

## حذف نویز دومیکروفونه بر اساس جداسازی منابع \_ کاربرد در هدست های تلفنی

بابک ناصرشریف	احمد اکبری	محسن رحمانی	نیلوفر صدیقیان
دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر	دانشکده مهندسی کامپیوتر
دانشگاه علم و صنعت ایران	دانشگاه علم و صنعت ایران	دانشگاه علم و صنعت ایران	دانشگاه علم و صنعت ایران
nasser_s@iust.ac.ir	akbari@iust.ac.ir	m_rahmani@iust.ac.ir	n.sadighian@yahoo.com

سکوت گفتار نویزی، تخمین میزند. این روش ها زمانیکه نویز ایستان باشد، کارایی مورد قبولی دارند. از جمله روش های شناخته شده تک میکروفونی، تفاضل طیف [۶] و فیلتر وینر [۷] را می توان نام برد. این روشها عموماً سبب ایجاد نویز موسیقی و اعوجاج گفتار می شوند. برای غلبه بر محدودیتهای روش تک میکروفون، روشهای چند میکروفون بکار می روند که نقطه قوت آنها کاهش نویز بر اساس مکان منابع گفتاری است برخی از روشهای چند میکروفون عبارتند از: تحلیل مولفه های مجزا (ICA) [۸] و تابع همدموسی<sup>۱</sup> [۴].

از میان روشهای حذف نویز، حذف نویز چند کاناله با استفاده از جداسازی کور منابع بدلیل آنکه جداسازی نویز از سیگنال صحبت بدون دانستن مشخصات کانال انتقال انجام می گیرد مورد توجه قرار گرفته است [۳] [۸]. اغلب الگوریتمها و روشهای موجود جهت حذف نویز بر اساس جداسازی منابع، قادر به جداسازی ترکیبهای خطی نویز و سیگنال صحبت هستند. این وضعیت در محیط آزمایشگاهی کاربرد دارد و در محیطهای واقعی نمی توان از آنها بهره برد [۳]. بنابراین برای حذف نویز در کاربردهای عملی، نمی توان از الگوریتمهایی که صرفاً نویز را از ترکیبات خطی سیگنال و نویز جدا می کنند استفاده نمود. جهت رفع این مشکل در این مقاله روشی پیشنهاد می شود که با بهره گیری همزمان از الگوریتمهای تک میکروفونه و جداسازی کور منابع دو میکروفونه قادر به جداسازی سیگنال تمیز از مخلوط کانولوتیو سیگنال صحبت و نویز باشد.

ساختار ادامه مقاله اینگونه است. در بخش دوم اصول جداسازی منابع مورد بحث قرار می گیرد. بخش سوم شامل فیلتر وینر دو میکروفونه است. بخش چهارم حاوی بحث جداسازی کور منابع در حوزه زمان و روش پیشنهادی در این حوزه است. بخش پنجم به ارزیابی کارایی الگوریتم پیشنهادی و نتایج آن اختصاص دارد. در نهایت بخش ششم حاوی جمع بندی نهایی است.

### ۲- جداسازی منابع

تصور کنید در اتاقی قرار دارید که دو نفر همزمان صحبت می کنند و دو میکروفون در دو مکان مختلف از اتاق قرار دارد. میکروفونها بسته به

چکیده: روشهای مختلفی برای حذف نویز صوتی از سیگنال گفتار ارائه شده اند. روشهای حذف نویز را از حیث تعداد میکروفون می توان به دو گروه تک میکروفونه و چند میکروفونه تقسیم کرد. در این میان روش های دو میکروفونی به عنوان حد وسط روش های دو میکروفونی و چند میکروفونی مطرح می شوند. یکی از روشهای حذف نویز در سیستمهای دو میکروفونه روش جداسازی منابع است. کارایی روشهای جداسازی، در محیط های غیرآزمایشگاهی که سیگنال های دریافتی به صورت ترکیب خطی نویز و گفتار نیستند، کاهش می یابد. به این منظور در مقاله حاضر، یک روش ترکیبی از فیلتر وینر و الگوریتم جداسازی منابع بر اساس معیار قابلیت تخمین در حوزه زمان پیشنهاد گردیده است. در این روش با بکارگیری فیلتر وینر ابتدا سطح نویز در سیگنال صحبت کاهش می یابد و سپس با استفاده از روش جداسازی منابع نویز باقیمانده از مرحله قبل از سیگنال صحبت جدا می گردد. نتایج ارزیابی با معیار PESQ و همچنین ارزیابی شنوایی بیانگر بهبود کیفیت گفتار با بکارگیری روش پیشنهادی است.

واژه های کلیدی: حذف نویز، جداسازی کور منابع، معیار قابلیت تخمین، الگوریتم MTP، فیلتر وینر

### ۱- مقدمه

با گسترش سیستمهای ارتباطی در چندین دهه اخیر، بهبود گفتار بیش از پیش اهمیت یافته است. نمونه این امر سیستمهای هدست در تلفنهای همراه است که در آن به علت فاصله میان دهان گوینده و میکروفون، سیگنال گفتار دریافتی به نویز آغشته می گردد. در این موارد یک سیستم کاهش و حذف نویز به افزایش کیفیت سیگنال گفتار نویزی کمک می کند.

روشهای حذف نویز از سیگنالهای گفتار به دو دسته کلی روشهای تک میکروفونی و چند میکروفونی تقسیم می شوند. الگوریتمهای تک میکروفونی بر اساس اطلاعات زمان-فرکانسی سیگنالهای ضبط شده عمل می کنند و میزان نویز را با استفاده از سیگنال نویزی، مثلاً محدوده

<sup>1</sup> Coherence

۳-۱ استفاده از فیلتر وینر به عنوان پیش پردازش

سیگنال های دریافتی در محیط نویزی را می توان به صورت زیر فرض کرد:

$$x_1(m) = s_1(m) + n_1(m) \quad (5)$$

$$x_2(m) = s_2(m) + n_2(m)$$

رابطه بین  $s_1(m)$  و  $s_2(m)$  را به صورت زیر می توان نمایش داد:

$$s_2(m) = s_1(m) * h(m) \quad (6)$$

که در آن  $h(m)$  پاسخ ضربه اتاق می باشد.

در صورتیکه تبدیل فوریه  $x_1(m)$ ,  $s_1(m)$ ,  $a_1(m)$ ,  $x_2(m)$ ,  $n_2(m)$  و  $h(m)$  را به ترتیب با  $X_1(f)$ ,  $S_1(f)$ ,  $N_1(f)$ ,  $X_2(f)$ ,  $N_2(f)$  و  $H(f)$  نشان دهیم، خواهیم داشت:

$$X_1(f) = S_1(f) + N_1(f) \quad (7)$$

$$X_2(f) = S_1(f) H(f) + N_2(f)$$

یکی از محدودیتهای موجود در جداسازی منابع، تعداد منابعی است، به این صورت که تعداد منابع نباید از تعداد میکروفونها بیشتر باشد. برای رفع این مشکل قبل از اقدام به جداسازی منبع صدا، تعداد منابع نویزی را کاهش و نسبت سیگنال به نویز را افزایش می دهیم. به این منظور به عنوان مرحله پیش پردازش یک فیلتر وینر تک میکروفونه روی هر کانال با هدف حذف نویز اعمال می شود. با اعمال این فیلتر به هریک از دو سیگنال مخلوط داریم (برای سادگی  $f$  نوشته نشده است):

$$\hat{S}_1 = |X_1| \times \frac{P_{S1}}{P_{S1} + P_{N1}} = (S_1 + N_1) \times \frac{P_{S1}}{P_{S1} + P_{N1}} = \frac{P_{S1}}{P_{S1} + P_{N1}} \times S_1 + \frac{P_{S1}}{P_{S1} + P_{N1}} \times N_1 \quad (8)$$

$$\hat{S}_2 = |X_2| \times \frac{P_{S2}}{P_{S2} + P_{N2}} = (S_2 + N_2) \times \frac{P_{S2}}{P_{S2} + P_{N2}} = \frac{P_{S2}}{P_{S2} + P_{N2}} \times S_2 + \frac{P_{S2}}{P_{S2} + P_{N2}} \times N_2 \quad (9)$$

در این رابطه  $P_S$  و  $P_N$  به ترتیب طیف توان سیگنال تمیز و طیف توان نویز تخمین زده می باشند. با استفاده از تعریف رابطه زیر

$$\alpha = \frac{P_{S1}}{P_{S1} + P_{N1}} \quad \beta = \frac{P_{S2}}{P_{S2} + P_{N2}} \quad (10)$$

فاصله از گوینده ها سیگنالها را با دامنه های متفاوت و با اختلاف زمانی دریافت می کنند. سیگنال های دریافتی با  $x_1(t)$  و  $x_2(t)$  نشان داده می شوند. در مورد سیگنال های دریافتی می توان نوشت:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \quad (1)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \quad (2)$$

در این رابطه  $s_1(t)$  و  $s_2(t)$  منابع صوت و  $a_{11}$ ,  $a_{12}$ ,  $a_{21}$  و  $a_{22}$  پارامترهایی هستند که مقدار آنها به فاصله میکروفونها از گویندگان بستگی دارد.

هدف از جداسازی، تخمین سیگنالهای صحبت اصلی  $s_1(t)$  و  $s_2(t)$  با استفاده از سیگنالهای ضبط شده  $x_1(t)$  و  $x_2(t)$  می باشد. یکی از روشهای حل این مسئله جداسازی کور منابع است، تاکنون روشها و الگوریتمهای متفاوتی برای حل مسئله جداسازی کور منابع تحت شرایط و مدلهای مختلف در حوزه های گوناگون پیشنهاد و معرفی شده اند، مدلهایی را که براساس نوع سیستم ترکیب کننده تقسیم بندی شده اند، می توان با توجه به ویژگی های زیر تقسیم بندی کرد [1]:

- خطی یا غیرخطی

- تغییرپذیر یا تغییرناپذیر با زمان

- حافظه دار یا بدون حافظه (لحظه ای)<sup>۲</sup>

سیگنالهای مخلوط، خطی تغییرناپذیر با زمان (LTI) و بدون حافظه به شکل زیر در نظر گرفته می شود:

$$X(t) = A.S(t) \quad (3)$$

$$x_j(t) = \sum_{i=1}^d a_{ij} s_i(t) \quad j = 1, \dots, m$$

که در آن  $A = [a_{ij}]$  یک ماتریس ثابت می باشد. مدل واقعی تر مدل کونوالو کننده می باشد که در آن سیگنال مشاهده شده در هر گیرنده را می توان بفرم زیر نمایش داد:

$$x_j(t) = \sum_{i=1}^d a_{ij}(t) * s_i(t) \quad (4)$$

جداسازی مخلوط کانوالو شده سیگنالها می تواند در هریک از حوزه زمان یا فرکانس انجام شود در این مقاله جداسازی منابع در دامنه زمان انجام شده است.

۳- جداسازی کور منابع در حوزه زمان

<sup>2</sup> -Memory less (Instantaneous)

آن را MTP (Maximize Temporal Predictability) نامیده‌ایم، - استفاده خواهیم نمود.

فرض کنید دو فیلتر داریم که یکی از آنها تغییرات سریع یا تغییرات کوتاه مدت سیگنال،  $x_S$  و دیگری میانگین یا تغییرات بلند مدت سیگنال،  $x_L$  را نشان می‌دهند. پردازش دوره کوتاه، تغییرات سریع را در خروجی نشان می‌دهد، درحالیکه پردازش دوره بلند، میانگین تغییرات را در خروجی نشان می‌دهد. اگر بخواهیم از روابط ریاضی استفاده کنیم، خروجی فیلتر پائین گذر کوتاه مدت برای سیگنال  $x$  در زمان  $\tau$  برابر است با:

$$\tilde{x}_\tau = \lambda_S \tilde{x}_{(\tau-1)} + (1 - \lambda_S) x_\tau \quad 0 \leq \lambda_S < 1 \quad (15)$$

خروجی فیلتر پائین گذر بلند مدت برای سیگنال  $x$  در زمان  $\tau$  برابر است با:

$$\bar{x}_\tau = \lambda_L \bar{x}_{(\tau-1)} + (1 - \lambda_L) x_\tau \quad 0 < \lambda_L \leq 1 \quad (16)$$

بطوریکه

$$\lambda_L \rightarrow 1, \lambda_S \rightarrow 0 \quad (17)$$

اگر  $V$  و  $U$  به صورت زیر تعریف شوند [۳]:

$$V = \sum_{\tau=1}^n (\bar{x}_\tau - x_\tau)^2 \quad (18)$$

$$U = \sum_{\tau=1}^n (\tilde{x}_\tau - x_\tau)^2$$

آنگاه:

$$\frac{V}{U} = \frac{\sum_{\tau=1}^n (\bar{x}_\tau - x_\tau)^2}{\sum_{\tau=1}^n (\tilde{x}_\tau - x_\tau)^2} = \frac{C + \sum_{\tau=1}^n x_\tau^2}{\sum_{\tau=1}^n \Delta x_\tau^2} \quad (19)$$

و با فرض داشتن میانگین صفر برای  $x$ ، میانگین دوره بلند  $\bar{x}_\tau \approx 0$  خواهد بود و بنابراین:

$$V = \sum_{\tau=1}^n (\bar{x}_\tau - x_\tau)^2 = \sum_{\tau=1}^n x_\tau^2 \quad (20)$$

در مورد نسبت  $V$  به  $U$  می‌توان نوشت:

$$F_{LOW\_SNR} = \frac{V}{U} = \frac{\sum_{\tau=1}^n x_\tau^2}{\sum_{\tau=1}^n \Delta x_\tau^2} \quad (21)$$

$$F_{High\_SNR} = \frac{V}{U} = \frac{\sum_{\tau=1}^n x_\tau^2}{\varepsilon}$$

و با فرض آنکه تابع تبدیل بین دو سیگنال،  $H$ ، قابل محاسبه است، و با

$$\text{نامگذاری } PRE_1 = \hat{S}_1 \text{ و } PRE_2 = \frac{\hat{S}_2}{H} \text{ داریم:}$$

$$\begin{aligned} PRE_1 &= \alpha \times S_1 + \alpha \times N_1 \\ PRE_2 &= \beta \times S_1 + \frac{\beta}{H} \times N \end{aligned} \quad (11)$$

$$\begin{aligned} PRE_1 &= S_1 + d S_1 + \alpha \times N_1 \\ PRE_2 &= S_1 + d' S_1 + \frac{\beta}{H} \times N_2 \end{aligned} \quad (12)$$

$d'$  و  $d$  نشان دهنده مقدار حذف نویز فیلتر وینر هستند هرچقدر میزان حذف نویز بیشتر باشد، مقادیر  $d$  و  $d'$  بزرگتر می‌شوند. در عمل برای

این الگوریتم، فیلتر وینر دوم سخت گیر<sup>۳</sup> بوده و  $\frac{\beta}{H}$  سریعتر از  $\alpha$  بسمت صفر میل می‌کند. در نتیجه  $pre_2$  و  $pre_1$  در حوزه زمان به ترکیبی خطی از  $s_1$  و اعوجاج<sup>۴</sup>  $s_1$  و مقدار کمی از نویز  $n_1$  تبدیل می‌شود:

$$\begin{aligned} pre_1 &= s_1 + d \cdot s_1 + \varepsilon \cdot n_1 \\ pre_2 &= s_1 + d' \cdot s_1 \end{aligned} \quad (13)$$

در صورتی که این رابطه را به صورت ماتریسی نمایش دهیم، داریم:

$$\begin{bmatrix} pre_1 \\ pre_2 \end{bmatrix} = \begin{bmatrix} 1+d & \varepsilon \\ 1+d' & 0 \end{bmatrix} \cdot \begin{bmatrix} s_1 \\ n_1 \end{bmatrix} = A \cdot \begin{bmatrix} s_1 \\ n_1 \end{bmatrix} \quad (14)$$

که  $A$  ماتریس مخلوط است. با مقایسه این رابطه با رابطه (۳) می‌توان دید که این رابطه یک مساله جداسازی منابع خطی است. بنابراین با هدف جدا سازی نویز باقیمانده و سیگنال گفتار میتوان  $pre_2$  و  $pre_1$  را بعنوان ورودی به یک الگوریتم جداسازی منابع داد.

### ۳-۲ معیار قابلیت تخمین سیگنال

یکی از روشهایی که در سالهای اخیر بصورت موفق جداسازی منابع صدا از سیگنالهای مخلوط را انجام داده است، الگوریتم مرجع [۳] می‌باشد. در روش مطرح شده، هر یک از منابع سیگنال درحوزه زمان و براساس بیشینه نمودن قابلیت تخمین سیگنال از سیگنالهای مخلوط جدا می‌شوند. در این بخش ابتدا جزئیات مربوط به معیار قابلیت تخمین سیگنال را بررسی می‌کنیم و سپس از این معیار در روشی که به اختصار

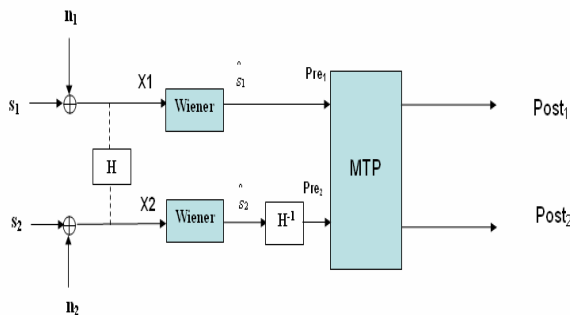
3 - strict

4 - distortion

مخلوط خطی سیگنال صحبت و نویز داشته باشند، پس آن مقدار از  $Wi$  که قابلیت تخمین سیگنال استخراج شده  $y_i$  را بیشینه کند، سیگنال منبع صحبت را بدست می‌آورد [۳].

الگوریتم MTP برای حذف نویز براساس جداسازی کور منابع در ترکیبهای خطی سیگنال صحبت و نویز بخوبی عمل می‌نماید، اما در محیطهای واقعی هنگامیکه تعداد منابع نویزی افزایش یافته و در واقع ترکیب کانولوتیوی از سیگنال صحبت و نویز وجود دارد، نتیجه خوبی ندارد. برای رفع این مشکل چنانچه در زیربخش قبل تشریح شد، باید به نحوی تعداد منابع نویزی را کاهش داده و عبارتی سطح سیگنال به نویز را افزایش دهیم و سپس از الگوریتم MTP برای جداسازی نویز باقیمانده استفاده نماییم. به نظر می‌رسد که برای کاهش نویز محیط ساده ترین و کم هزینه ترین روش استفاده از الگوریتمهای حذف نویز تک میکروفونه باشد که بدین منظور از فیلتر وینر استفاده می‌نماییم.

به این ترتیب خروجیهای مرحله پیش پردازش یعنی نتیجه  $pre_1$  و  $pre_2$  به عنوان ورودی الگوریتم MTP عمل کرده و سیگنال صحبت و نویز بطور جداگانه در خروجیهای الگوریتم MTP (با نامگذاری  $POST1$  و  $POST2$ ) ظاهر خواهند شد. بلوک دیاگرام مربوط به طرح پیشنهادی در شکل (۱) آمده است.



شکل (۱): بلوک دیاگرام سیستم بهبود سیگنال صحبت با ترکیبی از فیلتر وینر و MTP

#### ۴ - نتایج و آزمایشها

نتایج آزمایش این بخش برای گفتار پیوسته بیان شده در محیط دفتر کار گزارش شده اند. نویز های مورد استفاده نویز همهمه و اتومبیل می باشد که با نسبتهای سیگنال به نویز متفاوت به سیگنال تمیز اضافه شده اند. گویندگان در این گفتار پیوسته یک زن و یک مرد بوده اند. در کلیه آزمایشهای انجام شده از فیلتر وینر با تخمین نویز پیوسته استفاده شده است. برای نمایش میزان بهبود سیگنال صحبت در الگوریتم پیشنهاد شده نسبت به الگوریتمهای فیلتر وینر و MTP، خروجی سه الگوریتم یاد شده را با معیار PESQ سنجیدیم که نتایج آن در شکل ۲ نشان داده است.

چنانچه میزان سیگنال به نویز در سیگنال  $x$  کم باشد، آنگاه نسبت  $F_{Low\_SNR}$  بدلیل تغییرات شدید سیگنال نویزی مقدار کوچکی است. اما اگر میزان  $SNR$  زیاد باشد، با توجه به آنکه  $\Delta \bar{x}_p \rightarrow \epsilon$  نسبت  $F_{High\_SNR}$  عددی بزرگ خواهد بود، بعبارت دیگر رابطه  $F_{Low\_SNR} < F_{High\_SNR}$  برقرار است. پس می توان نتیجه گرفت که به بیشینه شدن معیار  $F = \text{Log}(U/V)$  به معنای افزایش سطح سیگنال به نویز در سیگنال است.

بنابراین با توجه به بحثها و استدلالهای انجام گرفته می‌توان معیار قابلیت تخمین را به عنوان ابزاری جهت افزایش سیگنال به نویز بکار برد.

#### ۳-۳ الگوریتم MTP

در زیربخش قبل، معیاری بنام قابلیت تخمین سیگنال  $F$ ، معرفی شد و نشان داده شد که این معیار برای سیگنالهای بدون نویز مقدار بیشتری دارد. الگوریتم MTP با استفاده از بیشینه نمودن معیار  $F$  در مرحله پس پردازش، نویز باقیمانده و اعوجاج ایجاد شده در مرحله پیش پردازش را از سیگنال صحبت جدا می‌کند. رابطه سیگنالهای نویزی بصورت زیر نشان داده می‌شود:

$$\begin{cases} x_1 = a_{11}s + a_{12}n \\ x_2 = a_{21}s + a_{22}n \end{cases} \quad (22)$$

این رابطه را بصورت زیر میتوان بازنویسی کرد:

$$\mathbf{x} = \mathbf{A} \begin{bmatrix} s \\ n \end{bmatrix} \quad (23)$$

که در آن بردار تصادفی  $X$  نشان دهنده سیگنالهای مشاهده شده و ماتریس  $A$  با عناصر  $a_{ij}$  نشان دهنده ماتریس مخلوط است. برای مطابقت نمودن این رابطه با روش نوشتاری عمومی از آنجائیکه  $s$  و  $n$  هر یک دارای منبع جداگانه هستند، از این پس هر دوی آنها را بشکل  $s_i$  نشان می‌دهیم.

بنابراین معادله (۲۲) بصورت زیر بازنویسی می‌شود:

$$x = A \cdot s \quad (24)$$

اگر سطرهای  $A$  بطور خطی مستقل باشند، در اینصورت هر منبع سیگنال  $s_i$  میتواند از  $x$  با یک ماتریس  $1 \times 2$ ، به شکل  $s_i = W_i x$  بازیابی شود. پس مسئله به پیدا نمودن ماتریس غیر مخلوط  $W = \{W_1 | W_2\}^t$  تبدیل می‌شود که هر سطر بردار  $W_i$  یک سیگنال متفاوت  $y_i$  را بازیابی می‌نماید بطوریکه  $y_i$  تخمینی از یک منبع سیگنال  $s_i$  است.

این دیدگاه برای معیار معرفی شده  $F(W_i, x)$  که برای قابلیت تخمین نسبی یک سیگنال  $y_i$  که از رابطه  $y_i = W_i x$  بازیابی شده است، استفاده می‌شود. اگر سیگنالهای منبع صحبت قابلیت تخمین بیشتری از هر

یکی از الگوریتمهای حذف نویز محیط استفاده از تابع همدوسی است که در مرجع [۴] مورد نظر قرار گرفته است. تابع همدوسی بین دو سیگنال  $x_1$  و  $x_2$  به صورت زیر تعریف می شود:

$$\Gamma_{x_1x_2}(f, n) = \frac{P_{x_1x_2}(f, n)}{\sqrt{P_{x_1x_1}(f, n)P_{x_2x_2}(f, n)}} \quad (25)$$

در این رابطه  $P_{x_1x_1}(f, n)$ ،  $P_{x_2x_2}(f, n)$  و  $P_{x_1x_2}(f, n)$  به ترتیب چگالی طیف توان  $x_1$ ، چگالی طیف توان  $x_2$  و چگالی طیف توان متقابل  $x_1$  و  $x_2$  هستند. ایده اصلی در روشهای مبتنی بر همدوسی آن است که در فرکانس- فریم های مختلف با توجه به زیاد و کم بودن همدوسی، مؤلفه فرکانسی کاهش یا افزایش یابند. به عبارت دقیقتر نزدیک بودن تابع همدوسی به یک، نشانگر نزدیک بودن دو سیگنال به همدیگر است. با توجه به وابسته بودن سیگنال گفتار دریافتی و مستقل بودن نویز دریافتی نشانگر کم بودن نویز نسبت به سیگنال گفتار است. به عکس نزدیک بودن همدوسی به صفر، نشانه کمتر مرتبط بودن دو سیگنال و یا به عبارت دیگر نویزی تر بودن سیگنال است.

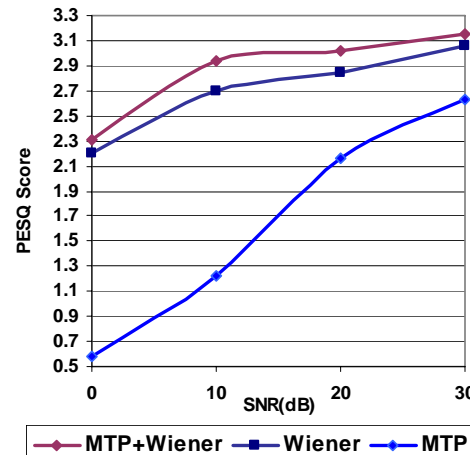
سیگنال بهبود یافته توسط الگوریتم پیشنهاد شده را با سیگنال بهبود داده شده توسط روش همدوسی با استفاده از معیارهای شنوایی و PESQ در حضور نویزهای مختلف، نویز همهمه (babble) نویز سفید (white)، نویز آگزوز اتوبوس (Volvo) و نویز ایجاد شده توسط اتومبیل سمند (car) مقایسه نمودیم. از لحاظ معیار شنوایی خروجی حاصل از روش همدوسی با وجود اینکه نویز را تا حد قابل قبولی حذف نموده بود، دارای اعوجاج بود. در صورتیکه سیگنال بهبود یافته توسط الگوریتم پیشنهاد شده ضمن کاهش نسبی بیشتر نویز، اعوجاج نیز نداشت.

نتایج خروجی دو الگوریتم از لحاظ معیار PESQ در شکل (۴) آمده است. با مقایسه نتایج مشاهده می شود که روش پیشنهاد شده از نظر این معیار به روش همدوسی ارجحیت دارد.

در تست شنوایی انجام شده از شنوندگان درخواست شد که نمرات مربوط به نویز باقیمانده و اعوجاج را به طور جداگانه اعلام کنند. شکل ۵ و ۶ به ترتیب نمرات داده شده از دیدگاه اعوجاج و نویز باقیمانده را نشان می دهند.

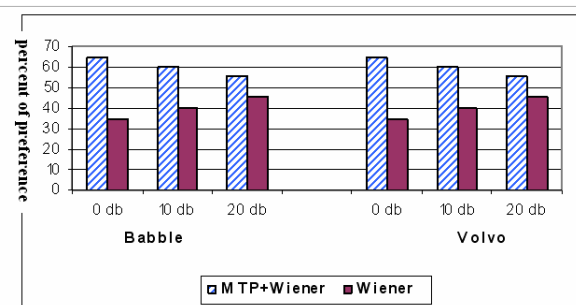
در تست دوم تست ترجیحی را براساس میزان نویز باقیمانده بر روی سیگنال بهبود یافته مقایسه کرده ایم. خروجی این تست در شکل (۶) نمایش داده شده است.

دومین معیار ارزیابی که برای الگوریتم پیشنهاد شده در نظر گرفته شده است تست شنوایی می باشد. برای معیار تست شنوایی از تست ترجیحی استفاده نموده ایم. در این روش، دو سیگنال حاصل از فیلتر وینر و روش ترکیبی وینر با MTP برای شنونده پخش می شوند



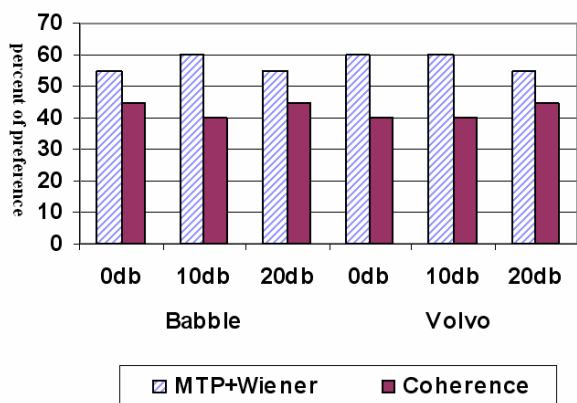
شکل (۲): مقایسه سیگنال بهبود یافته در الگوریتم پیشنهادی، فیلتر وینر و الگوریتم MTP بر حسب معیار PESQ در حضور نویز babble

. شنونده بگونه‌ای به این دو سیگنال امتیاز میدهد که مجموع امتیازها ۱۰۰ شود. هر چه عدد در نظر گرفته شده برای یک سیگنال بزرگتر باشد، یعنی شنونده آن سیگنال را از لحاظ شنوایی بیشتر ترجیح داده است. در آزمایش حاضر شنونده امتیازدهی را بر مبنای میزان اعوجاج و نویز بوجود آمده بر روی سیگنال گفتار به سیگنالهای خروجی انجام داده است. شکل (۳) نتایج تست شنوایی را وقتی که نویزهای همهمه و ولوو با سیگنال جمع شده اند، نشان می دهد. نتایج بیانگر برتری سیستم حذف نویزی است که از ترکیب دو الگوریتم بهره می‌گیرد. این مساله تایید کننده آن است که اعوجاج و نویز باقیمانده در سیگنال خروجی مرحله پیش پردازش، توسط الگوریتم MTP در مرحله پردازش ثانویه<sup>۵</sup> از بین می‌رود.



شکل (۳): درصد ترجیح سیگنالهای بهبود یافته توسط الگوریتم پیشنهاد شده و فیلتر وینر غیر ایده‌آل در حضور نویزهای Babble و Volvo

<sup>5</sup> - Post processing



شکل (۶): درصد ترجیح سیگنالهای بهبود یافته در تست دوم توسط الگوریتم پیشنهاد شده و حذف نویز مبتنی بر همدوسی در حضور نویزهای Babble و Volvo

#### ۵- نتیجه گیری

در این مقاله یک روش ترکیبی از دو روش فیلتر وینر و MTP جهت حذف نویز از سیگنال نویزی با استفاده از دو میکروفن پیشنهاد شد. الگوریتم جدید ابتدا نویز محیط را کاهش داده و سپس با جداسازی اعوجاج و نویز باقیمانده از سیگنال صحبت، کیفیت سیگنال بهبود یافته را افزایش می‌دهد.

در روش پیشنهادی از معیار قابلیت تخمین سیگنال به عنوان معیار جداسازی سیگنال صوتی و نویز پیشنهاد گردید و پس از محاسبه این معیار توانایی آن در جداسازی سیگنال و نویز نشان داده شد. روش پیشنهاد شده را با الگوریتمهای وینر، MTP و بهبود سیگنال مبتنی بر همدوسی مقایسه نمودیم. خروجی روش ها با معیارهای شنوایی و PESQ سنجیده شد و در هر دو معیار، روش پیشنهاد شده دارای امتیاز بهتری بود.

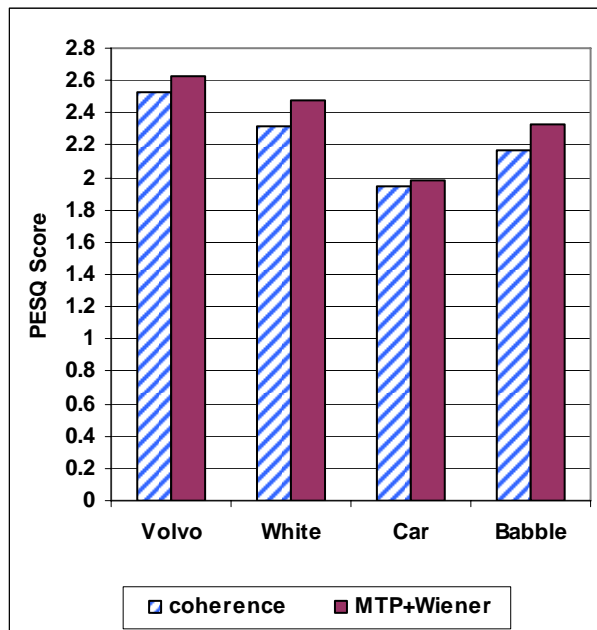
#### مراجع

[1] مسعود رضا آقا بزرگی صحاف ، "جداسازی کور منابع همبسته مشترکاً ایستان با اجزاء قابل پیش‌بینی و کانال لحظه‌ای خطی و تغییر ناپذیر با زمان " ، پایان نامه دکترای مهندسی برق - مخابرات ، دانشگاه صنعتی اصفهان ، ۱۳۸۲ .

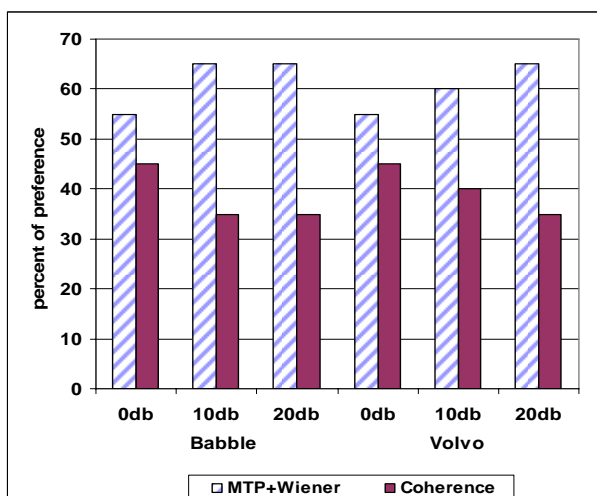
[2] J. Rosca et al. (Eds.): "Separating Underdetermined Convolutional Speech Mixtures " ICA 2006, LNCS 3889, pp. 674–681, 2006.

[3] James V. Stone ., "Blind Source Separation Using Temporal Predictability". Neural Computation 13, 1559–1574 (2001) © 2001 Massachusetts Institute of Technology

[4] R. Bouquin, A. Akbari, Azirani, G. Faucon, , "Enhancement of speech degraded by coherent and incoherent noise using a cross- spectral estimator", IEEE Transactions on Speech and



شکل (۴): مقایسه خروجیهای طرح پیشنهاد شده و الگوریتم همدوسی برای میکروفونهای ۲ و ۱ در حضور نویزهای مختلف. فاصله بین میکروفن ها 18 cm است.



شکل (۵): درصد ترجیح سیگنالهای بهبود یافته در تست اول توسط الگوریتم پیشنهاد شده و حذف نویز مبتنی بر همدوسی در حضور نویزهای Babble و Volvo

Audio Processing, Volume 5, Issue 5, Sept. 1997 Page(s):484 – 487

[5]. Pedersen, M.S., Wang, D.L., Larsen, J., Kjems, U.: "Overcomplete blind source separation by combining ICA and binary time-frequency masking" . In: Proceedings of the MLSP workshop, Mystic, CT, USA (2005)

[6] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. on ASSP*, vol. 27, no. 2, pp. 113-120, 1979

[7] X. Huang, A. Acero, and H. W. Hon, Spoken Language Processing: A Guide to Theory, Algorithm, and System Development, Prentice Hall PTR, Upper Saddle River, N.J., 2001.

[8] A. hyvarinen, E. Oja, "Independent Component Analysis: Algorithms and Applications", *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, 2000