

حذف نویز از سیگنال گفتار با اصلاح طیف فرکانسی سیگنال نویزی - بررسی روش‌های تخمین سیگنال به نویز محلی و کنترل و شکل دهی سطح نویز باقیمانده

محسن رحمانی، بابک ناصرشریف، روح‌الله عبدی‌پور و احمد اکبری
دانشکده‌ی مهندسی کامپیوتر، دانشگاه علم و صنعت ایران
نویسنده‌ی عهده‌دار مکاتبات: محسن رحمانی

چکیده:

دسته‌ی مهمی از روش‌های بهبود کیفیت گفتار، طیف فرکانسی دوره‌ی کوتاه سیگنال نویزی را تغییر داده و اصلاح می‌نمایند. اصلاحات طیفی این روش‌ها به صورت یک فیلتر در حوزه‌ی فرکانسی بیان می‌شود و بهره‌ی فیلتر می‌تواند به صورت تابعی از سیگنال به نویز نمایش داده شود. در این مقاله، روش‌های حذف نویز از سیگنال گفتار مورد بررسی و ارزیابی قرار می‌گیرد و بهره‌ی فیلتر برای برخی از روش‌های حذف نویز شناخته شده به صورت تابعی از سیگنال به نویز نمایش داده می‌شود. روش‌های تخمین سیگنال به نویز بررسی می‌شود و یک تکنیک جدید برای تخمین سیگنال به نویز پیشنهاد شده و ارزیابی می‌گردد و برتری آن بر روش‌های قبلی نشان داده می‌شود. در این مقاله اصلاح روش‌های حذف نویز به منظور کم کردن اعوجاج سیگنال گفتار مطرح شده است. در این روش‌ها اثرات اصلاح فرکانسی، طوری تنظیم می‌گردد که با باقی گذاشتن مقداری از نویز اولیه، بهبود مشخصی در نسبت سیگنال به نویز به دست آید و در عوض اعوجاج سیگنال گفتار حداقل گردد. برای این کار با اعمال یک پارامتر، بهره‌ی چند روش حذف نویز برای دستیابی به سطح مشخصی از حذف نویز تنظیم می‌گردد. همچنین برای خوشایند کردن نویز باقیمانده، از شکل دادن نویز باقیمانده استفاده شده است. به منظور ارزیابی این روش‌ها تست‌های شنوایی برای بررسی جنبه‌های مختلف گفتار، از جمله نویز موسیقی، نویز باقیمانده، پژواک و آثار غیر طبیعی نویز باقیمانده، طراحی و اجرا شده است. نتایج این تست‌ها در مقاله‌ی حاضر برای چند روش حذف نویز گزارش شده است و روش‌های حذف نویز با یکدیگر مقایسه شده‌اند.

کلمات کلیدی: بهبود گفتار، کاهش نویز، نویز موسیقی، اعوجاج و نویز باقی‌مانده، سیگنال به نویز پیشین و سیگنال به نویز پسین

۱- مقدمه

استفاده از بهبود گفتار، در بسیاری از کاربردهایی که صدا در یک محیط نویزی دریافت می‌گردد و یا میکروفن از گوینده فاصله دارد، ضروری است. از جمله‌ی این موارد می‌توان به دریافت صدا از طریق میکروفن تلفن‌های بدون گوشی^۱ و نیز میکروفن‌های رومیزی همه جهت در یک کنفرانس صوتی اشاره کرد. گوشی‌های تلفن همراه یا هدست^۲ که روی گوش قرار می‌گیرند و با فناوری بی‌سیم دندان آبی^۳ به تلفن همراه وصل می‌شوند نیز اخیراً بازار مناسبی پیدا کرده‌اند. در این دستگاه‌ها به علت فاصله‌ی چند سانتیمتری میکروفن با دهان گوینده، نویز محیط اطراف نیز دریافت می‌گردد و موجب افت کیفیت صوت می‌شود. سر و صدای اضافی مانند

^۱ - hands free

^۲ - headset

^۳ - bluetooth

همه‌مهی افراد در محیط کار یا نویز مربوط به ورزش باد در محیط آزاد نیز از جمله عوامل افت کیفیت صوت در سیستم‌های همدست می‌باشد. نظر به اهمیت این محصولات، لزوم استفاده از الگوریتم‌های کارا برای ارتقای کیفیت صوت به طور جدی مطرح می‌باشد. بهبود کیفیت گفتار در بازشناسی گفتار نیز مطرح است. یک سیستم بازشناسی گفتار ممکن است در محیطی غیر از محیط آموزش عمل کند. در این حالت برای جلوگیری از پایین آمدن دقت بازشناسی، نیاز به حذف نویز و بهبود گفتار احساس می‌شود. یک کاربرد دیگر که در آن بهبود کیفیت گفتار مطرح است، در مورد کدکننده‌هایی هستند که برای کد کردن گفتار تمیز، طراحی شده‌اند، ولی در عمل می‌بایست روی سیگنال‌های دریافت شده در محیط نویزی اعمال گردند. در این حالت نیز بهبود گفتار و حذف نویز می‌تواند صحت عملکرد کدکننده را افزایش دهد.

دسته‌ای عمده‌ای از روش‌های بهبود گفتار، روش‌هایی هستند که عملیات خود را روی طیف دوره‌ی کوتاه سیگنال انجام می‌دهند. در این روش‌ها یک فیلتر فرکانسی روی دامنه‌ی طیف اعمال می‌شود. پس از اعمال فیلتر، خروجی فیلتر به‌عنوان طیف تخمینی تلقی می‌شود. سیگنال بهبودیافته در حوزه‌ی زمان با استفاده از اندازه‌ی طیف فرکانسی تخمینی و مرحله‌ی سیگنال نویزی بازسازی می‌گردد. بهره‌ی فیلترهایی که در این روش‌ها به‌کار می‌روند، می‌توانند به صورت تابعی از نسبت توان سیگنال به نویز (SNR) بیان شوند. نسبت سیگنال به نویز، بیان‌کننده‌ی نسبت توان مقدار مؤلفه‌ی سیگنال مفید در یک فرکانس خاص و در یک فریم داده شده از سیگنال به مقدار توان مؤلفه‌ی نویز می‌باشد؛ لذا فیلترهای فرکانسی در مؤلفه‌های فرکانسی که نسبت سیگنال به نویز بزرگ است، مؤلفه مربوطه را به‌عنوان سیگنال مفید در نظر می‌گیرند و کمتر اصلاح می‌کنند. برعکس در مؤلفه‌های فرکانسی که نسبت سیگنال به نویز ضعیف باشد، مؤلفه‌های فرکانسی به‌عنوان نویز تلقی شده و تضعیف می‌گردد.

فرآیند تخمین نویز از مهمترین بخش‌های یک سیستم اصلاح فرکانسی است. اگرچه روش‌های قدیمی‌تر نویز را در نواحی سکوت تخمین می‌زنند (مثلاً در [۱ و ۲])، در روش‌های جدیدتر تخمین نویز با دنبال کردن کمینه‌ی طیف سیگنال نویزی به دست می‌آید. (به عنوان مثال در [۳، ۴، ۵ و ۶]) در این روش فرض بر این است که سطح انرژی سیگنال آغشته به نویز در صورت عدم وجود گفتار به سطح انرژی نویز باز می‌گردد؛ در این صورت می‌توان با دنبال کردن کمینه‌های طیف توان سیگنال آغشته به نویز، طیف توان نویز را به دست آورد. لازم به ذکر است که هدف روش‌های تخمین نویز، تخمین میانگین توان نویز در هر مؤلفه‌ی فرکانسی و دنبال کردن تغییرات آن با کم‌ترین تأخیر ممکن است. طیف تخمین زده شده از روی کمینه‌ها همواره متمایل به سمت مقادیر کم‌تر از مقدار میانگین است؛ در نتیجه در روش‌های تخمین نویز مبتنی بر کمینه‌های طیف کوتاه مدت، همیشه از یک فیلتر هموارکننده^۱ استفاده می‌شود تا با کاهش انحراف از میانگین نمونه‌های طیف، خطای تخمین کاهش بیابد. در نهایت نیز طیف تخمین زده شده در یک ضریب جبران‌کننده، ضرب می‌شود تا تخمین به مقادیر واقعی نزدیک‌تر شود.

^۱ - smoothing filter

روش‌های مختلفی برای تخمین سیگنال به نویز برای استفاده در فیلترهای فرکانسی ارائه شده است. در [۷] فیلترهای فرکانسی به‌صورت تابعی از دو تخمین متفاوت از سیگنال به نویز معرفی شده است. این سیگنال به نویزها به‌عنوان سیگنال به نویز پسین^۲ و سیگنال به نویز پیشین^۳ شناخته شده‌اند. در [۲] دو روش برای محاسبه‌ی سیگنال به نویز پیشین بر اساس سیگنال به نویز پسین ارائه شده است. در [۸] نشان داده شده‌است که سیگنال به نویز پیشین در نواحی مختلف گفتاری چگونه سیگنال به نویز پسین را دنبال می‌کند. در [۹] تأثیر استفاده از سیگنال به نویز پیشین به جای سیگنال به نویز پسین در چند فیلتر شناخته شده مورد بررسی قرار گرفته است. در [۱۰، ۱۱ و ۱۲] روش‌های ترکیبی که از تخمین سیگنال به نویز پیشین و پسین برای تخمین استفاده می‌کند، ارائه شده است.

فرآیند حذف نویز از سیگنال گفتار می‌تواند به از دست رفتن قسمت‌هایی از سیگنال مفید نیز منجر شود که به آن اعوجاج سیگنال گفته می‌شود. از بین رفتن قسمت‌های مفید سیگنال یا اعوجاج منجر به افت کیفیت گفتار و همچنین پایین آمدن قابلیت فهم گفتار می‌شود. یکی از روش‌های کاهش میزان اعوجاج، استفاده از فیلترهایی است که نویز زمینه را به طور کامل حذف نمی‌کنند.

اگرچه نویز موجود در سیگنال صوتی اثرات نامطلوبی از قبیل پایین آوردن کیفیت صوت و قابلیت فهم صوت دارد و موجب خستگی شنونده می‌گردد ولی در عوض موجب می‌گردد، که اطلاعاتی در رابطه با شرایط دریافت صوت و محیط اطراف گوینده‌ی اصلی به سیگنال صوتی اضافه گردد و در مواردی نقایص موجود در سیگنال اصلی کمتر مورد توجه شنونده قرار می‌گیرد. عموماً حذف کامل نویز به دلایل متعدد امکان‌پذیر نمی‌باشد که از آن جمله به در دسترس نبودن تخمین دقیق آمارگان نویز، ثابت نبودن آمارگان مربوطه با زمان، ضعف روش‌های پردازش سیگنال در به دست آوردن تخمین‌های دقیق و وجود خطا در روش‌های تخمین در حوزه‌ی فرکانس می‌توان اشاره نمود. در صورتی که حذف نویز صوتی اضافه شده به سیگنال اصلی به‌طور کامل صورت نگیرد از یک طرف مقداری نویز باقیمانده در سیگنال وجود دارد که به دلایل اشاره شده از نظر شنونده غیر طبیعی بوده و با نویز اولیه تفاوت قابل توجهی دارد؛ از سوی دیگر قسمتی از مؤلفه‌های جدید نویز

^۲ - a posteriori signal to noise ratio

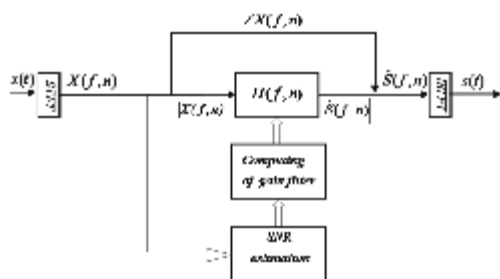
^۳ - a priori signal to noise ratio

۲- حذف نویز به روش اصلاح طیف فرکانسی

در روش های اصلاح طیف فرکانسی، سیگنال نویزی با اعمال تبدیل فوریه زمان-کوتاه^۱ به حوزه فرکانس انتقال می یابد. در این حالت در شرایطی که نویز جمع پذیر^۲ است و با این فرض که نویز با سیگنال تمیز ناهمبسته^۳ است، رابطه ی زیر را می توان نوشت.

$$X(f, n) = S(f, n) + B(f, n) \quad (1)$$

که در آن $X(f, n)$ ، $S(f, n)$ و $B(f, n)$ به ترتیب سیگنال نویزی، سیگنال تمیز و نویز در حوزه STFT هستند. در روش های بهبود گفتار هدف آن است که با در دست داشتن سیگنال نویزی تخمین مناسبی از سیگنال تمیز به دست آید. شکل (۱) بلوک دیاگرام روش های بهبود گفتار مورد بحث را نشان می دهد.



شکل ۱. بلوک دیاگرام عمومی روش های حذف نویز با اصلاح طیف فرکانسی

چنانچه در شکل دیده می شود فیلتر با مشخصه ی فرکانسی $H(f, n)$ بر طیف فرکانسی سیگنال نویزی، $|X(f, n)|$ ، اعمال شده و دامنه ی طیف سیگنال بهبود یافته، $\hat{S}(f, n)$ ، را نتیجه داده است. پس از اعمال فیلتر، از ترکیب دامنه ی $|\hat{S}(f, n)|$ با مرحله ی سیگنال نویزی و محاسبه ی عکس تبدیل فوریه سیگنال تخمینی در حوزه ی زمان به دست آمده است. در صورتی که تبدیل فوریه، دوره ی کوتاه گفتار تخمینی در هر فریم n ، $\hat{S}(f, n)$ باشد، تأثیر فیلتر با مشخصه ی فرکانسی $H(f, n)$ بر طیف سیگنال نویزی را می توان به صورت رابطه ی زیر نشان داد.

$$\hat{S}(f, n) = H(f, n) \cdot X(f, n) \quad (2)$$

روابط مختلفی برای $H(f, n)$ پیشنهاد شده است که در ادامه تعدادی از این روش ها معرفی می شوند.

که در اثر خطا در پردازش گفتار حاصل می گردد، به سیگنال اولیه اضافه می گردد که در برخی موارد نامطلوب تر از نویز اولیه است (نویز موسیقی). یک اثر نامطلوب دیگر در این موارد، حذف قسمتی از مؤلفه های سیگنال اصلی است که به عنوان اعوجاج سیگنال اولیه شناخته شده است. به عبارت دیگر سیگنال به دست آمده قسمتی از مؤلفه ها و جزئیات خود را از دست داده است. گرچه از دست رفتن این جزئیات با اضافه شدن نویز در سیگنال نویزی حادث شده است، با این حال در سیگنال نویزی به دلیل اثرات ماسک فرکانسی گوش چندان مشخص نمی باشد. در حالی که با حذف مؤلفه های قوی، نویز این اثرات به طور واضح در سیگنال اصلاح شده شنیده می شود. به منظور مقابله با نویز باقیمانده ی غیر طبیعی و نیز از دست رفتن جزئیات سیگنال اصلاح شده، روش های حذف نویز با سطح نویز قابل تنظیم مطرح می گردند. ایده ی اصلی این روش ها، کم کردن اثر نویز به جای حذف نویز می باشد. به عبارت دیگر وجود نویز زمینه از یک سو موجب اطلاع شنونده از محیط اطراف و شرایط دریافت صوت می گردد و از سوی دیگر موجب پی بردن شنونده به نقایص سیگنال در اثر خاصیت ماسک شنوایی گوش انسان می شود. این نویز می تواند با سطح توانی که برای شنونده قابل قبول باشد در سیگنال بهبود یافته حضور داشته باشد. به تعبیر دیگر توان نویز اولیه ی اضافه شده به سیگنال اصلی به اندازه ی معینی (مثلاً 10dB) کاهش یابد. این روش ها به عنوان روش های حذف نویز با سطح نویز قابل تنظیم یا با سطح نویز معین، در این مقاله مطرح شده اند.

در ادامه ی مقاله در بخش دوم روش های حذف نویز با اصلاح طیف فرکانسی و نمایش روش های حذف نویز به عنوان یک فیلتر وابسته به نسبت سیگنال به نویز مورد بررسی قرار می گیرند. در بخش سوم روش های تخمین نسبت سیگنال به نویز محلی مورد بررسی قرار گرفته و یک روش جدید برای تخمین نسبت سیگنال به نویز در یک فرکانس مفروض، ارائه می گردد. در بخش چهارم روش هایی برای حذف نویز پیشنهاد می گردد که سطح نویز حذف شده را به طور دلخواه تنظیم کنند. در این بخش همچنین شیوه ای برای شکل دهی نویز باقیمانده پیشنهاد می شود. بخش پنجم به ارزیابی و مقایسه ی روش های مطرح شده اختصاص دارد. و نهایت در بخش ششم نتیجه گیری به عمل می آید.

¹- Short Time Fourier Transform (STFT)

²- additive

³- uncorrelated

به صورت تابعی از نسبت سیگنال به نویز در رابطه، نشان داده شده است.

$$H_w(f) = \frac{SNR(f)}{SNR(f)+1} \quad (8)$$

روش تخمین دامنه‌ی سیگنال گفتار با معیار حداقل میانگین مربعات خطا^۳ (MMSE): در این روش ضرایب فوریه سیگنال گفتار در یک فریم خاص را از روی سیگنال نویزی در همان فریم تخمین زده شود، به گونه‌ای که میانگین مربعات اختلاف دامنه‌ی طیف تخمینی با دامنه‌ی طیف سیگنال گفتار تمیز حداقل شود. بدین ترتیب فیلتر اصلاح فرکانسی زیر تعریف شده است [۲]:

$$H_{MMSE}(f) = \frac{\sqrt{\pi v(f)}}{2 R_{po}(f)} \exp\left(-\frac{v(f)}{2}\right) \quad (9)$$

در رابطه‌ی فوق v_f به صورت زیر تعریف می‌شود و $I_0()$ و $I_1()$ به ترتیب توابع بسل مرتبه‌ی صفر و یک می‌باشد.

$$v(f) = \frac{R_{pr}(f)}{1 + R_{pr}(f)} R_{po}(f) \quad (10)$$

در روابط فوق $R_{pr}(f)$ و $R_{po}(f)$ به ترتیب نسبت سیگنال به نویز پیشین و پسین می‌باشند. این دو نسبت در بخش بعد تشریح می‌شوند.

روش JMAP: فیلتر MMSE برای بهسازی گفتار از نظر حداقل میانگین مربعات خطا، بهینه است و از طرف دیگر نویز باقیمانده‌ی این روش از نوع نویز موسیقی نمی‌باشد [۸ و ۲]. اما محاسبه‌ی این تابع بهره، نیاز به محاسبه‌ی توابع بسل و نمایی دارد که محاسباتی هزینه‌بر هستند. در [۱۱ و ۱۲] فیلترهای جایگزین برای فیلتر MMSE پیشنهاد شده است که با داشتن هزینه‌ی کمتر، رفتاری مشابه فیلتر MMSE از خود نشان می‌دهد. در این روش‌ها با بیشینه کردن احتمال توزیع سیگنال نویزی با شرط داشتن توزیع توأمان دامنه و فاز طیف سیگنال تمیز و با انجام پاره‌ای محاسبات فیلتر به شیوه‌ی زیر محاسبه می‌شود.

$$H_{JMAP}(k) = \frac{R_{pr}(k) + \sqrt{R_{pr}^2(k) + 2(1 + R_{pr}(k)) \frac{R_{pr}(k)}{R_{po}(k)}}}{2(1 + R_{pr}(k))} \quad (11)$$

تفاضل طیف توان: در این روش [۱] با فرض ناهمبسته^۱ بودن نویز با سیگنال تمیز، و به علاوه جمع‌پذیر بودن آن، رابطه‌ی سیگنال تمیز به صورت رابطه‌ی زیر تخمین زده می‌شود.

$$|\hat{S}(f, n)|^2 = |X(f, n)|^2 - |B(f, n)|^2 \quad (3)$$

در این رابطه $|X(f, n)|^2$ ، $|S(f, n)|^2$ و $|B(f, n)|^2$ به ترتیب تخمین طیف توان دوره‌ی کوتاه سیگنال نویزی، سیگنال تمیز و نویز می‌باشد. در صورتی که رابطه‌ی تفاضل طیف توان در قالب رابطه‌ی^۲ نوشته شود، رابطه‌ی زیر حاصل می‌شود.

$$\hat{S}(f, n) = X(f, n) \left(\sqrt{1 - \frac{1}{1 + |S(f)|^2 / |B(f)|^2}} \right) = X(f, n) \left(\sqrt{\frac{|S(f)|^2 / |B(f)|^2}{1 + |S(f)|^2 / |B(f)|^2}} \right) \quad (4)$$

در صورتی که $SNR(f, n)$ را به عنوان نسبت توان سیگنال یا $|S(f)|^2$ به توان نویز یا $|B(f)|^2$ تعریف نماییم، تابع $H(f, n)$ به صورت زیر به دست می‌آید:

$$H(f, n) = \sqrt{\frac{SNR(f, n)}{1 + SNR(f, n)}} \quad (5)$$

تفاضل طیف دامنه: یکی دیگر از روش‌های ساده و در عین حال کارا برای بهسازی گفتار، روش تفاضل طیف دامنه است. این روش، با کم کردن دامنه‌ی طیف نویز از دامنه‌ی طیف سیگنال نویزی، دامنه‌ی طیف سیگنال تمیز را تخمین می‌زند. رابطه‌ی تفاضل طیف دامنه به صورت زیر است.

$$|\hat{S}(f, n)| = |X(f, n)| - |B(f, n)| \quad (6)$$

با ترکیب دامنه‌ی طیف توان به دست آمده برای سیگنال تخمینی با فاز سیگنال نویزی، سیگنال تخمینی به دست می‌آید. رابطه‌ی تفاضل طیف دامنه^۲ را نیز می‌توان به شکل ضربی، به صورت تابعی از نسبت سیگنال به نویز نوشت. شکل ضربی تفاضل طیف دامنه به صورت رابطه‌ی زیر نوشته می‌شود.

$$H_{as}(f) = \left(1 - \sqrt{\frac{1}{SNR(f)+1}} \right) \quad (7)$$

فیلتر وینر: روش متداول دیگر در بهبود سیگنال گفتار، استفاده از فیلتر وینر می‌باشد [۱۳]. رابطه‌ی فیلتر وینر

^۱ - uncorrelated

^۲ - amplitude spectral subtraction

[۱۰] با استفاده از توزیع سوپر گوسی، رابطه‌ی زیر برای تابع بهره‌ی فیلتر به‌دست‌آمده است.

$$H_{MAP}(f) = u + \sqrt{u^2 + \frac{\beta - 0.5}{2R_{po}}} \quad u = \frac{1}{2} - \frac{\mu}{4\sqrt{R_{po}R_{pr}}} \quad (12)$$

$$\beta = 1, \quad \mu = 2.5$$

جدول ۱) روش‌های پیاده‌سازی و نام اختصاری انتخاب شده برای آن‌ها را نشان می‌دهد.

$$R_{po}(f, n) = \frac{X^2(f, n)}{E\{D^2(f)\}} - 1 \quad (13)$$

جدول ۱: روش‌های پیاده‌سازی شده

نام اختصاری	توضیح روش	فیلتر پیاده‌سازی شده
<i>PssRpo</i>	تفاضل طیف توان با استفاده از سیگنال به نویز پسین	$H(f, n) = \sqrt{\frac{R_{po}(f, n)}{1 + R_{po}(f, n)}}$
<i>PssRpr</i>	تفاضل طیف توان با استفاده از سیگنال به نویز پیشین	$H(f, n) = \sqrt{\frac{R_{pr}(f, n)}{1 + R_{pr}(f, n)}}$
<i>AssRpo</i>	تفاضل طیف دامنه با استفاده از سیگنال به نویز پسین	$H(f, n) = 1 - \sqrt{\frac{1}{1 + R_{po}(f, n)}}$
<i>AssRpr</i>	تفاضل طیف دامنه با استفاده از سیگنال به نویز پیشین	$H(f, n) = 1 - \sqrt{\frac{1}{1 + R_{pr}(f, n)}}$
<i>WnRpo</i>	وینر با استفاده از سیگنال به نویز پسین	$H(f, n) = \frac{R_{po}(f, n)}{1 + R_{po}(f, n)}$
<i>WnRpr</i>	وینر با استفاده از سیگنال به نویز پیشین	$H(f, n) = \frac{R_{pr}(f, n)}{1 + R_{pr}(f, n)}$
<i>MMSE</i>	روش <i>MMSE</i> [۲]	$H(f) = \frac{\sqrt{\pi}}{2} \frac{\sqrt{V}}{R_{po}(f) + 1} \exp\left(-\frac{V}{2}\right) \left[(1+V)I_0\left(\frac{V}{2}\right) + VI_1\left(\frac{V}{2}\right) \right]$ $V = \frac{R_{pr}(f)}{1 + R_{pr}(f)} (R_{po}(f) + 1)$
<i>JMAP</i>	روش <i>JMAP</i> [۱۸ و ۱۹]	$H(f) = \frac{R_{pr}(f) + \sqrt{R_{pr}^2(f) + 2(1 + R_{pr}(f)) R_{pr}(f) / R_{po}(f)}}{2(1 + R_{po}(f))}$
<i>MAP</i>	روش <i>MAP</i> [۱۰]	$H_{MAP}(f) = u + \sqrt{u^2 + \frac{\alpha - 0.5}{2R_{po}}} \quad u = \frac{1}{2} - \frac{\mu}{4\sqrt{R_{po}R_{pr}}}, \quad \beta = 1, \quad \mu = 2.5$

روش MAP: بسیاری از روش‌های اصلاح طیف فرکانسی، به‌خصوص روش‌هایی که یک تخمین‌زننده با حداقل میانگین مربعات خطا ایجاد می‌کنند، از یک توزیع گوسی با میانگین صفر برای مدل کردن قسمت‌های حقیقی و موهومی سیگنال‌های گفتار و نویز استفاده می‌کنند. هرچند مدل گوسی، تخمین خوبی برای ضرایب *DFT* نویز می‌باشد، ولی قسمت‌های حقیقی و موهومی ضرایب *DFT* سیگنال گفتار با تابع چگالی سوپر گوسی بهتر مدل می‌شوند. در مرجع

که در آن $E\{\}$ میانگین روی فریم‌های مختلف می‌باشد. سیگنال به نویز پسین فقط به طیف سیگنال نویزی در همان لحظه بستگی دارد. درمقابل، سیگنال به نویز پیشین به صورت رابطه‌ی زیر تعریف می‌گردد و این نسبت به میانگین توان سیگنال تمیز $E\{S^2(f)\}$ وابسته است. به عبارت دیگر این نسبت به توان سیگنال نویزی در لحظات دیگر نیز بستگی دارد.

معمولاً توان نویز از لحظات سکوت گفتار تخمین زده می‌شود و با توجه به آنکه توان سیگنال نویزی در هر قاب

۳- تخمین سیگنال به نویز محلی

چنانچه گفته شد، از آنجا که فیلترهایی که در بهبود گفتار مورد استفاده قرار می‌گیرد، به صورت توابعی از نسبت توان سیگنال به نویز در هر قاب هستند، تخمین دقیق این نسبت اهمیت زیادی دارد و دقت بهبود گفتار مستقیماً به دقت تخمین این نسبت وابسته است. یک تعریف برای سیگنال به نویز در یک قاب می‌تواند نسبت توان لحظه‌ای سیگنال گفتار در هر قاب به توان نویز باشد. این تخمین به نام سیگنال به نویز پسین یا لحظه‌ای $R_{po}(f, n)$ شناخته می‌شود. رابطه‌ی ۱۳ بیان کننده‌ی این تخمین است.

صفر می‌باشد به‌طوری‌که وزن بیشتری به نسبت سیگنال به نویز پیشین داده شود. ایده‌ی اصلی این انتخاب زیاد بودن واریانس تخمین نسبت سیگنال به نویز پسین می‌باشد. در مراجع [۹، ۱۵ و ۱۸] با تجزیه و تحلیل در مورد نسبت سیگنال به نویز پیشین و پسین و نیز با شبیه‌سازی آن‌ها این مسأله مورد تأیید قرار گرفته است. نسبت سیگنال به نویز پسین تغییرات سیگنال را بهتر دنبال می‌کند و لذا ترکیب آن‌ها می‌تواند به کارایی بیشتر منجر شود.

شکل ۲) تغییرات تابع بهره‌ی MMSE و JMAP (به ترتیب روابط ۱۵ یا ۱۶) را برحسب R_{po} و R_{pr} نشان می‌دهد. محور افقی نشان دهنده‌ی اندازه SNR پیشین برحسب dB و محور عمودی SNR پسین برحسب dB می‌باشد. در واقع جهت رسم هر یک از این منحنی‌ها اندازه‌ی فیلتر مربوطه (از روابط ۱۵ یا ۱۶) به ازای مقادیر مختلف سیگنال به نویزهای پیشین و پسین محاسبه شده و بر حسب آن‌ها نمایش داده شده است. جهت نمایش از سطح خاکستری استفاده شده است، بدین ترتیب که نقاط روشن‌تر بهره‌ی بیشتر و نقاط تاریک‌تر بهره‌ی کمتر را نشان می‌دهند. نقاط کاملاً سفید نشان دهنده‌ی بهره‌ی صفر dB است.

مشخص است، محاسبه‌ی سیگنال به نویز پسین مطابق رابطه‌ی ۱۳ انجام پذیر است.

$$R_{pr}(f) = \frac{E\{S^2(f)\}}{E\{B^2(f)\}} \quad (14)$$

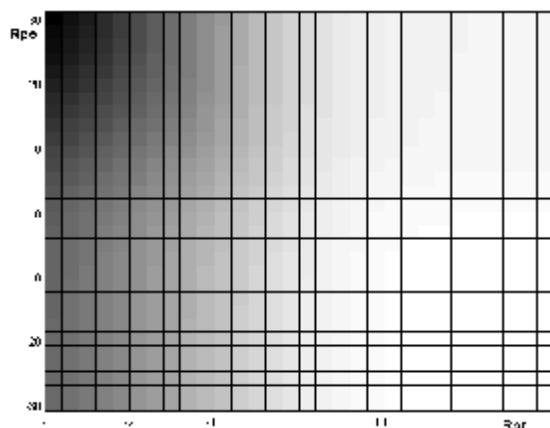
جهت تخمین سیگنال به نویز پیشین روش‌های مختلفی پیشنهاد شده است. یکی از روش‌های مورد استفاده جهت تخمین سیگنال به نویز پیشین روش Decision (DD) Directed می‌باشد. رابطه‌ی بازگشتی که برای این تخمین زننده پیشنهاد شده است، به صورت رابطه‌ی ۱۵ می‌باشد [۲]:

$$R_{pr}(f, r) = \lambda H^2(\hat{R}_{pr}(f, r-1), R_{po}(f, r-1)) R_{po}(f, r-1) + (1-\lambda) \max\{R_{po}(f, r), 0\} \quad (16)$$

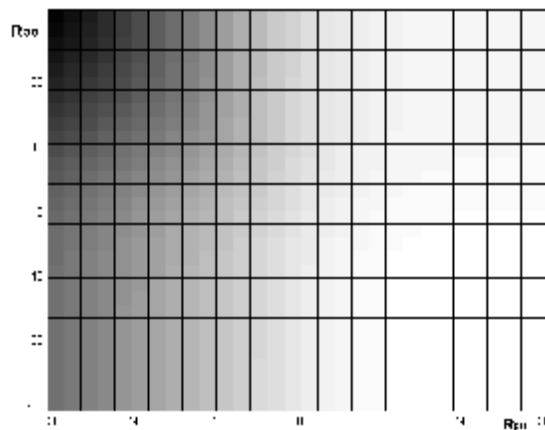
نسبت به صورت زیر تعریف می‌شود:

$$R_{comb}(f) = R_{po}^\lambda(f) R_{pr}^{(1-\lambda)}(f)$$

که در آن λ میزان نزدیکی SNR ترکیبی به $R_{pr}(f)$ و $R_{po}(f)$ را کنترل می‌کند. در صورتی‌که λ برابر ۰/۵ انتخاب گردد، تخمین جدید میانگین هندسی دو نسبت قبلی است. شرایط حدی انتخاب اعداد صفر و یک، برای λ تخمین جدید را به ترتیب معادل سیگنال به نویز پیشین و پسین قرار می‌دهد. پیشنهاد ما برای λ مقادیر نزدیک به



ب: روش Wolfe



الف: روش MMSE

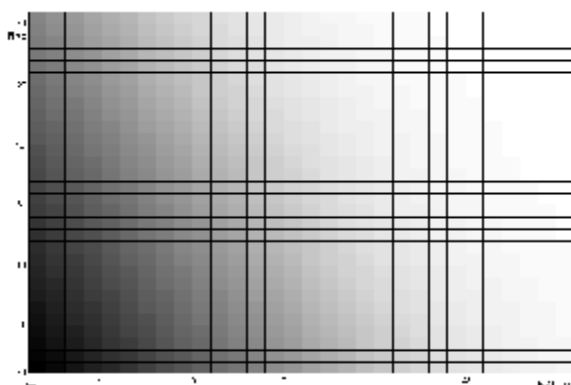
شکل ۲. تغییرات تابع بهره برحسب سیگنال به نویز پیشین و پسین. نواحی روشن‌تر مقادیر بالاتر تابع بهره را نشان می‌دهد. محورهای افقی و عمودی به ترتیب R_{pr} و R_{po} را نشان می‌دهند.

کوچک است و R_{po} مقداری بزرگ دارد، بهره‌ی اندکی ارائه نموده است. این رفتار را می‌توان یک ویژگی مثبت تلقی کرد، چرا که با توجه به اینکه R_{po} و R_{pr} دو تخمین متفاوت از یک کمیت می‌باشند، عدم توافق در این دو تخمین نشان

همان‌طور که در این دو شکل مشاهده می‌شود، به طور کلی با افزایش مقادیر سیگنال به نویزها اندازه‌ی فیلتر افزایش می‌یابد (شکل روشن‌تر می‌شود). به علاوه چنانچه در شکل دیده می‌شود، این دو فیلتر در مواردی که R_{pr}

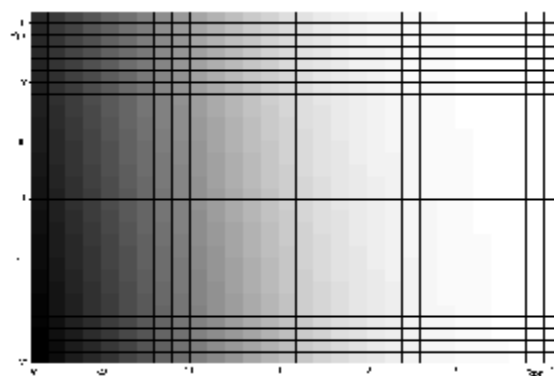
فیلترها کم باشد [۸]. در حالی که چنانچه دیده می شود این مقادیر نسبتاً زیاد است که این مسأله یک ویژگی منفی برای این دو روش می باشد.

بهره ی فیلتر تفاضل دامنه ی طیف با استفاده از SNR ترکیبی به ازای مقادیر $\lambda = 0.1$ و $\lambda = 0.3$ در شکل (۳) رسم شده اند.



ب: روش پیشنهادی با $\lambda = 0.3$

دهنده ی وجود خطا در یکی از دو تخمین است. تخمین Rpr از روی چند فریم متوالی به دست آمده است در حالی که Rpo تنها از روی یک فریم سیگنال نویزی محاسبه شده است، لذا می توان وزن بیشتری به تخمین Rpo داده و بهره فیلتر را اندک در نظر گرفت. بنابراین، این رفتار فیلترها توجیه مناسبی دارد. از سوی دیگر، در مواقعی که هر دو مقدار Rpo و Rpr کوچک بوده اند، انتظار این است که مقدار بهره ی این



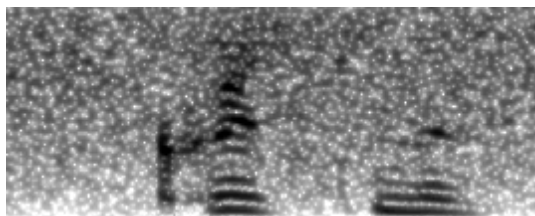
الف: روش پیشنهادی با $\lambda = 0.1$

شکل ۳. تغییرات تابع بهره روش تفاضل دامنه توان با تخمین پیشنهادی بر حسب سیگنال به نویز پیشین و پسین. نواحی روشن تر مقادیر بالاتر تابع بهره را نشان می دهد. محورهای افقی و عمودی به ترتیب Rpo و Rpr را نشان می دهند.

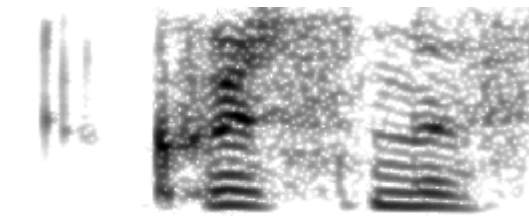
شکل (۴) اسپکتروگرام گفتار تمیز، گفتار نویزی و گفتار بهبود یافته را برای روش تفاضل دامنه ی طیف هنگام استفاده از تخمین های مختلف SNR نشان می دهد. در اینجا یک فیلتر اصلاح طیف معیار، یعنی تفاضل طیف دامنه ی مورد استفاده قرار گرفته است و هدف ما بررسی رفتار تخمین زننده SNR می باشد. با ثابت نگه داشتن فیلتر اصلاح طیف این اطمینان حاصل می گردد که تنها پارامتری که در آزمایش های مختلف تغییر کرده است، تخمین زننده SNR بوده و مقایسه ی رفتار تخمین زننده های مختلف امکان پذیر می گردد. در قسمت های مختلف "ج"، "د"، "ه" و "و" از شکل (۴) به ترتیب از تخمین SNR پسین (Rpo)، تخمین SNR پیشین (Rpr)، تخمین SNR پیشنهادی با $\lambda = 0.1$ و تخمین SNR پیشنهادی با $\lambda = 0.2$ استفاده شده است و اسپکتروگرام سیگنال بهسازی شده رسم شده است. با توجه به شکل ها، فیلتری که از تخمین سیگنال به نویز پسین استفاده می کند، نویز باقیمانده ی زیادی دارد. استفاده از سیگنال به نویز پیشین، نویز باقیمانده را کم می کند؛ اما بخشی از سیگنال را نیز تخریب می کند. شکل های "ه" و

همان طور که در شکل فوق مشاهده می شود، در مواردی که Rpr کوچک است و Rpo مقدار بزرگی دارد، مقدار بهره کوچک می باشد. به علاوه، در مواقعی که هر دو مقدار Rpo و Rpr کوچک بوده اند، مقدار بهره ی این فیلترها بر خلاف شکل ۲ اندک می باشد؛ در حالتی که هر دو تخمین Rpo و Rpr کوچک هستند؛ این دو تخمین همدیگر را تأیید می نمایند و با اطمینان بیشتری می توان گفت که مؤلفه های سیگنال ورودی نویزی بوده و تضعیف آن ها در فرکانس داده شده و فریم داده شده توسط فیلتر توجیه پذیر است. این رفتار به عنوان مزیت روش ترکیبی محسوب می شود. در شکل (۳) الف وزن کمتری به Rpo داده شده است، در حالی که در شکل های (ب) وزن بیشتری به Rpo داده شده است. با بیشتر شدن وزن Rpo در مواقع عدم توافق Rpo و Rpr ، کمیت آخر یعنی Rpo ملاک اصلی قضاوت فیلتر برای تعیین بهره قرار می گیرد که چندان مطلوب نمی باشد و مزایای ذکر شده در مورد شکل ۲ را از دست می دهد. لذا λ های بزرگ مطلوب نخواهد بود.

بیشتری خواهد بود. درعوض مؤلفه‌های سیگنال تمیز نیز کمتر حذف می‌شوند. به عبارت دیگر با تغییر پارامتر λ تنظیم مطلوب دو اثر اعوجاج اندک سیگنال و نویز باقیمانده امکان‌پذیر شده است.



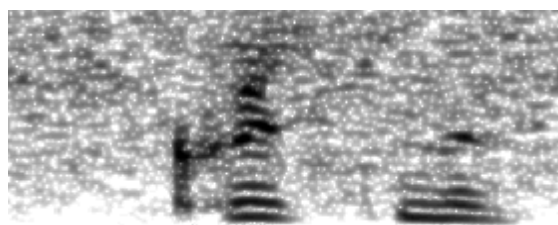
ب: سیگنال نویزی



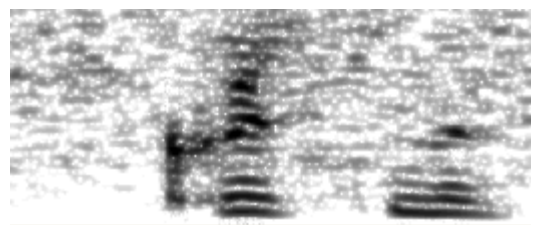
الف: سیگنال تمیز



د: سیگنال بهسازی شده با تخمین پسین (روش AssRpr جدول ۱)



ج: سیگنال بهسازی شده با تخمین پسین (روش AssRpo جدول ۱)



و: سیگنال بهسازی شده با روش پیشنهادی با $\lambda = 0.2$



ه: سیگنال بهسازی شده با روش پیشنهادی با $\lambda = 0.1$

شکل ۴. اسپکتروگرام سیگنال‌های تمیز، نویزی و بهسازی شده با روش‌های مختلف برای ارزیابی روش پیشنهادی تخمین SNR

می‌یابد [۱۶]. در الگوریتم پیشنهادی سطح نویز باقیمانده در حد مطلوبی قرار می‌گیرد. این امر موجب می‌شود که سیگنال توسط الگوریتم حذف کمتر آسیب ببیند و به عبارت دیگر اعوجاج کمتری پیدا کند در عین حال نویز دریافتی توسط شنونده نیز در حد نویز محیط باشد.

در این مقاله برای دستیابی به سطح نویز باقیمانده‌ی معین در روش‌های حذف نویز یا به تعبیر دیگر برای کنترل سطح توان نویز قسمتی از نویز اولیه به‌عنوان سیگنال مطلوب لحاظ می‌شود. در این روش فیلترهای معرفی شده علاوه بر SNR توابعی از یک پارامتر دیگر، α ، هستند که این

۴- پیشنهاد روش‌های حذف نویز با سطح نویز قابل تنظیم

در این بخش نحوه‌ی اصلاح تخمین سیگنال به نویز به‌منظور دستیابی به سطح نویز باقیمانده معین بررسی می‌شود. در روش‌های معمول حذف نویز، فیلترها چنان تنظیم می‌شوند که تمام نویز زمینه حذف شود؛ اما مسأله آن است که حذف کامل نویز در محیط‌هایی که سطح نویز معینی وجود دارد چندان مفید نخواهد بود. علاوه بر آن در سیستم حذف نویز با افزایش حذف نویز اعوجاج نیز افزایش

$$H(f, n) = \sqrt{\frac{SNR(f, n) + \alpha^2}{1 + SNR(f, n)}} \quad (19)$$

$$H(f, n) = 1 - \sqrt{\frac{1 + (\alpha^2 - 2\alpha)}{1 + SNR(f, n)}} \quad (20)$$

$$H(f, n) = \frac{SNR(f, n) + \alpha}{1 + SNR(f, n)} \quad (21)$$

جهت ارزیابی کارایی روش پیشنهادی در زیربخش ۳-۱ برای تخمین SNR و مقایسه آن با فیلترهای دیگر در هنگام کنترل میزان کاهش نویز، این روش به همراه فیلتر تفاضل دامنه طیف در این قسمت بکار گرفته شده است. فیلتر به کار رفته در این قسمت به صورت زیر تعریف می شود.

$$H_{comb}(f, n) = 1 - \sqrt{\frac{1 - \alpha}{1 + R_{comb}(f, n)}} \quad (22)$$

که در آن $R_{comb}(f, n)$ از رابطه ی (۱۶) به ازای $\lambda = 0.1$ به دست می آید.

$$R_{comb}(f, n) = R_{po}^{0.1}(f, n).R_{pr}^{0.9}(f, n) \quad (23)$$

نتایج پیاده سازی روش های فوق الذکر در بخش پنجم، گزارش می گردد.

۴-۱ شکل دهی نویز باقی مانده

در روش های کنترل سطح نویز باقیمانده که تاکنون بررسی شد، فیلتر حذف نویز به گونه ای عمل می کنند که نویز باقیمانده ضریبی از نویز زمینه باشد. کنترل سطح نویز به شیوه ای که بیان شد، (ضریب α)، اثرات مطلوبی در کاهش میزان اعوجاج سیگنال دارد، اما در پاره ای موارد برای برخی از انواع نویزها با توجه به آن که شکل طیفی نویز تغییر نمی کند، از میزان آزاردهندگی نویز چندان کاسته نمی شود. به عبارت دیگر در مورد بعضی نویزها، علاوه بر توان نویز شکل طیفی نویز در شنوایی و میزان آزاردهندگی تأثیر دارد. در این بخش روش هایی معرفی می شود که نویز باقیمانده را شکل دهی کند. در این روش ها هدف آن است که نویز باقیمانده از نظر شنونده خوشایند باشد. جهت دستیابی به این امر نویز باقیمانده با یک نویز مطلوب جایگزین می شود. در صورتی که بخواهیم در روش های کنترل سطح نویز امکان شکل دهی نویز به شکل مطلوب را داشته باشیم، در رابطه ی (۱۹) تا (۲۱)، پارامتر α تابعی از f می شود. در این حالت مثلاً برای تفاضل طیف دامنه، رابطه به صورت زیر نوشته می شود:

پارامتر سطح نویز باقیمانده را تعیین می کند. در صورتی که سطح مشخصی از حذف نویز، مورد نظر باشد باید بخشی از نویز زمینه نیز در $\hat{S}$ باقی بماند. به عبارت دیگر در این حالت سیگنال تخمینی مطلوب، به صورت زیر می باشد:

$$\hat{S}(f, n) = S(f, n) + \alpha B(f, n) \quad (17)$$

که در آن α میزان نویز باقیمانده را تعیین می کند و $S(f, n)$ ، $B(f, n)$ و $\hat{S}(f, n)$ به ترتیب دامنه ی طیف سیگنال تمیز، نویز و سیگنال تخمینی مطلوب می باشند؛ در این حالت مقدار افزایش نسبت سیگنال به نویز در سیگنال تخمینی یا بهبود یافته نسبت به سیگنال نویزی با G نشان داده می شود و برابر خواهد بود.

(۱۸)

$$G = 10 \log \frac{SNR(enhanced)}{SNR(noisy)} = 10 \log \frac{S^2 / \alpha^2 B^2}{S^2 / B^2} = -20 \log \alpha$$

که در آن $SNR(enhanced)$ نسبت سیگنال به نویز در سیگنال بهسازی شده و $SNR(noisy)$ نسبت سیگنال به نویز در سیگنال نویزی اولیه می باشد. به عنوان مثال برای دستیابی به بهبودی برابر 10db در سیگنال به نویز یا به عبارت دیگر کاهش 10db در نویز اولیه، مقدار α برابر 0.32 تعیین می شود.

در روش هایی که تمامی نویز زمینه را حذف می کنند، α برابر صفر خواهد بود. تخمین سیگنال به نویز، به منظور تخمین سیگنال با کاهش مشخص سطح نویز باقیمانده برای چند فیلتر ارائه شده است.

در بخش دوم، فیلتر اصلاح طیف فرکانسی برای چند روش متداول با فرض حذف کامل نویز برحسب SNR محاسبه گردید. در این بخش همان روابط برای حذف قسمتی از نویز محاسبه خواهد شد. به طوری که با مقدار مشخص برای G کاهش نویز، فیلتر اصلاح طیف فرکانسی تعیین شود. در مورد فیلترهای فرکانسی مطرح شده در بخش ۲، در صورتی که در این فیلترها حذف کامل نویز مورد نظر نباشد، بلکه باقی ماندن α برابر نویز اولیه در سیگنال مطلوب مطابق رابطه (۱۷) مورد نظر باشد، روابط فیلترها با اضافه کردن پارامتر α اصلاح می شوند. به این ترتیب روابط (۱۹)، (۲۰)، (۲۱) برای روش های تفاضل طیف توان، تفاضل طیف دامنه و فیلتر وینر محاسبه می شوند. مقدار پارامتر α بیانگر قسمتی از نویز است که در سیگنال مطلوب باقی مانده است. α برابر صفر همان فیلتر اولیه را می دهد، در حالیکه مقدار α برابر یک معادل عدم اعمال فیلتر اصلاح طیف فرکانسی است.

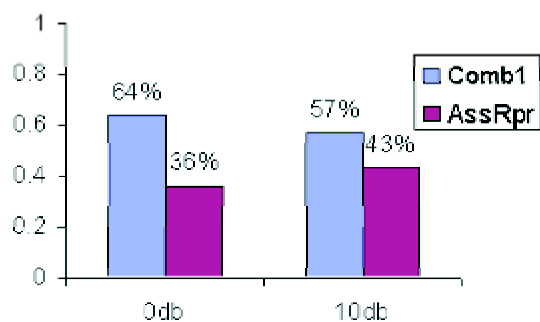
ضریب فراموشی است که برابر ۰/۹۸ در نظر گرفته شده است.

روش‌های پیاده‌سازی شده در سه زیربخش با هم مقایسه شده‌اند. در زیر بخش اول نتایج مربوط به روش‌های تخمین سیگنال به نویز ارائه شده است، زیربخش دوم مربوط به روش‌های کنترل سطح نویز می‌باشد و زیربخش سوم به ارزیابی شکل‌دهی نویز اختصاص دارد.

۵-۱ ارزیابی روش‌های تخمین سیگنال به نویز محلی

جهت انجام مقایسه‌ی روش‌های مختلف تخمین نویز، بهره‌ی سیگنال به نویز گفتار بهبود یافته برای روش‌های مختلف و نویزهای متفاوت در جدول (۲) نشان داده شده است. داده‌های جدول با میانگین‌گیری روی نتایج حاصله از آزمایش روی چهار جمله به‌دست آمده است. در این جدول منظور از Comb1، Comb2 و Comb3 همان روش پیشنهادی برای تخمین SNR (رابطه‌ی (۱۶) به همراه فیلتر تفاضل دامنه‌ی طیف می‌باشد که به‌ترتیب برای $\lambda = 0.1$ ، $\lambda = 0.2$ و $\lambda = 0.3$ محاسبه شده‌اند. نتایج به‌دست‌آمده نشان می‌دهد که روش پیشنهادی از نظر نسبت سیگنال به نویز گفتار پردازش شده با روش‌های دیگر قابل مقایسه است.

برای مقایسه‌ی روش پیشنهادی تخمین نسبت سیگنال به نویز با روش decision-directed، یک تست شنوایی غیر رسمی با چهار شنونده انجام شده است. در این آزمایش فیلتر تفاضل دامنه‌ی طیف به‌همراه هر یک از روش‌های تخمین SNR فوق به‌کار برده شد. آزمایش (۵) روی چهار جمله از دو گوینده مرد انجام شد. فایل‌های انتخابی با نویزهای سفید، ماسکین و همهمه به‌ازای SNRهای 0db و 10db ترکیب شده بودند. در این آزمایش از شنوندگان درخواست شد این دو روش تخمین SNR را با هم مقایسه و درصد ترجیح هر یک را مشخص نمایند. شکل (۶) درصد انتخاب هر روش را توسط شنوندگان در دو نسبت سیگنال به نویز نشان می‌دهد.



شکل ۶. نتایج تست شنوایی برای مقایسه Comb1 و AssRpr

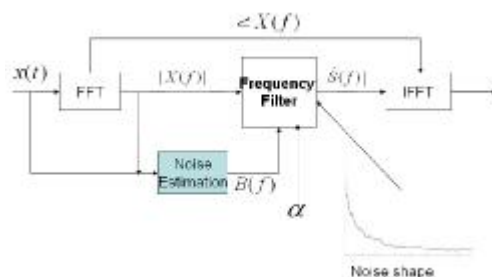
$$H(f) = \sqrt{\frac{SNR(f) + \alpha(f)^2}{1 + SNR(f)}}, \alpha(f) = \frac{|D(f)|}{|B(f)|} \quad (24)$$

در رابطه‌ی فوق $D(f)$ نویز مطلوب با شکل طیفی از پیش تعیین شده است. چنانچه گفته شد نویز مطلوب به‌گونه‌ای انتخاب می‌شود که از نظر شنوایی خوشایند باشد. در عمل $\alpha(f)$ در هر فریم محاسبه می‌شود.

$$\alpha(f, n) = \begin{cases} \frac{|D(f, n)|}{|B(f, n)|} & |D(f)| \leq |B(f)| \\ 1 & |D(f)| > |B(f)| \end{cases} \quad (25)$$

در رابطه‌ی فوق در فرکانس‌هایی که سطح نویز مطلوب از نویز باقیمانده بیشتر است، سطح نویز باقیمانده تغییری نمی‌کند.

شکل (۵) بلوک دیاگرام پیشنهادی برای سیستم بهبود گفتار را نشان می‌دهد. پارامتر α چنان‌که پیشتر گفته شد، می‌تواند با توجه به سطح نویز محیط تعیین شود. شکل نویز انتخاب شده جهت شکل‌دهی چنان انتخاب می‌شود که برای گوش خوشایند باشد. انتخاب نویزهای پایین‌گذر برای این منظور مناسب به نظر می‌رسد.



شکل ۵. بلوک دیاگرام پیشنهادی برای شکل‌دهی نویز در بخش بعد روش‌های شکل‌دهی نویز ارزیابی می‌شود.

۵-۲ ارزیابی

جهت پیاده‌سازی و ارزیابی فیلترهای اصلاح شده، طول پنجره‌ی پردازش برابر ۳۲ میلی‌ثانیه و فرکانس نمونه‌برداری برابر ۸۰۰۰ هرتز، همپوشانی ۵۰٪ و پنجره‌ی هنینگ انتخاب شده است. تخمین توان نویز از روی فایل نویز صورت گرفته است. هدف از این عمل، کاهش اثرگذاری خطای تخمین نویز در ارزیابی و مقایسه‌ی روش‌ها بوده است. طیف توان نویز به‌کمک رابطه‌ی زیر از روی فایل نویزی میانگین‌گیری شده است:

$$P_{noise}(f, n) = \lambda P_{noise}(f, n-1) + (1-\lambda)|D(f, n)|^2 \quad (26)$$

که در آن $P_{noise}(f, n-1)$ طیف توان نویز تخمین زده شده در فریم قبلی می‌باشد و $|D(f, n)|^2$ تخمین دوره‌ی کوتاه طیف نویز در فریم کنونی است که از روی ۱۰ فریم درست قبل از شروع هر جمله مورد تست تخمین زده شده است. λ یک

جدول ۲ بهبود در نسبت سیگنال به نویز گفتار پردازش شده بر حسب dB.

الف) نویز سفید			ب) نویز همه‌په			ج) نویز اتومبیل		
0db	10db		0db	10db		0db	10db	
6.7	3.7	AssRpr	5.1	2.1	AssRpr	13.5	10.4	AssRpr
9.7	4.9	AssRpo	8.7	4.4	AssRpo	11.3	6.2	AssRpo
8.3	4.0	2-step	5.4	1.4	2-step	14.4	11.4	2-step
7.3	4.2	Comb1	5.6	2.5	Comb1	14	10.8	Comb1
9	5.5	Comb2	7.5	4	Comb2	13.2	8.5	Comb2
9.3	5.6	Comb3	7.9	4.2	Comb3	12.7	7.8	Comb3

۲-۵ ارزیابی روش‌های حذف نویز با کنترل سطح نویز باقی مانده

در روش‌های ارزیابی به عمل آمده برای روش‌های حذف کنترل شده نویز از دو معیار افزایش سیگنال به نویز^۱ و کاهش میزان نویز باقیمانده^۲ استفاده شده است. این معیارها برای تعیین دقت عملکرد سیستم‌های بهبود گفتار، مورد استفاده قرار گرفته‌اند [۱۷] و مورد تأیید سازمان ETSI نیز هستند. برای محاسبه‌ی این معیار ابتدا سیگنال، به سه سطح مختلف نویز تقسیم می‌شود، سپس در هر سطح نویز، نسبت سیگنال به نویز به طور جداگانه محاسبه می‌شود. آنگاه از سیگنال به نویزهای به دست آمده به طور وزن دار (با توجه به تعداد فریم‌های هر سطح) میانگین‌گیری می‌شود. در مورد NPLR، میزان کاهش توان سیگنال در نواحی غیر گفتاری محاسبه می‌شود. جزئیات بیشتر در مرجع [۱۷] موجود است. در آزمایش‌های انجام شده در این زیربخش فیلترهای اصلاح طیف برای دستیابی به یک میزان کاهش نویز (NPLR) مشخص تنظیم می‌شوند، سپس میزان SNRI به دست آمده در آن NPLR برای روش‌های مختلف با هم مقایسه می‌شود.

آزمایش‌های انجام شده روی چهار جمله از دو گوینده‌ی زن و دو گوینده‌ی مرد انجام شده است. جملات از پایگاه داده‌ی TIMIT انتخاب شده است. به هرکدام از جملات سه نوع نویز سفید، همه‌په^۳ و اتومبیل^۴ با نسبت‌های سیگنال به نویز مختلف اضافه شده است. نویزها از پایگاه داده‌ی Noise92 انتخاب شده اند.

همان‌طور که توضیح داده شد، معیار NPLR بیانگر کاهش سطح نویز موجود در یک فرآیند بهبود گفتار است

که با توجه به اطلاعات لحظات سکوت محاسبه می‌شود. از سوی دیگر، SNRI بیانگر افزایش SNR پس از انجام بهبود گفتار است که از روی کل سیگنال محاسبه می‌شود. در آزمایش‌های انجام شده توانایی روش‌ها در دستیابی به SNRI بیشتر با NPLR یکسان مقایسه شده است. همان‌طور که مشاهده می‌شود، مقدار SNRI به دست آمده برای هر SNR ورودی و هر، با مقدار NPLR مربوطه تقریباً یکسان می‌باشد و کمتر از 3db با آن تفاوت دارد که قابل چشم‌پوشی می‌باشد.

برای ارزیابی شهودی^۵ روش‌های پیشنهادی یک تست شنوایی با چهار شنونده صورت گرفته است. ویژگی‌های مورد توجه در این ارزیابی، وجود پژواک^۶، نویز موسیقی، ایجاد هرگونه اثر غیر طبیعی در نویز زمینه^۷ علاوه بر نویز و ایجاد اثرات غیر طبیعی در گفتار^۸ بوده است. از شنوندگان خواسته شده بود که اعداد ۱ و ۲ و ۳ را به ترتیب به غیر قابل درک بودن^۹، قابل درک بودن ولی غیر آزاردهنده و آزاردهنده بودن به هر یک از ویژگی‌های مذکور نسبت بدهند. باید توجه کرد که برخلاف روش نمره‌دهی MOS در روش نمره-دهی حاضر نمره‌ی کمتر به معنی بهتر بودن روش است. در نهایت، نمرات چهار شنونده، میانگین‌گیری شده‌اند. فایل‌های بهسازی شده، برای هر یک از نویزهای سفید، ماشین و همه‌په و برای یک SNR ورودی و NPLR مشخص، با هم مقایسه شده‌اند که نتایج آنها در شکل‌های ۸ تا ۱۰ آورده شده‌اند.

⁵ - subjective

⁶ - echo

⁷ - background artifact

⁸ - speech artifact

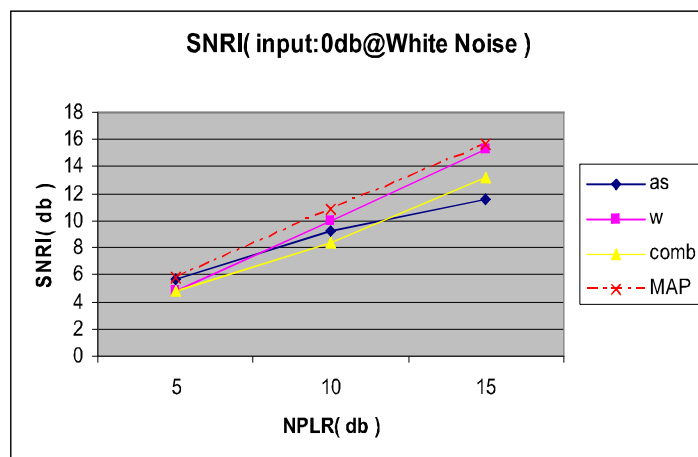
⁹ - not noticeable

¹ - Signal to Noise Ratio Improvement

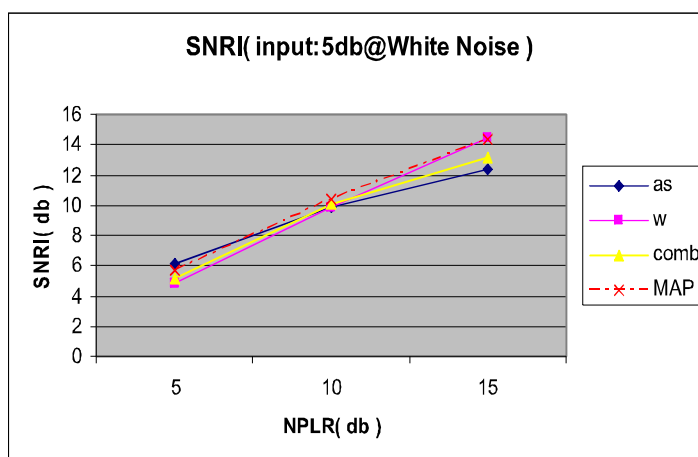
² - Noise Power Level Reduction

³ - babble

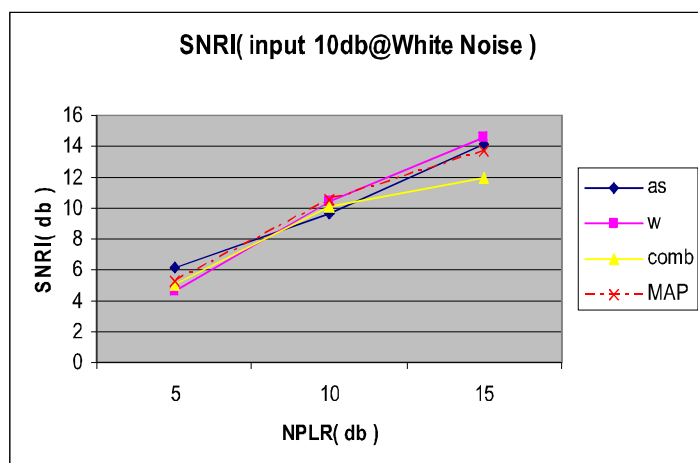
⁴ - Volvo



input SNR = 0db (الف)

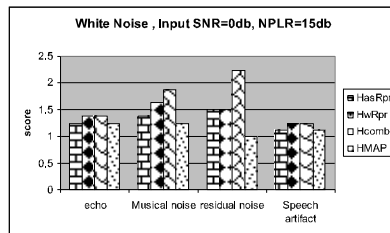


input SNR = 5db (ب)

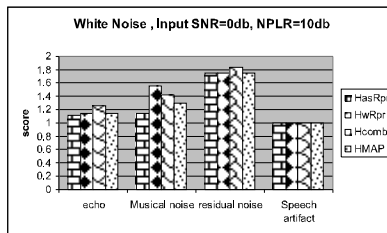


input SNR = 10db (ج)

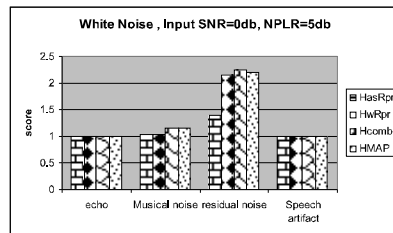
شکل ۷. SNRI بدست آمده برای چهار روش انتخابی به ازای NPLRهای مختلف برای نویز سفید و SNRهای ورودی مختلف



NPLR = 15db (ج)

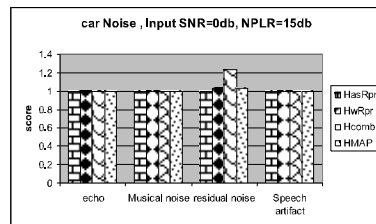


NPLR = 10db (ب)

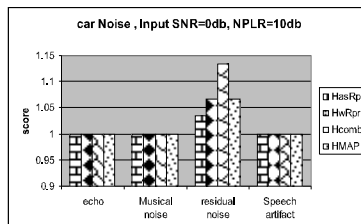


NPLR = 5db (الف)

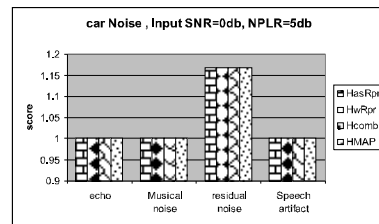
شکل ۸. نتایج تست شنوایی برای چهار روش انتخابی به ازای NPLRهای مختلف برای نویز سفید با input SNR=0db



NPLR = 15db (ج)

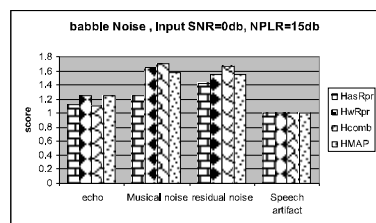


NPLR = 10db (ب)

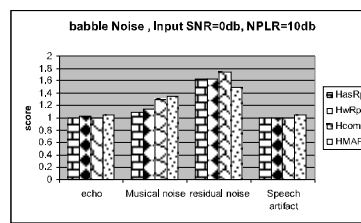


NPLR = 5db (الف)

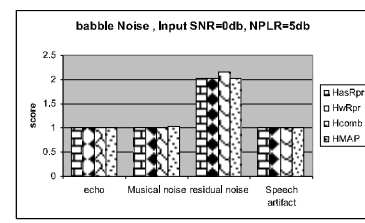
شکل ۹. نتایج تست شنوایی برای چهار روش انتخابی به ازای NPLRهای مختلف برای نویز ماشین با input SNR=0db



NPLR = 15db (ج)



NPLR = 10db (ب)



NPLR = 5db (الف)

شکل ۱۰. نتایج تست شنوایی برای چهار روش انتخابی به ازای NPLRهای مختلف برای نویز همهمه با input SNR=0db

باقیمانده برای کاهش 10db در نویز ورودی آزمایش شده‌اند. در ابتدا میزان حذف نویز 10db بدون شکل‌دهی نویز مورد نظر قرار گرفته است، سپس آزمایش فوق با شکل‌دهی نویز باقی‌مانده تکرار شده است. نویزهای ورودی واقعی مورد نظر قرار گرفته‌اند. در این حالت توان نویز باقی‌مانده برابر توان نویز باقی‌مانده در حالت قبلی است؛ به عبارت دیگر نویز باقی‌مانده با یک نویز پایین‌گذر جایگزین شده است. در نویز پایین‌گذر انتخاب شده، فرکانس قطع 4KHz و شکل طیف به صورت خطی است، توان نویز در 0khz برابر 30db و در 4khz برابر 0db است.

شکل‌های (۱۱) و (۱۲) نسبت بهبود سیگنال به نویز را زمانی که از شکل‌دهی استفاده می‌کنیم و زمانی که از شکل‌دهی استفاده نمی‌کنیم، نشان می‌دهند. نتایج آزمایش برای ۵ سطح نویز مختلف (0db, 5db, 10db, 15db و 20db) برای ۵ نوع نویز مختلف گزارش شده است. نویزهای بالاگذر و پایین‌گذر موجود در جدول به طور مصنوعی ایجاد شده‌اند. فیلتر مورد استفاده در این آزمایش یک فیلتر تفاضل

مقایسه‌های انجام شده، به صورت نسبی است؛ بدین معنی که در هر NPLR و هر نوع نویز، چهار روش بهبود گفتار با هم مقایسه شده‌اند. به عبارت دیگر هر نمودار به تنهایی معنا دارد و مقادیر دو نمودار قابل مقایسه نیست. همان‌گونه که مشاهده می‌شود، بیشترین اثرات^۱ مربوط به نویز موزیکال و آثار غیر طبیعی نویز زمینه^۲ می‌باشند و تفاوت‌ها از نظر تخریب اندک می‌باشد. همچنین مشاهده می‌شود که کمترین میزان نویز موزیکال و نویز باقیمانده مربوط به روش تفاضل طیف دامنه می‌باشد. پس از آن روش‌های MAP و وینر از نویز موزیکال و باقیمانده‌ی یکسانی برخوردارند و در نهایت Hcomb بیشترین میزان نویز موزیکال و باقیمانده را در بین چهار روش مورد مقایسه به خود اختصاص داده است.

۳-۵ ارزیابی روش‌های شکل‌دهی نویز

هدف این زیربخش ارزیابی شکل‌دهی نویز باقیمانده است. برای این منظور روش‌های حذف نویز با کنترل سطح نویز

¹ artifact

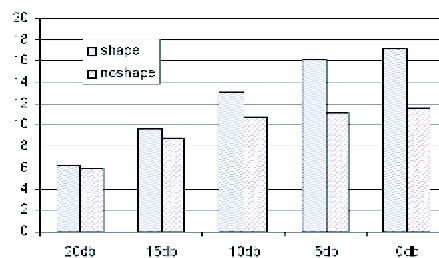
² background artifact

افزودن پارامتر کنترل بهره‌ی فیلتر به فیلترهای اصلاح طیفی می‌توان به کاهش سطح نویز دلخواه دست یافت.

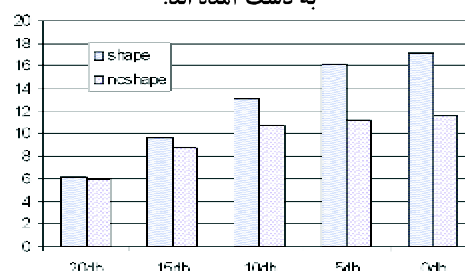
برای ارزیابی از دو معیار NPLR و SNRI استفاده شده است. این دو معیار جهت اندازه‌گیری بهره‌ی سیگنال به نویز و میزان نویز باقیمانده در سیگنال به کار رفته است و آزمایش‌های نشان دهنده‌ی هم‌جهت بودن نتایج حاصل از این معیارها و تست‌های شنیداری می‌باشد.

در تخمین سیگنال به نویز، نتایج حاصله از تست شنیداری و همچنین شکل طیف‌های رسم شده، نشان دهنده‌ی مفید بودن روش ترکیبی بوده است. آزمایش‌های مربوط به کنترل سطح نویز نشان دادند که روش تفاضل طیف دامنه می‌تواند، با تنظیم پارامتر کنترل بهره، محدودده‌ی وسیعی از کاهش سطح نویز (از 5db تا بیش از 20db) را در پی داشته باشد. انجام تست‌های شنوایی نیز قابل قبول بودن نتایج حاصله را از دو جنبه نویز زمینه و کیفیت سیگنال صوتی تأیید نمود.

طیف دامنه بوده و فیلتر به‌گونه‌ای تنظیم شده است که سطح نویز در سیگنال بهبود یافته 10db کاهش یابد.



شکل ۱۱. افزایش سیگنال به نویز در سیگنال به نویزهای مختلف با استفاده از شکل دهی نویز و بدون استفاده از شکل دهی نویز. نتایج با میانگین گیری روی انواع نویزهای مختلف به دست آمده اند.



شکل ۱۲. افزایش سیگنال به نویز با حضور نویزهای مختلف با استفاده از شکل دهی نویز و بدون استفاده از شکل دهی نویز. نتایج با میانگین گیری روی سیگنال به نویزهای مختلف به دست آمده اند.

۸- مراجع

- [1] S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 113-120, Apr. 1979.
- [2] Y. Ephraim, D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 32, pp. 1109-1121, Dec. 1984.
- [3] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. 7th Eur. Signal Process. Conf. (EUSIPCO'94)*, pp. 1182-1185, Edinburgh, UK, 13-16 September, 1994.
- [4] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.* vol. 9, pp. 504-512, July 2001.
- [5] I. Cohen, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Process. Lett.*, vol. 9, no. 1, pp. 12-15, January 2002.
- [6] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, pp. 466 - 475, September 2003.
- [7] R. J. McAulay, M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Transactions on*

۷- نتیجه‌گیری

فیلترهای اصلاح طیفی را می‌توان به‌صورت تابعی از سیگنال به نویز نشان داد. نسبت سیگنال به نویز برای یک نمونه فرکانسی در یک فریم داده شده از یک سیگنال گفتار محاسبه می‌شود و نقش مهمی در محاسبه‌ی فیلتر اصلاح فرکانس دارد. روش‌های قبلی برای تخمین این نسبت در مقاله‌ی حاضر بررسی و پیاده‌سازی شده‌اند و نقش تخمین سیگنال به نویز در روش‌های حذف نویز مورد بررسی قرار گرفته است. به‌علاوه یک روش ترکیبی برای تخمین سیگنال به نویز پیشنهاد شده است. در روش پیشنهادی تخمین نویز تابعی از سیگنال به نویز پیشین و پسین بوده و نسبت به تغییرات هر دو حساس است. با استفاده از روش پیشنهادی می‌توان میزان نویز باقیمانده و اغتشاش را به نحو مطلوبی تنظیم کرد. در این مقاله همچنین، استفاده از حذف مقدار کنترل شده‌ی نویز از سیگنال نویزی به‌عنوان روشی برای کاهش اعوجاج مطرح شده است و نشان داده شد که با



محسن رحمانی دوره‌ی کارشناسی را در رشته‌ی مهندسی کامپیوتر-سخت افزار از دانشگاه صنعتی شیراز در سال ۱۳۷۸ به پایان رساند. وی درجه‌ی کارشناسی ارشد و دکترا را در رشته مهندسی کامپیوتر- معماری کامپیوتر از دانشگاه علم و صنعت ایران به ترتیب در سال های ۱۳۸۰ و ۱۳۸۷ دریافت کرد. زمینه‌ی تحقیقاتی وی پردازش و بهسازی سیگنال گفتار است. او در حال حاضر عضو هیئت علمی دانشکده‌ی مهندسی دانشگاه اراک می باشد.

نشانی رایانامک (پست الکترونیکی) ایشان عبارت است از:

m_rahmani@iust.ac.ir



روح ا. عبدی‌پور دوره‌ی کارشناسی را در رشته‌ی مهندسی کامپیوتر-نرم افزار و دوره‌ی کارشناسی ارشد را در رشته‌ی مهندسی کامپیوتر- هوش مصنوعی از دانشگاه علم و صنعت ایران به ترتیب در سال های ۱۳۸۰ و ۱۳۸۲ دریافت کرد. زمینه‌ی تحقیقاتی وی پردازش و بهسازی سیگنال گفتار و پایگاه داده است. او در حال حاضر عضو هیئت علمی دانشگاه ملایر می باشد.

نشانی رایانامک (پست الکترونیکی) ایشان عبارت است از:

r_abdipour@iust.ac.ir



بابک ناصرشریف متولد تهران در سال ۱۳۵۳ است. وی دوره‌ی کارشناسی را در رشته‌ی مهندسی کامپیوتر- سخت افزار از دانشگاه صنعتی امیرکبیر در سال ۱۳۷۶ به پایان رساند. وی درجه‌ی کارشناسی ارشد را در رشته‌ی مهندسی کامپیوتر- هوش مصنوعی در سال ۱۳۷۹ از دانشگاه علم و صنعت ایران و درجه‌ی دکترا را در همین شاخه در سال ۱۳۸۶ و از همین دانشگاه دریافت کرد. رشته‌ی تخصصی وی پردازش و بازشناسی سیگنال گفتار است. او در حال حاضر عضو هیئت علمی گروه مهندسی دانشگاه گیلان می باشد.

نشانی رایانامک (پست الکترونیکی) ایشان عبارت است از:

nasser_s@iust.ac.ir

- Acoustics, Speech, and Signal Processing*, vol.28, no. 2, pp. 137-145, Apr. 1980
- [8] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Transactions on Speech and Audio Processing*, Vol. 2, pp 345 - 349, Apr 1994.
- [9] P. Scalart, J.V. Filho, "Speech enhancement based on a priori signal to noise estimation," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp 629 - 632, May 1996.
- [10] T. Lotter and P. Vary, "Noise Reduction by Joint Maximum a Posteriori Spectral Amplitude and Phase Estimation with Super-Gaussian Speech Modeling," *European Signal Processing Conference*, Sep 2001.
- [11] P.J. Wolfe, S.J. Godsill, "simple alternative to the Ephraim and Malah suppression rule for speech enhancement," *11th IEEE Workshop on Statistical Signal Processing*, pp 496-499, Aug 2001.
- [12] P. J. Wolfe and S. J. Godsill. "Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement," *EURASIP Journal on Appl. Sig. Processing*, pp 1043-1051, 2003.
- [13] X. Huang, A. Acero, and H. W. Hon, *Spoken Language Processing: A Guide to Theory, Igorithm, and System Development*, Prentice Hall PTR, Upper Saddle River, N.J., 2001.
- [14] Y. Ephraim, H.L. Ari, W.J.J. Roberts, *A brief survey of speech enhancement*, in *The Electronic Handbook*, Second edition, CRC Press, pp.2088-2097, 2003.
- [15] I. Cohen, "On the Decision-Directed Estimation Approach of Ephraim and Malah," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.1, pp 293-296. May 2004.
- [16] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Proc.*, vol. 3, pp. 251-266, July 1995.
- [17] E. Paaanen, B. Ayad, V.V. Mattila, "New Objective Measures for Characterization of Noise Suppression Algorithms," *IEEE Workshop on Speech Coding*, pp.23-25, 2000.

[۱۸] ا. اکبری، "حذف نویز صوتی از سیگنال گفتار بر اساس استفاده از نسبت سیگنال به نویز پیشین"، سومین کنفرانس بین‌المللی سالیانه انجمن کامپیوتر ایران، دی ۱۳



احمد اکبری دوره‌ی کارشناسی را در رشته‌ی مهندسی الکترونیک و کارشناسی ارشد را در رشته‌ی مخابرات از دانشگاه صنعتی اصفهان به‌ترتیب در

سال ۱۳۶۵ و ۱۳۶۸ به پایان رساند. وی درجه‌ی DHET در شبکه‌های کامپیوتری را از ENSEEIHT در شهر تولوس فرانسه کسب کرد. ایشان همچنین درجات DEA و PhD را در ارتباطات و پردازش سیگنال از دانشگاه رنس (Renes 1, France) در سال‌های ۱۳۷۱ و ۱۳۷۴ اخذ کرد. وی در سال ۱۳۷۵ به دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران ملحق شد. در حال حاضر ایشان دانشیار و رئیس مرکز فن‌آوری اطلاعات این دانشگاه می‌باشد. زمینه‌های تحقیقاتی مورد علاقه‌ی وی شامل مدلسازی اکوستیکی گفتار، بهسازی گفتار و بازشناسی گفتار می‌باشد.

نشانی رایانامک (پست الکترونیکی) ایشان عبارت است از:

akbari@iust.ac.ir