

طراحی تخمین گر بیشینه در ستمنمایی در بهسازی گفتار مبتنی بر کتاب کد با نسبت سیگنال به نویز منفی

رقیه دوست^۱

^۱استادیار پژوهشکده فناوری ارتباطات، پژوهشگاه ارتباطات و فناوری اطلاعات (مرکز تحقیقات مخابرات ایران) doost@itrc.ac.ir

چکیده

در این مقاله تخمین گر جدیدی بر اساس بیشینه درست‌نمایی برای بهسازی گفتار با روش سنتز مبتنی بر کتاب کد ارائه می‌شود. در روش بهسازی گفتار مبتنی بر کتاب کد، با انتخاب بهینه اندیس‌های کتاب کد گفتار، سیگنال گفتار بهسازی شده سنتز می‌شود و به این ترتیب در واقع جداسازی نویز و گفتار از یکدیگر صورت می‌گیرد. از این رو با این روش می‌توان گفتارهای نویزی، با نسبت سیگنال به نویز کمتر از صفر دسیبل را بهسازی نمود. البته در این روش انتخاب صحیح اندیس‌های کتاب کد بسیار مهم است. از این رو در این مقاله تخمین گر بیشینه درست‌نمایی با اعمال وزن‌های بهبود دهنده کیفیت شنیداری، برای گفتار و نویز طراحی می‌شود. رابطه به دست آمده برای این تخمین گر به عنوان تابع فاصله در طراحی کتاب‌های کد نیز استفاده می‌شود. این روش برای گونه‌های مختلف و نویزهای گوناگون شبیه‌سازی شد. نتایج نشان می‌دهد که گفتار بهسازی شده با استفاده از تخمین گر بیشینه درست‌نمایی با وزن‌های کیفیت شنیداری نسبت به تخمین گر فاصله اقلیدسی، کیفیت شنیداری بهتری دارد. همچنین روش ارائه شده در برخورد با نویزهای غیرایستاد یا ایستاد و نسبت سیگنال به نویز منفی (یا مثبت) موفق‌تر از روش‌های دیگر عمل می‌کند. هزینه بهسازی با کیفیت برتر در این روش، نیاز به زمان نسبتاً طولانی برای بهسازی است.

کلیدواژه

بهسازی گفتار، کتاب کد، نسبت سیگنال به نویز (SNR)، سنتز گفتار.

مقدمه

می‌کنند، روشهای پارامتری می‌باشند [۱ و ۲]. در روش‌ها و مدل‌های پارامتری ساختارهای قابل پیش بینی استفاده می‌شود. اگر اطلاعات اولیه‌ای از سیگنال گفتار در دسترس باشد با استفاده از الگوریتم‌هایی که مبتنی بر مدل انتخابی هستند، می‌توان نتایج بهتری از بهسازی گفتار به دست آورد. گرچه با انتخاب نامناسب یک مدل، نتایج نامطلوب به دست می‌آید. پس انتخاب مدل در این سیستم‌ها یک رکن اساسی و اثرگذار محسوب می‌شود.

در روش‌های تفریق طیفی تطبیقی [۳ و ۴]، فیلتر وینر تطبیقی [۴ و ۵]، زیرفضا [۶ و ۷] و فیلتر کالمن [۸ و ۹] نیاز به استفاده از روش‌های تخمین نویز می‌باشد. این در حالی است که روش‌های تخمین نویز در SNRهای پایین و نویزهای غیرایستاد کارایی خوبی ندارد [۱۰ و ۱۱]. در روش HMM،

با رشد روز افزون استفاده از سیستم‌های گفتاری در کاربردهای عملی و روزمره، نیاز به حفظ کیفیت گفتار امری اجتناب ناپذیر و ضروری می‌باشد. بدون استفاده از سیستم‌های بهسازی گفتار، عملکرد بسیاری از این سیستم‌ها به شدت پایین می‌آید. به این ترتیب مقوله بهسازی گفتار یکی از مباحث بسیار مهم پردازش گفتار می‌باشد.

یک شیوه از طبقه بندی سیستم‌های بهسازی گفتار بر اساس پارامتری یا غیر پارامتری بودن آن می‌باشد. روش‌هایی مثل تفریق طیفی^۱ که با پردازش بر روی خود گفتار نویزی به بهسازی آن می‌پردازند غیر پارامتری می‌باشند. روش‌هایی مثل HMM^۲ که از مدل‌های آماری برای گفتار و نویز استفاده

^۳ subspace

^۴ Signal to Noise Ratio

^۱ spectral subtraction

^۲ Hidden Markov Model

است. در نهایت کیفیت گفتار بهسازی شده در SNR های 0dB ، 5dB و 10dB مورد ارزیابی قرار گرفته است.

با توجه به بررسی‌های صورت گرفته، عمده روش‌های بهسازی گفتار برای مواجهه با SNRهای بیشتر از صفر دسیبل به کار رفته‌اند. در [۱۷] این مقوله مورد توجه قرار گرفته است و روش سنتز گفتار بهسازی شده مبتنی بر کتاب کد، برای بهسازی گفتارهای نویزی با نسبت سیگنال به نویز کمتر از صفر دسیبل پیشنهاد و در محیط شبیه‌سازی پیاده‌سازی شد. در این روش به ازای هر قاب گفتار نویزی و با استفاده از یک تخمین‌گر طراحی شده، سرچ بر روی اندیس‌های کتاب‌های کد گفتار و نویز انجام می‌شود. اندیس‌های مناسب کتاب کد گفتار انتخاب می‌شوند و گفتار بهسازی شده سنتز می‌شود. این شیوه‌ی بهسازی گفتار در واقع جداسازی دو منبع گفتار و نویز از یکدیگر می‌باشد [۱۸].

در [۱۹]، از PW_ED^{۱۲} به عنوان تابع فاصله در طراحی کتاب کد استفاده شد. این تابع فاصله در واقع تابع فاصله اقلیدسی وزن داده شده با ضرایب کیفیت شنیداری می‌باشد. در [۱۷] از این تابع فاصله در طراحی کتاب کد و نیز به عنوان تخمین‌گر، در پیدا کردن اندیس‌های بهینه‌ی کتاب کدها استفاده شده است.

در مقاله حاضر نیز از روش سنتز مبتنی بر کتاب کد برای بهسازی گفتار استفاده می‌کند. با این تفاوت که در آن تخمین‌گر پیشینه درست‌نمایی^{۱۳} با وزن‌های شنیداری^{۱۴} برای بخش حقیقی و موهومی^{۱۵} STFT طراحی و بررسی می‌شود. در روش سنتز مبتنی بر کتاب کد، تخمین‌گر مورد استفاده در جستجوی اندیس‌های کتاب کدها بسیار حائز اهمیت می‌باشد. چرا که اندیس‌های انتخاب شده از کتاب کد گفتار به منظور سنتز گفتار بهسازی شده مورد استفاده قرار می‌گیرد. به این ترتیب دقت تخمین‌گر در یافتن اندیس‌های صحیح و بهینه از کتاب کد گفتار موجب بهبود سنتز گفتار بهسازی شده با کیفیت بالاتر می‌شود. هم‌چنین تابع فاصله متناظر با این تخمین‌گر در طراحی کتاب کد گفتار و کتاب کد نویز به کار می‌رود. این خود مبین اهمیت دو چندان تخمین‌گر می‌باشد. چرا که بردارهای کتاب کد گفتار به طور مستقیم در سنتز گفتار بهسازی شده استفاده می‌شوند. پس با طراحی بهتر کتاب کد گفتار توسط تابع فاصله مناسب گفتار بهسازی شده با کیفیت بهتر سنتز می‌شود. در تخمین‌گر پیشینه درست‌نمایی از

ضرایب LP^۵ مربوط به گفتار و ضرایب LP نویز به عنوان اطلاعات ذخیره شده در کتاب کد مورد استفاده قرار گرفته‌اند [۱۱]. منتخبی از این ضرایب و واریانس‌هایی که با استفاده از آنها تخمین زده شد، برای تشکیل فیلتر وینر و بهسازی گفتار در قاب^۶ مورد نظر به کار گرفته شد. فیلترهای تشکیل شده تنها حاوی اطلاعات پوش می‌باشند و فاقد اطلاعات دقیق‌تری مثل پیچ می‌باشند. بنابراین برای SNRهای پایین کارایی لازم را ندارند.

روش بازسنتز مبتنی بر فهرست^۷ [۱۲] نیز در SNRهای زیر صفر دسیبل عملکرد خوبی از خود نشان نمی‌دهد، چرا که در SNRهای خیلی پایین در انتخاب فهرست دچار مشکل می‌شود. هم‌چنین در این روش در بخش پیش‌پردازش از فیلتر وینر استفاده کرده است که در SNRهای منفی عملکرد مطلوبی ندارد. از محدودیت‌های قابل توجه این روش این است که برای هر گفتار نویزی و برای هر نسبت سیگنال به نویز می‌بایست سیستم به طور جداگانه آموزش داده شود [۱۲].

در [۱۳] به ارایه روشی مبتنی بر مدل برای بهسازی گفتار پرداخته است که در آن بهبود قابلیت فهم^۸ سمعک‌های دوتایی مورد توجه بوده است. بهسازی در این مقاله بر اساس فیلتر کالمن می‌باشد. این روش برای SNRهای بیشتر از 0dB عملکرد خوبی از خود نشان داده است.

در [۱۴] روشی مبتنی بر یادگیری دیکشنری^۹ و تجزیه ماتریس مرتبه پایین^{۱۰}، ارایه داده است. با روش پیشنهاد شده کیفیت و قابلیت فهم گفتار بهسازی شده نسبت به روش‌های مورد بررسی دیگر در SNRهای بالاتر از 0dB بهبود داشته است.

بهسازی گفتار با استفاده از شبکه‌های عصبی نیز در [۱۵] برای گفتارهای نویزی با SNRهای مثبت مورد بررسی قرار گرفته است.

در [۱۶] با استفاده از روش فاکتورگیری ماتریس غیرمنفی^{۱۱} (NMF) به بهسازی گفتار پرداخته است. در این مقاله برای مدلسازی دامنه طیف گفتار تمیز از کتاب کد استفاده کرده است. الگوریتم کلاسیک تخمین نویز نیز به منظور تخمین چگالی طیف توان نویز مورد استفاده قرار گرفته

^{۱۲} Perceptually Weighted Euclidean Distance
^{۱۳} Maximum Likelihood Estimator
^{۱۴} Perceptually Weighted
^{۱۵} Short Time Fourier Transform

^۵ Linear Prediction
^۶ frame
^۷ Inventory Style Speech Resynthesis
^۸ Intelligibility.
^۹ dictionary learning
^{۱۰} low-rank matrix
^{۱۱} Nonnegative matrix factorization

$$\sigma_w = \sqrt{E\{w^2(n)\}} \quad (4)$$

در روابط فوق تابع $E\{\cdot\}$ ، معرف امید ریاضی می‌باشد. سپس گفتار و نویز را به صورت زیر نرمالیزه می‌کنیم:

$$\hat{x}(n) = \frac{1}{\sigma_x} x(n). \quad (5)$$

$$\hat{w}(n) = \frac{1}{\sigma_w} w(n). \quad (6)$$

در این روابط $\hat{x}(n)$ و $\hat{w}(n)$ مقادیر نرمالیزه شده می‌باشند. پس:

$$E\{\hat{x}^2(n)\} = 1. \quad (7)$$

$$E\{\hat{w}^2(n)\} = 1. \quad (8)$$

با استفاده از روابط بالا می‌توان رابطه‌ی زیر را برای گفتار نویزی نوشت:

$$y(n) = \sigma_y \left(\frac{\sigma_x}{\sigma_y} \hat{x}(n) + \frac{\sigma_w}{\sigma_y} \hat{w}(n) \right). \quad (9)$$

با فرض اینکه گفتار و نویز مستقل از هم و دارای میانگین صفر می‌باشند، خواهیم داشت:

$$\sigma_y^2 = \sigma_x^2 + \sigma_w^2. \quad (10)$$

با استفاده از رابطه‌ی (10) و تعریف $\alpha = \frac{\sigma_x}{\sigma_y}$ رابطه‌ی (9) به صورت زیر بازنویسی می‌شود:

$$y(n) = \sigma_y (\alpha \hat{x}(n) + \sqrt{1-\alpha^2} \hat{w}(n)). \quad (11)$$

در رابطه‌ی بالا $\sigma_y \alpha \hat{x}(n)$ همان گفتار می‌باشد.

با توجه به روابط بالا مبرهن است که در قاب‌های سکوت که تنها از نویز تشکیل شده‌اند، $\alpha = 0$ خواهد بود. به همین شکل $\alpha = 1$ مربوط به قاب‌هایی است که شامل گفتار و بدون نویز می‌باشند. با استفاده از رابطه (11) می‌توان نوشت:

$$Y_R(k) = \sigma_y (\alpha X_R(k) + \sqrt{1-\alpha^2} W_R(k)). \quad (12)$$

که در آن X_R ، Y_R ، W_R و α به ترتیب بخش حقیقی پارامترهای STFT مربوط به $y(n)$ ، $\hat{x}(n)$ و $\hat{w}(n)$ می‌باشند. با فرض دسترسی به کتاب کد گفتار و کتاب کد نویز که حاوی بردارهای بخش حقیقی STFT هستند، $Y_R(k)$ را مدل می‌نماییم.

خواص آماری سیگنال گفتار بهره برده می‌شود. این در حالی است که در تخمین گر فاصله اقلیدسی، صرفاً فاصله برداری ارزیابی می‌شود و هیچ گونه خواص آماری گفتار مورد استفاده قرار نمی‌گیرد. همانطور هم که در [20 و 21] بررسی شده است، به کارگیری خواص آماری گفتار در دست‌یابی به معیارهای شباهت موجب بهبود عملکرد پردازش گفتار می‌شود. هم‌چنین در روش ارائه شده در این مقاله از به کارگیری فازضرایب STFT بی‌نیاز بوده‌ایم. البته به این منظور فرآیند جستجو در کتاب کدها دو بار (بخش حقیقی و بخش موهومی ضرایب STFT به طور جداگانه) صورت گرفت. در نتیجه مدت زمان لازم برای بهسازی به بیش از دو برابر روش [17] افزایش یافت. نتایج شبیه‌سازی‌ها و بررسی‌ها، حاکی از بهبود عملکرد بهسازی گفتار با روش ارائه شده در این مقاله، نسبت به [17] می‌باشد.

از آنجایی که روش پیشنهادی سنتز مبتنی بر کتاب کد به جداسازی گفتار و نویز از یکدیگر می‌پردازد، پیش‌بینی می‌شود که با کاهش همبستگی بین نویز و گفتار میزان بهسازی افزایش یابد. این موضوع در انتهای این مقاله بررسی و شبیه‌سازی شد. نتایج شبیه‌سازی حاکی از افزایش کیفیت گفتار بهسازی شده با کاهش همبستگی نویز و گفتار می‌باشد.

در بخش‌های بعدی در ابتدا به طراحی تخمین گر بیشینه درست‌نمایی با وزن‌های شنیداری پرداخته می‌شود. سپس مطابق روش‌های مبتنی بر کتاب کد، از این تخمین گر به عنوان تابع فاصله در طراحی کتاب‌های کد استفاده می‌شود. پس از بیان الگوریتم بهسازی با روش ارائه شده، به شبیه‌سازی و مقایسه نتایج آن با روش‌های دیگر پرداخته می‌شود. سپس در ادامه، اثر همبستگی متقابل^{۱۶} نویز و گفتار در بهسازی، بررسی می‌شود.

طراحی تخمین گر بیشینه درست‌نمایی با وزن‌های شنیداری

گفتار نویزی $y(n)$ عبارت است از [17]:

$$y(n) = x(n) + w(n). \quad (1)$$

در این رابطه $x(n)$ و $w(n)$ به ترتیب، گفتار و نویز جمعی می‌باشند. هم‌چنین پارامتر n اندیس نمونه‌ی زمانی می‌باشد. انحراف معیار گفتار نویزی، گفتار، و نویز عبارتند از:

$$\sigma_y = \sqrt{E\{y^2(n)\}}. \quad (2)$$

$$\sigma_x = \sqrt{E\{x^2(n)\}}. \quad (3)$$

به این ترتیب با گرفتن لگاریتم از رابطه‌ی بالا تخمین‌گر MLRD^{۱۸} به صورت زیر در خواهد آمد:

$$MLRD(i_R, j_R, \alpha) = \sum_{k=1}^K (\ln |Z_{R_\alpha}^{i_R j_R}(k)| + \frac{Y_R^2(k)}{(Z_{R_\alpha}^{i_R j_R}(k))^2}) \quad (20)$$

می‌توان با اعمال وزن‌های شنیداری، تخمین‌گر فوق را بهینه کرد. به این ترتیب خواهیم داشت.

$$PW_MLRD(i_R, j_R, \alpha) = \sum_{k=1}^K V^Y(k) \cdot (\ln |Z_{R_\alpha}^{i_R j_R}(k)| + \frac{Y_R^2(k)}{(Z_{R_\alpha}^{i_R j_R}(k))^2}) \quad (21)$$

در این رابطه V^Y بردار وزن‌های شنیداری است که مؤلفه‌های آن متناسب با اهمیت شنیداری گفتار در آن مؤلفه فرکانسی می‌باشد. این وزن شنیداری طبق [۱۹] به دست می‌آید. به این ترتیب در دو حالت اعمال یا عدم اعمال وزن‌های شنیداری می‌توان از یکی از روابط زیر را به عنوان تخمین‌گر استفاده کرد:

$$\{i_R^*, j_R^*, \alpha^*\} = \arg \min_{i_R, j_R, \alpha} MLRD(i_R, j_R, \alpha) \quad (22)$$

$$\{i_R^*, j_R^*, \alpha^*\} = \arg \min_{i_R, j_R, \alpha} PW_MLRD(i_R, j_R, \alpha) \quad (23)$$

که در این رابطه PW_MLRD^{۱۹} تخمین‌گر بیشینه درست‌نمایی مربوط به بخش حقیقی ضرایب STFT و با وزن‌های شنیداری می‌باشد. به طور مشابه تمام روابط فوق برای بخش موهومی نیز برقرار می‌باشد. تنها کافی است اندیس‌های R تبدیل به I شوند. علاوه بر این مقدار بهینه‌ی α ، یعنی α^* ، در گام قبل تعیین شده است. پس خواهیم داشت:

$$\{i_I^*, j_I^*\} = \arg \min_{i_I, j_I} PW_MLID(i_I, j_I, \alpha^*) \quad (24)$$

که در آن PW_MLID^{۲۰} تخمین‌گر بیشینه درست‌نمایی مربوط به بخش موهومی ضرایب STFT و با وزن‌های شنیداری می‌باشد. آن‌گاه:

$$\hat{x}_C^*(n) = IFFT(X_{R_C}^{i_R^*} + jX_{I_C}^{j_I^*}); \quad j = \sqrt{-1} \quad (25)$$

به این ترتیب قاب سنتز شده عبارت است از

$$x^*(n) = \sigma_y \cdot \alpha^* \cdot \hat{x}_C^*(n) \quad (26)$$

$$Z_{R_\alpha}^{i_R j_R}(k) = \sigma_y (\alpha X_{R_C}^{i_R}(k) + \sqrt{1 - \alpha^2} W_{R_C}^{j_R}(k)) \quad (13)$$

کتاب کد نویزی می‌باشد. $Z_{R_\alpha}^{i_R j_R}$ هم بخش حقیقی مدل گفتار نویزی توسط مقدار α ، $X_{R_C}^{i_R}$ و $W_{R_C}^{j_R}$ می‌باشد. در اینجا هدف این است که بهترین مدل را برای Y_R به دست آوریم.

حال به طراحی تخمین‌گر بیشینه درست‌نمایی (ML)^{۱۷} برای بخش حقیقی گفتار نویزی می‌پردازیم. بخش حقیقی و موهومی $Y(k)$ را می‌توان به صورت متغیرهای تصادفی گوسی با میانگین صفر مدل نمود [۴]. با این فرض تابع چگالی احتمال $Y_R(k)$ این‌گونه نوشته می‌شود:

$$f_{Y_R(k)}(Y_R(k)) = \frac{1}{\sqrt{\pi \cdot E\{Y_R^2(k)\}}} e^{-\frac{Y_R^2(k)}{E\{Y_R^2(k)\}}} \quad (14)$$

سپس خواهیم داشت:

$$f_{Y_R(k)}(Y_R(k) | Z_{R_\alpha}^{i_R j_R}(k)) = \frac{1}{\sqrt{\pi \cdot E\{(Z_{R_\alpha}^{i_R j_R}(k))^2\}}} e^{-\frac{Y_R^2(k)}{E\{(Z_{R_\alpha}^{i_R j_R}(k))^2\}}} \quad (15)$$

با توجه به اینکه $Z_{R_\alpha}^{i_R j_R}$ یک مقدار معین است نه تصادفی، می‌توان نوشت:

$$f_{Y_R(k)}(Y_R(k) | Z_{R_\alpha}^{i_R j_R}(k)) = \frac{1}{\sqrt{\pi} |Z_{R_\alpha}^{i_R j_R}(k)|} e^{-\frac{Y_R^2(k)}{(Z_{R_\alpha}^{i_R j_R}(k))^2}} \quad (16)$$

سپس تعریف می‌کنیم:

$$Y_R = [Y_R(1), \dots, Y_R(K)] \quad (17)$$

$$Z_{R_\alpha}^{i_R j_R} = [Z_{R_\alpha}^{i_R j_R}(1), \dots, Z_{R_\alpha}^{i_R j_R}(K)] \quad (18)$$

از آنجایی که تمام مؤلفه‌های Y_R (یعنی $Y_R(1), \dots, Y_R(K)$) را می‌توان از یکدیگر مستقل فرض کرد [۲۸]، پس می‌توان نوشت:

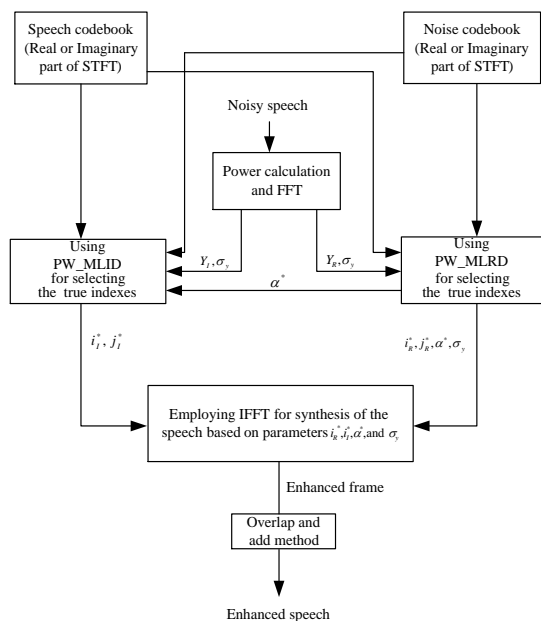
$$f_{Y_R}(Y_R | Z_{R_\alpha}^{i_R j_R}) = \prod_{k=1}^K f_{Y_R(k)}(Y_R(k) | Z_{R_\alpha}^{i_R j_R}(k)) = \frac{1}{\prod_{k=1}^K |Z_{R_\alpha}^{i_R j_R}(k)|} e^{-\sum_{k=1}^K \frac{Y_R^2(k)}{(Z_{R_\alpha}^{i_R j_R}(k))^2}} \quad (19)$$

^{۱۸} Maximum Likelihood Real Distance

^{۱۹} Perceptually Weighted Maximum Likelihood Real Distance

^{۲۰} Perceptually Weighted Maximum Likelihood Imaginary Distance

^{۱۷} Maximum Likelihood



شکل ۱: بهسازی گفتار با استفاده از کتاب کد و تخمین گر بخش حقیقی و موهومی STFT

نتایج شبیه‌سازی

در این بخش در ابتدا کتاب کد گفتار تمیز و نیز نویز طراحی می‌شود. طراحی کتاب کدها وابسته به گوینده و وابسته به نویز می‌باشد. از این رو در این مقاله کتاب کد گفتار برای ۲ گوینده به طور جداگانه و نیز برای نویزهای مختلف به طور مستقل طراحی گردید. با استفاده از این کتاب کدها و تخمین‌گر طراحی شده، بهسازی گفتار در شرایط گوناگون و SNRهای مختلف شبیه‌سازی شد و روش پیشنهادی این مقاله باروش‌های دیگر مقایسه گردید.

طراحی کتاب کد گفتار

کتاب کد گفتار وابسته به گوینده با اندازه‌های مختلف و برای هر یک از دو گوینده به طور جداگانه با استفاده از دیتابیس Cooke [۲۴ و ۲۵] طراحی گردید. این طراحی برای تابع فاصله‌های مختلفی بوده است. به ازای اندازه‌های مختلف کتاب کد، متوسط PESQ به دست آمد. نتایج حاصل برای گوینده‌های spk1 و spk2 در شکل‌های ۲ و ۳ مشاهده می‌شود. این نتایج برای تابع فاصله‌های ERD^{22} ، PW_ERD^{23} ، $MLRD^{24}$ ، PW_MLRD^{25} می‌باشد. ERD ، و PW_ERD تابع فاصله‌های اقلیدسی [۱۹] بر روی بخش حقیقی (و موهومی) پارامترهای STFT می‌باشند. این تابع فاصله‌ها جایگزین تابع فاصله‌های

طراحی کتاب‌های کد گفتار و نویز

در طراحی کتاب کد بخش حقیقی پارامترهای STFT، ابتدا تمام نمونه‌های زمانی قاب‌های آموزشی به صورت زیر نرمالیزه می‌شوند.

$$\hat{x}^r(n) = \frac{x^r(n)}{\sqrt{\frac{1}{FL} \sum_{n=1}^{FL} x^{r^2}(n)}}; \quad r=1, \dots, M \quad (27)$$

سپس بخش حقیقی و موهومی تبدیل FFT سیگنال $\hat{x}^r(n)$ استخراج می‌شود. حال با استفاده از LBG و تابع فاصله‌ی به دست آمده در (۲۲) یا (۲۳) به طراحی کتاب کد بخش حقیقی گفتار و نیز نویز می‌پردازیم. از آنجایی که تابع چگالی بخش موهومی و حقیقی گفتار و نیز تابع چگالی بخش حقیقی و موهومی نویز یکسان است [۲۳]، پس فقط کتاب کد بخش حقیقی طراحی می‌شود و هنگام جستجو در بخش موهومی از همان کتاب کد بخش حقیقی استفاده می‌شود.

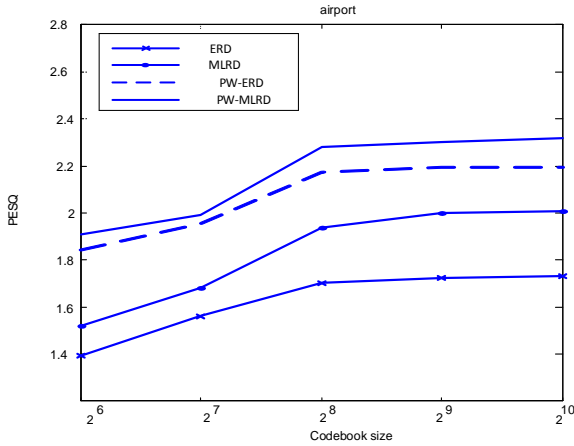
الگوریتم بهسازی گفتار با روش سنتز مبتنی بر کتاب کد و تخمین گر پیشینه درست‌نمایی

بلوک دیاگرام شکل ۱ روش بهسازی مبتنی بر کتاب کد با استفاده از تخمین‌گر رابطه ۲۳ و با در دست داشتن کتاب‌های کد طراحی شده را نشان می‌دهد. در گام اول توان و تبدیل فوریه هر قاب از گفتار نویزی محاسبه می‌شود. سپس با به کارگیری تخمین‌گر PW_MLRD و با استفاده از توان و بخش حقیقی تبدیل فوریه قاب گفتار نویزی، جستجو و انتخاب اندیس‌های بهینه‌ی کتاب کد گفتار تمیز، کتاب کد نویز و نیز مقدار بهینه α^* صورت می‌گیرد. بعد از انتخاب اندیس‌های مناسب بخش حقیقی تبدیل فوریه گفتار و مقدار بهینه α^* و با به کارگیری تخمین‌گر PW_MLRD، جستجو به منظور انتخاب اندیس بهینه بخش موهومی تبدیل فوریه گفتار صورت می‌گیرد. بعد از انتخاب اندیس‌های بخش حقیقی و موهومی قاب گفتار از کتاب‌های کد مربوط و با داشتن توان فرم نویزی و α^* ، قاب مورد بررسی گفتار بهسازی شده طبق روابط (۲۵) و (۲۶) به دست می‌آید. بعد از انجام این الگوریتم برای تمام قاب‌های گفتار نویزی، قاب‌های گفتار بهسازی شده به دست می‌آیند. با کنار هم قرار دادن تمام این قاب‌ها با روش همپوشانی و جمع α^* ، گفتار بهسازی شده سنتز می‌شود.

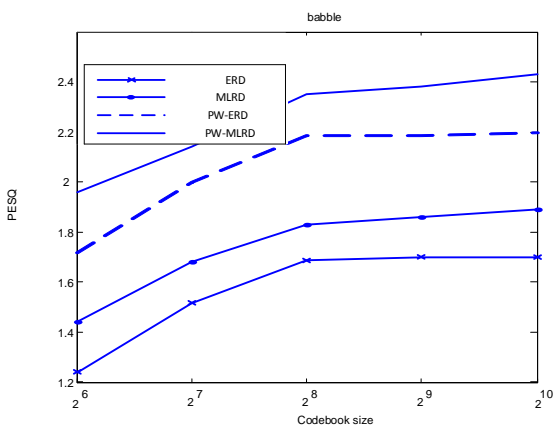
^{۲۲} Euclidian Real Distance
^{۲۳} Perceptually Weighted_ Euclidian Real Distance
^{۲۴} Maximum Likelihood Real Distance
^{۲۵} Perceptually Weighted_ Maximum Likelihood Real Distance

^{۲۱} Overlap and Add Method

آمده برای نویزهای فرودگاه و همهمه^{۲۶} در شکل‌های ۴ و ۵ مشاهده می‌شود. برای شش نویز دیگر نیز شبیه‌سازی و نتایج مشابه به دست آمد. با توجه به نحوه تغییرات PESQ نسبت به اندازه کتاب کد نویز، اندازه مناسب برای کتاب کد هر نویز مطابق جدول ۱ می‌باشد.



شکل ۴: تغییرات PESQ بر حسب اندازه کتاب کد برای نویز فرودگاه



شکل ۵: تغییرات PESQ بر حسب اندازه کتاب کد برای نویز همهمه

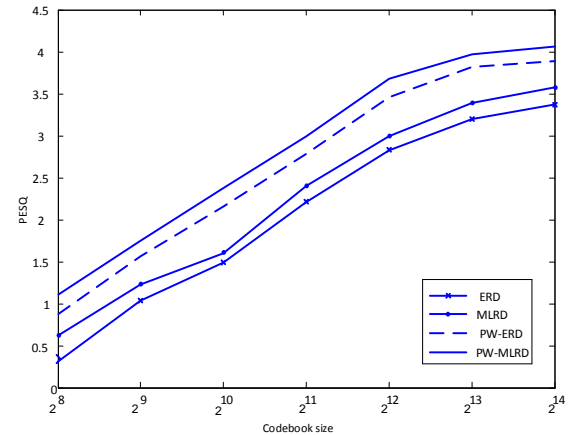
جدول ۱: اندازه مناسب برای کتاب‌های کد نویز (تعداد بردارهای هر کتاب)

noise	codebook size
airport	256
babble	256
car	128
factory	128
restaurant	512
street	512
volvo	8
white	128

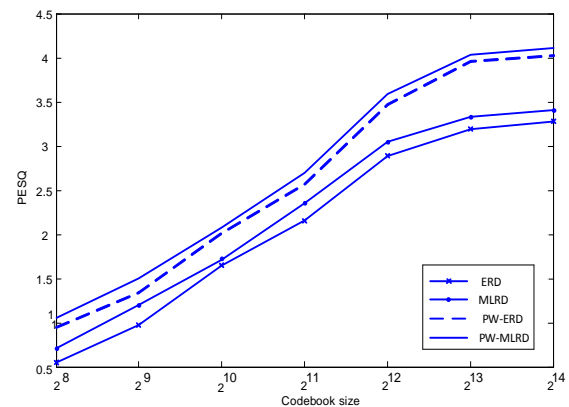
بهسازی گفتار با روش ارائه شده و مقایسه با روش‌های دیگر

در این قسمت با چندین روش گفتار نویزی مورد بهسازی قرار می‌گیرد. ابتدا روش پیشنهادی این مقاله که از تابع فاصله‌ی

بیشینه درست‌نمایی در الگوریتم توضیح داده شده می‌باشند. با توجه به این‌که PESQ برابر با ۳.۸ نشان‌دهنده‌ی کیفیت مناسب برای کتاب کد گفتار می‌باشد [۲۶]، و نیز با در نظر داشتن منحنی تغییرات PESQ بر حسب اندازه کتاب، اندازه کتاب هر گوینده برابر با ۸۱۹۶ (۱۳ بیتی) مناسب می‌باشد. انتخاب این اندازه برای کتاب کد از این رو است که با افزایش این مقدار، نرخ افزایش PESQ به طور قابل توجهی کند می‌شود.



شکل ۲: تغییرات PESQ بر حسب اندازه کتاب کد برای گوینده‌ی spk1



شکل ۳: تغییرات PESQ بر حسب اندازه کتاب کد برای گوینده‌ی spk2

طراحی کتاب کد نویز

کتاب کد نویز برای هشت نویز فرودگاه، همهمه، خیابان، رستوران، و ماشین از دیتابیس aurora و نویزهای کارخانه، ولوو و سفید از دیتابیس NATO RSG-10 به طور جداگانه طراحی گردید. این طراحی برای اندازه‌های مختلف و هر یک از تابع فاصله‌های یادشده انجام شد. در مورد هر تابع فاصله کتاب کد گفتار طراحی شده در بالا با تمام اندازه‌های کتاب کد نویز از همان تابع فاصله، برای بهسازی گفتارهای نویزی با SNR برابر با -۱۰dB به کار می‌روند. به ازای هر اندازه کتاب کد نویز، متوسط PESQ گفتار بهسازی شده به دست آمد. نتایج به دست

بهسازی ارائه شده تقریباً ده برابر روش INV_RESYN خواهد شد. هزینه‌ی عمده‌ای که به منظور به دست آوردن کیفیت شنیداری خوب پرداخت شده است صرف حافظه و زمان زیاد نسبت به روش‌های مورد مقایسه می‌باشد. حافظه‌های مورد نیاز برای ذخیره بردارهای بخش حقیقی STFT هر گوینده به اندازه ۱۸۹۶ بردار ۱۳ بیتی و کتاب‌های کد هر نویز با اندازه‌های متناظر با جدول ۱ مؤید این موضوع است. هم‌چنین برای بهسازی هر فریم نویزی دو بار جستجوی کامل در کتاب کد گفتار با اندازه یاد شده و کتاب کد نویز به طور جداگانه به ازای بخش حقیقی و موهومی بردار STFT صورت می‌گیرد. این امر موجب طولانی شدن زمان بهسازی نسبت به سایر روش‌ها می‌شود. روش ارائه شده کیفیت شنیداری خوبی را هم در مورد نویزهای شدیداً غیرایستاد و هم در مورد نویزهای ایستاد در SNRهای زیر صفر دسیبل (و نیز SNRهای مثبت) از خود نشان می‌دهد.

به منظور ارزیابی کیفی گفتارهای بهسازی شده از معیار کیفی MOS^{۲۸} استفاده شده است. به این منظور گفتارهای نویزی و بهسازی شده توسط پنج نفر شنیده شد و امتیازی بین ۱ تا ۵ به هر یک از آن گفتارها داده شد. میانگین نتایج حاصل برای نویز همهمه در جدول ۴ مشاهده می‌شود. لازم به ذکر است که چون PESQ گفتار بهسازی شده بسیار نزدیک به MOS مربوط به آن و یک جایگزین کمی مناسب برای آن است [۲۹] و نیز به دلیل آن‌که به دست آوردن مقادیر MOS برای این تعداد زیاد سیگنال گفتار، بسیار وقت‌گیر می‌باشد، در این مقاله به ارائه مقادیر PESQ بسنده شده است و فقط به عنوان نمونه برای نویز همهمه مقادیر MOS ارائه گردیده است. طبق نتایج جدول ۴ و همانطور هم که با استفاده از نتایج جداول مربوط به PESQ پیش بینی می‌شد، روش ارائه شده دارای MOS بهتری نسبت به دیگر روش‌ها می‌باشد

با گوش کردن به گفتارهای بهسازی شده با روش PW_MLRD، این نکته دریافت می‌شود که در SNRهای زیر صفر دسیبل (-۱۰dB) نیز گفتار بهسازی شده کاملاً قابل فهم می‌باشد. هم‌چنین اعوجاج موجود در گفتار بهسازی شده با کاهش SNR ورودی بیشتر می‌شود. این اعوجاج، حاصل از سنتز می‌باشد. در گفتار بهسازی شده از نویز اولیه که مخرب گفتار بوده است اثری نمی‌باشد. علت افزایش اعوجاج با کاهش SNR نیز افزایش خطا در تشخیص اندیس‌های کتاب کد گفتار می‌باشد.

وزن داده شده‌ی رابطه‌های ۲۳ استفاده می‌کند. سپس به منظور مقایسه، با استفاده از روش [۱۷] بهسازی صورت می‌گیرد. این روش را STFT_CB می‌نامیم. هم‌چنین روش سوم استفاده از فیلتر وینر بوده است [۲۷]. این روش را با نماد WF می‌شناسیم. به این ترتیب روش ارائه شده در این مقاله با یک روش پایه‌ای مقایسه می‌شود. روش چهارم، استفاده از [۲۸] بوده که در آن از کتاب کد LPC استفاده کرده است. این روش را تحت عنوان روش LPC_CB می‌شناسیم. از آنجایی که این روش مبتنی بر کتاب کد می‌باشد، به منظور مقایسه در این بخش به کار گرفته شد. البته این روش مستقل از گوینده می‌باشد. روش پنجم استفاده از بازسنتز مبتنی بر فهرست^{۲۷} [۱۲] می‌باشد. این روش را با نماد INV_RESYN بیان می‌کنیم. از لحاظ محدودیت‌های اعمال شده، روش INV_RESYN مشابه روش ارائه شده در این مقاله می‌باشد. میانگین نتایج به دست آمده در SNRهای ورودی برابر با -۱۰dB، -۵dB، ۰dB، ۵dB و ۱۰dB، در جدول‌های ۲ و ۳ مشاهده می‌گردد.

با توجه به نتایج جدول‌های ۲ و ۳ این گونه برمی‌آید که PESQ گفتارهای بهسازی شده حاصل از روش ارائه شده‌ی PW_MLRD از روش STFT_CB [۱۷] بهتر می‌باشد. علت اصلی استفاده از تخمین گر بیشینه درست‌نمایی در جستجوی اندیس‌های کتاب کد در روش PW_MLRD می‌باشد. به این منظور فرآیند جستجو در کتاب کدها دو بار (برای هر یک از بخش‌های حقیقی و موهومی) صورت می‌گیرد. هزینه آن مدت زمان لازم برای بهسازی توسط روش PW_MLRD است که بیش از دو برابر روش STFT_CB می‌باشد. روش ارائه شده در این مقاله از بقیه‌ی روش‌های مورد مقایسه نیز، PESQ بسیار بهتری از خود نشان می‌دهد. در مقابل روش ارائه شده در مقاله حاضر، روش WF هیچگونه آموزشی نمی‌بیند. روش LPC_CB از کتاب کد پوش طیف استفاده کرده است و وابسته به گوینده نمی‌باشد. در روش INV_RESYN آموزش نویز صورت نمی‌گیرد. بلکه در یک SNR معین و برای یک گوینده و نویز مشخص، گفتار نویزی و گفتار تمیز آموزش داده می‌شود. هم‌چنین در تابع فاصله مورد استفاده در چندی کردن بردارهای MFCC و نیز معیار شباهت به کار رفته، خواص شنیداری گوش لحاظ نشده است. علاوه بر آن همانطور که پیش‌تر بیان شد، در گام‌های میانی این روش از فیلتر وینر نیز استفاده می‌شود که افت کیفیت خروجی در SNRهای خیلی پایین را در برخواهد داشت. به این ترتیب منطقی است که بهسازی حاصل از روش ارائه شده، از روش INV_RESYN و نیز دو روش مقایسه‌ای دیگر بهتر باشد. گرچه در مقابل، زمان صرف شده در روش

جدول ۲: متوسط PESQ گفتار بهسازی شده توسط روش‌های گوناگون و SNR ورودی -5dB و -10dB

noise	input SNR(dB)	PESQ					
		Noisy	PW-MLRD	STFT-CB	WF	LPC-CB	INV-RESYN
airport	-10	0.67	2.10	1.80	0.70	0.97	0.96
	-5	0.85	2.67	2.23	0.81	1.27	1.37
babble	-10	0.49	2.11	1.75	0.66	0.90	0.98
	-5	0.67	2.47	1.96	0.84	1.36	1.44
car	-10	0.49	2.10	1.74	0.65	0.89	0.96
	-5	0.66	2.44	2.01	0.90	1.25	1.39
factory	-10	0.61	2.09	1.62	0.73	0.94	0.97
	-5	0.88	2.41	2.18	1.01	1.37	1.62
restaurant	-10	0.50	2.18	1.65	0.65	0.69	0.81
	-5	0.80	2.56	2.17	0.97	0.91	1.13
street	-10	0.49	2.14	1.63	0.63	0.83	1.07
	-5	0.67	2.50	2.06	0.98	1.34	1.59
volvo	-10	1.88	3.42	2.92	1.94	2.12	2.34
	-5	2.29	3.49	3.01	2.33	2.67	2.79
white	-10	0.50	2.10	1.86	0.76	0.85	0.91
	-5	0.71	2.49	2.22	1.02	1.45	1.54

جدول ۳: متوسط PESQ گفتار بهسازی شده توسط روش‌های گوناگون و SNR ورودی 0dB، 5dB و 10dB

noise	input SNR(dB)	PESQ					
		Noisy	PW-MLRD	STFT-CB	WF	LPC-CB	INV-RESYN
airport	0	1.07	3.28	2.66	1.42	1.58	1.73
	5	1.46	3.43	2.94	1.87	1.93	2.18
	10	1.89	3.66	3.17	2.22	2.25	2.59
babble	0	0.93	3.02	2.53	1.34	1.52	1.68
	5	1.30	3.28	2.77	1.70	1.87	2.21
	10	1.76	3.48	3.06	2.13	2.22	2.63
car	0	0.89	2.90	2.38	1.55	1.63	1.78
	5	1.27	3.22	2.64	1.64	1.77	2.17
	10	1.68	3.41	2.92	2.15	2.26	2.61
factory	0	1.27	3.08	2.72	1.82	1.87	1.89
	5	1.73	3.39	2.80	1.91	2.02	2.34
	10	2.16	3.57	3.17	2.34	2.53	2.78
restaurant	0	0.99	3.24	2.52	1.59	1.75	1.85
	5	1.28	3.53	2.71	1.90	1.98	2.27
	10	1.66	3.64	3.07	2.15	2.44	2.60
street	0	0.95	3.17	2.60	1.85	1.94	1.97
	5	1.37	3.37	2.76	2.07	2.30	2.46
	10	1.86	3.56	3.04	2.34	2.68	2.79
volvo	0	2.66	3.60	3.07	3.10	3.28	3.38
	5	3.02	3.66	3.23	3.21	3.41	3.52
	10	3.21	3.74	3.36	3.36	3.52	3.72
white	0	0.88	3.17	2.56	1.79	2.07	۲.۱۶
	5	1.13	3.30	2.80	2.01	2.22	2.37
	10	1.46	3.41	2.94	2.24	2.61	2.80

جدول ۴: متوسط MOS گفتار بهسازی شده توسط روش‌های گوناگون و SNR ورودی -۵dB، -۱۰dB، ۰dB، ۵dB و ۱۰dB برای نویز همهمه

input SNR(dB)	MOS					
	Noisy	PW-MLRD	STFT-CB	WF	LPC-CB	INV-RESYN
-10	1.00	2.08	1.88	1.02	1.10	1.21
-5	1.00	2.55	2.15	1.12	1.33	1.41
0	1.15	3.07	2.65	1.42	1.54	1.83
5	1.45	3.36	2.94	1.77	1.88	2.36
10	1.92	3.45	3.18	2.21	2.28	2.72

جدول ۵: متوسط بهبود (افزایش) PESQ حاصل از بهسازی و متوسط همبستگی متقابل نویز و گفتار

noise	airport	babble	car	factory	restaurant	street	volvo	white
C_{XW}	0.4016	0.4169	0.3965	0.5068	0.3270	0.3711	0.5123	0.2924
PESQ increment	1.84	1.84	1.82	1.58	1.98	1.88	0.97	1.97

در این مقاله به بهسازی گفتار نویزی با روش سنتز مبتنی بر کتاب کد، با استفاده از تخمین گر بیشینه درست نمایی با وزن‌های کیفیت شنیداری پرداخته شده است. در روش سنتز مبتنی بر کتاب کد، انتخاب هرچه صحیح‌تر اندیس‌های کتاب کد به منظور سنتز بهینه گفتار بهسازی شده، از اهمیت ویژه‌ای برخوردار است. از این رو در این مقاله طراحی تخمین‌گر بیشینه درست‌نمایی (ML) با اعمال وزن‌های بهبود دهنده کیفیت شنیداری، برای بخش‌های حقیقی و موهومی STFT گفتار و نویز صورت گرفت. رابطه این تخمین‌گر به عنوان تابع فاصله در طراحی و تشکیل کتاب کد بخش حقیقی و موهومی گفتار وابسته به گوینده و نیز کتاب نویز وابسته به نوع نویز استفاده شد. به کارگیری بخش حقیقی و موهومی و استفاده از کتاب کد آن‌ها باعث شد که نیازی به بررسی فاز نباشد. نتایج شبیه‌سازی حاکی از بهبود کیفیت شنیداری گفتار بهسازی شده نسبت به روش‌های قبلی می‌باشد. روش ارائه شده در برخورد با نویزهای غیرایستاد و نیز ایستاد موفق‌تر از روش‌های دیگر عمل می‌کنند. علت هم آن است که روش ارائه شده قاب به قاب و مستقل از قاب‌های قبلی به جداسازی گفتار از نویز می‌پردازد. در روش بهسازی با استفاده از سنتز مبتنی بر کتاب کد به دلیل برخورد یکسان با نویز و گفتار در واقع به جداسازی نویز و گفتار از یکدیگر می‌پردازد و با انتخاب بهینه اندیس‌های کتاب کد گفتار، سیگنال گفتار بهسازی شده را سنتز می‌کند. از این رو با این روش می‌توان گفتارهای نویزی، با نسبت سیگنال به نویز کمتر از صفر دسیبل را بهسازی نمود. از نتایج شبیه‌سازی این‌گونه استنباط می‌شود که تخمین گر بیشینه درست نمایی با وزن‌های کیفیت شنیداری نسبت به تخمین‌گر فاصله اقلیدسی با وزن‌های شنیداری بهتر عمل می‌کند. تخمین‌گر طراحی شده با انتخاب مناسب‌تر از اندیس‌های کتاب کد گفتار، گفتار بهسازی شده را با کیفیت شنیداری بهتری سنتز می‌کند. مزیت بهسازی گفتار با کیفیت بیشتر در SNRهای

اثر همبستگی متقابل نویز و گفتار در بهسازی

در این جا به بررسی اثر همبستگی متقابل^{۲۹} طیف‌های نویز و گفتار در میزان بهسازی پرداخته می‌شود. به منظور در نظر داشتن تمام طیف‌های ممکن نویز و گفتار از دامنه‌های STFT کتاب کدهای طراحی شده استفاده می‌شود. با توجه به تعریف ارائه شده برای همبستگی متقابل بین دو سیگنال [۳۰] رابطه زیر برای همبستگی متقابل بین گفتار X و نویز W پیشنهاد می‌شود.

$$C_{XW} = \frac{1}{N_x N_w} \sum_{i=1}^{N_x} \sum_{j=1}^{N_w} \frac{X_c^i W_c^j}{|X_c^i| |W_c^j|}, \quad i=1 \dots N_x, \quad j=1 \dots N_w \quad (28)$$

که در آن:

$$|X_c^i| = \sqrt{\sum_{k=1}^K X_c^i(k)^2}, \quad |W_c^j| = \sqrt{\sum_{k=1}^K W_c^j(k)^2} \quad (29)$$

متوسط همبستگی متقابل هر یک از نویزها با گفتار گوینده‌های مختلف محاسبه شد. علاوه بر آن متوسط افزایش PESQ گفتار بهسازی شده نسبت به گفتار نویزی اولیه در SNRهای مختلف و گوینده‌های موجود به دست آمد. حاصل در جدول ۵ مشاهده می‌گردد.

همانطور که مشاهده می‌شود نویز سفید و نویز رستوران که کمترین همبستگی متقابل را با گفتار دارند، بیشترین افزایش PESQ را در بردارند. در مقابل نویزهای ولوو و کارخانه که بیشترین همبستگی متقابل را با گفتار دارند، کمترین افزایش PESQ را در پی دارند.

جمع‌بندی

- Transactions on Speech and Audio Processing, vol. 18, No. 6, pp. 1243-1257, 2010.
- [13] Mathew Shaji Kavalekalam, Jesper Kjær Nielsen, Jesper B. Unsow Boldt, Mads Græsbøll Christensen, "Model-based Speech Enhancement for Intelligibility Improvement in Binaural Hearing Aids", IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol: 27, Issue: 12, pp. 2150 - 2161, Dec. 2019.
- [14] Yunyun Ji, Wei-Ping Zhu, Benoit Champ Agne, "Speech Enhancement Based on Dictionary Learning and Low-Rank Matrix Decomposition", IEEE Access, vol: 7, pp. 4936 – 4947. Dec 2018.
- [15] Ahmet E. Bulut, Kazuhito Koishida, "Low-Latency Single Channel Speech Enhancement Using U-Net Convolutional Neural Networks", IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2020, Barcelona, Spain, 4-8 May 2020.
- [16] Zhigang Bai, Changchun Bao, Bofang Yan, "Speech Enhancement Based on Codebook Constrained Nonnegative Matrix Factorization", International Conference on Audio, Language and Image Processing, ICALIP 2018, Shanghai, China, 16-17 July 2018.
- [17] Roghayeh Doost, and Abolghasem Sayadian, "Speaker-Dependent Speech Enhancement Using Codebook-based Synthesis for Low SNR Applications", International Journal of Information & Communication Technology Research, vol. 5, Number 2, Spring 2013.
- [18] Wai Lok Woo, S.s. Dlay, "Single channel audio source separation", Wseas Transaction on Signal Processing, Issue 4, vol. 4, April 2008.
- [19] Roghayeh Doost, Abolghasem Sayadian, and Hossein Shamsi, "A New Perceptually Weighted distance Measure for Vector Quantization of the STFT Amplitudes in the Speech Application", IEICE Electronic Express, vol.6, no.12, pp.824-830, June 25, 2009.
- [20] Elior Hadad, and Sharon Gannot, "Maximum Likelihood Multi-Speaker Direction of Arrival Estimation Utilizing a Weighted Histogram", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 14 May 2020.
- [21] Tuomas Virtanen, and Marko Hel'én, "Probabilistic Model Based Similarity Measures for Audio Query-by-Example", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, 21-24 Oct. 2007.
- [22] Israel Cohen, "Relaxed Statistical Model for Speech Enhancement and a Priori SNR Estimation", IEEE Transaction on Speech and Audio Processing, vol. 13, No. 5, pp. 870-881, September 2005.
- [23] Rainer Martin, "Speech Enhancement Based on Minimum Mean-Square Error Estimation and Supergaussian Priors", IEEE Transaction on Speech and Audio Processing, vol. 13, No. 5, pp. 845_856, September 2005.
- [24] M. Cooke, J.R. Hershey, and S.J. Rennie, "Monaural speech separation and recognition challenge", Elsevier Computer Speech and Language, vol. 24, No. 1, pp. 1–15, 2010.
- [25] Cooke, M.P., Lee, T.W., 2006. Speech separation challenge website.
- خیلی پایین (یا بالا) نسبت به روش‌های مورد مقایسه، صرف زمان طولانی‌تر به منظور جستجو در کتاب‌های کد حقیقی و موهومی گفتار و نویز، و نیز به‌کارگیری حافظه نسبتاً زیاد برای ذخیره بردارهای کتاب‌کد می‌باشد.

مراجع

- [1] Amol Chaudhari ; and S. B. Dhonde, " A review on speech enhancement techniques ", International Conference on Pervasive Computing (ICPC), Pune, India, 8-10 Jan 2015.
- [2] Benesty, and Jacob, Fundamentals of Speech Enhancement, SpringerBriefs in Electrical and Computer Engineering, Springer International Publishing, Switzerland, pp. 1-44, 2018.
- [3] .Navneet Upadhyay, and Abhijit Karmakar, "Speech Enhancement using Spectral Subtraction-type Algorithms: A Comparison and Simulation Study", Eleventh International Multi-Conference on Information Processing (IMCIP), Bangalore, pp. 574–584, India, 21-23 August 2015.
- [4] Siddala Vihari, A. Sreenivasa Murthy, Priyanka Soni and D. C. Naik, "Comparison of Speech Enhancement Algorithms", Twelfth International Multi-Conference on Information Processing (IMCIP), pp. 666–676, Bangalore, India, 2016.
- [5] Navneet Upadhyay, and Rahul Kumar Jaiswala, "Single Channel Speech Enhancement: using Wiener Filtering with Recursive Noise Estimation", 7th International conference on Intelligent Human Computer Interaction, IHCI, Procedia Computer Science , pp.22–30, 2016.
- [6] Xuzheng Dai, Baoxian Yu, and Xianhua Dai, "An Improved Signal Subspace Algorithm for Speech Enhancement", IFIP International Federation for Information Processing, Springer, Berlin, Heidelberg, vol 445, pp. 104–114, 2014.
- [7] Jacob Benesty, Jesper Rindom Jensen, and Mads Graesboll Christensen, Speech Enhancement-A Signal Subspace Perspective, Academic Press, Kindle Edition, pp. 43-95, 2014.
- [8] Mariyadasu Mathel, Siva Prasad Nandyala, "Speech Enhancement Using Kalman Filter for white, random and color noise", International Conference on Devices, Circuits and Systems (ICDCS), pp. 1-4, Coimbatore, India , 15-16 March 2012.
- [9] M. Brookesand, and N.Dionelis, "Phase aware single channel speech enhancement with modulation domain kalman filtering," IEEE Transactionson Audio, Speech and Language Processing, volume 26, no.5, pp.937-950, 2018.
- [10] Saeed V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction, Second Edition, Wiley, pp.333-354, ch11, 2000.
- [11] Hossein Sameti, Hamid Sheikhzadeh, Li Deng, and Robert L. Brennan , " HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise ", IEEE Transactions on Speech and Audio Processing, Volume 6, No. 5, pp. 445-455, 1998.
- [12] X. Xiao, and R. M. Nickel, "Speech Enhancement with Inventory Style Speech Resynthesis", IEEE

- Audio Processing, vol. 14, No. 1, pp. 163-176, January 2006.
- [29] J. G. Beerends, A. P. Hekstra, A. W. Rix, and M. P. Hollier, "Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part II – Psychoacoustic model ", Journal of the Audio Engineering Society, vol. 50, No. 10, pp.765-778, October 2002.
- [30] Alan V. Oppenheim, Ronald W. Schaffer, and John R. Buck, Discrete-Time Signal Processing, Second Edition, Upper Saddle River, NJ : Prentice Hall, pp. 811-820, 1999.
- <http://www.dcs.shef.ac.uk/~martin/SpeechSeparationChallenge.htm>.
- [26] Thiago de M. Prego and Sergio L. Netto, "Efficient Search in the Adaptive Codebook for ITU-T G.729 Codec", IEEE Signal Processing Letters, vol. 16, No.10, pp.881-884, October 2009.
- [27] Saeed. V. Vaseghi, Advanced Signal Processing and Digital Noise Reduction, ch. 6, Wiley, 1998.
- [28] Sriram Srinivasan, Jonas Samuelsson, and W. Bastiaan Kleijn, "Codebook Driven Short-Term Predictor Parameter Estimation for Speech Enhancement ", IEEE Transactions on Speech and