

## بهبود کیفیت گفتار با استفاده از روش نوین ترکیبی فیلتر کالمن و فیلتر ذره‌ای

سید یاسر نصیب<sup>۱</sup>، رضا جاویدان<sup>۲</sup>

<sup>۱</sup>دانشجوی کارشناسی ارشد دانشگاه آزاد دزفول ، sy.nasib@gmail.com

<sup>۲</sup>استادیار دانشگاه صنعتی شیراز ، javidan@sutec.ac.ir

چکیده - در سال‌های اخیر روش‌های متنوعی برای بهبود کیفیت گفتار پیشنهاد شده است در میان این روش‌ها، روش فیلتر کالمن و فیلتر ذره‌ای به دلایل متعددی از جمله توانایی بالا در حذف نویز و سرعت بالای الگوریتم از اهمیت زیادی برخوردار هستند. مهمترین چالش در روش کالمن و نیز فیلتر ذره‌ای، نداشتن تخمین درست از پارامترهای صوت و نویز می‌باشد. که به دلیل تخمین نادقیق مقداری نویز باقی می‌ماند. در این مقاله یک فیلتر ذره‌ای بر اساس جداسازی فریم‌های صوتی از غیرصوتی و نسبت مقادیر متفاوت تعداد ذره و مرتبه برازش خودکار به این فریم‌ها ارائه شده است که حجم محاسبات را به شدت کاهش داده، در حالی که کیفیت به دست آمده به صورت محسوس تغییر نمی‌کند. در این مقاله، از فیلترهای کالمن و ذره‌ای برای کاهش نویزهای صوتی مخصوصاً غیرسفید استفاده شده است. نتایج شبیه سازی، کارایی روش پیشنهادی را نشان می‌دهد.

کلید واژه- بهبود گفتار، فیلتر کالمن، فیلتر ذره‌ای و نویز سفید.

کاربردهای متنوع بهسازی گفتار، اهداف متفاوتی برای فرایند بهسازی ایجاد کرده است.

### ۱- مقدمه

تکنیک‌های مختلفی برای بهبود سیگنال گفتار وجود دارد. سیستم‌های تک‌کاناله، متداول‌ترین نوع الگوریتم‌های بلادرنگ می‌باشد چرا که به سادگی قابل پیاده‌سازی هستند و به طور نسبی هزینه آنها کمتر از سیستم‌هایی با چند کانال ورودی می‌باشد [۴]. در میان روش‌های تک‌کاناله حذف نویز، روش فیلتر کالمن و فیلتر ذره‌ای به دلایل متعددی از جمله توانایی بالا در حذف نویز و سرعت بالای الگوریتم از اهمیت زیادی برخوردار هستند [۵]. مهمترین چالش در روش کالمن و نیز فیلتر ذره‌ای، نداشتن تخمین درست از پارامترهای صوت و نویز می‌باشد. که دلیل تخمین نادقیق مقداری نویز باقی می‌ماند [۶].

فیلتر ذره‌ای به دسته ای از فیلترها اطلاق می‌گردد که در آن‌ها از روش‌های مونت کارلو برای تخمین توزیع پسین استفاده می‌شود در واقع فیلتر ذره ای دسته از الگوریتم های مونت کارلو دنباله‌ای محسوب می‌شود. ساختار فیلترهای ذره‌ای در زمان حاضر شباهت زیادی به فیلتر های بوت استرپ دارد. ایده اولیه چنین روش هایی به سال های قبل و دهه های ۱۹۶۰ و ۱۹۷۰ میلادی بر می‌گردد. اما در آن زمان به دلیل نبودن قدرت

یکی از موضوعات مهم پردازش سیگنال گفتار، کاهش و حذف نویز ناخواسته از سیگنال اصلی و بهبود گفتار است [۱]. بهسازی گفتار به فرایند بازبایی گفتار تمیز از سیگنال نویزی گفته می‌شود و در مواردی به کار می‌رود که سیگنال گفتار تحت تاثیر نویز، انعکاس یا سایر عوامل تخریب قرار گرفته است. فرایند بازبایی می‌تواند به شکل بازبایی خود سیگنال گفتار برای افزایش سطح ادراک در سیستم شنوایی انسان یا به شکل بهسازی ویژگی‌های استخراج شده از گفتار برای استفاده در سیستم‌های بازشناسی گفتار یا گوینده مقاوم در برابر نویز باشد [۲]. اعوجاج می‌تواند ناشی از نویز جمع شونده محیطی یا حاصل از اعوجاج خطی یا غیرخطی در مسیر انتقال باشد. ورود نویز به سیستم ممکن است در محل منبع سیگنال، در مسیر مخابره و یا در محل گیرنده رخ دهد؛ عوامل تخریبی وارد شونده در مسئله می‌تواند ترکیبی از این سه حالت باشد؛ نویز ممکن است با سیگنال اصلی جمع، ضرب و حتی در حالت کلی پیچیده شود و ممکن است نویز مستقل و یا وابسته به سیگنال اصلی باشد [۳].

در سال ۲۰۱۴ آقای چان و همکارش از یک مدل چند کاناله برای بهبود کیفیت گفتار استفاده کرد، در این کار از مفهوم پردازش آرایه‌ای استفاده نمود و نشان داد به پیدا کردن مکان منبع اصلی می‌توان نویزها را از صدای بدست آمده بهتر حذف کرد [۸]. در سال ۲۰۱۴ آقای مشتاق و همکارش با ترکیب مدل مخفی مارکوف و فیلترهای ذره‌ای یک روش جدید برای بهبود کیفیت گفتار ارائه داد. در این کار ابتدا یک مدل دینامیکی برای سیگنال صحبت تمیز و سیگنال نویز بدست آمده است و سپس به کمک فیلتر ذره‌ای سیگنال صحبت ردیابی می‌شود و توزیع نویز به کمک مدل مخفی مارکوف محاسبه می‌شود. در این مقاله یک راهکار بهینه مبتنی بر جداسازی فریم‌های صدا دار و بی‌صدا پیشنهاد شده است که به خوبی نویز باند پهن و هارمونیک‌های اصلی فرکانس پایین غیر ایستای آنها را حذف می‌کند و سیگنال پردازش شده دارای حداقل اعوجاج و نویز خواهد شد. این روش در مقایسه با سایر روش‌های مرسوم تخمین گفتار تمیز از کارایی بسیار بالاتری برخوردار است. برای این منظور ابتدا فیلتر کالمن و فیلتر ذره‌ای به طور مختصر توضیح داده شده است و در ادامه روش پیشنهادی ارائه می‌گردد و در نهایت نتایج شبیه‌سازی روش پیشنهادی نشان داده شده است.

### ۳- تئوری فیلتر کالمن برای کاهش نویز گفتار

فیلتر کالمن یک فیلتر تطبیقی حداقل کننده مجذور خطا می‌باشد که یک راه حل بازگشتی محاسباتی بهینه برای تخمین سیگنال با فرض نویز گوسی می‌دهد. این الگوریتم استفاده بهینه از اطلاعات غیردقیق در یک سیستم خطی یا غیرخطی که سیستم دارای خطای گوسی است، به صورت پیوسته سیستم را به روزرسانی می‌کند و بهترین  $X(k)$  برای حالت فعلی سیستم بهترین تخمین را می‌زند. تئوری فیلتر کالمن برپایه روش فضای حالت که در آن یک معادله حالت به عنوان مدل‌های دینامیکی تولید سیگنال می‌باشند و معادله مشاهده که  $Y(k)$  سیگنال مشاهده شده و نویز را مدل می‌کند عمل می‌کند. برای سیگنال  $X(k)$  که  $Y(k)$  سیگنال نویزی آن است، معادله فضای حالت مدل شده و مدل مشاهده شده به صورت زیر تعریف می‌شود:

محاسباتی لازم و کامپیوترهای قوی قابلیت چنین روش‌هایی از نظر دور مانده بود. با ساخت کامپیوترهای قوی و ارزان در دو دهه اخیر این روش‌ها مورد توجه بسیاری از محققان در زمینه‌های مختلف قرار گرفته است. در نتیجه این تحقیقات پیشرفت‌های زیادی در این زمینه بدست آمده که نتایج روش‌های اولیه را بسیار بهبود داده‌اند.

هدف از این مقاله، کاهش و حذف انواع نویز مخصوصاً نویزهای رنگی با استفاده از فیلترهای کالمن و ذره‌ای می‌باشد. این نویزها اغلب شامل یک بخش باند پهن و تعداد محدودی از سیگنال‌های سینوسی غالب با بیشترین انرژی نویز در فرکانس‌های پایین می‌باشد که به شدت کیفیت سیگنال گفتاری را پایین می‌آورد. در این مقاله یک راهکار بهینه مبتنی بر جداسازی فریم‌های صدا دار و بی‌صدا پیشنهاد شده است که به خوبی نویز باند پهن و هارمونیک‌های اصلی فرکانس پایین غیر ایستای آنها را حذف می‌کند و سیگنال پردازش شده دارای حداقل اعوجاج و نویز خواهد شد. این روش در مقایسه با سایر روش‌های مرسوم تخمین گفتار از کارایی بسیار بالاتری برخوردار است.

### ۲- فیلتر ذره‌ای

به دسته‌ای از فیلترها اطلاق می‌گردد که در آن‌ها از روش‌های مونت کارلو برای تخمین توزیع پسین استفاده می‌شود در واقع فیلتر ذره‌ای دسته‌ای از الگوریتم‌های مونت کارلو دنباله‌ای محسوب می‌شود [۷]. ساختار فیلترهای ذره‌ای در زمان حاضر شباهت زیادی به فیلترهای بوت استرپ دارد. لازم به ذکر است که ایده اولیه چنین روش‌هایی به سال‌های قبل و دهه‌های ۱۹۶۰ و ۱۹۷۰ میلادی بر می‌گردد. اما در آن زمان به دلیل نبودن قدرت محاسباتی لازم و کامپیوترهای قوی قابلیت چنین روش‌هایی از نظر دور مانده بود. با ساخت کامپیوترهای قوی و ارزان در دو دهه اخیر این روش‌ها مورد توجه بسیاری از محققان در زمینه‌های مختلف قرار گرفته است. در نتیجه بسیار بهبود داده‌اند. هدف بسیاری از این تحقیقات اخیر نیز ارائه قضایا و شروط همگرایی برای این دسته از روش‌ها بوده است. قدرت این روش‌ها به توانایی آن‌ها در توصیف مدل‌های غیرخطی-ناگوسی مربوط می‌شود؛ سادگی پیاده‌سازی آن‌ها نیز از دیگر محبوبیت این دسته از روش‌ها می‌باشد.

شامل یک کسر است که فرم تخمین از مقادیر قبلی می باشد و شامل اطلاعات جدید و نه مشاهدات جدید نمی باشد.

استفاده از فیلتر کالمن در بهبود گفتار در فرم امروزی ابتدا به وسیله پالیوال معرفی شد. این روش مناسب ترین روش برای حذف نویز سفید با توجه به فروض فیلتر کالمن می باشد. در مشتقات معادلات کالمن، فرض می شود که نویز فرآیند ناهمبسته است و یک توزیع نرمال دارد. این فرض به این ختم می شود که نویز خاصیت سفید بودن داشته باشد. اگر چه روش های متفاوت از کالمن برای حذف نویز رنگی نیز ارائه شدند [۸]. فرض می کنیم که سیگنال گفتار در طول هر فریم ایستان است. بنابراین مدل گفتار در طول هر گفتار به همان صورت باقی می ماند. برای پردازش تک بعدی سیگنال گفتار با استفاده از مدل فضای حالت، بردار حالت را تعریف می کنیم:

$$x(k) = [x(k-p+1)x(k-p+2)x(k-p+x(k))] \quad (7)$$

که  $x(k)$  سیگنال گفتار در زمان  $k$  است. سیگنال گفتار به وسیله نویز سفید  $n(k)$  آلوده می شود.

$$y(k) = x(k) + n(k) \quad (8)$$

سیگنال گفتار را می توان با استفاده از فرایند AR با مرتبه  $p$  مدل نمود.

$$x(k) = \sum a_i x(k-i) + u(k) \quad i = 1 \dots p \quad (9)$$

که  $a_i$  ضرایب و  $u(k)$  خطا تخمین است.  $n(k)$  دارای توزیع گوسی می باشد. روش های زیادی برای بدست آوردن پارامترهای مدل از اطلاعات نویزی وجود دارد. این پارامترها اگر چه فرض می شوند که از قبل داده شده اند. روش های دیگر سعی می کنند که ابتدا پارامترهای مدل را محاسبه کنند و سپس برای حذف نویز سیگنال گفتار از آنها استفاده کنند. روش های دیگری وجود دارند که به صورت متوالی تخمین می زنند و این مقادیر را تصحیح می کنند و به این وسیله گفتار بهبود می یابد. یک بلوک پیش تمیز کننده برای استخراج تخمین این مقادیر استفاده می -

$$X(k) = AX(k-1) + W(k-1) \quad (1)$$

$$Y(k) = HX(k) + n(k) \quad (2)$$

که  $X(k)$  بردار سیگنال با ابعاد  $P$  می باشد که به آن پارامتر حالت در زمان  $k$  نیز می گویند.  $A$  ماتریس حالت گذار  $P \times P$  می باشد که نسبت حالت های فرآیند در زمان های  $K$  و  $K-1$  می باشد.  $W(k)$  (نویز فرآیند) ورودی که بردار تحریک ورودی با ابعاد  $P$  در معادله تحریک می باشد. فرض می شود  $W(k)$  یک فرآیند (گوسی) نرمال می باشد.  $Q$  ماتریس  $P \times P$  کواریانس  $W(k)$  یا کواریانس نویز فرآیند می باشد.  $Y(k)$  بردار مشاهده نویزی با ابعاد  $M$  می باشد و  $H$  نیز ماتریس  $M \times P$  می باشد.  $n(k)$  بردار نویز به ابعاد  $M$  که به آن نویز اندازه گیری می گویند. فرض می شود که  $n(k)$  توزیع نرمال داشته که  $R$  ماتریس کواریانس برای ماتریس  $n(k)$  می باشد. ما  $X(k|k-1)$  را به عنوان تخمین قیاسی در زمان  $k$  که از زمان های قبل تر شروع می شود، تعریف می کنیم.  $X(k|k)$  را تخمین حالت استقرایی می گوئیم که در زمان  $k$  اندازه  $Y(k)$  را به ما می دهد. خاطرنشان می شویم که  $X(k|k-1)$  (1) تخمین مقدار  $X(k)$  بر پایه مقادیر قبلی و نه مقدار مشاهداتی فعلی در زمان  $k$  می باشد.  $X(k|k)$  در سمت دیگر از اطلاعات در زمان حال مشاهده استفاده می کند. یک تخمین خطای استقرایی و قیاسی به صورت زیر تعریف می شود:

$$e^-(k) = X(k) - X(k|k-1) \quad (3)$$

$$e(k) = X(k) - X(k|k) \quad (4)$$

کواریانس خطا تخمین قیاسی به صورت زیر است:

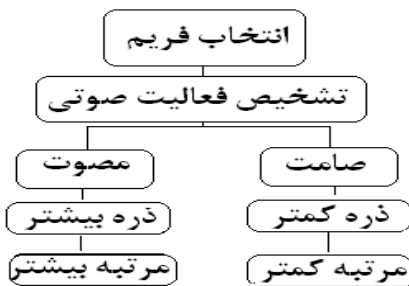
$$P^-(k) = E\{e^-(k)e^{-T}(k)\} \quad (5)$$

کواریانس خطا تخمین استقرایی به صورت زیر است:

$$P(k) = E\{e(k)e^T(k)\} \quad (6)$$

در فرمول نتیجه گرفته شده از فیلتر کالمن، هدف یافتن یک معادله است که تخمین حالت استقرایی را محاسبه می کند. به صورتی که مجموعه خطی از تخمین های قیاسی و تفاوت وزن بین اندازه گیری اصلی و اندازه گیری تخمین باشد. بنابراین تخمینی

الگوریتم می‌شوند. فرض ما این است که این ضرایب سبب پایداری سیستم می‌شوند. از سمت دیگر ما همیشه برای برازش مدل نویز از تعداد کمی ضریب استفاده می‌کنیم (۲ یا ۳) زیرا این تعداد ضریب برای تخمین نویز ما کافی هست و به تعداد بیشتری ضریب نیاز نداریم. حال برای بهینه سازی میزان محاسبات از این اصل استفاده می‌کنیم که می‌توان قسمت‌های صوتی و غیرصوتی یک گفتار را از هم تفکیک نمود. حال به قسمت‌های صوتی مرتبه برازش بالاتر اختصاص داده و برای قسمت‌های غیرصوتی و یا سکوت میزان این ضرایب را برابر تعداد ضرایب اختصاص داده به نویز در نظر می‌گیریم. با احتساب این عمل و نیز وارد کردن این اصل که برای مدل کردن نویز نیاز به تعداد کمی ضریب داریم روش RBPF را از لحاظ حجم محاسبات بهینه می‌کنیم. ابتدا کل گفتار را به قسمت های ۲۵ میلی ثانیه ای (۲۰۰ نمونه‌ای) تقسیم نموده و با استفاده از الگوریتم‌های تشخیص واج‌های صوتی از غیرصوتی وضعیت هر قسمت را تعیین نموده و سپس با توجه به وضعیت هر قسمت تعداد ضرایب AR و تعداد ذرات را مشخص می‌کنیم.



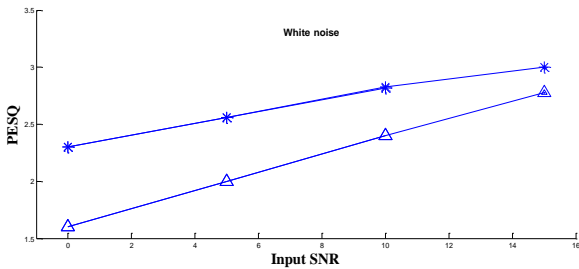
شکل ۱: نمودار روش پیشنهادی برای فیلتر ذره‌ای.

شکل ۱ نحوه عملکرد این الگوریتم پیشنهادی را نشان می‌دهد. ما در این الگوریتم به هیچ عنوان به دنبال بهبود کیفیت گفتار نیستیم و تنها به دنبال این هستیم که با کاهش حجم محاسبات این روش را از لحاظ زمان محاسبه بهینه‌تر کنیم. یک مدل GAR به کار برده می‌شود که قسمت جدید آن از توزیع معمولی نمایشی پیروی کرده و مشخصات غیرگوسی را منعکس می‌کند. در عوض مدل در این مورد نتیجه‌گیری احتمال بسیار سخت می‌شود. توزیع قیاسی حول متغیرهای مخفی به وسیله کنار هم گذاشتن فیلتر ذره‌ای و فیلتر کالمن حاصل می‌شود. استنتاج احتمال به وسیله RBPF حاصل می‌شود. همچنین ما یک EM

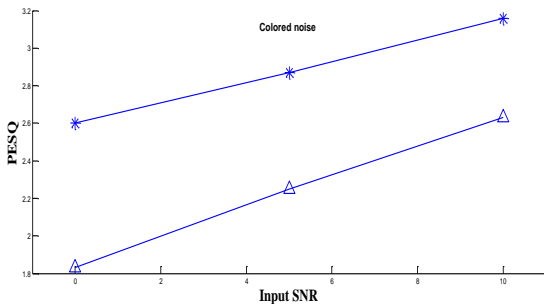
شود (مثل روش تفریق طیفی). مقدار اولیه از اطلاعات نویزی استفاده می‌شود که یک ماتریس کواریانس خطای تخمین استقرایی به صورت قطری که مقدار قطر اصلی آن است، تشکیل می‌شود. ضرایب برای قسمت‌هایی که هم‌پوشانی دارند و یا ممکن است هم‌پوشانی نداشته باشند، محاسبه می‌شود. در انتها باید مواظب بود که پیوستگی پارامترهای فیلتر حفظ شود (به عنوان مثال مطمئن شد که پارامترهای فیلتر ذخیره شده در نیمه فریم باشند). در هم‌پوشانی استفاده از فیلتر کالمن تاخیر یافته نتایج بهتری از لحاظ عملکرد نسبت به فیلتر کالمن به صورت همزمان دارد.

#### ۴- روش پیشنهادی

محاسبات در فیلتر ذره‌ای ارتباط مستقیم با تعداد ذرات و نیز مرتبه ضرایب برازش دارد. به این معنا که اندازه ماتریس گذر حالت و همچنین بقیه ماتریس‌ها تابع مجموع مرتبه ضرایب AR هستند. درجه پیچیدگی مسئله را نیز همین مشخص می‌کند یعنی اگر یک ماتریس به اندازه  $N \times N$  داشته باشیم میزان محاسبات  $N^2$  و اگر ابعاد ماتریس را به اندازه  $1$  درایه در هر بعد افزایش دهیم، این میزان محاسبه برای به اندازه  $(N+1)^2$  می‌شود. ذرات نیز رابطی مستقیم با میزان محاسبات دارد. بسته به اینکه ما چه پارامترهایی را در فیلتر ذره‌ای پردازش کنیم، میزان محاسبات ما تغییرات زیادی می‌کند. به عنوان مثال اگر تنها ضرایب برازش خودکار و واریانس نویز تحریک مورد بررسی قرار گیرد میزان پردازش به اندازه  $M \times 2$  که  $M$  تعداد ذرات ما است لحاظ می‌شود. در هر پله زمانی این مقدار محاسبه باید انجام شود. حال اگر پارامترهای دیگری نیز به پارامترهای ما اضافه شود، میزان پردازش نیز به صورت خطی افزایش می‌یابد. مسئله اصلی در مورد مرتبه ضرایب برازش این است که هر چه این مرتبه افزایش یابد، میزان پایداری نیز کاهش می‌یابد. به زبان ساده‌تر در هر ذره ضریب AR را باید کمی تغییر دهیم و این میزان تغییر را می‌توان به عنوان یک نویز گوسی و یا غیر گوسی در نظر گرفت. حال اگر تعداد ضرایب زیاد باشد مثل ۱۵ یا ۱۲ باید پس از هر جانشانی باید حتماً تست کرد که آیا این ضرایب، تابع تمام قطب را پایدار نگه می‌دارد و یا اینکه از حالت پایدار به سمت ناپایداری پیش برده و سبب تخمین نادرست در



شکل ۲: مقایسه بین حالت نویزی و ایده‌آل در برابر نویز سفید (\* حالت ایده‌آل و Δ برای حالت نویزی).



شکل ۳: مقایسه بین حالت نویزی و ایده‌آل در برابر نویز رنگی (\* حالت ایده‌آل و Δ برای حالت نویزی).

در ادامه به سراغ مقایسه ۳ روش اصلی که در پی بررسی آن هستیم می‌رویم. برای روش KEM نیز تعداد تکرار را برابر ۲ و ضرایب AR گفتار را برابر ۱۳ در نظر می‌گیریم. نمودارهای زیر برای ۳ نوع نویز مورد نظر به صورت زیر به دست می‌آید. نمودارهایی که در ادامه می‌آیند روش کالمن متوالی را با کالمن ترتیبی و نیز RBPF مقایسه کرده و معیار آنها نیز در این میان PESQ می‌باشد. کمترین اثر بخشی در میان نویزها مربوط به نویز صنعتی می‌باشد زیرا از لحاظ فرکانسی تداخل فرکانسی بالایی با گفتار انسان دارد و از دید زمانی نیز به سبب اعوجاج زیادی میشوند.

روش پیشنهادی ما در برابر نویز سفید در SNR های پایینتر نسبت به روش های دیگر کارایی بهتری دارد و همچنین مزیت اصلی روش پیشنهادی کاهش زمان و هزینه محاسبات نسبت به دیگر روش ها می باشد.

متوالی تحمیلی پیشنهاد می‌کنیم که پارامترهای مدل را به صورت بازگشتی با استفاده از فشار مثبت برای تعدادی از پارامترها بدست می‌آورد. این کار سبب از کار درآمدن EM تحمیلی متوالی با استفاده از فیلتر کالمن می‌شود که در KGSD انجام شده است. هر لحظه از زمان، قدم E محاسبه زیر را انجام می‌دهد.

$$\mathbb{E}\{f(X_t)|Y_{1:t}\} = \int f(X_t)p(X_t|Y_{1:t})dX_t \quad (9)$$

که  $f(X_t)$  یک تابع آماری تصادفی است و می‌تواند  $f(X_t) = \log p(X_t; \theta)$  در مسئله ما باشد. محاسبه احتمال با فرض غیرگوسی بودن گفتار کاری سخت است. ما از فیلتر ذره‌ای رانو بلک ولیدز برای محاسبه متوالی رابطه (۱۰) استفاده می‌کنیم. خاطر نشان می‌شویم که  $X_t$  حالت شامل حالت گفتار  $s_t$  و نیز حالت نویز  $n_t$  می‌باشد.

یک راه حل جبری برای تخمین حالت نویز دارد. رانوبلک ویدز برای نمونه، برخی از متغیرها را حاشیه‌بندی می‌کند تنها نمونه‌هایی را می‌گیرد که از متغیرهایی باقی مانده‌اند (حالت گفتار) پس ما می‌توانیم ناحیه نمونه برداری را کوچک کرده و کیفیت تخمین را افزایش دهیم.

## ۵- نتایج شبیه سازی

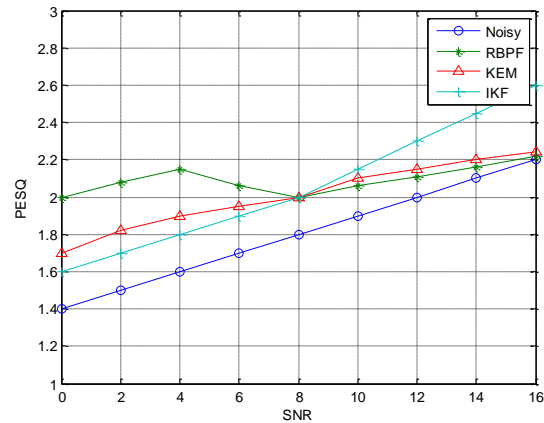
در این قسمت الگوریتم پیشنهادی را بر روی صداهایی که از پایگاه داده NOUZEOUSE انتخاب شده است اعمال می‌کنیم و نتایج حاصل را با یکدیگر مقایسه می‌کنیم. در ابتدا برای فیلتر کالمن متوالی ۲ حالت مختلف فرض کردیم حالت اول که آن را نویزی نام نهادیم این بود که پارامترهای گفتار با استفاده از سیگنال آلوده به نویز و نیز تخمین پارامترهای نویز با استفاده از قسمت‌های سکوت به دست آید. حالت دوم حالتی بود که پارامترهای سیگنال تمیز را با استفاده از ضرایب AR خود سیگنال تمیز به فیلتر کالمن دهیم. در زیر با توجه به SNR ورودی و نیز نوع نویز نتایج را مشاهده می‌کنید. تعداد تکرار در هر SNR برابر ۳ مرتبه می‌باشد. مرتبه ضرایب AR را برای گفتار ۱۳ در نظر می‌گیریم.

احتمال غیرگوسی را نیز پشتیبانی کند زیرا از یک روش نمونه- برداری اهمیت متوالی برخوردار است و احتمال آماری را نسبت به توزیع احتمال های پیچیده محاسبه می کند. اخیراً فیلتر ذره ای برای بهبود گفتار نیز استفاده شده است

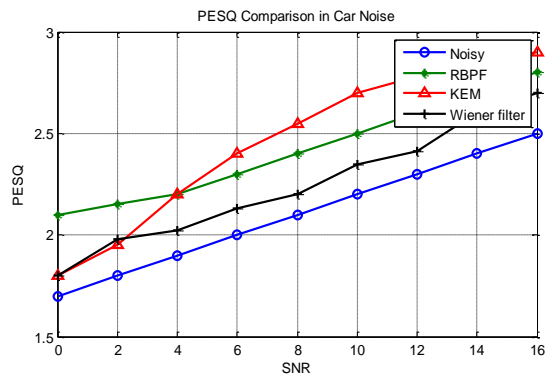
در این مقاله برای حل مسئله بهبود گفتار، سیگنال گفتار براساس استنتاج احتمال فرموله شده است و یک مدل فضای حالت استفاده می شود. در این چهارچوب وظیفه بهبود گفتار متوالی شامل مجتمع سازی بر پایه توزیع معین حول متغیرهای مخفی (حالت های برداری) بوده و یک سیگنال مشاهده یا اندازه گیری را دریافت می کند. در عوض بیشتر روش های موجود ما از مدل عمومی شده برازش خودکار AR برای بررسی ساخت گفتار استفاده می کند، که مشخصات غیرگوسی را در برگیرد. فرض می کنیم نویز رنگی یک سیگنال تصادفی گوسی است و مدل پردازش خودکار معمولی را به علاوه یک قسمت گوسی در نظر می گیرد. همچنین رانویک ویدز برای توزیع احتمال قیاسی حول متغیرهای مخفی به کار می رود که به ۲ قسمت تجزیه می شوند. یکی که به وسیله فیلتر کالمن محاسبه آماری می شود و دیگری که به وسیله فیلتر ذره ای تخمین زده می شود. در این مقاله از فیلتر رانویک ویدز برای استنتاج احتمال استفاده شده است که محاسبات متوالی مجتمع سازی نسبت به توزیع احتمال قیاسی را انجام می دهد. روش پیشنهادی ما در برابر نویز سفید در SNR های پایینتر نسبت به روش های دیگر کارایی بهتری دارد و همچنین مزیت اصلی روش پیشنهادی کاهش زمان و هزینه محاسبات نسبت به دیگر روش ها می باشد. در نهایت، نتایج شبیه سازی کارایی و موثر بودن روش پیشنهادی را در حل مسئله بهبود گفتار نشان می دهد.

## مراجع

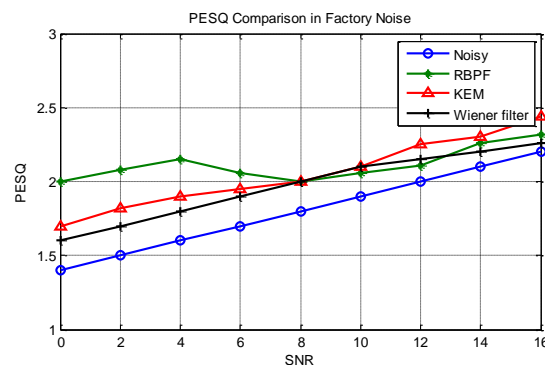
- [1] Lim, J.S. and Oppenheim, A.V., "Enhancement and bandwidth compression of noisy speech", Proceeding IEEE, Vol. 67, No.12, pp. 1586-1604, December 1979.
- [2] Virag, N., "Speech enhancement based on masking properties of the human auditory system", Master thesis, Swiss Federal Institute of Technology, 1996.
- [3] Srinivasan, S., "Knowledge-Based Speech Enhancement", Doctoral thesis, KTH-Royal Institute of Technology, Stockholm, October 2005.
- [4] Ephraim, Y., Lev-Ari, H. and Roberts, W., "A Brief Survey of Speech Enhancement", the Electronic Handbook August, 2003.



شکل ۴: مقایسه بین الگوریتم های مختلف در برابر نویز سفید



شکل ۵: مقایسه بین الگوریتم های مختلف در برابر نویز رنگی



شکل ۶: مقایسه بین الگوریتم های مختلف در برابر نویز صنعتی

## ۶- نتیجه گیری

دلیل اینکه فیلتر ذره ای به جای فیلتر کالمن استفاده شده است این است که فیلتر کالمن از لحاظ مدل کردن گفتار بر پایه توزیع گوسی محدودیت دارد؛ در عوض فیلتر ذره ای می تواند توزیع های



چهارمین کنفرانس ملی ایده های نو در مهندسی برق



۲۰۲۱ آبان ماه ۱۳۹۴ - دانشگاه آزاد اسلامی واحد اصفهان (خوراسگان)

- [5] Mustiere, F., Bouchard, Bolic, M., "Low-cost modifications of Rao-Blackwellized particle filters for improved speech denoising", Signal Processing, 2008.
- [6] Berouti, M., Schwartz, R. and Makhoul, J., "Enhancement of speech corrupted by acoustic noise", Proceeding IEEE, April 2002.
- [7] Paliwal, K. K. and Basu, A., "A speech enhancement method based on Kalman filtering", in Proceeding IEEE Speech, Signal Processing, pp. 177-180, 2010.
- [8] Gibson, J. D., Koo, B. and Gray, S. D. "Filtering of colored noise for speech enhancement and coding", IEEE Transaction Acoustics, Speech, Signal Processing, vol. 39, pp. 1732-1742, 2014.