطه ای. (a) و (b): سیگنالهای نویزی ورودی. (c) و (d): سیگنالهای جداسازی شده خروجی در ادامه به توصیف روشی برای حذف نویز زمینه می پردازیم.

از آنجایی که در حوزه فرکانس عمل کانولوشن به ضرب تبدیل می شود، در الگوریتم MAD با پنجره بندی سیگنال با پنجره هایی به طول T، محاسبه فیلتر جداکننده در پنجره kام، در حوزه فرکانس و از روی مشاهدات زیر انجام می گیرد:

$$X(\omega, t) \approx A(\omega)s(\omega, t) + n(\omega, t), \quad \text{for } P \ll T \quad (3)$$

در این معادله، T طول تبدیل فوریه زمان کوتاه T نقطه ای در پنجره kام و با شروع از لحظه  $t_k = kT$  بوده که در آن T خیلی بزرگتر از P؛ طول فیلتر ترکیب کننده؛ انتخاب می شود تا بتوان کانولوشن چرخشی را با کانولوشن خطی جایگزین کرد.

اگر  $\hat{R}_x(\omega, t_k) = E[X(\omega, t_k) X^H(\omega, t_k)]$  نشان دهنده ماتریس همبستگی متقابل سیگنالهای میکروفونها و  $\hat{\Lambda}_s(\omega, t_k) = E[s(\omega, t_k) s^H(\omega, t_k)]$  ماتریس همبستگی متقابل منابع باشد، برای قطری سازی ماتریس همبستگی مشاهدات ناشی از میکروفونها  $W(\omega)$  را می توان با استفاده از معیار زیر به دست آورد [۱]:

$$J = \arg \min_{W, \hat{\Lambda}_s} \sum_k \sum_{\omega=1}^T \left\| W \hat{R}_x(\omega, t_k) W^H - \hat{\Lambda}_s(\omega, t_k) \right\|^2$$

$$W(\tau) = 0 \quad \forall \tau > Q, Q \ll T \quad (4)$$

$$W_{ii}(\omega) = 1$$

در کاربردهایی که فیلتر ترکیب کننده سیگنالها، A، به جای عمل کانولوشن، سیگنالها را بدون تأخیر و فقط با تغییر دامنه با هم مخلوط می کند ممکن است ترتیب سیگنالهای بازسازی شده با سیگنالهای اصلی متفاوت بوده و به اصلاح عمل جداسازی همراه با جایگشت صورت گیرد.

در الگوریتم MAD و با در نظر گرفتن فیلتر کانوالو کننده، برای هر مؤلفه فرکانسی بطور جداگانه عمل جداسازی انجام گرفته و در نهایت مؤلفه های فرکانسی هر منبع به طور جداگانه با هم ترکیب شده و منبع بازسازی شده را در حوزه زمان خواهند ساخت. حال اگر در اجرای الگوریتم در هر فرکانس جایگشتی صورت گیرد، برای هر منبع و در بعضی از فرکانسها مؤلفه هایی از منابع دیگر را خواهیم داشت و در این صورت سیگنال بازسازی شده در حوزه زمان،

### ۳- حذف نویز زمینه

$$x(n+1) = Fx(n) + Gu(n+1) \quad (۸)$$

$$y(n) = Hx(n) + v(n) \quad (۹)$$

که  $x(n) = [s(n-p+1), \dots, s(n)]^T$  بردار حالت  $P \times 1$  بعدی،  $F$  ماتریس انتقال  $P \times P$  به صورت

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ a_p & a_{p-1} & \cdot & \dots & a_1 \end{bmatrix} \quad (۱۰)$$

و  $H = G^T = [0 \ 0 \ \dots \ 0 \ 1]$  به ترتیب بردار سطری مشاهدات و بردار ورودی می باشند.

#### ۴-۱- تخمین پارامترهای گفتار:

به منظور اجرای الگوریتم، ابتدا باید ماتریس  $F$  ساخته شود. برای این کار باید پارامترهای  $AR$  گفتار روی هر فریم به دست آید. معادلات Yule-Walker این کار را انجام میدهند.

در این صورت پارامترهای گفتار به صورت زیر تخمین زده می شود:

$$\begin{bmatrix} \hat{a}_p \\ \cdot \\ \cdot \\ \hat{a}_1 \end{bmatrix} = \begin{bmatrix} r_{yy}(0) & \dots & r_{yy}(p-1) & r_{yy}(p) \\ \cdot & \dots & \cdot & \cdot \\ \cdot & \dots & \cdot & \cdot \\ r_{yy}(p-1) & \dots & r_{yy}(1) & r_{yy}(0) \end{bmatrix}^{-1} \begin{bmatrix} r_{yy}(1) \\ \cdot \\ \cdot \\ r_{yy}(p) \end{bmatrix} \quad (۱۱)$$

که  $r_{yy}(k) = E[y(n)y(n-k)]$  تابع همبستگی مشاهدات می باشد.

#### ۵- فیلتر کالمن

هدف ما رسیدن به سیگنال گفتار تمیز از روی سیگنال نویزی می باشد. به این منظور از الگوریتم ارائه شده در [۴] استفاده می کنیم. این الگوریتم به صورت زیر بیان می شود:

$$e(n) = y(n) - \hat{x}(n|n-1)$$

$$\hat{x}(n+1|n) = \hat{F} \hat{x}(n|n-1) + \hat{F} K(n) e(n)$$

در مرجع [۲] یک روش تک کاناله حذف نویز زمینه با استفاده از تبدیل موجک به صورت فیلتر بانکی نشان داده شده است که در آن با استفاده از یک واحد (Voice- VAD Activity Detector) و با تشخیص بخشهای گفتار و نویز اقدام به بهسازی گفتار حاصل از روش BSS می شود. اما نتایج تجربی حاکی از آن است که استفاده از واحد VAD برای جداسازی نویزهای نایستان چندان موفق نمی باشد. علاوه بر این در چنین موقعیتهایی که هر سیگنال خروجی روش BSS شامل مؤلفه های ضعیفی از سیگنال دیگر می باشد، استفاده از VAD موجب حذف مؤلفه های ضعیفتر و باقی ماندن مؤلفه های قویتر خواهد شد. در این صورت در سیگنال بهسازی شده، پس زمینه های منقطع از صدای گوینده دیگر هم خواهیم داشت که کیفیت صدا را به شدت پایین می آورد.

ایده پیشنهادی ما استفاده از یک فیلتر کالمن اصلاح شده می باشد. بر خلاف سایر روشهای متداول، در این روش هیچ نیازی به VAD به منظور تخمین واریانس نویز وجود ندارد.

#### ۴- مدل گفتار نویزی و فیلتر کالمن

سیگنال گفتار  $s(n)$  با یک مدل  $AR$  مرتبه  $P$  به صورت زیر مدل میشود:

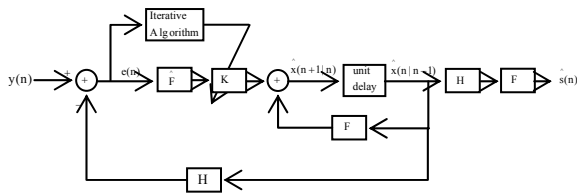
$$s(n) = \sum_{k=1}^p a_k s(n-k) + u(n) \quad (۶)$$

که  $s(n)$ ،  $n$  امین نمونه گفتار و  $a_i$ ،  $i$  امین پارامتر  $AR$  و  $u(n)$  نویز سفید گوسی است. سیگنال مشاهده شده  $y(n)$  به صورت زیر در نظر گرفته می شود:

$$y(n) = s(n) + v(n) \quad (۷)$$

که  $v(n)$  نویزی سفید گوسی و نا همبسته با  $u(n)$  است. نویزهای  $u(n)$  و  $v(n)$  را با میانگین صفر و واریانس به ترتیب  $\sigma_u^2$  و  $\sigma_v^2$  در نظر می گیریم.

سیستم را با معادلات حالت زیر مدل می کنیم:



شکل (۳) بهسازی گفتار با استفاده از فیلتر کالمن پیشنهادی

سیگنال نویزی  $y(n)$  وارد سیستم شده، پارامترهای AR در یک فریم آن تخمین زده شده و ماتریس  $\hat{F}$  ساخته می شود. به این ترتیب نمونه نوآوری  $e(n)$  به دست می آید. بهره  $K$  از روی دنباله نوآوری بهینه شده و این کار مرتب تکرار می شود. خروجی سیستم به طور مرتب نمونه های  $s(n)$  گفتار تمیز خواهد بود.

همانگونه که قبلا بیان شد در تمام روشهای قبلی، واریانس نویز از بخش سکوت گفتار تخمین زده می شود، در حالی که در الگوریتم حاضر تخمین واریانس نویز، قسمتی از تخمین بهره کالمن می باشد. در واقع بهره کالمن و به تبع آن واریانس نویز با استفاده از همبستگی دنباله نوآوری تخمین زده می شود.

#### ۷- شبیه سازی کامپیوتری

برای ارزیابی عملکرد سیستم از سیگنال گفتار یک گوینده زن و یک گوینده مرد از دادگان TIMIT و یک نویز همهمه که هر سه دارای فرکانس نمونه برداری ۱۶KHZ هستند، استفاده نموده ایم. فاصله بین دومیکروفون ورودی BSS را حدود ۱۰ cm (۵ نمونه)، درجه  $p$  پارامترهای AR برابر ۱۰ و طول فریم های گفتار را ۲۰ ms (معادل ۳۲۰ نمونه) در نظر گرفته ایم.

در همه آزمایشها فرض براین است که منابع گفتار به میکروفونهای مربوط به خود نزدیک بوده و نویز زمینه اثر یکسانی روی هر کدام از آنها داشته باشد. فیلترهای بین منابع و میکروفونها را به صورت فیلترهای FIR در نظر می گیریم.

که در این رابطه  $\hat{x}(n|n-1)$  تخمین mmse بردار حالت  $x(n)$  از روی مشاهدات  $y(1)$  و... و  $y(n-1)$  بوده،  $e(n)$  دنباله نوآوری،  $k(n)$  بهره کالمن در مرحله  $n$  ام و  $\hat{F}$  تخمین ماتریس  $F$  با معادلات Yule-Walker می باشد. معیار همگرایی این الگوریتم رسیدن به دنباله نوآوری سفید است. معیار سفید بودن را می توان به صورت زیر بیان کرد [۴]:

$$|\hat{r}_{ee}(l)| \leq 1.95 \frac{|\hat{r}_{ee}(0)|}{\sqrt{N}} \quad \text{for } l \geq 0 \quad (13)$$

که در آن  $\hat{r}_{ee}(l)$  تخمین همبستگی  $e(n)$  به صورت زیر بوده:

$$\hat{r}_{ee}(l) = \frac{1}{N} \sum_{i=1+l}^N e(i) e(i-1) \quad \text{for } 0 \leq l \leq p \quad (14)$$

و  $N$  طول فریم گفتار مورد استفاده است.

در این صورت بهره  $K$  به صورت زیر به روز می شود:

$$(15)$$

$$K(i+1) = K(i) + \begin{bmatrix} H \hat{F} \\ H(\hat{F}(I - K(i)H)) \hat{F} \\ \dots \\ H(\hat{F}(I - K(i)H))^{p-1} \hat{F} \end{bmatrix}^{-1} \times \begin{bmatrix} \hat{r}_{ee}(1) \\ \hat{r}_{ee}(2) \\ \dots \\ \hat{r}_{ee}(p) \end{bmatrix} / \hat{r}_{ee}(0)$$

به این ترتیب بردار حالت تخمین زده شده  $\hat{x}$  با استفاده از مقدار جدید  $K$  طبق معادله (۱۲) به روز می شود. در نهایت سیگنال گفتار تمیز با استفاده از معادله زیر به دست می آید:

$$\hat{s}(n) = H \hat{x}(n|n) = H \hat{F} \hat{x}(n|n-1) \quad (16)$$

در بخش بعدی طریقه اعمال این الگوریتم را روی سیگنال گفتار بیان خواهیم نمود.

#### ۶- بهسازی گفتار با استفاده از فیلتر کالمن

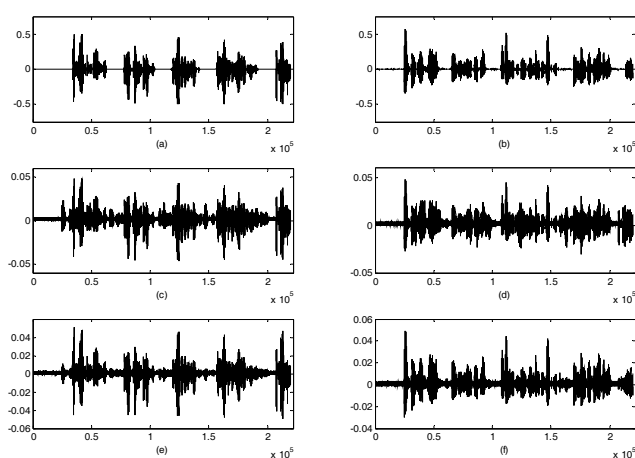
##### پیشنهادی

روش پیشنهادی تک کاناله بهسازی گفتار با استفاده از فیلتر

می شود خروجی روش BSS علاوه بر نویز زمینه، مؤلفه هایی از سیگنال گوینده دیگر را هم در بر دارد که در خروجی سایر روش ها نیز می توان آن را مشاهده کرد.

جدول ۲: بهبود SNR در روشهای مختلف با تغییر SNR ورودی BSS

Input SNR (dB)	SNR improvement (dB)				
	BSS	BSS+ Kalman	BSS+ Wiener	BSS+ MMSE	BSS+ SS.
0	5.7	11	10.6	9.8	9.7
5	4.5	9.4	9	8.2	8.2
10	3.9	8	7.9	7	6.9
15	3.2	6.2	6.2	5.5	5.5



شکل ۴: مقایسه روشهای مختلف بهسازی گفتار (a) گفتار تمیز (b) خروجی BSS (c) BSS+Kalman (d) BSS+Wiener (e) BSS+MMSE (f) روش تفریق طیفی BSS+ روش تفریق طیفی BSS+ به منظور ارزیابی بیشتر عملکرد سیستم، از معیار LAR distance استفاده می کنیم. این معیار به نوعی فاصله طیفی سیگنال نویزی را نسبت به سیگنال بدون نویز اولیه نشان می دهد. نشان داده شده که در میان روشهای مختلف اندازه گیری فاصله طیفی، معیار LAR بیشترین تطابق را با آزمونهای شخص گرا دارد [۵].

معیار LAR distance را می توان بصورت زیر محاسبه نمود:

با در نظر گرفتن  $K(m, q)$  ( $q=1, \dots, Q$ ) به عنوان  $Q$  ضریب انعکاسی فریم  $m$  ام، فاکتور AR بصورت زیر تعریف می شود:

$$AR(m, q) = \frac{1 + K(m, q)}{1 - K(m, q)} \quad (17)$$

برای نشان دادن عملکرد خوب الگوریتم در مرحله اول با تغییر SNR در یکی از میکروفونهای ورودی واحد BSS، بهبود SNR در خروجی فیلتر کالمن محاسبه شد. این نتایج در جدول ۱ نشان داده شده است.

جدول ۱: بهبود SNR در فیلتر کالمن

Input SNR (dB)	0	5	10	15
SNR Improvement (dB)	11	9.4	8	6.2

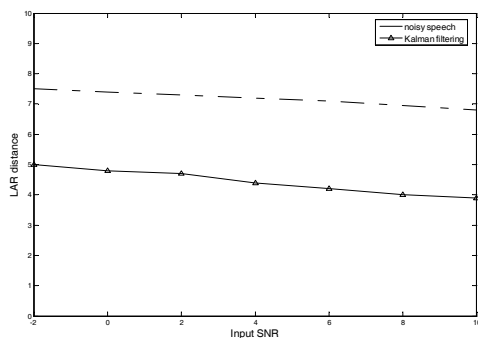
مشاهده می شود با وجود اینکه مقدار SNR ورودی BSS عدد بزرگی بوده، ولی هنوز هم کاهش قابل توجه نویز را در فیلتر کالمن داریم. علاوه بر این انتظار می رود که با افزایش SNR ورودی و بهبود کیفیت سیگنال ورودی، فیلتر نویز کمتری را حذف کند که مقادیر محاسبه شده SNR خروجی نیز همین تغییر را نشان می دهد.

در جدول ۲ بهبود SNR را برای سیستم فقط شامل BSS، همراه با فیلتر کالمن و BSS همراه با روشهای تک کاناله فیلتر وینر، MMSE و روش تفریق طیفی، با تغییر مقدار SNR ورودی بلوک BSS نشان داده ایم.

مشاهده می شود که به غیر از روش فیلتر وینر، سایر روشهای عملکرد بدتری نسبت به سیستم BSS همراه با فیلتر کالمن دارند. اگرچه به نظر می رسد استفاده از فیلتر وینر در مواقعی که نسبت سیگنال به نویز ورودی بیشتر شود عملکرد بهتری داشته باشد، اما همانطور که قبلاً اشاره شد، مشکل روشهای تک کاناله با استفاده از VAD این است که بعضی از مؤلفه های سیگنال گوینده دیگر را حذف کرده، در حالیکه مؤلفه های دیگر را باقی می گذارند. در این صورت صدای منقطع باقیمانده از گوینده دیگر در گفتار اصلی، آزاردهنده خواهد بود. تستهای شنوایی انجام شده با حضور افراد مختلف هم کیفیت بهتر سیگنال بهسازی شده با استفاده از روش فیلتر کالمن نسبت به سایر روشهای ذکر شده را تأیید می کند.

شکل ۴، شکل موجهای خروجی حاصل از هر پنج روش ذکر شده در یک محیط با انعکاس نسبتاً زیاد را همراه با شکل حاصل از سیگنال اصلی نشان می دهد. همانطور که مشاهده

مؤلفه های فرکانسی است. اخیراً برای حل این مشکل روشهایی پیشنهاد شده که در اکثر آنها سعی بر آن است تا با مرتب کردن مؤلفه های فرکانسی مجاور که بیشترین همبستگی را نسبت به هم دارند، به نوعی از جایگشت این مؤلفه ها جلوگیری شود [6]. اما چنین روشهایی به علت زمان محاسباتی زیاد برای الگوریتم های on-line مناسب نبوده و تحقیق در این زمینه همچنان ادامه دارد. روش پیشنهادی ما اگرچه برای حذف مؤلفه های جایگشت شده مناسب نیست اما بازده خوبی در حذف نویزهای زمینه از خود نشان می دهد.



شکل ۵: LAR distance گفتار نویزی و خروجی فیلتر کالمن

اگر  $AR_s(m, q)$  و  $AR_z(m, q)$  به ترتیب پارامترهای گفتار تمیز اولیه و سیگنال گفتار خروجی سیستم AR باشد LAR distance برای  $m$  امین فریم بصورت زیر قابل محاسبه خواهد بود:

(۱۸)

$$LAR_{sz}(m) = \left\{ \frac{1}{Q} \sum_{q=1}^Q 20 \log_{10} \left[ \frac{AR_s(m, q)}{AR_z(m, q)} \right] \right\}^2$$

برای

اینکه فریم های با LAR های خیلی بزرگ و غیر حقیقی را حذف کنیم، LAR distance نهایی را با حذف فریم هایی که در آنها فاصله LAR از ۹۵٪ مقدار ماکزیمم بیشتر باشد و متوسط گیری روی بقیه فریم ها بدست می آوریم. در این تحقیق برای محاسبه LAR از پنجره های همینگ با طول ۳۲۰ نمونه استفاده شد.

شکل (۵) مقدار LAR distance نهایی بین گفتار تمیز و خروجی فیلتر کالمن را برای نویز سفید نشان می دهد. برای مقایسه دقیقتر، در این شکل LAR distance بین گفتار نویزی ورودی فیلتر و گفتار تمیز هم رسم شده است.

## ۸- نتیجه گیری

در این تحقیق، ساختاری مرکب از یک واحد جداسازی کورمنابع (BSS) و فیلتر کالمن را معرفی نمودیم. نشان داده شد که در حضور نویز زمینه ساختار BSS همراه با فیلتر کالمن عملکرد به مراتب بهتری از یک سیستم BSS ساده دارد. علاوه بر این عملکرد این سیستم از سیستم های BSS همراه با سایر روشهای تک کاناله حذف نویز نیز بهتر می باشد. علت این امر عدم استفاده از VAD در این سیستم است. در واقع در این سیستم با مشکل مواجهه با نویزهای نایستان که در سایر الگوریتمهای تک کاناله حذف نویز وجود دارد کمتر روبرو خواهیم شد.

الگوریتم MAD که در این مقاله به عنوان روش جداسازی کورمنابع مورد استفاده قرار گرفت علاوه بر عملکرد خوبی که در محیط های با انعکاس نسبتاً زیاد دارد به علت اجرا در حوزه فرکانس، به مراتب از روشهای BSS حوزه زمان از پیچیدگی محاسباتی کمتری برخوردار می باشد.

اما مسأله اصلی در مورد این الگوریتم، مشکل جایگشت

## مراجع

- [1] L. Parra, C. Spence, "Convolutional blind separation of non-stationary sources," *IEEE Trans. on Signal and Audio Processing*, vol. 8, pp. 320-327, 2000.
- [2] E. Visser, T. W. Lee, "Speech enhancement using blind source separation and two-channel energy based speaker detection," in *Proc. of ICASSP*, pp. 836-839, 2003.
- [3] M.Z. Ikram, D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signal in a reverberant environment" in *Proc. of ICASSP*, 2000.
- [4] E. Grivel, M. Gaba and M. Najim, "Speech enhancement as a realization issue," *Signal Processing*, vol. 82, pp. 1963-1978, Dec 2002.
- [5] H. R. Abutalebi, H. Sheikhzadeh, R. L. Brennan, and G. H. Freeman, "Affine projection algorithm for oversampled subband adaptive filters," in *Proc. of ICASSP*, Hong Kong, China, Apr. 2003.
- [6] N. Murata, S. Ikeda, A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol.4, Issue 1-4, pp. 1-24, Oct. 2001.