بهسازی گفتار با استفاده از سیستم ترکیبی جداساز کور منابع و حذف‌کننده وفقی نویز

حمیدرضا ابوطالبی[i]؛ مجید پوراحمدی[ii]

چکیده

در این مقاله، یک روش جدید بهسازی گفتار در محیط‌های شامل دو منبع صوتی و همراه با نویز زمینه را معرفی خواهیم کرد. روش ما شامل ترکیب یک واحد جداسازی کور منابع و یک ساختار جدید حذف کنندهٔ وفقی نویز است. برای این سیستم دو ورودی در نظر گرفته شده است. سیگنال‌های دریافت شده به وسیله این ورودی‌ها شامل ترکیب دو منبع نقطه‌ای سیگنال به همراه نویز زمینه‌ای است که در هردو ورودی یکسان در نظر گرفته می‌شود. واحد اول یا همان جداساز کور، منابع نقطه‌ای (مانند سیگنال‌های صوتی دو گوینده نزدیک به هم) را جدا می‌کند و در ادامه، حذف کنندهٔ وفقی نویز، نویز زمینه را از این سیگنال‌ها حذف خواهد کرد. به منظور جداسازی کور منابع در واحد اول، در مراجع مختلف از الگوریتم‌های متنوعی استفاده شده است که ما در این تحقیق از الگوریتم Multiple Adaptive Decorrelation با قابلیت اجرا در حوزه فرکانس استفاده می‌کنیم. در بلوک دوم هم از یک ساختار بهبودیافته حذف کننده وفقی نویز استفاده می‌شود. بررسی‌های کمی و کیفی، برتری سیستم پیشنهادی را بر ترکیب جداسازی کور منابع با روش‌های تک‌کاناله بهسازی گفتار نشان‌می‌دهد.

کلمات کلیدی:

بهسازی گفتار، جداسازی کور منابع، گفتارنشتی، فیلتر وفقی.

Speech Enhancement Using the Hybrid Blind Source Separator-Adaptive Noise Canceller

H. R. Abutalebi; M. Pourahmadi

ABSTRACT

In this paper, we present a new robust speech enhancement method for the environments consist of two point sources and background noise. The method is based on the combination of two blocks; Blind Source Separation and a new Adaptive Noise Canceller structure. The first block separates point sources such as two nearby speaker voice signals whereas the second one eliminates the background noise from the desired speaker signal. In our experiments, we have employed the frequency domain Multiple Adaptive Decorrelation algorithm as the blind separation method. In addition, we show that conventional adaptive noise canceller is not proper for our goal and we have to employ a new structure that is called Asymmetric Cross-Talk Resistant Adaptive Noise Canceller. This system works well when the noise signal has speech leakage and the structure is not causal. Both objective and subjective experiments demonstrate the superiority of the proposed hybrid method over the combination of blind source separator with conventional single-channel speech enhancement methods.

KEYWORDS:

speech enhancement, blind source separation, cross talk, adaptive filter.

[i] استادیار، دانشکده مهندسی برق، دانشگاه یزد: habutalebi@yazduni.ac.ir

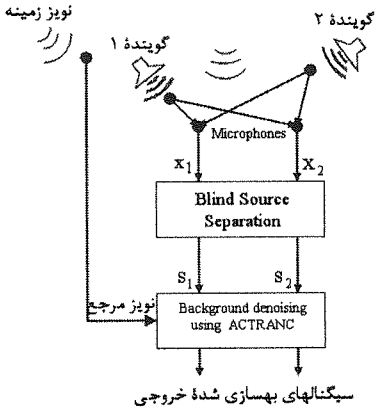
[ii] کارشناس ارشد، دانشکده مهندسی برق، دانشگاه یزد: m_pourahmad82@yahoo.com

۱- مقدمه

از آنجا که عملکرد مناسب بسیاری از سیستم‌های پردازش گفتار مستلزم کارکرد آنها در شرایط نویزی است، از این رو نیاز به یک واحد حذف نویز در آنها احساس می‌شود. در میان روش‌های مختلف، روش وفقی حذف نویز یا ANC (Adaptive Noise Canceller) پراستفاده ترین روش دوکاناله بهسازی گفتار است. در این روش، در کنار میکروفون اصلی (primary) از یک میکروفون مرجع (reference) نیز برای ضبط نویز استفاده می‌شود. با استفاده از الگوریتم‌های وفقی، ANC تابع تبدیل بین این دو میکروفون را تخمین می‌زند و در نهایت در خروجی خود، گفتار پاکسازی شده از نویز را ارایه می‌دهد. البته، عملکرد ANC (و روش‌های وابسته به آن) در مواقعی که انعکاس سیگنال در محیط نسبتاً زیاد باشد تا حد قابل توجهی کاهش خواهد یافت [۱]- [۶]. در چنین مواردی استفاده از روش جداسازی کور منابع یا BSS (Blind Source Separation) به منظور تبدیل منابع مخلوط شده به مؤلفه‌های سازنده می‌تواند به عنوان جایگزینی برای روش ANC مطرح باشد. در روش‌های متداول BSS هیچ اطلاع قبلی در مورد منابع یا کانال ترکیب کننده در اختیار نیست؛ اما فرض براین است که منابع نقطه‌ای است و هیچ نویز زمینه‌ای هم در محیط موجود نیست [۱]. با این وجود، در محیط‌های واقعی، منابع نویز زیادی موجود است که در ترکیب با هم نویز زمینه پیوسته‌ای ایجاد می‌کنند که با سیگنال منابع نقطه‌ای ترکیب می‌شود [۱]؛ در چنین شرایطی BSS قادر به حذف این گونه نویزها نخواهد بود. در [۷] و [۸] نشان داده شده است که با افزایش تعداد میکروفون‌ها، از نظر تئوری روش BSS می‌تواند نویز زمینه را نیز پاکسازی کند؛ اما نتایج تجربی نشان می‌دهد که در چنین شرایطی هم سیستم عملکرد قابل قبولی ندارد.

روش پیشنهادی ما، که شامل ترکیب BSS با یک بلوک حذف نویز زمینه است، در شکل (۱) نشان داده شده است. در این بلوک از یک ساختار نامتقارن CTRANC (Cross-Talk Resistant Adaptive Noise- Canceller که ACTRANC (Asymmetric CTRANC) نامیده می‌شود، استفاده کرده‌ایم.

فرض بر این است که محیط شامل دو منبع نقطه‌ای گفتار همراه با یک نویز زمینه (مثلاً نویز همهمه) باشد. اثر این نویز زمینه را در همه میکروفون‌ها یکسان در نظر می‌گیریم. در این سیستم از سه میکروفون استفاده شده است که دوتای آنها به عنوان ورودی‌های BSS و در فاصله‌ای نزدیک به منابع نقطه‌ای قرار گرفته است و سومی در فاصله‌ای دورتر از آنها و به عنوان میکروفون مرجع برای ضبط نویز زمینه استفاده می‌شود. میکروفون سوم به عنوان ورودی نویز مرجع برای سیستم ACTRANC عمل می‌کند.



شکل (۱): سیستم پیشنهادی بهسازی گفتار

۲- جداسازی کور منابع

در این تحقیق، برای جداسازی منابع از روش MAD (Multiple Adaptive Decorrelation) [۷]،[۸] استفاده شده است. این الگوریتم برای جداسازی منابع غیر ایستان کانوالو شده به صورت فریم به فریم مورد استفاده قرار می‌گیرد. ما این الگوریتم را به علّت سرعت بالا و نتایج مطلوب آن در مورد سیگنال‌های گفتار انتخاب کرده‌ایم. در روش MAD با m سیگنال x(t) حاصل از m میکروفون، که شامل سیگنال‌های m منبع نقطه‌ای s(t) و نویز زمینه جمع شونده n(t) است، داریم:

$$x(t)=\sum_{\tau=0}^{p}A(\tau)s(t-\tau)+n(t) \qquad (۱)$$

که P درجه کانولوشن و A(t) یک ماتریس ترکیب کننده m×m است. با این فرض که انعکاس محیط خیلی زیاد نباشد، منابع نقطه‌ای با دامنه و تأخیرهای مختلف به میکروفون‌ها می‌رسند. وظیفه سیستم جداکننده، پیداکردن این تأخیرها و اعمال آن روی سیگنال‌های مخلوط به منظور جداسازی آنهاست؛ اما نویز زمینه، که از برایند تعداد زیادی منبع ناشی می‌شود، تقریباً به طور یکسان و بدون تأخیر به همه میکروفون‌ها می‌رسد [۹]. در این صورت سیستم BSS قادر به جداسازی آن از سیگنال‌های گفتار نیست و در نتیجه نویز زمینه را در همه سیگنال‌های جداشده خواهیم داشت. وظیفه واحد ACTRANC جداسازی این نویز زمینه است که در بخش (۳) به تفصیل در مورد آن توضیح داده خواهد شد.

۲-۱- روش MAD

در سال‌های اخیر، روش‌های مختلفی برای جداسازی کور منابع با استفاده از همبستگی‌زدایی آمارگان‌های درجه دو یا

آمارگان‌های با درجه بالاتر، معرفی شده اند [۷]. روش MAD، که در پژوهش حاضر از آن بهره گرفته شده است، به منظور همبستگی‌زدایی آمارگان‌های درجه ۲ سیگنال‌های غیرایستان مورد استفاده قرار می‌گیرد. در این الگوریتم، هدف یافتن یک دنباله از ماتریس‌های جداکنندهٔ W(t) از روی سیگنال‌های ترکیب شده x(t) معادله (۱) به گونه‌ای است که:

$$\hat{s}(t) = \sum_{\tau=0}^{Q} W(\tau)\, x(t-\tau) \qquad (۲)$$

در این رابطه Q طول فیلتر جداکننده و $\hat{s}(t) = [\hat{s}_1(t),...,\hat{s}_m(t)]^T$ بردار تخمین منابع در لحظه t است.

از آنجا که در حوزه فرکانس عمل کانولوشن به ضرب تبدیل می‌شود، در الگوریتم MAD با پنجره بندی سیگنال با پنجره‌هایی به طول T، محاسبه فیلتر جداکننده در پنجره kاُم، در حوزه فرکانس و از روی مشاهدات زیر انجام می‌گیرد:

$$X(\omega,t) \approx A(\omega)s(\omega,t) + n(\omega,t), \quad \text{for } P << T \qquad (۳)$$

در این رابطه، T طول تبدیل فوریه زمان کوتاه T نقطه‌ای x(t) در پنجره kاُم و با شروع از لحظه $t_k = kT$ است که در آن T خیلی بزرگ‌تر از P (طول فیلتر ترکیب کننده) انتخاب می‌شود 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(ω) را می‌توان با استفاده از معیار زیر به دست آورد [۷]:

$$J = \arg\min_{W,\hat{\Lambda}_S} \sum_{k}\sum_{\omega=1}^{T} \left\| W\hat{R}_x(\omega,t_k)W^H - \hat{\Lambda}_s(\omega,t_k) \right\|^2 \qquad (٤)$$
$$W(\tau) = 0 \qquad \forall \tau > Q \ , \ Q << T$$
$$W_{ii}(\omega) = 1$$

در کاربردهایی که فیلتر ترکیب کننده سیگنال‌ها، A، به جای عمل کانولوشن، سیگنال‌ها را بدون تأخیر و فقط با تغییر دامنه با هم مخلوط می‌کند، ممکن است ترتیب سیگنال‌های بازسازی شده با سیگنال‌های اصلی متفاوت باشد و به اصطلاح، عمل جداسازی همراه با جایگشت صورت گیرد [۱۵].

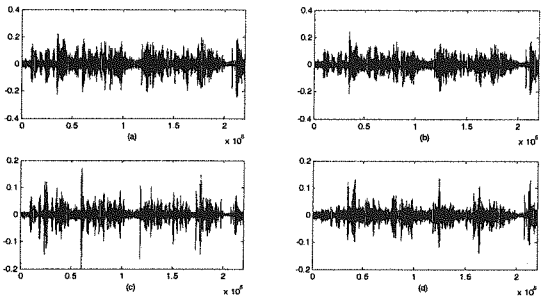
در الگوریتم MAD و با درنظر گرفتن فیلتر کانوالوکننده، برای هر مؤلفه فرکانسی به طور جداگانه عمل جداسازی انجام می‌گیرد و در نهایت، مؤلفه‌های فرکانسی هر منبع به طور جداگانه با هم ترکیب می‌شود و منبع بازسازی شده را در حوزه زمان خواهند ساخت. حال اگر در اجرای الگوریتم در هر فرکانس جایگشتی صورت گیرد، برای هر منبع و در بعضی از فرکانس‌ها مؤلفه‌هایی از منابع دیگر را خواهیم داشت و در این صورت سیگنال بازسازی شده در حوزه زمان، مؤلفه‌هایی از منابع دیگر را به همراه خواهد داشت [۷].

شرط اول معادله (٤) (خیلی کوچک‌تر بودن طول فیلتر جداکننده، Q، از T) برای این در نظر گرفته شده است که مؤلفه‌های زمانی این فیلتر هموارتر شود و مسأله جایگشت فرکانسی مرتفع گردد [۷]. البته باید توجه داشت که این شرط همیشه نمی‌تواند مانع از جایگشت شود[۹]. شرط دوم معادله (٤) نیز برای جلوگیری از تغییر مقیاس سیگنال‌های بازسازی شده نسبت به منابع اصلی اعمال می‌شود.

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

$$\Delta W^*(\omega) \sim \sum_k (W\hat{R}_x(\omega,t_k)W^H - \hat{\Lambda}_s(\omega,t_k))W(\omega)\hat{R}_x(\omega,t_k) \qquad (٥)$$

آزمایش‌های مختلف عملکرد بسیار مؤثر این الگوریتم در شرایط مختلف نویزی را نشان می‌دهد. شکل(۲) سیگنال‌های ورودی و خروجی حاصل از اجرای الگوریتم MAD در یک محیط با انعکاس نسبتاً زیاد را نشان می‌دهد. سیگنال‌های ورودی شامل گفتار دو گوینده است که همراه با نویز زمینه همهمه ضبط شده است. مشاهده می‌شود که در سیگنال‌های بازسازی شده نویز زمینه اصلی هنوز هم وجود دارد. علاوه بر این، به علت انعکاس بالای سیگنال‌های گفتار در محیط، هر سیگنال خروجی هنوز هم شامل مؤلفه‌هایی از سیگنال دیگر است که الگوریتم BSS نتوانسته آنها را جدا کند. در ادامه به توصیف روشی برای حذف نویز زمینه می‌پردازیم.



شکل(۲): جداسازی کورمنابع نقطه‌ای. (a) و(b): سیگنال‌های نویزی ورودی. (c) و(d): سیگنال‌های جداشدهٔ خروجی

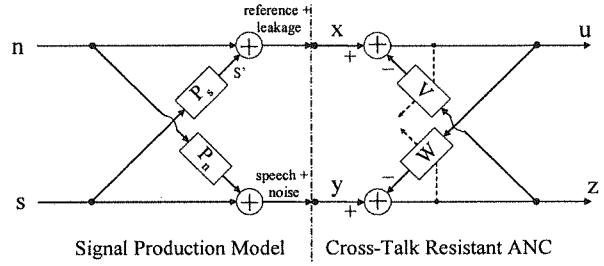
۳- حذف نویز زمینه

در مراجع [۹] و [۱۰]، روشی تک‌کاناله برای حذف نویز زمینه با استفاده از تبدیل موجک نشان داده شده است که در آن با استفاده از یک واحد VAD (Voice Activity Detector) و با تشخیص بخش‌های گفتار و نویز به بهسازی گفتار حاصل

از روش BSS اقدام می‌شود؛ ولی نتایج تجربی حاکی از آن است که استفاده از واحد VAD برای جداسازی نویزهای ناایستان چندان موفق نیست. علاوه براین، در چنین موقعیتی که هر سیگنال خروجی روش BSS شامل مؤلفه‌های ضعیفی از سیگنال دیگر است، استفاده از VAD موجب حذف مؤلفه‌های ضعیف‌تر و باقی ماندن مؤلفه‌های قوی‌تر خواهد شد. در این صورت در سیگنال بهسازی شده، پس‌زمینه‌های منقطعی از صدای گوینده دیگر هم خواهیم داشت که کیفیت صدا را به شدت پایین می‌آورد.

ایده پیشنهادی ما استفاده از یک سیستم جدید حذف نویز (ANC) است. همان طور که قبلاً نیز بیان شد در سیستم کلی پیشنهادی (شکل ۱) سه میکروفون موجود است که دوتای آن متعلق به واحد BSS و سومی به عنوان میکروفون مرجعِ واحد ANC کار می‌کنند. نشان داده شده است که بازده روش ANC به شدت به استقلال ورودی مرجع از گفتار وابسته است؛ اما در محیط‌های واقعی قرار گرفتن دو میکروفون در فاصله‌ای نزدیک به هم موجب نشت گفتار به سیگنال مرجع می‌شود (نیمه سمت چپ شکل ۳) و در نهایت، موجب حذف قسمت‌هایی از گفتار در خروجی سیستم خواهد شد.

برای غلبه بر مشکل گفتار نشتی، ساختار CTRANC که در نیمه راست شکل(۳) نشان داده شده، معرفی شده است [۳]. در این ساختار یک فیلتر وفقی ثانویه، که در شکل با V نشان داده شده است، برای حذف گفتارنشتی از ورودی مرجع استفاده می‌شود. وظیفه این فیلتر ثانویه مدلسازی تابع تبدیل بین گفتار در میکروفون اصلی (primary) و گفتارنشتی در میکروفون مرجع است که ضرائب هر دو فیلتر وفقی به طور جداگانه و با الگوریتم NLMS (Normalized Least Mean Square) تنظیم می‌شود. در ساختار CTRANC با فیدبک (شکل ۳)، خروجی هر فیلتر ورودی فیلتر دیگر است. این ساختار تا زمانی که منبع گفتار نزدیک میکروفون اصلی و منبع نویز نزدیک میکروفون مرجع باشد بخوبی کار می‌کند [۳]. به علت وجود شرط علّی بودن فیلترها، ساختار CTRANC در هر کانال، سیگنالی را که دیرتر به میکروفون برسد، حذف می‌کند و سیگنال مقدم (پیشین) را نگه می‌دارد [۲].



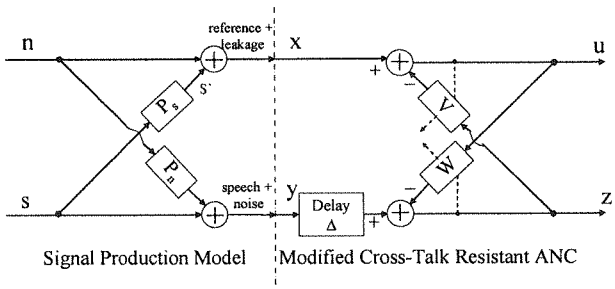
شکل(۳): ساختار CTRANC

در تحقیق ما، نویز زمینه به طور تقریباً یکنواخت و همزمان در کل محیط وجود دارد و بنابراین با توجه به قید علّی بودن فیلترهای وفقی، عملکرد سیستم مطلوب نخواهد بود. برای کاهش این محدودیت و با الهام از مفاهیم مربوط به شکل‌دهی بیم (Beamforming)، یک بلوک تأخیر سر راه روی ورودی اصلی (y در شکل ۳)، قرار داده می‌شود تا در فیلتر وفقی حذف کننده نویز (W)، نویز موجود در میکروفون اصلی از نویز موجود در میکروفون مرجع عقب تر باشد. ما این ساختار را CTRANC نامتقارن (ACTRANC) نامیده و با نتایج تجربی، عملکرد برتر آن را نسبت به سیستم CTRANC نشان داده‌ایم.

علاوه بر این، افزایش سرعت همگرایی سیستم، به جای الگوریتم NLMS از الگوریتم AP (Affine Projection) برای تنظیم ضرائب فیلترها استفاده شده است. در ادامه، پس از معرفی کلی ساختار ACTRANC، در مورد الگوریتم AP، که در کاربرد برای ساختار با دو فیلتر وفقی آن را اصطلاحاً DAP (Double Affine Projection) می‌نامیم، صحبت خواهیم کرد.

۳-۱- ساختار ACTRANC

به منظور رهایی از قید علّی بودن ساختارهای CTRANC متداول، با قرار دادن یک بلوک تأخیر روی کانال اصلی ورودی، ساختار جدیدی به نام ساختار نامتقارن CTRANC یا ACTRANC (Asymmetric CTRANC) پیشنهاد می‌شود که در شکل (٤) نشان داده شده است:



شکل(۴): ساختار ACTRANC

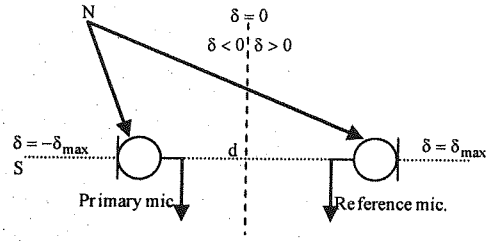
تأثیر قرار دادن واحد تأخیر و مقدار مناسب این تأخیر با توجه به شکل (٥) قابل تشریح است: δ را به عنوان تفاضل بین تأخیر انتشار از منبع نویز تا میکروفون اصلی و تا میکروفون مرجع در نظر می‌گیریم. با فرض ثابت بودن منبع گفتار در محل S و جابجایی محل منبع نویز (N)، مقدار δ بین δ_{max} و $-\delta_{max}$ تغییر می‌کند. داریم:

$$\delta_{max} = \frac{d}{c} f_s \qquad (٦)$$

در این رابطه، d فاصله بین میکروفون‌های اصلی و مرجع بر حسب متر، c سرعت صوت بر حسب متر بر ثانیه و f_s فرکانس

نمونه‌برداری است.

با فرض علّی بودن فیلتر وفقی W، برای عملکرد مناسب ANC مؤلفه نویز در ورودی مرجع باید زودتر از مؤلفه نویز در ورودی اصلی باشد؛ به بیان دیگر، با توجه به شکل (۵) باید $\delta > 0$ باشد. با تعبیر نویز زمینه به صورت برایندی از نویزهای نقطه‌ای در کل فضا، می‌توان گفت که ساختار CTRANC فقط قادر به پاکسازی نویزهای با منشأ نیمه راست (شکل ۵) (یا به عبارتی با $\delta > 0$) خواهد بود.



شکل(۵): بررسی فاصله نسبی منبع نویزتا میکروفون‌های اصلی و مرجع

استفاده از بلوک تأخیر (Δ) برای امکان حذف نویزهایی با منشأ نیمه چپ (شکل ۵) است. با به کارگیری بلوک تأخیر (Δ) بر سر راه ورودی اصلی، تفاضل تأخیر میان مؤلفه نویز رسیده به میکروفون مرجع و به مؤلفه نویز میکروفون اصلی عملاً با $\delta_{eq} = \delta + \Delta$ برابر می‌شود؛ بنابراین با تنظیم مناسب مقدار Δ، در نقاط بیشتری از فضا $\delta_{eq} > 0$ برقرار است و یا به بیان دیگر، به میزان نویز قابل حذف افزوده خواهد شد. هر چقدر منبع نویزی به سمت S برود، باید مقدار Δ را به حدی افزایش داد تا شرط $\delta_{eq} > 0$ محقق شود. از سوی دیگر، با توجه به محدودیت علّی بودن فیلترV، مؤلفه سیگنال گفتار در ورودی اصلی (بعد از واحد تأخیر) باید زودتر از مؤلفه گفتار نشتی در ورودی مرجع باشد؛ بنابراین Δ باید کمتر از δ_{max} باشد. با در نظر گرفتن این دو جنبه، مقدار مناسب برای Δ (برای دستیابی به حداکثر میزان حذف نویز) برابر خواهد بود با:

$$\Delta = \delta_{max}\text{-}1 \qquad (۷)$$

بر پایه بحث شکل‌دهی بیم، قرار دادن بلوک تأخیر به صورت هدف‌گیری میکروفون اصلی به سمت منبع گفتار (S) تعبیر می‌شود که باعث حذف اثر نویز نشتی در میکروفون اصلی خواهد شد. بر طبق این تعبیر، با افزایش Δ از صفر به سمت δ_{max} ، بیم مجموعه دو میکروفون، در جهت منبع گفتار (S) متمرکز می‌شود و بخش‌های عمده‌تری از محیط اطراف، که محل حضور نویز زمینه است، در خارج از بیم مجموعه قرار می‌گیرد. این به نوبه خود قدرت حذف نویز را در ACTRANC نسبت به CTRANC بهبود می‌بخشد.

۳-۲- الگوریتم Double Affine Projection

در این مقاله از یک الگوریتم DAP به منظور تنظیم ضرائب فیلترهای وفقی استفاده می‌شود. در مقایسه با NLMS، DAP سرعت همگرایی بیشتر و خطای باقیمانده کمتری خواهد داشت. در الگوریتم DAP(p) تصویرسازی در p بعد صورت می‌گیرد و با افزایش بعد تصویرسازی DAP (p)، سرعت همگرایی ضرائب وزنی در ازای افزایش هزینه پیچیدگی محاسباتی الگوریتم افزوده خواهد شد. در حالی که بار محاسباتی الگوریتم NLMS برای دو فیلتر وفقی (با طول M) تقریباً 4M است، تکنیک DAP(p) حجم محاسباتی برابر با 2(2M+20p) دارد [۱۲]،[۱۳].

الگوریتم DAP(p) برای سیستم ACTRANC و با تأخیر Δ به صورت زیر توصیف می‌شود:

– فرایند فیلترکردن:

$$\begin{aligned} \underline{u}(k) &= \underline{x}(k-\Delta) - Z'(k)\underline{V}^*(k) \\ \underline{z}(k) &= \underline{y}(k) - U'(k)\underline{W}^*(k) \end{aligned} \qquad (۸)$$

که درآن:

$$\begin{aligned} \underline{x}(k) &= [x(k),....,x(k-P+1)] \\ \underline{y}(k) &= [y(k),....,y(k-P+1)] \\ Z(k) &= [\underline{z}(k),....,\underline{z}(k-P+1)] \end{aligned} \qquad (۹)$$

و

$$\begin{aligned} U(k) &= [\underline{u}(k),...,\underline{u}(k-P+1)] \\ \underline{V}(k) &= [V_0(k),....,V_{L1}(k)] \\ \underline{W}(k) &= [W_0(k),....,W_{L2}(k)] \end{aligned} \qquad (۱۰)$$

و علامت ' مبین عمل ترانهاده (Transposition) است.

– به‌روز رسانی ضرائب فیلتر:

$$\begin{aligned} \underline{V}(k+1) &= \underline{V}(k) + \mu_1 Z(k)\left[Z^H(k)Z(k) + \delta_1 I\right]^{-1}\underline{u}^*(k) \\ \underline{W}(k+1) &= \underline{W}(k) + \mu_2 U(k)\left[U^H(k)U(k) + \delta_2 I\right]^{-1}\underline{z}^*(k) \end{aligned} \qquad (۱۱)$$

در این رابطه، ماتریس I ماتریس واحد است. مقادیر δ_i $(i = 1,2)$ پارامترهای تنظیم ماتریس معکوس خودهمبستگی $\left[U^H(k)U(k) + \delta_2 I\right]^{-1}$ یا $\left[Z^H(k)Z(k) + \delta_1 I\right]^{-1}$ است که به منظور بهبود رفتار سیستم در شرائط مغکوس‌ناپذیری $Z^H(k)Z(k)$ یا $U^H(k)U(k)$ مورد استفاده قرار می‌گیرد. در چنین شرایطی می‌توان گفت برای مقادیر به قدر کافی بزرگ δ_1 و δ_2، $Z^H(k)Z(k) + \delta_1 I$ و $U^H(k)U(k) + \delta_2 I$ معکوس‌های بهتری خواهند داشت. پارامتر اندازه گام، μ، نیز مقدار متغیری است که همانند الگوریتم NLMS برای پایداری

باید $2 > \mu > 0$ انتخاب شود.

٤- شبیه سازی

برای ارزیابی عملکرد سیستم کلی (شکل ۱)، آزمایش‌های متعددی در اتاقی به ابعاد تقریبی چهار متر انجام گرفته است. به منظور ایجاد یک نویز زمینه مناسب در فضای میان اتاق، چهار بلندگو در چهار گوشه اتاق، سیگنال نویز همهمه را به طور همزمان پخش کرده‌اند. سیگنال گفتار مربوط به دو گوینده نیز، از دو گوینده زن و مرد از دادگان TIMIT استخراج شده است. همگی این سیگنال‌ها، فرکانس نمونه‌برداری ١٦kHz دارند. میکروفون‌ها در وسط اتاق با آرایش خطی مطابق شکل (۱) به گونه‌ای هستند که فاصله بین دومیکروفون ورودی BSS برابر با ۱۰cm است و میکروفون مرجع نیز در فاصله بین ۳۰-۵۰cm از آنها (با فاصلهٔ متغیر نسبت به میکروفون‌های BSS برای دستیابی به وضعیت‌های مختلف هندسی و SNR ورودی) قرار گرفته است. دو منبع گفتار نیز در مجاورت میکروفون مربوط به خود قرار گرفته‌اند.

در این تحقیق، از الگوریتم DAP(4) برای تنظیم فیلترهای وفقی استفاده کرده‌ایم. در الگوریتم DAP، $\mu_1 = \mu_2 = 0.1$ و $\delta_1 = \delta_2 = 0.01$ انتخاب می‌شود.

برای نشان دادن عملکرد خوب الگوریتم در مرحلهٔ اول با تغییر SNR در میکروفون مرجع و ثابت نگه داشتن SNR ورودی واحد BSS، بهبود SNR، که از کم کردن SNR ورودی نویزی از SNR خروجی بهسازی شده بدست می‌آید، در خروجی‌های CTRANC و ACTRANC محاسبه شد. ذکر این نکته لازم است که برای فراهم سازی امکان محاسبه SNR، از یک واحد VAD برای تشخیص بخش‌های گفتار و سکوت استفاده شده است. این نتایج برای وقتی که در ورودی بلوک BSS، SNR=10dB باشد، در جدول(۱) نشان داده شده است.

جدول(۱): بهبود SNR در CTRANC و ACTRANC

Ref-Input SNR (dB)	SNR improvement (dB)	
	ACTRANC	CTRANC
-5	12	1.4
0	9.8	1.2
5	8.0	0.9
10	7.6	0.7

مشاهده می‌شود با وجود اینکه مقدار SNR ورودی BSS عدد بزرگی است، ولی هنوز هم کاهش قابل توجه نویز را در ACTRANC داریم؛ ولی این مقدار در مورد CTRANC اصلاً رضایت بخش نیست. علاوه بر این، انتظار می‌رود با افزایش SNR ورودی و بهبود کیفیت سیگنال ورودی، ACTRANC نویز کمتری را حذف کند که مقادیر محاسبه شده SNR خروجی نیز همین تغییر را نشان می‌دهد.

برای ارزیابی بیشتر عملکرد سیستم، از معیار Log Area-Ratio (LAR) distance استفاده می‌کنیم. نشان داده شده است که در میان روش‌های مختلف اندازه گیری فاصله طیفی، معیار فاصله LAR بیشترین تطابق را با آزمون‌های شخص گرا دارد [۱۴].

با در نظر گرفتن K(m,q) (q=1,...,Q) به عنوان Q ضریب انعکاسی فریم m اُم، فاکتور AR به‌صورت زیر تعریف می‌شود:

$$AR(m,q) = \frac{1 + K(m,q)}{1 - K(m,q)} \qquad (۱۲)$$

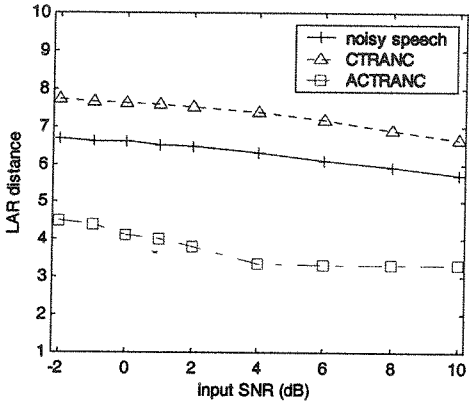
اگر $AR_s(m,q)$ و $AR_z(m,q)$ به ترتیب پارامترهای AR گفتار تمیز اولیه و سیگنال گفتار خروجی سیستم باشد LAR-distance برای mاُمین فریم به صورت زیر قابل محاسبه خواهد بود:

$$LAR_{sz}(m) = \left\{ \frac{1}{Q} \sum_{q=1}^{Q} \left| 20 \log_{10} \left[\frac{AR_s(m,q)}{AR_z(m,q)} \right] \right|^2 \right\}^{\frac{1}{2}} \qquad (۱۳)$$

برای اینکه فریم‌های با LAR های خیلی بزرگ و غیر حقیقی را حذف کنیم، LAR-distance نهایی را با حذف فریم‌هایی که در آنها فاصله LAR از ٩٥٪ مقدار ماکزیمم بیشتر باشد و متوسط گیری روی بقیه فریم‌ها به دست می‌آوریم [۱۴]. برای محاسبه LAR از پنجره‌های همینگ با طول ۲۰۰ نمونه استفاده شده است.

شکل(۶) مقدار LAR-distance نهایی بین گفتار تمیز و خروجی‌های روش‌های CTRANC و ACTRANC را نشان می‌دهد. برای مقایسه دقیق‌تر، در این شکل LAR–distance بین گفتار تمیز و گفتار نویزی ورودی کانال اصلی هم رسم شده است. از این شکل می‌توان عملکرد به مراتب بهتر ACTRANC را نسبت به CTRANC مشاهده کرد. دلیل این بهبود عملکرد، قرار دادن واحد تأخیر و عملکرد آزادتر سیستم نسبت به محل منبع نویز است. علاوه براین، می‌توان مشاهده کرد که LAR–distance خروجی CTRANC از گفتار نویزی هم بیشتر است که حساسیت روش نسبت به محل منبع نویز را نشان می‌دهد.

در جدول(۲) بهبود SNR را برای سیستم فقط شامل BSS ، BSS همراه با بلوک ACTRANC (با میزان SNR =-10dB در ورودی مرجع) و BSS همراه با یکی از سه روش تک کاناله فیلتر وینر، MMSE و روش تفریق طیفی، بر حسب تغییرات SNR در ورودی بلوک BSS نشان داده‌ایم.



شکل(۶): LAR-distance برای گفتار نویزی، خروجی CTRANC و ACTRANC

جدول(۲): بهبود SNR در روش‌های مختلف با تغییر SNR ورودی

Input SNR (dB)	SNR improvement (dB)				
	BSS	BSS+ ACTRANC	BSS+ Wiener	BSS+ MMSE	BSS+ SS.
0	4.4	12	11.6	9.8	9.7
5	5	12.8	11.9	9.82	9.8
10	7.4	13.5	13.5	11.8	11.9
15	11.3	16.3	17.1	15.5	15.5

مشاهده می‌شود که به غیر از روش فیلتر وینر، سایر روش‌ها عملکرد بدتری نسبت به سیستم BSS همراه با واحد ACTRANC دارند. اگرچه به نظر می‌رسد 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 این است که بعضی از مؤلفه‌های سیگنال گوینده دیگر را حذف می‌کند، در حالی که مؤلفه‌های دیگر را باقی می‌گذارند؛ در این صورت، صدای منقطع باقیمانده از گوینده دیگر در گفتار اصلی، آزاردهنده خواهد بود.

برای بررسی مقایسه‌ای دو روش BSS+ACTRANC و BSS+Wiener به انجام آزمون Mean Opinion Score (MOS) [۱۴] با حضور ۳۰ شنونده و بر روی ۴۸ نمونه گفتار (با جایگشت‌های مختلف گوینده زن و مرد و در SNR های ورودی مختلف) اقدام شد. نتایج این آزمون، که در جدول (۳) نشان داده شده است، برتری کیفیت سیگنال بهسازی شده با استفاده از روش BSS+ACTRANC را (نسبت به روش BSS+Wiener) تأیید می‌کند.

توضیح این مطلب لازم است که اگرچه پیاده‌سازی فعلی ACTRANC در حوزه زمان به بار محاسباتی بیشتر درمقایسه با فیلتر وینر (در حوزه فرکانس) منجر می‌شود، ولی با پیاده‌سازی ACTRANC در حوزه فرکانس (و در قالب فیلترهای وقفی بلوکی) می‌توان دید که دو تکنیک BSS+ACTRANC و BSS+Wiener بار محاسباتی تقریباً یکسانی دارند [۱۵]. زمان‌های اندازه‌گیری شده در شبیه‌سازی برای محاسبات در دو روش، مؤید این موضوع است.

شکل (۷) شکل موج‌های نمونه خروجی حاصل از پنج روش ذکر شده در فوق را به همراه با شکل موج سیگنال اصلی نشان می‌دهد. همان طور که مشاهده می‌شود خروجی روش BSS علاوه بر نویز زمینه، مؤلفه‌هایی از سیگنال گوینده دیگر را هم در بر دارد که در خروجی سایر روش‌ها نیز می‌توان آن را مشاهده کرد.

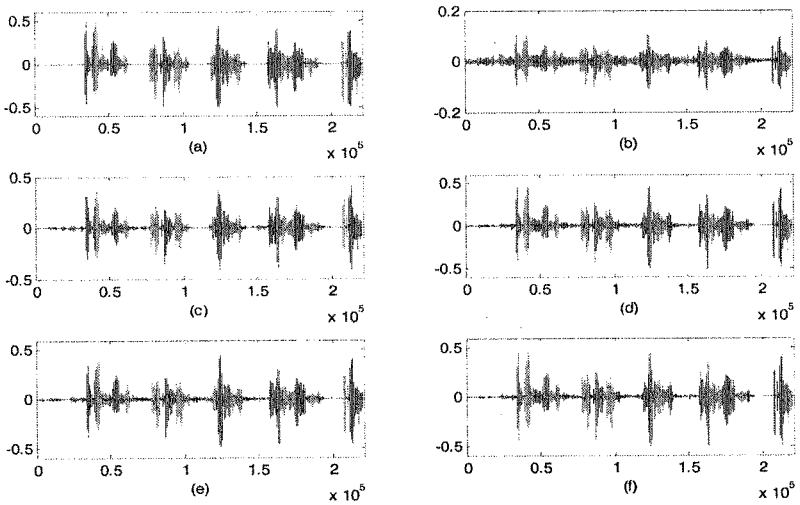
جدول (۳): نتایج آزمون MOS برای خروجی روش‌های BSS+ACTRANC و BSS+Wiener

Input SNR (dB)	MOS Result	
	BSS+ACTRANC	BSS+Wiener
0	3.8	3.3
5	3.9	3.3
10	3.9	3.3
15	4.0	3.4

۵- نتیجه‌گیری

در این تحقیق، ساختاری مرکب از یک واحد جداسازی کورمنابع (BSS) و یک واحد ACTRANC معرفی شد. نشان داده شد که در حضور نویز زمینه ساختار BSS همراه با CTRANC نامتقارن عملکرد به مراتب بهتری از یک سیستم BSS ساده دارد. علاوه بر این، کیفیت گفتار خروجی این سیستم از سیستم‌های BSS همراه با روش‌های تک کاناله حذف نویز نیز بهتر است.

الگوریتم MAD، که در این مقاله به عنوان روش جداسازی کورمنابع مورد استفاده قرار گرفت، علاوه بر عملکرد خوبی که در محیط‌های با انعکاس نسبتاً زیاد دارد، به علّت اجرا در حوزه فرکانس، به مراتب از روش‌های BSS حوزه زمان از پیچیدگی محاسباتی کمتری برخوردار می‌باشد [۱۶]؛ اما مسأله اصلی در مورد این الگوریتم، مشکل جایگشت مؤلفه‌های فرکانسی است. اخیراً برای حل این مشکل روش‌هایی پیشنهاد شده است که در اکثر آنها سعی بر آن است تا با مرتب کردن مؤلفه‌های فرکانسی مجاور، که بیشترین همبستگی را نسبت به هم دارند، به نوعی از جایگشت این مؤلفه‌ها جلوگیری شود [۱۷]؛ اما چنین روش‌هایی به علّت زمان محاسباتی زیاد برای کاربردهای بلادرنگ (real-time) مناسب نیست و تحقیق در این زمینه همچنان ادامه دارد. روش پیشنهادی ما اگرچه برای حذف مؤلفه‌های جایگشت شده مناسب نیست، اما بازده خوبی در حذف نویزهای زمینه از خود نشان می‌دهد.



شکل(۷): مقایسه روش‌های مختلف بهسازی گفتار (a) گفتار تمیز (b) خروجی BSS (c) BSS+ACTRANC (d) BSS+Wiener (e) BSS+MMSE (f) BSS+SS

۶- مراجع

[۱] ابوطالبی، ح.ر.؛ پوراحمدی، م. و آقابزرگی، م.ر.؛ "بهسازی گفتار با استفاده از جداسازی کور منابع در حضور نویز زمینه"، مجموعه مقالات چهاردهمین کنفرانس مهندسی برق ایران ، دانشگاه صنعتی امیرکبیر ، اردیبهشت ۱۳۸۵.

[۲] Abutalebi, H. R., Pourahmadi, M., and Aghabozorgi, M. R., "A new dual-microphone speech enhancement method for oriented noises", in Proc. of International Conf. on Spoken Language Processing (ICSLP), Pensilvania, USA, pp 2150-2153, 2006

[۳] Weinstein, E., Feder, M., and Oppenheim, A. V., "Multi-channel signal separation by decorrelatio,", IEEE Trans. on Speech and Audio Processing, vol. 1, no 4, pp 405-413, 1993

[۴] Mirčhandani, G., Zinser, Jr. R. L., and Evans, J. B., "A new adaptive noise cancellation scheme in the presence of Crosstalk", IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing, vol. 39 no. 10 pp. 681-694, Oct. 1992.

[۵] Gerven, S. V., and Compernole, D. V., "Signal separation by symmetric adaptive decorrelation: Stability, convergence and uniqueness", IEEE Trans. on Signal Processing, vol. 43, no. 7, pp 1602-1612, July 1995.

[۶] Compernolle, D. V., and Gerven, S. V., "Signal separation in a symmetric adaptive noise canceler by output decorrelation", in Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 221-224, 1992

[۷] Parra, L., and Spence, C., "Convolutive blind separation of nonstationary sources", IEEE Trans. on Speech and Audio Processing, vol. 8, no. 3, pp.320-327, May 2000.

[۸] Parra, L., Spence, C., and Vries, B. D., "Convolutive blind source separation based on multiple decorrelation", IEEE Workshop Neural Networks Signal Processing, UK, pp. 2332 , Sept. 1998

[۹] Visser, E. , and Lee, T. W., "Speech enhancement using blind source separation and two channel energy based speaker detection", in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp 836-839, 2003

[۱۰] Visser, E., Otsuka, M., and Lee, T. W.,"A spatio-temporal speech enhancement scheme for robust speech recognition in noisy environments", Speech Communication, vol.41 pp. 393-407, Dec. 2002

[۱۱] Kou, S. M., and Peng, W. M., "Principle and application of asymmetric crosstalk resistant adaptive noise canceller", in Proc. of the IEEE Workshop on Signal Processing Systems, pp. 605-614, Oct. 1999.

[۱۲] Abutalebi, H. R., Sheikhzadeh, H., Brennan, R. L., and Freeman, G. H., "Affine projection algorithm for oversampled subband adaptive filters", in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp 209-212, Hong Kong, China, Apr. 2003.

[۱۳] Gabrea, M., "Double affine projection algorithm-based speech enhancement algorithm", in Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp 856-859, Hong Kong, China, Apr. 2003.

[۱۴] Quackenbush, S. R., Banwell T. P., and Clements, M. A., Objective Measures of Speech Quality, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[۱۵] S. Haykin, Adaptive Filter Theory, 4th edition, Prentice-Hall, 2002.

[۱۶] Ikram, M. Z., and Morgan, D. R., "Exploring permutation inconsistency in blind separation of speech signal in a reverberant environment", in Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 1041-1044, 2000.

[۱۷] Murata, N., Ikeda, S., and Ziehe, A.,"An approach to blind source separation based on temporal structure of speech signals", Neurocomputing, vol. 4, pp. 1-24, Oct. 2001