

کاهش نویز در گفتار با استفاده از کدگذاری تنک و تابع انقباض

علیرضا شیرخدائی^۱، محمدمهدی همایون پور^۲

^۱ کارشناس ارشد، مهندسی کامپیوتر گرایش هوش مصنوعی، دانشگاه صنعتی امیرکبیر، تهران
alireza.shirkhodaei@gmail.com

^۲ دانشیار، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران
homayoun@aut.ac.ir

چکیده

در این مقاله به روشی برای حذف و کاهش نویز از گفتار می پردازیم که بر کدگذاری تنک و اصول انقباضی بنا شده است. کل عملیات در حوزه زمان انجام می گیرد. از مهمترین کاربردهای این روش می توان به استفاده در سیستم های تشخیص گفتار و نیز در ساخت سمک برای کم شنوایان، اشاره نمود که کاهش نویز در این سیستم ها حیاتی هستند. نکته ی مهم در روش کدگذاری تنک چگونگی تعیین پارامترهای لازم برای تابع انقباض و تخمین واریانس نویز است. در این روش تخمین واریانس نویز با روشی ابتکاری در زمان آزمایش انجام می پذیرد، و تعیین سایر پارامترهای لازم در زمان آموزش از روی دادگان تمیز به خوبی انجام می پذیرد و سبب می گردد که نه تنها نویز بسیار خوب کاهش یابد، بلکه ساختار کلیدی گفتار نیز حفظ گردد. کارایی روش فوق با الگوریتم های شناخته شده ای چون فیلترگذاری وینر و تفریق طیفی با سه مقیاس نسبت سیگنال به نویز قطعه ای، ارزیابی ادراکی کیفیت گفتار و میانگین امتیاز نظرات ارزیابی و مقایسه می گردد و نتیجه گرفته می شود که برتری قابل توجهی نسبت به سایر روش ها ارائه داده است. به عنوان نمونه در مقیاس نسبت سیگنال به نویز قطعه ای، کدگذاری تنک انقباضی در نسبت سیگنال به نویز ۱۰ دسی بل ۶/۲ بوده در حالی که فیلترگذاری وینر ۲/۸ و تفریق طیفی نتیجه بهتری از ۲/۲ ارائه نداده است و در شدت های دیگر نویز نیز به همین منوال می باشد.

کلمات کلیدی

کاهش نویز، کدگذاری تنک، تابع انقباض، بهسازی خودکار گفتار

توسط اعمال فیلتری که مبتنی بر معیار کمینه نمودن مربع میانگین خطا است، بهسازی می کند.

۱- مقدمه

برخی از فن آوری های بهسازی تلاش می کنند تا خصوصیات پوششی سیستم شنوایی انسان را به دست آورند زیرا این سیستم سبب می شود تا نویز جمع شونده، با بزرگی زیر آستانه ی پوششی، توسط گوش انسان قابل شنیدن نباشد. همین مساله در چهارچوب تفریق طیفی تعمیم یافته توسط ویراگ^۴ [4] در سال ۱۹۹۹ و در الگوریتم بهسازی گفتار مرتبه ی بتای کمینه ی، کمینه سازی مربع میانگین خطا^۵ توسط یو و همکاران^۵ [5] در سال ۲۰۰۶ مورد بررسی قرار گرفته است. با استفاده از وزن دهی ادراکی ارائه شده توسط هو و لوویزو^۶ [6] در سال ۲۰۰۳ نسخه ی تصحیح شده ای از فیلترگذاری وینر عرضه شد.

امواج صوتی در کاربردهای واقعی، اغلب توسط کانال های مخبراتی یا نویزهای پس زمینه جمع شونده، خراب می شوند. هدف بهسازی خودکار، تخمین سیگنال گفتار تمیز از سیگنال مشاهده شده ی نویزی است. تفریق طیفی^۱ [1] و فیلترگذاری وینر^۲ [2,3] جزو اولین فن آوری های بهسازی گفتار محسوب می گردند و هنوز به دلیل کارایی قابل توجه و پیچیدگی کم محاسباتی، پرمصرف هستند. تفریق طیفی با کسر تخمین بزرگی طیف نویز از بزرگی طیف گفتار نویزی به دست می آید. الگوریتم فیلترگذاری وینر سیگنال گفتار نویزی را

۲-۲- تحلیل مولفه مستقل

تحلیل مولفه مستقل را می‌توان برای تبدیل کردن داده به گونه‌ای که تا جای ممکن از نظر آماری از یکدیگر مستقل باشند، استفاده نمود. با فرض خطی بودن تحلیل مولفه مستقل، مدل داده در موضوع ما را می‌توان به صورت $x=As$ نوشت که x بردار پنجره‌ی K نمونه‌ی مشاهده شده‌ی گفتار تمیز و s بردار منابع تحتانی (مستقل) و A ماتریس مخلوط‌ساز است. هدف از تحلیل مولفه مستقل، تخمین A و s از سیگنال مشاهده شده‌ی x است به طوری که منابع تخمین زده شده s به طور آماری مستقل باشند.

الگوریتم تحلیل مولفه مستقل معمولاً شامل دو مرحله است: پیش سفیدسازی^{۱۱} و جداسازی^{۱۲}. مرحله‌ی سفیدسازی، مخلوط ورودی x را با به کارگیری ماتریس تبدیل مبتنی بر تحلیل مولفه اصلی^{۱۳}، که با U نشان داده می‌شود، ناهمبسته^{۱۴} و نرمال‌سازی می‌کند. پس از آن، مرحله‌ی جداسازی، سیگنال سفید شده را با استفاده از ماتریس تبدیل متعامد یک‌به‌یک^{۱۵} تبدیل می‌کند تا مستقل باشد. ماتریس متعامد یک‌به‌یک B را می‌توان با برخی شروط تخمین زد. یکی از پرمصرف‌ترین الگوریتم‌های تحلیل مولفه مستقلی که مبتنی بر آنتروپی منفی^{۱۶} است، تحلیل مولفه مستقل سریع [16] است. تحلیل مولفه مستقل سریع مبتنی بر الگوریتمی تکراری است که ماتریس وزن B را بدست می‌آورد.

برای پنجره‌ی سیگنال داده شده‌ی x ، مولفه‌های مستقل s به شکل زیر به دست می‌آیند:

$$s=Wx \quad (2)$$

که W ماتریس جداساز تحلیل مولفه مستقل می‌باشد (یعنی معکوس ماتریس مخلوط‌ساز A)، که از $W=BU$ به دست می‌آید.

۲-۳- قانون بهسازی

الگوریتم کدگذاری تنک انقباضی در حوزه‌ی تبدیل تحلیل مولفه مستقل متعامد شده^{۱۷}، انجام می‌شود. هر پنجره از سیگنال x با استفاده از معادله (۲) با W تبدیل می‌گردد. قبل از آن W به شکل زیر متعامد می‌شود:

$$W=W(W^TW)^{-1/2} \quad (3)$$

در حوزه‌ی تحلیل مولفه مستقل، مولفه‌های پنجره‌ی گفتار تبدیل شده‌ی s از یکدیگر مستقل خواهند بود و در نتیجه، بهسازی مولفه‌های انفرادی به صورت جداگانه قابل انجام می‌باشد.

در این تحقیق از مدل تابع چگالی احتمالی مولفه مستقل s به کار برده شده در [8] استفاده کرده‌ایم:

$$p(s)=\frac{1}{2d}\frac{(\alpha+2)[\alpha(\alpha+1)/2]^{(\alpha/2+1)}}{[\sqrt{\alpha(\alpha+1)/2}+|s/d|]^{(\alpha+3)}} \quad (4)$$

روش بهسازی گفتار مبتنی بر کمینه نمودن مربع میانگین خطا بر گوسی بودن مدل نویز و گفتار تکیه دارند. درحالی که برای بسیاری از کاربردها، مولفه‌های طیفی نویز می‌تواند با متغیرهای تصادفی گوسی مدل شود، در همین حال نشان داده شده است که مولفه‌های طیفی گفتار تمیز با توزیع‌های غیرگوسی، موثرتر مدل می‌گردد [7]. همین مساله، فن‌آوری‌های اخیر بهسازی گفتار که بر چهارچوب بیشینه‌سازی پسین^{۱۸} متکی هستند و با مدل غیرگوسی گفتار سروکار دارند را بوجود آورده است. الگوریتم بهسازی گفتار مبتنی بر بیشینه‌سازی پسین که به آن کدگذاری تنک انقباضی^{۱۹} می‌گویند در سال‌های ۱۹۹۹ و ۲۰۰۱ توسط هاپوارینن^۸ ارائه شد. الگوریتم کدگذاری تنک انقباضی، بهسازی را در حوزه‌ی تبدیل یافته‌ی به دست آمده از تحلیل مولفه مستقل^{۱۱} انجام می‌دهد. این الگوریتم برای بهسازی تصویر ارائه شده بود و بعدها برای بهسازی سیگنال گفتاری، به صورت موفق به کار گرفته شد [9-11]. اخیراً الگوریتم‌های بهسازی گفتار مبتنی بر بیشینه‌سازی پسین دیگری نیز از سال‌های ۲۰۰۳ تا ۲۰۱۳ ارائه شده‌اند [7,12-15].

۲- بهسازی گفتار آلوده به نویزهای جمع‌شونده

مسئله بهسازی سیگنال گفتاری که با نویز جمع‌شونده مستقل آلوده شده باشد را در نظر می‌گیریم. $x(n)$ نمونه‌های سیگنال گفتار تمیز و $v(n)$ نمونه‌های سیگنال نویز را نشان می‌دهند و سیگنال مشاهده شده‌ی نویزی $y(n)$ است:

$$y(n)=x(n)+v(n) \quad (1)$$

$x(n)$ متغیر تصادفی غیرگوسی فرض می‌شود و $v(n)$ نویز گوسی با میانگین صفر و واریانس σ^2 است. بهسازی گفتار را در حوزه تبدیل یافته انجام می‌دهیم، یعنی سیگنال گفتار به قطعات کوچک زمانی (پنجره‌ها) تقسیم می‌شود و سپس با تبدیل (خطی) تبدیل می‌شوند. در اینجا y و x را بلوک‌های K نمونه‌ای از سیگنال نویزی و تمیز z و s را معادل نسخه تبدیل یافته آن‌ها در نظر می‌گیریم. واریانس سیگنال و نویز در حوزه تبدیل یافته با σ_s^2 و σ_v^2 نشان داده می‌شوند.

۲-۱- کدگذاری تنک انقباضی

الگوریتم کدگذاری تنک انقباضی، بهسازی گفتار نویزی را در حوزه‌ی تبدیل مبتنی بر تحلیل مولفه مستقل انجام می‌دهد که ماتریس تبدیل تحلیل مولفه مستقل از داده‌ی گفتار تمیز، قبل از پروسه‌ی بهسازی به دست می‌آید. در ادامه، پروسه‌ی به دست آوردن ماتریس تبدیل تحلیل مولفه مستقل و قانون بهسازی کدگذاری تنک انقباضی را شرح می‌دهیم.

نمونه برداری ۲۲۰۵۰ هرتز باشد، تعداد نمونه ها در یک پنجره ۱۷۶ نمونه خواهد بود.

بعد از تغییر شکل، گفتار نویزی اولیه به شکل زیر نوشته می شود:

$$X = B + V \quad (12)$$

فرض می کنیم که ماتریس متعامد W را به دست آورده ایم،

می توانیم چنین بنویسیم:

$$Y = WX = WB + WV = S + V' \quad (13)$$

که Y ماتریسی است به شکل $[y_1, y_2, \dots, y_n]$ و S ماتریسی است به شکل $[s_1, s_2, \dots, s_n]$ و مولفه خروجی s_i تنک است. چون W متعامد است و V ماتریس نویز سفید است و

$$R_{V'} = V'V'^T = WR_V W^T \quad (14)$$

$R_{V'}$ ماتریس کواریانس نویز سفید اصلی و R_V ماتریس کواریانس نویز سفید هستند. این معادله نشان می دهد که V' نیز ماتریس نویز سفید است با همان ماتریس کواریانس مانند V . بنابراین در فضای کدگذاری تنک، مولفه تمیز نیز با نویز گوسی سفید خراب شده است. تابع انقباضی کدگذاری تنک را می توان به هر مولفه y_i اعمال کرد تا مولفه تمیز s_i را تخمین بزنیم.

$$\hat{s}_i = g(y_i) \quad (15)$$

\hat{S} که $\hat{S} = [\hat{s}_1, \hat{s}_2, \dots, \hat{s}_n]$ ماتریس تخمین زده شده ی تمیز در فضای کدگذاری تنک است. تبدیل معکوس ماتریس تمیز تخمینی به ماتریس گفتار اصلی از طریق تبدیل متعامد W به دست می آید:

$$\hat{X} = W^T \hat{S} \quad (16)$$

در پایان شکل موج گفتار بهسازی شده \hat{X} با تغییر شکل \hat{X} به شکل برداری با چسباندن تمام ستون های ماتریس کنار هم به دست می آید.

۲-۵- تبدیل کدگذاری تنک

از آن جایی که تبدیل تنک ساز W از مهم ترین بخش های کار است، در این بخش بر آن هستیم تا چگونگی به دست آوردن آن را شرح دهیم. در این روش W را از داده های آموزشی که همان مشخصات آماری گفتار آزمایشی را دارند، به دست می آوریم. می توان W را به عنوان یک دیکشنری که حاوی بردارهای پایه است در نظر گرفت. همان طور که قبلا بحث شد چون تحلیل مولفه مستقل، ماتریس تبدیل بسیار تنکی را ارائه می دهد، در به دست آوردن W به کار گرفته شده است و سپس تغییرات لازم جهت متعامد کردن آن طبق فرمول (۳) انجام می شود.

برای محاسبه W از دادگان فارسی 18 که حاوی ۵۰ مورد گفته های ۲۹ مرد و ۲۱ زن است، استفاده می کنیم. هر مورد شامل ۱۰ جمله ادا شده است. برای ساخت W بنابر مشخصات سخت افزاری (حافظه و سرعت) می توان از تعداد ۱۰ مورد تا تمام آنها استفاده نمود. در آزمایشات از ۱۳ نمونه که شامل ۷ صدای مرد و ۶ صدای زن است

که α و d پارامترهای مدل توزیع هستند که با استفاده از فرمول های زیر قابل محاسبه هستند [11,17]:

$$p(s = 0) = \frac{v\eta}{2\Gamma(\frac{1}{v})}, \quad v = F\left(\frac{\mu|s|}{\sqrt{\sigma_s^2}}\right) \quad (5)$$

$$F(u) = \frac{\Gamma(2/u)}{\sqrt{\Gamma(1/u)\Gamma(3/u)}}, \quad \eta = \frac{1}{\sqrt{\sigma_s^2}} \left(\frac{\Gamma(3/v)}{\Gamma(1/v)}\right)^{\frac{1}{2}} \quad (6)$$

$$d = \sqrt{\sigma_s^2}, \quad K = d^2 p(s = 0)^2 \quad (7)$$

$$\alpha = \frac{2-K+\sqrt{K(K+4)}}{2K-1}, \quad a = \sqrt{\frac{\alpha(\alpha+1)}{2}} \quad (8)$$

که $\mu|s|$ میانگین $|s|$ و $\Gamma(\cdot)$ تابع گاما می باشد. تخمین \hat{S} از سیگنال تمیز در حوزه ی تبدیل تحلیل مولفه مستقل از سیگنال نویزی z با استفاده از قانون تخمین کدگذاری تنک انقباضی [8] به دست می آید:

$$\hat{S} = \text{sign}(z) \times \max\left(0, \frac{|z| - ad + \sqrt{(|z| + ad)^2 - 4\sigma^2(\alpha + 3)}}{2}\right) \quad (9)$$

سپس پنجره ی سیگنال گفتار بهسازی شده در حوزه ی زمان \hat{X} توسط $\hat{X} = W^T \hat{S}$ به دست می آید.

قانون بهسازی کدگذاری تنک انقباضی بر فرض آن که سیگنال غیرگوسی بوده و نویز گوسی باشد، استوار است [8]، گرچه این برای نویزهای با چگالی غیر گوسی هم قابل استفاده بوده و کارایی خوبی ارائه داده است [15].

۲-۴- پیاده سازی

فرض می کنیم سیگنال گفتار نویزی x با خراب شدن رشته ی گفتار اصلی b با نویز گوسی v تولید شده است:

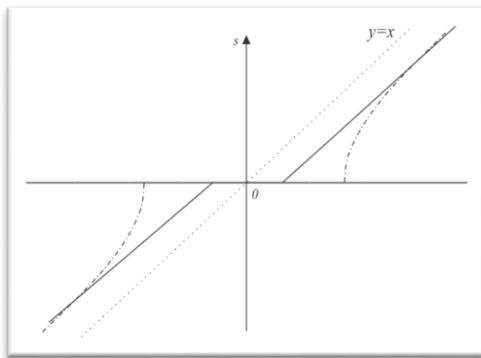
$$x = b + v \quad (10)$$

ماتریس گفتار را با تغییر شکل x بدون هم پوشانی (یا با هر میزان هم پوشانی) می سازیم که در آن طول پنجره ها به اندازه ی $n=1$ میلی ثانیه باشد و در نتیجه تعدادی پنجره خواهیم داشت که به شکل زیر خواهند بود:

$$X = \begin{bmatrix} x_1 & x_{n+1} & x_{2n+1} & \dots & x_{(m-1)n+1} \\ \vdots & & & \ddots & \vdots \\ x_n & x_{2n} & x_{3n} & \dots & x_{mn} \end{bmatrix} \quad (11)$$

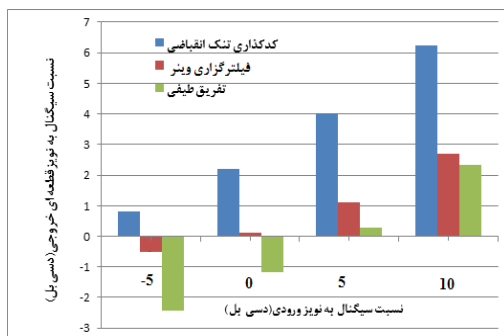
و ماتریسی $m \times n$ خواهیم داشت که m تعداد پنجره ها و n تعداد نمونه ها در یک پنجره ۸ میلی ثانیه ای خواهد بود که اگر نرخ

نوع گوسی هستند و تاثیر کاهش نویز با کدگذاری تنک را با روش‌های دیگر و مطرحی چون فیلترگذاری وینر و تفریق طیفی، مقایسه کرده‌ایم. برای پیاده‌سازی فیلترگذاری وینر از الگوریتم اسکالارت و همکاران^{۲۱} و برای پیاده‌سازی تفریق طیفی از الگوریتم بل^{۲۲} استفاده نموده‌ایم.



شکل (۱): ترسیم تابع انقباضی. تاثیر توابع، کاهش مقدار مطلق آرگومان خود می‌باشد که به سطح نویز وابسته است و آرگومان‌های کوچک، صفر می‌شوند. [8]

در شکل (۲) سه روش از نظر پیشرفت، با نسبت سیگنال به نویز قطعه‌ای^{۲۳} با یکدیگر مقایسه شده‌اند. برای هر یک از چهار نسبت سیگنال به نویز مقادیر ۱۰، ۵، ۰، ۵- دسی‌بل برای نویز افزوده شده استفاده شده است.



شکل (۲): مقایسه‌ی خروجی نسبت سیگنال به نویز قطعه‌ای سه روش بهسازی گفتار

مقادیر بدست آمده متوسط ۱۰ مورد آزمایش شده با نمونه‌های دیده نشده از دادگان می‌باشد. با توجه به شکل (۲) روش کدگذاری تنک انقباضی، بالاترین نسبت سیگنال به نویز قطعه‌ای را در هر چهار سطح نسبت سیگنال به نویز ورودی را دارا می‌باشد و اختلاف آن با دو روش دیگر بسیار زیاد و مشهود است به عنوان نمونه با نویز ۱۰ دسی‌بل، نسبت سیگنال به نویز روش کدگذاری تنک انقباضی ۶/۲ می‌باشد، درحالی‌که این مقدار در روش فیلترگذاری وینر ۲/۸ و در روش تفریق طیفی ۲/۲ می‌باشد.

استفاده شد و برای آزمایش از چندین تک نمونه گفته‌های آقا و خانم بهره گرفته شد. هر نمونه آزمایشی با چهار شدت نویز سفید متفاوت از ۱۰ دسی‌بل تا ۵- دسی‌بل آلوده می‌شود. جملات آموزشی آلوده شده به نویز مانند یک بردار به یکدیگر چسبانده می‌شوند و سپس به فرم ماتریس مانند (۱۱) تغییر شکل داده می‌شوند.

برای محاسبه دو پارامتر α و d در (۷) و (۸) به محاسبه مقدار توزیع در نقطه‌ی صفر نیاز داریم که این امر دقت زیادی را می‌طلبد. از بین مقادیر مختلف، این مقادیر را با اندکی تغییر از [11] و [17] استخراج می‌نماییم:

$$\sigma_s^2(k, 0) = S_k(0)^2 \quad (17)$$

$$\begin{aligned} \sigma_{|S|}(k, 0) &= |S_k(0)| \\ \sigma_s^2(k, m) &= \zeta_{s^2} \sigma_s^2(k, m-1) + (1 - \zeta_{s^2}) S_k(m)^2 \\ \sigma_{|S|}(k, m) &= \zeta_{|S|} \sigma_{|S|}(k, m-1) + (1 - \zeta_{|S|}) |S_k(m)| \end{aligned} \quad (18)$$

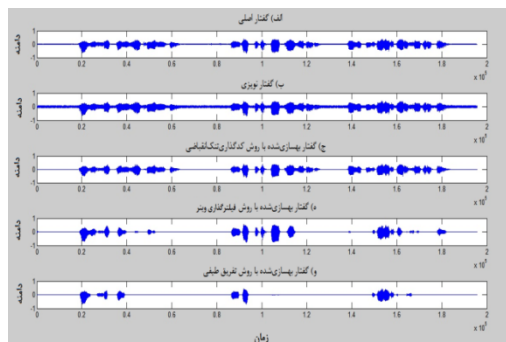
در فرمول‌های بالا ζ_{s^2} و $\zeta_{|S|}$ ثابت‌هایی در بازه [۰، ۱] هستند. از (۱۷) و (۱۸) استفاده می‌شود تا d به‌دست آید و سپس K محاسبه می‌گردد. قبل از محاسبه K لازم است که مقدار توزیع را در صفر محاسبه نماییم که این کار بواسطه فرمول‌های (۵) الی (۸) و با کمک (۱۷) و (۱۸) انجام می‌پذیرد. پس از محاسبه‌ی پارامترهای لازم d و α در فضای تنک سیگنال تمیز، می‌توانیم از تابع انقباضی (۹) برای نویززدایی از مولفه‌های سیگنال آلوده z استفاده کنیم.

این تابع انقباضی، شکلی مانند نقطه خط در شکل (۱) خواهد داشت. در مقایسه چگالی لاپلاس به شکل خط و تابع $y=x$ بصورت نقطه چین نمایش داده شده است. آن‌چه که در این تابع مشاهده می‌شود دانستن واریانس نویز σ_N^2 است که در زمان آزمایش باید محاسبه گردد. برای محاسبه‌ی این پارامتر مهم روش‌های مختلفی ارائه شده است که مورد آزمایش قرار گرفت اما نتایج خوبی ارائه ندادند، در نهایت با روشی ابتکاری، ۱۰ فریم حدود ۸۰ میلی ثانیه از سیگنال مورد آزمایش ورودی را که معمولاً حاوی سکوت است برای تخمین این مقدار استفاده نموده و انحراف مطلق میانگین^{۱۹} آن را بدست می‌آوریم که باعث شد نتایج بسیار خوبی حاصل گردد.

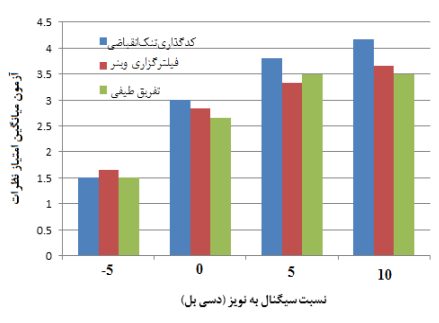
پس از اعمال تابع انقباضی و کاهش یافتن نویز، ماتریس حاصل را با تبدیل معکوس (W^T) به حوزه‌ی زمان برمی‌گردانیم و سپس با کنارهم قرار دادن ستون‌های ماتریس، گفتار تمیزشده به‌دست می‌آید.

۳- آزمایش‌ها و نتایج

در آزمایش‌ها دادگان فارس‌دات مورد استفاده قرار گرفته است و نویز سفید گوسی به آن افزوده شده است. نویزهای مذکور با چهار درجه نسبت سیگنال به نویز^{۲۰} ۱۰، ۵، ۰ و ۵- دسی‌بل تولید شده‌اند. نویزها از



شکل (۴): شکل موج سه جمله از نمونه S138 از (الف) گفتار اصلی (ب) گفتار نویزی شده با نرخ ۱۰ دسی بل (ج) گفتار بهسازی شده با روش کدگذاری تنک انقباضی (د) گفتار بهسازی شده با روش فیلترگذاری وینر (ه) گفتار بهسازی شده با روش تفریق طیفی.



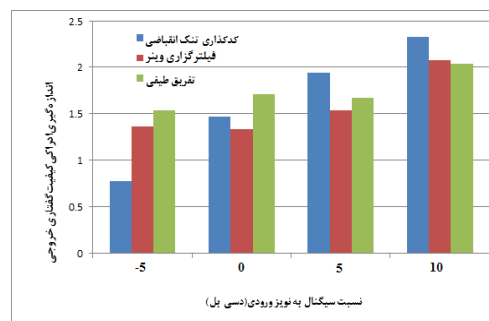
شکل (۵): نمودار مقایسه‌ای سه روش در آزمون میانگین امتیاز نظرات

۴- نتیجه

بیش از دو دهه، کاهش نویز یکی از سرفصل‌های تحقیقاتی پرتوجه در پردازش گفتار و سیگنال‌های صوتی بوده است. کاهش نویز در تحقیقات کدگذاری گفتار، بازنمایی گفتار و کمک شنوایی اجتناب ناپذیر است. به علاوه برای موفقیت در محصولات تجاری، از ارتباطات بی‌سیم نظیر تلفن‌های همراه، رایانه‌های جیبی گرفته تا ابزار مخابراتی برپایه اینترنت نظیر اسکایپ و پیام‌رسان‌ها، نیاز ضروری می‌باشد. از بین چهار گروه الگوریتم‌های شناخته شده بهسازی گفتار، یعنی الگوریتم‌های مبتنی بر تفریق طیفی، فیلترگذاری وینر، روش‌های مبتنی بر مدل‌های آماری و الگوریتم‌های زیرفضا^{۲۵}، در دهه اخیر فن‌آوری کاهش نویز که بسیار مطرح بوده و توجه بسیاری از محققان را به خود جلب نموده تا برای بهبود آن تلاش نمایند، کدگذاری تنک انقباضی است که با نگاهی به لیست مقالات مطرح در سال جاری در زمینه نوپزدایی چندین مقاله ارائه و ثبت نام شده‌اند.

در کدگذاری تنک انقباضی، با استفاده از ایده‌های آماری، نظیر استفاده از بیشینه سازی پسین و به کمک تبدیل مشهوری چون تحلیل مولفه مستقل که پیشتر در زمینه جداسازی کور منابع گفتاری

دومین اندازه‌گیری، اندازه‌گیری ادراکی کیفیت گفتاری^{۲۴} است که در حال حاضر استاندارد صنعتی جهانی به کار گرفته شده برای آزمایش کیفیت صدا می‌باشد. در شکل (۳) خروجی محاسبه‌ی اندازه‌گیری ادراکی کیفیت گفتاری هر سه روش مقایسه شده‌اند. در دو نرخ نویز ۱۰ و ۵ دسی بل، کدگذاری تنک انقباضی کیفیت کاملاً بهتری را ارائه می‌دهد و در نرخ ۰ دسی بل کیفیتی در حد دو روش دیگر دارد و در نرخ ۵- دسی بل کیفیت صدای خروجی آن از دو روش دیگر کمتر است.



شکل (۳): مقایسه‌ی خروجی اندازه‌گیری ادراکی کیفیت گفتار با سه روش متفاوت بهسازی گفتار

در شکل (۴)، شکل موج یک نمونه از دادگان فارسات را در حوزه‌ی زمان با ورودی ۱۰ دسی بل نسبت سیگنال به نویز با سه روش کاهش نویز می‌بینیم. گفتار تمیز و نویزی نیز آورده شده‌اند. شکل موج‌ها به ترتیب: (الف) گفتار اصلی (ب) گفتار نویزی (ج) گفتار بهسازی شده با روش کدگذاری تنک انقباضی (د) گفتار بهسازی شده با روش فیلترگذاری وینر (ه) گفتار بهسازی شده با روش تفریق طیفی، می‌باشند. از شکل (۴) می‌توان فهمید که روش بهسازی گفتار کدگذاری تنک انقباضی توانسته است به بهترین وجهی نویز را کاهش دهد و در عین حال شکل موج کاملاً به شکل موج اصلی شبیه است.

در بخش آزمون ادراکی، یا همان امتیازات میانگین نظرات، از ۶ نفر شنونده با شنوایی عادی شامل ۴ مرد و ۲ زن برای انجام آزمایش استفاده شده است. از قبل آموزشی به آن‌ها داده نشده و در محلی آزمایش انجام شد که از نظر صدا ایزوله باشد و از هدفون کمک گرفته شده است. بلندی صدا برای همه یکسان بوده و نمونه‌های بهسازی شده با سه روش نامبرده برای ایشن پخش شده است. در معیار میانگین امتیاز نظرات از شنوندگان خواسته می‌شود که به جملات گوش کنند و به هر جمله بنا به کیفیت شنیداری آن، عددی بین ۵ (عالی) تا ۱ (بد) اختصاص دهند. امتیازات میانگین نظرات از میانگین ریاضی تمام امتیازهای افراد بدست می‌آید. سه روش و چهار نسبت سیگنال به نویز ۱۰، ۵، ۰ و ۵- آزمایش شده است. نتیجه بدست آمده در شکل (۵) ارائه شده است.

- Int. Conf. on Independent Component Analysis and Blind Signal Separation, 2001.
- [12] Martin, R., *Speech enhancement using MMSE short time spectral estimation with Gamma distributed speech priors*, ICASSP, Orlando, Florida 1, pp. 253-256, 2002.
- [13] Wolfe, P., Godsill, S., *Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement*. EURASIP J. Appl. Signal Process. 10, 1043-1051, 2003.
- [14] Gazor, S., Zhang, W., *Speech enhancement employing Laplacian-Gaussian mixture*, IEEE Trans. Speech Audio Process. 13 (5), 896-904, 2005.
- [15] Zou, X., Jancovic, P., Liu, J., Kokuer, M., *Speech signal enhancement based on MAP algorithm in the ICA space*, IEEE Trans. Signal Process. 56 (5), 1812-1820, 2008a.
- [16] Hyvarinen, A., Karhunen, J., Oja, E., *Independent Component Analysis*, John Wiley & Sons, Inc., 2001.
- [17] Jancovic, P., Zou, X., Kokuer, M., *Speech enhancement based on Sparse Code Shrinkage employing multiple speech models*, Speech Communication 54 (2012) 108-118, 2011.

زیر نویس ها

- ^۱ Spectral Subtraction
- ^۲ Wiener filtering
- ^۳ Virag, 1999
- ^۴ β -order minimum MSE
- ^۵ You et al., 2006
- ^۶ Hu and Loizou, 2003
- ^۷ Maximum A Posteriori (MAP)
- ^۸ Sparse Code Shrinkage (SCS)
- ^۹ Hyvarinem, 1999-2001
- ^{۱۰} Independent Component Analysis (ICA)
- ^{۱۱} Pre-Whitening
- ^{۱۲} Unmixing
- ^{۱۳} Principle Component Analysis (PCA)
- ^{۱۴} Decorrelate
- ^{۱۵} Orthonormal
- ^{۱۶} Negentropy
- ^{۱۷} Orthogonalised
- ^{۱۸} Farsdat
- ^{۱۹} Mean Absolute Deviation (MAD)
- ^{۲۰} Signal to Noise Ratio (SNR)
- ^{۲۱} Scalart et al, 1996
- ^{۲۲} Boll, 1979
- ^{۲۳} Segmental Signal to Noise Ratio (SSNR)
- ^{۲۴} Perceptual Evaluation of Speech Quality (PESQ)
- ^{۲۵} Subspace Algorithms

نتایج چشمگیری را ارائه داده است، سعی می‌گردد که گفتار را به فضایی انتقال دهد که به حداکثر تنگی برسد و با به‌کارگیری محاسباتی مناسب، و با تخمین‌زدن پارامترهای چگالی مولفه‌های بدست آمده که مستقل هستند، به تابع انقباضی برسد که توانایی حذف مولفه‌های ضعیف‌تر را که بیشتر نویز هستند فراهم کند.

از دلایلی که برای برتری این روش بر دو روش مرسوم دیگر می‌توان ذکر کرد، سادگی در فهم و اجرا و داشتن پشتوانه تئوری قوی است.

آزمایش‌های عملی که صورت گرفت نشان داد که تابع انقباضی در کاهش نویز بسیار کارآمد می‌باشد و نسبت به سایر روش‌های مطرح دیگر بهتر عمل می‌نماید. از طرفی آزمایش‌های ادراکی نشان داد که سطح ادراکی جملات نویززدایی شده نسبت به سایر روش‌ها بخصوص در نسبت‌سیگنال‌به‌نویزهای بالاتر، بهتر عمل نموده است.

در کارهای آتی می‌توان تاثیر روش کدگذاری تنک‌انقباضی را بر سایر انواع نویزها بررسی نمود و نیز می‌توان تاثیر گوسی بودن گفتار را نیز بر کارایی روش مورد بررسی قرار داد.

مراجع

- [1] Boll, S., *Suppression of acoustic noise in speech using spectral subtraction*, IEEE Trans. Acoust Speech Signal Process. 27 (2), 113-120, 1979.
- [2] Lim, J., Oppenheim, A., *Enhancement and bandwidth compression of noisy speech*, Proc. IEEE 67, 1586-1604, 1979.
- [3] McAulay, R., Malpass, K., *Speech enhancement using a minimum mean-square error short-time spectral amplitude enhancement*, IEEE Trans. Acoust. Speech Signal Process. 28 (2), 137-145, 1980.
- [4] Virag, N., *Single channel speech enhancement based on masking properties of the human auditory system*, IEEE Trans. Speech Audio Process. 7 (2), 126-137, 1999.
- [5] You, C., Koh, S., Rahardja, S., *Masking-based beta-order MMSE speech enhancement*, Speech Commun. 48, 50-70, 2006.
- [6] Hu, Y., Loizou, P., *A perceptually motivated approach for speech enhancement*. IEEE Trans. Speech Audio Process. 11 (5), 457-465, 2003.
- [7] Lotter, T., Vary, P., *Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model*, EURASIP J. Appl. Signal Process., 1110-1126, 2005.
- [8] Hyvarinen, A., *Sparse code shrinkage: Denoising of nonGaussian data by maximum likelihood estimation*, Neural Comput. 11 (7), 1739-1768, 1999.
- [9] Lee, J. H., Jung, H. Y., Lee, T. W., Lee, S. Y., *Speech enhancement with MAP estimation and ICA-based speech features*, Electronic Lett. 36, 1506-1507, 2000.
- [10] Potamitis, I., Fakotakis, N., Kokkinakis, G., *Speech enhancement using the sparse code shrinkage technique*, ICASSP, Salt Lake City, Utah, pp. 621-624, 2001.
- [11] Choi, C., Choi, S., Kim, S., *Speech enhancement using sparse code shrinkage and global soft decision*, In: Proc.