بازشناسی مقاوم گفتار با دادگان مفقود با استفاده از روشی مبتنی بر کوراریانس

حجت محمدنژاد۱، منصور ولی۲ ایمان اسمعیلی۳
[1] دانشکده فنی دانشگاه شاهد، mohammadnejad@shahed.ac.ir
[2] دانشکده فنی دانشگاه شاهد، vali@shahed.ac.ir
دانشکده فنی دانشگاه شاهد، iesmaili@shahed.ac.ir

چکیده - سیستم‌های بازشناسی گفتار در مواجهه با گفتارهایی که در اثر نویز تخریب شده‌اند عملکرد ضعیفی از خود نشان می‌دهند. روش‌های دادگان مفقود سعی دارند که تاثیر نویز را با حذف یا تخمین مؤلفه‌های بازنمایی نامعتبر تخریب شده توسط نویز، جبران کنند و بازشناسی را با استفاده از مؤلفه‌های معتبر باقیمانده و تخمین زده شده انجام دهند. این روش ها، بردارهای ویژگی ورودی به مدل بازشناسی را بدون اینکه در سیستم بازشناسی تغییری ایجاد کنند، بهبود می‌بخشند. در این تحقیق ما یک روش جبران ویژگی را پیشنهاد خواهیم داد که در آن، مؤلفه‌های نامعتبر در نمایش اسپکتروگرام (نمایش زمانی- فرکانسی) با مؤلفه‌های تخمینی بدست آمده از مؤلفه‌های معتبر و اطلاعات آماری که نسبت به دادگان تمیز وجود دارد، جایگزین می‌شوند. در حالت کلی، روش‌های جبران سازی در اثر استخراج ویژگی‌های کپستروم از اسپکتروگرام بازسازی شده، در مواجهه با نویز غیرایستان ، نرخ بهبود زیادی در صحت بازشناسی دارند. روش‌های بازسازی دادگان مفقود از لحاظ محاسباتی کم‌هزینه بوده، چرا که نیاز به آموزش چندباره مدل بازشناسی برای دادگان نویزی و اصلاح مدل را از بین می‌برد.

کلید واژه- بازشناسی مقاوم گفتار، جبران سازی مبتنی بر ویژگی، دادگان مفقود

۱- مقدمه

صحت بازشناسی، برای گفتاری که توسط نویز تخریب شده است کاهش چشمگیری می‌یابد مخصوصا اگر سیستم بازشناسی بر روی دادگان تمیز آموزش داده شده باشد. روش های متعددی برای جبران اثرات نویز بر روی سیستمهای بازشناسی گفتار مطرح شده اند. تقریبا بیشتر آنها نویز را ایستان فرض می کنند بنابراین در بازشناسی گفتارهای نویزی غیر ایستان، کارایی نخواهند داشت[۱].

روش های مبتنی بر ویژگی های مفقود بطور خاص این نوید را می‌دهند که قابلیت جبران اثرات نویز غیر ایستان را دارا هستند. روش‌های دادگان مفقود، الگوریتم جبران سازی هر گونه از نویز را شامل می‌شوند و قابلیت خوبی در بازشناسی مقاوم گفتار در حضور سطوح بالایی از نویز، از خود نشان داده‌اند.

این روش بر اساس این حقیقت است که چون انرژی سیگنال گفتار و نویز در باندهای فرکانسی متمایز، مختلف می باشند، نویز در نواحی متمایز اسپکتروگرام تاثیر متفاوتی می-گذارد. به عبارتی دیگر اسپکتروگرام گفتار نویزی، برخی نواحی با نسبت سیگنال به نویز (*SNR*) بالا و به تناسب، برخی نواحی دیگر با معیار سیگنال به نویز پایین را نمایش می‌دهد. [۲]

در این روش‌ها نواحی تخریب شده توسط نویز با معیار *SNR* پایین در نمایش اسپکتروگرام گفتار، به عنوان نواحی مفقود شناسایی شده و نامعتبر فرض می‌شوند و با نمایش ماسک از اسپکتروگرام پاک می‌شوند. سپس از نواحی معتبر با معیار *SNR* بالا، برای جبران نواحی مفقود استفاده می‌شود.

بیشتر تکنیک های بازشناسی گفتار که از روش ویژگی های مفقود بهره می گیرند مبتنی بر اصلاح شیوه‌ای است که در آن، سیستم بازشناسی، احتمال کلاس‌ها و یا حالت‌ها را برای دستیابی به ویژگی‌های نامعتبر محاسبه می‌کند[۳]. این روشها مدل بازشناسی را محدود می‌کنند که در حوزه لگاریتم طیف، جائیکه نواحی مفقود مشخص شده اند، عمل بازشناسی را صورت دهد و این یک اشکال جدی است زیرا اگر چه روش‌های ویژگی های مفقود، رویکردهای بسیار موثری هستند اما صحت بازشناسی، زمانی که از ویژگی های لگاریتم طیفی برای بازشناسی استفاده می شود ضعیف تر از سایر ویژگی های مورد استفاده نظیر ضرایب کپستروم می‌باشد.

روش‌های مبتنی بر ویژگی‌های مفقود، در عمل از یک دانش اولیه از خواص آماری گفتار تمیز که توسط سیستم بازشناسی گفتار، مدل شده اند استفاده می‌کنند. [۳]

در این تحقیق ما یک روش بازسازی نواحی تخریب شده از نمایش اسپکتروگرام توسط نویز را به‌عنوان یک مرحله پیش پردازش قبل از بازشناسی مطرح می‌کنیم که از آن به عنوان بازسازی مبتنی بر همبستگی (correlation-based reconstruction) و یا بازسازی مبتنی بر کواریانس (covariance-based reconstruction) یاد می‌شود و در نهایت نمایش اسپکتروگرام بازسازی شده می تواند به یک مجموعه ویژگی انتخابی، نظیر کپستروم تبدیل شده و توسط مدل بازشناسی استاندارد مورد استفاده قرار گیرد. این روش مزایای دیگری نیز دارد که نیاز به انواع مدل بازشناسی را از بین برده و امکان استفاده از اطلاعاتی که بطور آشکارا در مدل بازشناسی بکار گرفته نمی‌شوند، نظیر ارتباط زمانی بین مؤلفه‌های بردارهای لگاریتم طیفی، فراهم می شود.

الگوریتمی که در این تحقیق مطرح شده است، از ساده ترین مشخصات آماری توزیع لگاریتم طیف گفتار تمیز به‌منظور بازسازی ناحیه های تخریب شده از اسپکتروگرام استفاده می‌کند. نتایج ارزیابی نشان می‌دهد که این روشها زمانیکه نواحی تخریب شده از اسپکتروگرام، بطور کامل تشخیص داده شوند، می‌توانند بطور قابل توجهی در بازشناسی، موثر باشند.

در روش های بازسازی خودکار مبتنی بر ویژگی های مفقود، لازم است که نواحی تخریب شده از اسپکتروگرام هم بطور خودکار تشخیص داده شوند. در این تحقیق از دو معیار متفاوت انرژی منفی و همچنین *SNR* برای تشخیص نواحی مفقود استفاده شده است. ارزیابی ها نشان می‌دهد که استفاده از این دو معیار در ترکیب با روشهای مبتنی بر ویژگی های مفقود مطرح شده در این تحقیق، بهبود قابل توجهی در صحت بازشناسی گفتار آلوده به نویز سفید خواهد داشت.

در بخش۲ نحوه شناسایی مولفه‌های مفقود از روی اسپکتروگرام سیگنال مطرح می شود. در بخش۳ روش بازسازی مبتنی بر کواریانس، شرح داده می شود. در بخش۴ عملکرد این روش برای *SNR* های مختلف ارزیابی شده و درنهایت نتیجه گیری از مقاله در بخش۵ صورت خواهد گرفت.

۲- شناسایی مولفه‌های مفقود

برای اینکه روش‌های مبتنی بر ویژگی های مفقود در عمل قابل اجرا باشد نباید هیچ دانش اولیه‌ای برای شناسایی مولفه های نامعتبر بکار گرفته شود و به همین دلیل یکی از مشکل ترین قسمت‌های روش‌های مبتنی بر ویژگی‌های مفقود، شناسایی مولفه‌های معتبر و نامعتبر در نمایش اسپکتروگرام است. برای این منظور می‌بایست طیف نویز موجود در تک تک مولفه ها را تخمین زد و از این معیار برای شناسایی مولفه‌های نامعتبر استفاده کرد که به این رویکرد، روش مبتنی بر *SNR* گفته می‌شود.

برای تخمین *SNR* مولفه‌های طیفی، تخمینی از طیف توان نویز مخرب نیاز است. برای این منظور چند فریم اول هر گویش بعنوان سکوت فرض شده و میانگین طیف توان آنها، بعنوان طیف توان نویز اولیه در نظر گرفته می‌شود[۴]. در ادامه برای بدست آوردن تخمینی از تغییرات آهسته نویز برای هر مولفه در نمایش اسپکتروگرام، می‌توان از رابطه بازگشتی (۱) استفاده کرد [۵].

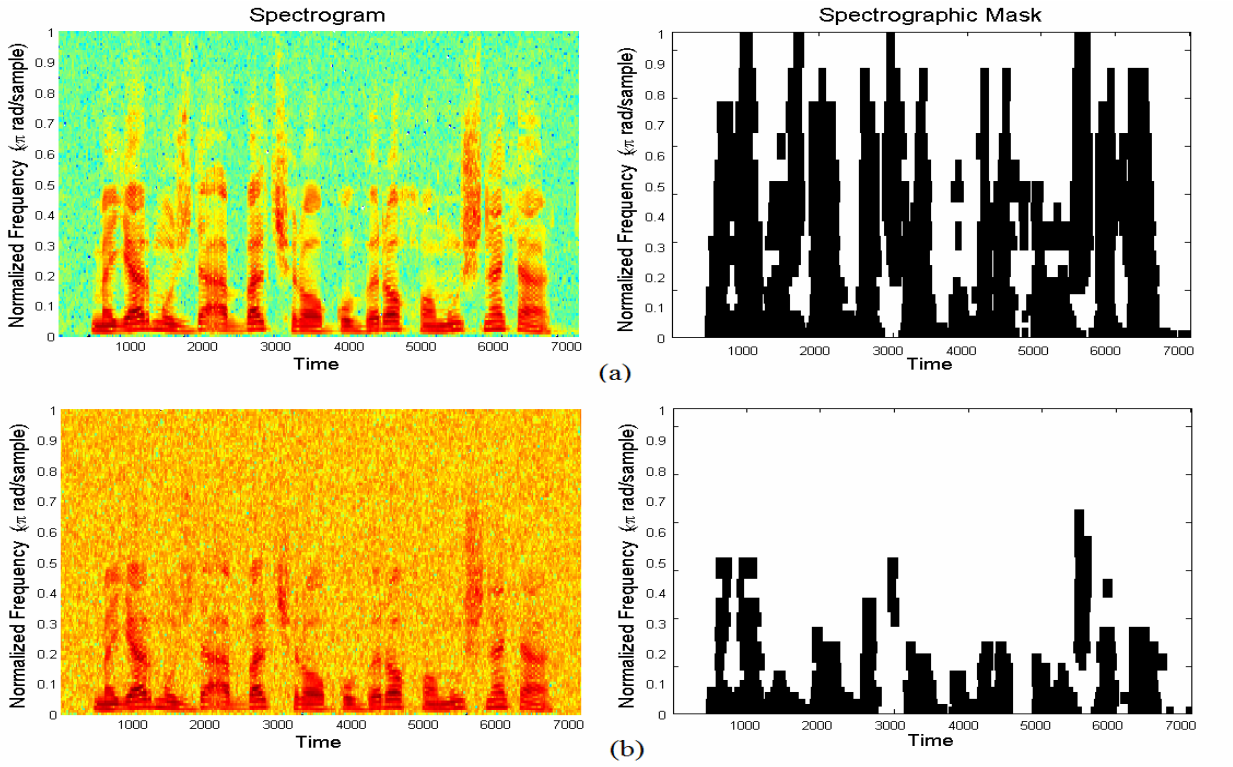
$$\hat{N}_p(m,k)=\begin{cases}(1-\lambda)\hat{N}_p(m-1,k)+\lambda Y_p(m,k) \\ \qquad if \quad Y_p(m,k)<\beta\hat{N}_p(m,k) \\ \hat{N}_p(m-1) \\ \qquad otherwise\end{cases} \qquad (۱)$$

در رابطه فوق، λ و β بترتیب برابر ۰/۹۵ و ۲ می‌باشند *Yp(m, k)* و *Np(m, k)* بترتیب نشان دهنده طیف kامین باند فرکانسی سیگنال گفتار نویزی و سیگنال نویز در m امین فریم می‌باشند. λ معیار دنبال کننده تغیرات کند و تند نویز می‌باشد.

معیارهای مختلفی برای شناسایی مولفه‌های نامعتبر از روی طیف تخمینی نویز پیشنهاد شده‌اند. یکی از این معیارها که توسط Al-Maleki ارائه شده، معیار انرژی منفی می‌باشد که تشخیص مولفه‌های نامعتبر، مبتنی بر این اساس است که انرژی در آن مولفه کمتر از انرژی تخمینی نویز باشد و یا بعبارت دیگر *Yp(m, k)* نامعتبر خواهد بود اگر رابطه (۲) برقرار باشد [۶]:

$$\left|Y_p(m,k)\right| \le \left|\hat{N}_p(m,k)\right| \qquad (۲)$$

از طرفی دیگر اگر *SNR* هر مولفه ای زیر صفر *dB* باشد معیار *SNR*، باندهای طیفی را نامعتبر تشخیص می‌دهد. برای تخمین *SNR*، تخمین طیف سیگنال تمیز نیز مورد نیاز است. این امر بوسیله کسر طیف نویز از طیف سیگنال نویزی میسر می‌شود که رابطه (۳) آن‌را نمایش می‌دهد[۷].



شکل ۱: (a) نمایش اسپکتروگرام و ماسک اسپکتروگرام یک گویش تمیز (b) نمایش اسپکتروگرام و ماسک اسپکتروگرام همان گویش تخریب شده توسط نویز سفید با SNR معادل ۱۰ dB

$$\hat{X}_p(m,k)=\begin{cases} Y_p(m,k)-\hat{N}_p(m,k) \\ if \quad Y_p(m,k)-\hat{N}_p(m,k)>\gamma Y_p(m,k) \\ \gamma Y_p(m,k) \\ \qquad otherwise \end{cases} \qquad (۳)$$

که در آن $Xp(m,k)$، تخمین k امین مولفه طیفی سیگنال تمیز را در m امین فریم نمایش می‌دهد.

طبق معیار SNR، اگر طیف تخمینی سیگنال تمیز از طیف تخمینی نویز کمتر باشد مولفه طیفی $Yp(m,\ k)$ نامعتبر فرض می‌شود که در رابطه (۴) نشان داده شده است.

$$\hat{X}_p(m,k)<\hat{N}_p(m,k) \qquad (۴)$$

در عمل، بهترین شناسایی ویژگی های مفقود زمانی حاصل می‌شود که دو معیار انرژی منفی و معیار SNR بصورت ترکیبی در شناسایی مولفه‌های نامعتبر بکار گرفته شوند.

شکل۱ نمونه‌ای از نمایش اسپکتروگرام و ماسک اسپکتروگرامی را که بوسیله تخمین نویز بدست آمده نشان می‌دهد.

بطور معمول، شناسایی ویژگی های مفقود براساس تخمین نویز، زمانی عملی و موثر است که نویز، ایستا و یا شبه ایستا باشد. برای نویزهای غیر ایستا و زودگذر این روش منجر به نتایج ضعیفتری درشناسایی نواحی مفقود اسپکتروگرام خواهد شد [۴].

۳- روش بازسازی مبتنی بر کواریانس

یکی دیگر از روش‌های مبتنی بر ویژگی‌های مفقود روش بازسازی مبتنی بر کواریانس می‌باشد. دراین روش بردارهای طیفی لگاریتمی استخراج شده از گفتار، نمونه‌های یک فرایند تصادفی گوسی ایستا فرض می‌شوند. دانش اولیه درباره سیگنال-های گفتار تمیز بصورت پارامترهای آماری نظیر مقدار میانگین بردارها و کواریانس بین مولفه‌های آنها، نمایش داده می‌شوند.

$\mu(k)$ نشان دهنده میانگین k امین باند فرکانسی از m امین فریم بردار طیفی $X(m,k)$، می باشد. کواریانس بین k_1 امین باند فرکانسی از m امین بردار طیفی $X(m,k_1)$ و k_2 امین باند فرکانسی از $m+\xi$ امین بردار طیفی $X(m+\xi,k_2)$، نیز با $c(\xi,k_1,k_2)$ نمایش داده می شود که مقدار نرمالیزه شده نظیر آن $r(\xi,k_1,k_2)$ می‌باشد که از طریق روابط زیر بدست می‌آیند:

$$\mu(k) = E[X(m.k)]$$
$$c(\xi, k_1, k_2) = E(X(m, k_1) - \mu_{k_1})(X(m+\xi, k_2) - \mu_{k_2}) \quad (۵)$$
$$r(\xi, k_1, k_2) = \frac{c(\xi, k_1, k_2)}{\sqrt{c(\xi, k_1, k_1)c(\xi, k2, k_2)}}$$

در رابطه فوق $E[.]$ عملگر امید ریاضی را نشان می‌دهد. این پارامترها از بردارهای طیفی استخراج شده از مجموعه دادگان تمیز بدست می‌آیند.

برای تخمین مولفه‌های طیفی نامعتبر در m امین بردار طیفی $X(m)$، همه‌ی آن‌ها را در بردار $X_u(m)$ گرد می‌آوریم سپس همه‌ی مولفه‌های طیفی معتبر در نمایش اسپکتروگرام را که دارای کواریانس نرمالیزه شده‌ی حداقل ۰/۵ با حداقل یکی از مولفه‌های بردار $X_u(m)$ باشند را شناسایی و در بردار $X_r^{(n)}(m)$ مرتب می کنیم. مقدار میانگین و کواریانس بین مولفه‌های $X_u(m)$ و $X_r^{(n)}(m)$ و همچنین کواریانس متقابل بین آنها بوسیله پارامترهای میانگین و کواریانس حاصل از دانش اولیه‌ی فرایند گوسی مورد نظر ساخته می‌شوند.

نظر به اینکه فرایند گوسی فرض شده است، مقدار صحیح مولفه‌های نامعتبر $X_u(m)$ را می‌توان بوسیله الگوریتم MAP تخمین زد که در زیر رابطه نهایی آن آورده شده است.

$$\hat{X}_u(m) = \mu_u + C_{ru}.C_{rr}^{-1}.(Y_r(m) - \mu_r) \quad (۶)$$

که در آن μ_u و μ_r بترتیب نشان دهنده مقدار میانگین مولفه‌های $X_u(m)$ و $X_r^{(n)}(m)$ و همچنین C_{rr} و C_{ru} بترتیب کواریانس و کواریانس متقابل بین مولفه‌های $X_r^{(n)}(m)$ و $X_u(m)$ می‌باشند. $\hat{X}_u(m)$ نیز معرف مقادیر بازسازی شده بردار ویژگی‌های مفقود نظیر $X_u(m)$ می‌باشد.

برای بازشناسی بهتر لازم است مقادیر بازسازی شده‌ی ویژگی های مفقود در یک حیطه خاص محدود شوند که آن را در ترکیب با رابطه فوق اصطلاحا Bounded MAP می‌نامند. رابطه (۷) نحوه اعمال این محدودیت را نشان می‌دهد.

$$\hat{X}_u(m,k) = \begin{cases} \hat{X}_u(m,k) & if \quad \hat{X}_u(m,k) \le X_u(m,k) \\ X_u(m,k) & otherwise \end{cases} \quad (۷)$$

از آنجایی که سیگنال نویزی در اثر اضافه شدن نویز مخرب به گفتار تمیز حاصل شده است در نتیجه مقادیر لگاریتم طیفی سیگنال گفتار نویزی را می‌توان به عنوان یک کران بالا برای مقادیر بازسازی شده در نظر گرفت.

۴- ارزیابی

در این تحقیق روش بازسازی ویژگی های مفقود مبتنی بر کواریانس بر روی بخشی از مجموعه دادگان فارس‌دات پیاده سازی شده است که شامل ۲۰۰ گویش از ۱۰۰ گوینده مختلف می باشد.

در این تحقیق، از ضرایب کپستروم فرکانس مل استخراج شده از ویژگی های لگاریتم طیفی و مشتقات آن که از مجموعه دادگان تمیز بدست آمده‌اند، برای آموزش مدل بازشناسی HMM استفاده شده است.

برای در نظر گرفتن محیط‌های نویزی مختلف، سیگنال گفتار تمیز را بصورت مجازی با *SNR*های ۲۰ الی صفر *dB* توسط نویز سفید گوسی تخریب کرده و روش جبران‌سازی مبتنی بر کواریانس روی آن اعمال شده است. دادگان مفقود در سیگنال های گفتار نویزی توسط ماسک اسپکتروگرام، شناسایی شده و توسط الگوریتم Bounded MAP بازسازی شده اند. بردارهای طیفی بازسازی شده که بیانگر ویژگی‌های لگاریتم طیفی فرکانس مل هستند به ضرایب کپستروم فرکانس مل (MFCC) تبدیل شده و با افزودن مشتقات اول و دوم آن‌ها، بردار بازنمایی برای بازشناسی دادگان نویزی فراهم می‌شود.

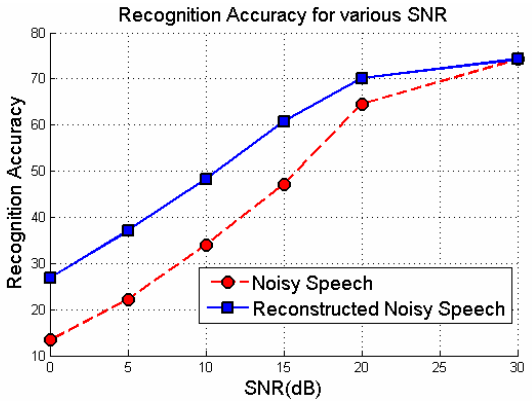
مدل بازشناسی HMM توسط ۱۴۰ گویش برچسب گذاری شده، برای هر واج از مجموعه دادگان فارس‌دات آموزش داده می‌شود. ۳۴ مدل مخفی مارکوف برابر با ۳۴ واج فارسی، برای هر دسته از بردارهای بازنمایی آموزش داده شده است. مدلهای مخفی مارکوف شامل سه حالت چپ به راست به همراه شانزده تابع چگالی گوسی برای هر حالت هستند. پارامترهای آماری نظیر میانگین، کواریانس متقابل و کواریانس نرمالیزه شده شرح داده شده در بخش ۳، برای بازسازی ویژگی‌های مفقود، از مجموعه دادگان تمیز استخراج می‌شوند.

مجموعه دادگان تست شامل ۶۰ گویش نویزی آلوده به نویز سفید گوسی با *SNR*های ۲۰ الی صفر *dB* می‌باشد. جدول ۱ و شکل ۲، صحت بازشناسی را برای دادگان نویزی و همچنین دادگان نویزی اصلاح شده توسط روش دادگان مفقود مبتنی بر کواریانس نمایش می‌دهد.

با توجه به جدول۱ مشهود است که جبرانسازی با روش مبتنی بر ویژگی‌های مفقود عملکرد قابل توجهی در محیط های نویزی با *SNR*های مختلف بر روی صحت بازشناسی مدل داشته است.

جدول ۱: صحت بازشناسی برای دادگان تمیز و نویزی اصلاح نشده و نویزی بازیابی شده در ازای *SNR* های ۲۰ الی صفر *dB* در حیطه لگاریتم طیفی

Recognition Accuracy	Clean	SNR 20	SNR 15	SNR 10	SNR 5	SNR 0
Noisy Speech	74.22	64.50	47.22	33..99	22.23	13.51
Noisy Reconstructed Speech	74.22	70.14	60.74	48.29	37.07	26.75



شکل۲ : صحت بازشناسی بدست آمده از روش جبران‌سازی مبتنی بر ویژگیهای مفقود در *SNR* های بین صفر الی ۲۰ dB درحیطه لگاریتم طیفی.

۵- نتیجه‌گیری

در این مقاله عملکرد روش جبران سازی مبتنی بر کواریانس که ویژگی های مفقود یا نامعتبر را از روی نمایش لگاریتم طیفی سیگنال گفتار نویزی بازیابی می‌کند، شرح داده شد. این روش از یک سری اطلاعات آماری بدست آمده از بردارهای لگاریتم طیفی دادگان تمیز برای بازسازی نواحی تخریب شده توسط نویز استفاده می‌کند.

در صورتی‌که ۱) خاصیت کران بالایی برای جایگزینی مقادیر بازسازی شده در نمایش اسپکتروگرام سیگنال گفتار نویزی در نظر گرفته شود و ۲) روند بازسازی، مبتنی بر جداسازی طیفی باشد، بازسازی بهتر و به تبع آن بازشناسی بهتری بدست خواهد آمد.

ما نشان دادیم که روش مبتنی بر ویژگی‌های مفقود، در بازسازی مقادیر مفقود در مواجهه با نویز ایستا، در حد بسیار بالایی موثر است. بطوریکه برای سیگنال با SNR=0 dB به میزان ۱۳/۲۵٪ افزایش در صحت بازشناسی گفتار حاصل شده است.

مراجع

[1] X. Xiong, "Speech Enhancement with Applications in Speech Recognition" *A First Year Report Submitted to the School of Computer Engineering of the Nanyang Technological University*. 2006

[2] R. P. Lippman, "Using Missing Feature Theory to Actively Select Features for Robust Speech Recognition with Interruptions, Filtering and Noise", *Proc. Eurospeech*,1997

[3] M. P. Cooke, A. Morris, and P. D. Green, "Recognizing Occluded Speech*", ESCA Tutorial and Workshop on Auditory Basis of Speech Perception, Keele University*, 1996.

[4] B. Raj, R. Stern "Improving recognition accuracy in noise by using partial spectrographic information: Missing Feature in ASR'' *IEEE SIGNAL PROCESSING MAGAZINE*, pp. 101-116, 2005.

[5] H.G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition,'' *in Proc. IEEE Conf. Acoustics, Speech Signal Processing, Detroit, Michigan*, pp. 153–156, 1995.

[6] M. El-Maliki and A. Drygajlo, "Missing features detection and handling for robust speaker verification,'' *in Proc. Eurospeech, Budapest, Hungary*, pp. 975–978, 1999.

[7] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction,'' *IEEE Trans. Acoust. Speech Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.