

کاهش اثر نوفه باد در گفتار ضبط شده با بهره گیری از مجموعه داده های تحلیل طیفی تکین (تطت-اس اس ای) و شیوه یادگیری ماشین

معصومه شفیعیان^{*}، فرشاد قیاسوند

دانشکده فنی و مهندسی رسانه، دانشگاه صدا و سیما

دریافت: ۱۴۰۰/۰۷/۰۴، پذیرش: ۱۴۰۰/۱۲/۲۶

چکیده

در حالت کلی ضبط صدا در محیط های مختلف با چالش های متفاوتی روبه رو می باشد. نوفه باد در فضای باز اغلب باعث افت شدید کیفیت علامت گفتار می شود. بنابراین باید الگوریتم هایی برای کاهش نوفه باد به کار برد. به دلیل ماهیت نایستا و پهن باند بودن نوفه باد، صافی کردن (فیلتر کردن) و حذف کردن آن بسیار مشکل است. تحلیل طیفی تکین (اس اس ای) یک روش قدرتمند تخمین طیفی است که از آن در کاربردهایی از جمله کاهش نوفه، پیش بینی سری زمانی و غیره استفاده می شود. الگوریتم اس اس ای علامت ها را به فضاهای مقدار ویژه تجزیه می کند، مولفه های اصلی را براساس سهم آن ها انتخاب و دسته بندی می کند و سرانجام مولفه های مطلوب را ساخته و به حوزه زمان برمی گرداند. در این پژوهش از روش اس اس ای و الگوریتم کی-مینز و صافی بول به شکل ترکیبی و به منظور کاهش نوفه باد در علامت های گفتار استفاده شده است. علامت گفتار مورد استفاده در روند پژوهش از دادگان تی آی ام آی تی استخراج شده است و نوفه باد استفاده شده نیز مربوط به دادگانی است که توسط دانشگاه فنی آخن ارائه گردیده اند. نتایج نشان می دهند که روش پیشنهادی قادر است تا مقدار ۵۱ درصد میزان کیفیت علامت گفتار را از لحاظ کیفیت ادراکی بهبود بخشد. روش ارائه شده در مقایسه با روش های مرسوم از قبیل وزن دهی طیفی و تفریق طیفی نتایج بهتری از خود نشان می دهند.

کلیدواژه ها: تحلیل طیفی تکین (اس اس ای)، کاهش نوفه باد، الگوریتم کی-مینز، کیفیت ادراکی، علامت گفتار.

۱. مقدمه

از آن جا که نوفه باد کیفیت صدا در ضبط صدا در فضای باز را کاهش می دهد، تلاش برای کاهش آن اجتناب ناپذیر می باشد. نوفه باد به دلیل اغتشاش هوا در ورودی میکروفون و لرزش غشاء میکروفون اتفاق می افتد و میزان قابلیت فهم علامت گفتار را تا حد زیادی کاهش می دهد. در جریان باد با سرعت کم تر از ۵ متر بر ثانیه میزان نوفه می تواند به حدود ۱۱۰ دسی بل (اس پی ال) برسد [۱]. علاوه بر ضبط علامت های صوتی، نوفه باد یکی از مهم ترین مشکلات افرادی که از سمعک استفاده می کنند نیز می باشد. چنین نوفه ای مکالمه و گفتگو را برای این افراد حتی در محیط هایی که باد آرامی می وزد، بسیار دشوار و غیرممکن می کند و باعث می شود که همه یا

بخشی از صحبت های طرف مقابل را متوجه نشوند. روش های بسیاری برای کاهش اثر نوفه باد ارائه شده است [۲] که البته هر کدام از روش ها شرایط مخصوص به خود را دارند. در حالت کلی می توان روش های کاهش نوفه باد را به دو دسته روش هایی که از ساختار فیزیکی برای کاهش نوفه باد استفاده می کنند [۳-۴] و روش های مبتنی بر پردازش علامت تقسیم بندی نمود. علاوه بر ساختار فیزیکی روش های پردازش علامت نیز به شکل گسترده ای مورد مطالعه قرار گرفته اند [۵-۷]. در آغاز، علامت نوفه باد دریافتی در چند نقطه ای مختلف مورد بررسی قرار گرفت و مشخص شد که برخلاف نوفه های متنوع دیگر هیچ ارتباطی بین نوفه باد در مکان های مختلف وجود ندارد؛ البته توجه به این نکته که در بسامدهای پایین تر قضیه کمی متفاوت خواهد شد، ضروری می باشد. برای پی بردن

^{*} نویسنده پاسخگو: masoume@gmail.com

طیفی تکین تجزیه می‌کند [۱۰]، که از نظر فیزیکی به‌عنوان مؤلفه‌های نوسانی و روند و غیره معنی‌دار هستند؛ در نتیجه با استفاده از چنین مبنایی می‌توان با جداکردن نوفه پس‌زمینه، گفتار بدون نوفه را بازسازی کرد.

۲-۱-۲. مرحله دوم

پس از تجزیه سری زمانی علامت توسط الگوریتم اس‌اس‌ای و ایجاد داده‌ها و مؤلفه‌های قابل استفاده، باید از میان این داده‌ها، آن دسته از داده‌هایی را که مربوط به علامت گفتار می‌باشند استخراج نموده و از کنار هم قرار دادن آن‌ها علامت گفتار بدون نوفه را بازسازی نمود. در این پژوهش از الگوریتم کی-مینز^۲ به‌منظور دسته‌بندی و استخراج داده‌های مربوط به گفتار استفاده شده است که در ادامه مقاله توضیح داده می‌شوند.

۲-۱-۳. مرحله سوم

به‌عنوان یک فرآیند کامل‌کننده و بهبوددهنده، در مرحله سوم از صافی (فیلتر) بول استفاده می‌شود. اگرچه این روش در جرگه روش‌های کاهش نوفه رایج می‌باشد و به علت ناپایداری نوفه باد استفاده از آن به‌تنهایی برای نوفه باد نتایج درخور توجهی را نشان نمی‌دهد، اما از آن‌جا که در مرحله‌های قبل تا مقدار زیادی ناپایداری نوفه باد از بین می‌رود، استفاده از این روش به‌صورت ترکیبی با الگوریتم ارائه‌شده منجر به بهبود نتایج می‌شود.

۲-۲. پیش‌پردازش

تبدیل‌های گوناگونی در پردازش، برای گرفتن اطلاعات و تجزیه علامت و اعمال مراحل بعدی پردازش مورد استفاده قرار می‌گیرد. یکی از تبدیلات مهم و کارآمد در این زمینه تبدیل کسینوسی گسسته (دی‌سی‌تی)^۳ است. تبدیل کسینوسی گسسته، شباهت بسیاری به تبدیل فوریه گسسته (دی‌اف‌تی)^۴ دارد، با این تفاوت که حاصل تبدیل فقط مقادیر حقیقی دارد. تبدیل کسینوسی گسسته، دنباله‌ای محدود از اعداد (داده‌ها) را به‌صورت مجموع توابع

به رابطه مکانی موجود بین نوفه باد از آرایه‌های میکروفونی استفاده می‌شود. پس از آزمایش‌های متعدد مشخص شد که می‌توان ارتباط کمی بین نوفه باد در میکروفون‌هایی که بسیار نزدیک به یکدیگر می‌باشند مشاهده نمود. بعد از تجزیه و تحلیل زمانی و مکانی علامت نوفه باد، محققان توانستند با استفاده از یک دسته‌بندی‌کننده‌ی الگوی ترکیبی گوسی، نوفه باد و علامت صوتی را از یکدیگر جدا کنند. در سال‌های اخیر روش‌های متنوع دیگری برای کاهش اثر نوفه باد استفاده شده‌اند که هرکدام از آن‌ها از ویژگی‌های مختلف نوفه باد و گفتار برای این منظور بهره‌برده‌اند [۶-۸]. به دلیل طبیعت متغیر با زمان و ناپایداری نوفه باد، بسیاری از روش‌های پردازش علامت برای کاهش نوفه باد کارایی محدودی نشان داده‌اند و دارای محدودیت‌های ذاتی اعوجاجی شدن علامت بهبودیافته و کمبود سطح انرژی مورد نیاز برای ضبط صدا پس از بهبود مواجه هستند. از آن‌جا که الگوریتم تجزیه و تحلیل طیفی تکین (اس‌اس‌ای)^۱ در کاربردهای مختلف با موفقیت به‌کار گرفته شده است، این مقاله به بررسی و تحقیق درباره اثر ترکیب الگوریتم اس‌اس‌ای و روش یادگیری ماشین در کاهش نوفه باند در علامت گفتار می‌پردازد.

۲. مواد و روش‌ها

۲-۱. الگوریتم پیشنهادی

پس از پیش‌پردازش که در بخش ۲-۲ توضیح داده می‌شود، روش پیشنهادی را می‌توان به شکل تجزیه‌ی علامت گفتار تخریب‌شده به‌نوعی مؤلفه‌های قابل تفکیک و در ادامه جداسازی مؤلفه‌های دلخواه به شکلی که نوفه باد به حداقل برسد بیان نمود، و در مرحله‌ی آخر از صافی (فیلتر) بول به منظور کاهش نوفه باقی‌مانده استفاده می‌شود [۹]. در حالت کلی روش پیشنهادی را می‌توان به سه مرحله اصلی تقسیم‌بندی نمود.

۲-۱-۱. مرحله اول

روش اس‌اس‌ای یک سری زمانی را به مؤلفه‌های حوزه‌ی

^۲ K-means

^۳ DCT

^۴ DFT

^۱ SSA; Singular Spectrum Analysis

۲-۳-۱. فرایند جای‌گذاری

۲-۳-۱-۱. نمایش برداری

همان‌طور که گفته شد اس‌اس‌ای یک روش طیفی می‌باشد و در آن سری زمانی به فضای برداری انتقال داده می‌شود. اگر X یک سری زمانی به طول N_t با مقادیر حقیقی در نظر گرفته شود، می‌توان سری زمانی مذکور را در یک بردار به شکل رابطه‌ی ۱ نمایش داد. لازم به ذکر است که مقادیر سری زمانی در این پژوهش، علامت خروجی مرحله‌ی پیش‌پردازش می‌باشند [۱۰].

$$x^T = (x_1, x_2, \dots, x_{N_t}) \quad (1)$$

۲-۳-۱-۲. ایجاد ماتریس تراچکتوری

برای ایجاد ماتریس تراچکتوری^۳ یک پنجره به طول m در نظر گرفته می‌شود و روی سری زمانی حرکت داده می‌شود. لازم به ذکر است انتخاب مقدار مناسب طول پنجره در عملکرد الگوریتم نقش بسزایی دارد. ماتریس تراچکتوری در سطر اول خود شامل بخشی (برابر طول پنجره) از سری زمانی می‌باشد که بدون تأخیر در سطر اول جای گرفته است، در باقی سطرها به ترتیب با ایجاد یک واحد تأخیر در هر سطر نسبت به سطر قبل سری زمانی جای‌گذاری می‌شود. بنابراین تعداد ستون‌های ماتریس تراچکتوری برابر طول پنجره یعنی m می‌باشد. همان‌طور که مشخص است طول بردارهای جای‌گذاری شده با یکدیگر برابر می‌باشند و درایه‌های مربوط به هر سطر نسبت به سطر بعدی به میزان یک واحد شیفت داده شده است. با دقت در ماتریس به وجود آمده مشخص می‌شود که عناصر روی قطرهای فرعی یکسان می‌باشند. در نهایت ماتریس تراچکتوری مشابه رابطه‌ی ۲ تولید می‌شود.

$$Y = \begin{bmatrix} x_1 & \cdots & x_m \\ \vdots & \ddots & \vdots \\ x_N & \cdots & x_{N_t} \end{bmatrix} \quad (2)$$

۲-۳-۱-۳. محاسبه ماتریس کواریانس

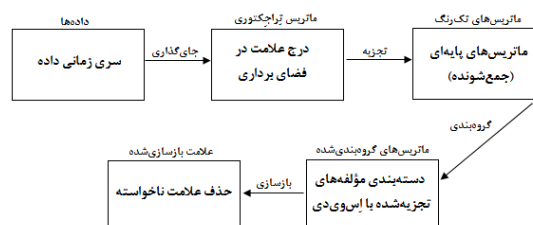
ماتریس کواریانس، ماتریسی است که همبستگی بین مقادیر $x(t)$ و $x(t+k)$ را محاسبه می‌کند که در واقع

کسینوسی با بسامدهای متفاوت نمایش می‌دهد. این تبدیل استفاده گسترده‌ای در علوم و مهندسی دارد؛ از فشرده‌سازی با اتلاف امواج صوتی (مانند ام‌پی۳^۱) و تصویر (مانند جی‌پی‌ای‌سی جی^۲) که مقادیر کوچک تبدیل در بسامدهای بالا قابل حذف هستند، تا روش‌های طیفی برای راه‌حل عددی معادله دیفرانسیل با مشتقات پاره‌ای، در گستره‌ی استفاده از دی‌سی‌تی‌ها قرار می‌گیرد.

همان‌طور که اشاره شد یکی از تفاوت‌های مهم تبدیل فوریه و تبدیل کسینوسی گسسته مربوط به وجود و یا عدم وجود مقادیر مختلط در نتایج تبدیل می‌باشد. با توجه به این که الگوریتم کی-مینز در دسته‌بندی داده‌های مختلط به مشکل برمی‌خورد، در این پژوهش برای پیش‌پردازش داده‌ها (علامت گفتار) از تبدیل کسینوسی گسسته استفاده می‌شود. در حالت کلی علامت را باید به گونه‌ای پیش‌پردازش نمود که الگوریتم اس‌اس‌ای توانایی بیش‌ترین تفکیک‌کنندگی را داشته باشد. به‌صورت تجربی مشخص می‌شود که دی‌سی‌تی این قابلیت را دارا می‌باشد.

۲-۳. الگوریتم اس‌اس‌ای

ایده‌ی اس‌اس‌ای، بر پایه‌ی تجزیه سری زمانی به تعدادی مؤلفه‌ی قابل تفسیر می‌باشد. هر کدام از مؤلفه‌های تجزیه شده را می‌توان در یک گروه خاص شرکت داد و آن‌ها را گروه‌بندی کرد. در حالت کلی اس‌اس‌ای دارای دو قسمت اساسی است که مکمل یکدیگرند و عبارتند از تجزیه و بازسازی. این الگوریتم عمدتاً براساس رویکردهای آماری می‌باشد و بسیاری از روابط موجود در این الگوریتم براساس جبر خطی هستند. مراحل مختلف الگوریتم اس‌اس‌ای در شکل ۱ مشاهده می‌شود [۱۰].



شکل ۱. بلوک دیاگرام الگوریتم اس‌اس‌ای.

¹ MP3

² JPEG

³ Trajectory

بردار ویژه نیز خواص همان بردار را دارا است. با فرض متقارن بودن ماتریس C (یعنی $C=C^T$) و حقیقی بودن مقادیر ویژه، در صورتی که این مقادیر با یکدیگر متفاوت باشند؛ بردارهای ویژه مربوط به هر کدام از آن‌ها بر یکدیگر عمود هستند. شکل کلی ماتریس‌های بردارهای ویژه و مقادیر ویژه به شکل رابطه‌ی ۶ می‌باشد.

$$E = \begin{bmatrix} e_1^1 & \cdots & e_1^m \\ \vdots & \ddots & \vdots \\ e_m^1 & \cdots & e_m^m \end{bmatrix} \quad (6)$$

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}$$

ماتریس مقادیر ویژه، یک ماتریس متقارن می‌باشد که مقادیر ویژه روی قطر اصلی آن قرار گرفته‌اند و مقدار آن‌ها هر چقدر به سمت انتهای قطر اصلی حرکت شود؛ کاهش می‌یابد. مرتبه‌ی^۳ ماتریس تراچکتوری برابر تعداد مقادیر ویژه غیر صفر است.

۲-۳-۲. تجزیه و تشکیل ماتریس‌های پایه‌ای

برای محاسبه‌ی مؤلفه‌های اساسی از بردارهای ویژه استفاده می‌شود؛ در حالت کلی می‌توان این‌گونه بیان نمود که از ضرب ماتریس تراچکتوری بردارهای ویژه مؤلفه‌های اساسی به دست می‌آیند. مؤلفه اساسی مربوط به هر مقدار و بردار ویژه با استفاده از رابطه‌ی ۷ زیر محاسبه می‌شود [۱۱].

$$V_i = \frac{\gamma_i^T E_i}{\sqrt{\lambda_i}} \quad i = 1 \dots d \quad (7)$$

برای محاسبه‌ی ماتریس‌های پایه‌ای نیز از مؤلفه‌های اساسی محاسبه شده استفاده می‌شود که رابطه‌ی آن به شکل رابطه‌ی ۸ می‌باشد.

$$\gamma_i = \sqrt{\lambda_i} E_i V_i^T \quad i = 1 \dots d \quad (8)$$

لازم به ذکر است که برای محاسبه‌ی مؤلفه‌های اساسی و ماتریس‌های پایه‌ای از همه‌ی مقادیر و بردارهای ویژه استفاده نمی‌شود و آن دسته از مقادیر ویژه که برابر صفر و یا مقداری نزدیک به صفر می‌باشند به علت این‌که حاوی اطلاعات خاصی نیستند در نظر گرفته نمی‌شوند. به عنوان

همان محاسبه‌ی همبستگی بین مقادیر تأخیر داده شده‌ی علامت اصلی می‌باشد. دو روش برای محاسبه ماتریس کوواریانس وجود دارد. یک روش محاسبه مستقیم ماتریس کوواریانس تأخیری از روی سری زمانی می‌باشد. ماتریس کوواریانس یک ماتریس $m \times m$ و به شکل تریپل‌ت^۱ می‌باشد. مقدار هر یک از درایه‌های این ماتریس به میزان تأخیر بین نمونه‌ها بستگی دارد. رابطه محاسبه ماتریس کوواریانس به شکل رابطه‌ی ۳ می‌باشد [۱۰].

$$c_{ij} = \frac{1}{N_t - |i - j|} \sum_{t=1}^{N_t - |i - j|} X(t) X(t + |i - j|) \quad (3)$$

که c_{ij} ، بیانگر درایه‌های ماتریس کوواریانس و N_t برابر طول سری زمانی می‌باشد.

درو روش دوم از ماتریس تراچکتوری و ترانهاده‌ی آن به منظور محاسبه ماتریس کوواریانس تأخیری استفاده می‌شود. در یک بیان ریاضی می‌توان محاسبه ماتریس کوواریانس را به شکل رابطه‌ی ۴ بیان نمود.

$$C = \sum_{i=1}^N \gamma_i \gamma_i^T \quad (4)$$

در رابطه‌ی ۴، C و N به ترتیب بیانگر ماتریس کوواریانس و تعداد سطرهای ماتریس تراچکتوری می‌باشند.

۲-۳-۱-۴. محاسبه ماتریس مقادیر ویژه و بردارهای ویژه

در این مرحله با استفاده از تجزیه اس‌وی‌دی^۲ بردارها و مقادیر ویژه ماتریس کوواریانس محاسبه می‌شود. ماتریس کوواریانس (C) یک ماتریس متقارن با بردارهای ویژه‌ی مستقل خطی است. اگر E ماتریسی باشد که ستون‌های آن شامل بردارهای ویژه‌ی ماتریس C باشد و هم‌چنین ماتریس Λ شامل مقادیر ویژه باشد، در این صورت رابطه‌ی ۵ بین بردارها و مقادیر ویژه و ماتریس C برقرار است [۱۰].

$$C = E \Lambda E^T \quad (5)$$

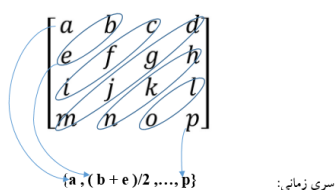
ماتریس Λ مربوط به مقادیر ویژه می‌باشد که یک ماتریس قطری است و تمامی درایه‌های روی قطر اصلی آن دارای مقادیری نامنفی می‌باشند. نکته قابل توجه در مورد ماتریس E منحصر به فرد نبودن آن است زیرا مضربی از هر

^۱ Toeplitz

^۲ SVD

^۳ Rank

مقادیر ویژه در نظر گرفته می‌شود. در نهایت تمامی ماتریس‌هایی که مربوط به گفتار می‌باشند با یکدیگر جمع می‌شوند و ماتریسی مشابه ماتریس تراچکتوری مربوط به گفتار بدون نوفه ساخته می‌شود. سپس ماتریس به وجود آمده با استفاده از میانگین‌گیری روی قطر فرعی، مشابه شکل ۲ به صورت تک‌بعدی بازنویسی می‌شود، و پس از معکوس کردن پیش‌پردازش (آدی‌سی‌تی) علامت گفتار بدون نوفه بازسازی می‌شود [۱۱].



شکل ۲ میانگین‌گیری قطری.

۴-۳-۲. مؤلفه‌های روش اس‌اس‌ای

در روش اس‌اس‌ای مؤلفه‌های مختلفی وجود دارد که عبارتند از: طول قاب، طول پنجره، آستانه مقادیر ویژه (مقادیر ویژه کم‌تر از این مقدار در نظر گرفته نمی‌شوند) و میزان تأثیر هر کدام از دسته‌های ماتریس‌های پایه‌ای در علامت گفتار تمیز؛ انتخاب مقادیر مناسب علاوه بر کاهش میزان محاسبات در نحوه عملکرد روش اس‌اس‌ای نیز تأثیرگذار است. جدول ۱ مقدار مناسب این مؤلفه‌ها را که در این پژوهش به صورت تجربی به دست آمده نشان می‌دهد.

جدول ۱ مؤلفه‌های اس‌اس‌ای.

ردیف	مؤلفه	مقدار مؤلفه
۱	قاب	۰/۳۸ ثانیه
۲	پنجره	۰/۰۱ ثانیه
۳	آستانه	10^{-4}
۴	ضریب	۰/۲۵

۴-۲. الگوریتم کی-مینز

در خوشه‌بندی کی-مینز از بهینه‌سازی یک تابع هدف استفاده می‌شود. پاسخ‌های حاصل از خوشه‌بندی در این روش، ممکن است به کمک کمینه‌سازی یا بیشینه‌سازی تابع هدف صورت گیرد. به این معنی که اگر ملاک «میزان

مثال از m مقدار ویژه، d مقدار آن در نظر گرفته می‌شود [۱۰].

ماتریس‌های γ_i ، ماتریس‌های رتبه ۱ می‌باشند به همین دلیل به آن‌ها ماتریس‌های پایه‌ای گفته می‌شود. برای انجام عملیات کاهش نوفه، دسته‌بندی و جداسازی این ماتریس‌های پایه‌ای بسیار مهم است. در واقع باید آن دسته از ماتریس‌های پایه‌ای که مربوط به علامت گفتار می‌باشند را جدا کرده و با استفاده از آن‌ها سری زمانی مربوط به گفتار بدون نوفه بازسازی شود. اگر همه‌ی ماتریس‌های پایه‌ای با هم جمع شوند ماتریس حاصل برابر ماتریس تراچکتوری که از روی علامت نوفه‌ای ایجاد شده است خواهد شد [۱۱].

۳-۳-۲. گروه‌بندی ماتریس‌های پایه‌ای

مرحله گروه‌بندی مربوط به قسمت دوم الگوریتم اس‌اس‌ای، یعنی قسمت بازسازی می‌باشد. در این مرحله ماتریس‌های پایه‌ای براساس شباهت به چندین گروه تقسیم می‌شوند. گروه‌بندی ماتریس‌ها با استفاده از الگوریتم کی-مینز صورت می‌پذیرد و در پایان گروه‌بندی ماتریس‌های مربوط به هر گروه با یکدیگر جمع می‌شوند. فرض کنید که اندیس‌های مربوط به گروه I-اُم به شکل $\{i_1, \dots, i_n\}$ باشند در این صورت ماتریس مربوط به گروه I، γ_I می‌باشد که از رابطه‌ی ۹ به دست می‌آید.

$$\gamma_I = \gamma_{i_1} + \dots + \gamma_{i_p} \quad (9)$$

سپس γ_I ‌هایی که مربوط به علامت بدون نوفه می‌باشند با یکدیگر جمع می‌شوند. فرض کنید m گروه از آن‌ها مربوط به علامت بدون نوفه باشند، در نتیجه ماتریس نهایی از رابطه‌ی ۱۰ به دست می‌آید.

$$\gamma = \gamma_{I_1} + \dots + \gamma_{I_m} \quad (10)$$

برای تشخیص این‌که یک گروه مربوط به علامت گفتار و یا نوفه باد می‌باشد، از مجموع مقادیر ویژه‌ی ماتریس‌های موجود در گروه‌ها استفاده می‌شود. اگر مجموع مقادیر ویژه‌ی ماتریس‌ها در هر گروه از یک آستانه مشخص بیش‌تر باشد، آن گروه مربوط به گفتار در نظر گرفته می‌شود. در غیر این صورت گروه مربوطه به نوفه اختصاص می‌یابد و در جمع‌بندی نهایی تأثیر داده نمی‌شود. آستانه موردنظر به صورت تجربی بین ۰/۲ الی ۰/۲۵ مجموع کل

استفاده می‌شود.

۲-۴-۱. مؤلفه‌های کی-مینز

یکی از مؤلفه‌های مهم که در تعیین عملکرد این الگوریتم نقش بسزایی دارد؛ تعیین معیار دسته‌بندی داده‌ها می‌باشد. در این پژوهش از معیار همبستگی برای دسته‌بندی داده‌ها استفاده می‌شود. مؤلفه دیگر مربوط به تعداد خوشه‌ها می‌باشد که در عملکرد سامانه بسیار تأثیرگذار است. در این پژوهش تعداد دسته‌های تخصیص داده شده برابر ۶ می‌باشد و تعداد دفعات تکرار الگوریتم بیش از ده مرتبه در نظر گرفته شده است.

۲-۵. صافی بول^۴

استفاده از این صافی (فیلتر) به صورت سری و عمده‌تاً به منظور بهبود نتایج می‌باشد، صافی بول یک روش چندمرحله‌ای است. صافی بول جزء روش‌های تفریق طیفی می‌باشد و می‌توان این صافی را به عنوان یک روش مناسب تفریق طیفی در عملیات کاهش نوفه به کار برد. صافی مذکور برای علامت‌های نوفه‌ای که پایداری محلی دارند روش مناسبی است. در علامت گفتار آغشته به نوفه باد، پس از انجام مراحل قبل تا مقدار زیادی از ناپایداری نوفه باد کاسته می‌شود در نتیجه صافی بول می‌تواند کمک‌کننده باشد. در یک بیان کلی در این صافی ابتدا نوع پنجره‌ی مناسب انتخاب می‌شود. در ادامه طیف دامنه علامت در هر پنجره محاسبه می‌شود و هم‌چنین طیف نوفه در زمان‌هایی که گفتار وجود ندارد، تخمین زده می‌شود و از طیف علامت کلی تفریق می‌شود. پس از تفریق طیفی پردازش‌های دیگری نیز روی علامت انجام می‌شود به عنوان مثال «یکسوسازی نیم‌موج» که در آن زمان‌هایی که مقدار طیف منفی می‌شود، مقدار طیف برابر صفر قرار می‌گیرد. سپس یک روش کاهش نوفه ثانویه نیز برای از بین بردن نوفه باقی‌مانده اعمال می‌شود و در نهایت در زمان‌هایی که گفتار وجود ندارد علامت با یک ضریب ثابت تضعیف می‌شود [۹]. در صافی بول این پژوهش، طول پنجره ۰/۲۵ میلی‌ثانیه، هم‌پوشانی ۰/۵ میلی‌ثانیه و نوع پنجره همینگ در نظر گرفته شده است.

فاصله بین اشیاء باشد، تابع هدف براساس کمینه‌سازی خواهد بود. پاسخ عملیات خوشه‌بندی، پیدا کردن خوشه‌هایی است که فاصله بین اشیاء هر خوشه کمینه باشد. در مقابل، اگر از تابع مشابهت برای اندازه‌گیری مشابهت اشیاء استفاده شود، تابع هدف را طوری انتخاب می‌کنند که پاسخ خوشه‌بندی مقدار آن را در هر خوشه بیشینه کند. معمولاً زمانی که هدف کمینه‌سازی باشد، تابع هدف را «تابع هزینه» نیز می‌نامند.

فرض کنید مشاهدات $(x_1, \dots, x_n)$ که دارای d -بعد هستند باید به k بخش یا خوشه تقسیم شوند. این بخش‌ها یا خوشه‌ها با مجموعه‌ای به نام $S = \{S_1, \dots, S_k\}$ شناخته می‌شوند. اعضای خوشه‌ها باید به شکلی از مشاهدات انتخاب شوند که تابع «مجموع مربعات درون خوشه‌ها (دابلایوسی‌اس‌اس)»^۱ که در حالت یک-بعدی شبیه واریانس است، کمینه شود. بنابراین، تابع هدف در این الگوریتم به صورت رابطه‌ی ۱۱ نوشته می‌شود.

$$\arg \min_{S_i} \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg \min_{S_i} \sum_{i=1}^k |S_i| \text{Var } S_i \quad (11)$$

در اینجا منظور از μ_i میانگین خوشه S_i و $|S_i|$ تعداد اعضای خوشه i -ام است. البته می‌توان نشان داد که کمینه کردن این مقدار به معنی بیشینه‌سازی «میانگین مربعات فاصله بین نقاط در خوشه‌های مختلف (بی‌سی‌اس‌اس)»^۲ است زیرا طبق قانون واریانس کل، با کم‌شدن مقدار دابلایوسی‌اس‌اس، مقدار بی‌سی‌اس‌اس افزایش می‌یابد، زیرا واریانس کل ثابت است.

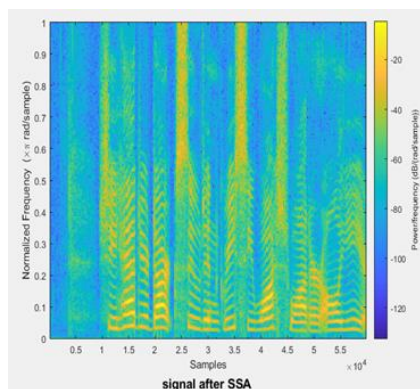
از آنجایی که در این روش خوشه‌بندی، محاسبه فاصله بین نقاط توسط تابع فاصله اقلیدسی انجام می‌شود، از این الگوریتم به صورت استاندارد، فقط برای مقدارهای عددی (و نه ویژگی‌های کیفی) می‌توان استفاده کرد. از طرف دیگر نسخه‌های تعمیم‌یافته از روش خوشه‌بندی کی-مینز نیز وجود دارد که با توابع دیگری مانند فاصله منهتن، گریشین^۳ و یا فاصله‌هایی که برای داده‌های باینری قابل استفاده است، مراحل خوشه‌بندی را انجام می‌دهد. در این پژوهش از الگوریتم‌های تعمیم‌یافته برای دسته‌بندی

^۱ WCSS; Within-Cluster Sum of Squares

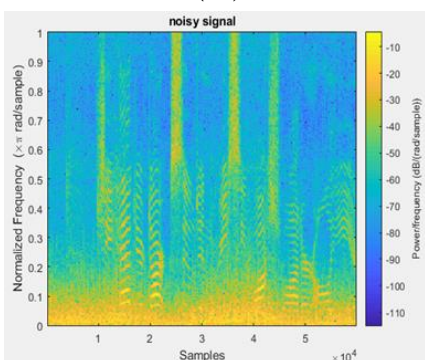
^۲ BCSS; Between Cluster Sum of Squares

^۳ Correlation

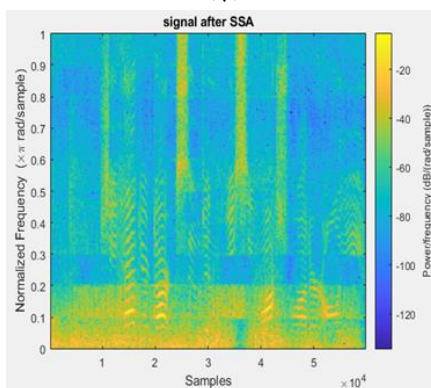
^۴ Boll



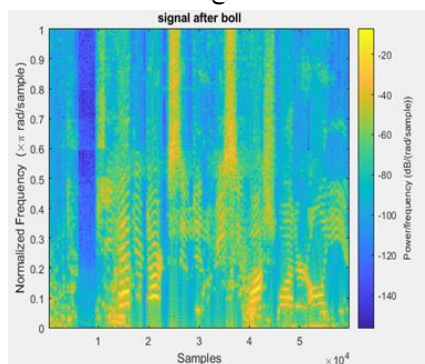
(الف)



(ب)



(ج)



(د)

شکل ۳ اسپکتروگرام یک علامت در شرایط مختلف: (الف) بدون نوفه، (ب) SNR=-5dB، (ج) پس از اعمال الگوریتم اس‌اس‌ای، (د) پس از اعمال الگوریتم اس‌اس‌ای و گذر از صافی بول.

جدول ۲ نتایج خروجی با معیار اس‌ان‌آر بر حسب دسی‌بل.

حالت	علامت نوفه‌ای	علامت در خروجی روش اس‌اس‌ای	علامت در خروجی صافی بول
الف	-۵	۳۶۱	۵۰۶
ب	۰	۵۳	۶۳
ج	۵	۷	۷۵۲

جدول ۳ نتایج خروجی با معیار پی‌ای‌اس‌کیو.

حالت	علامت نوفه‌ای	علامت در خروجی روش اس‌اس‌ای	علامت در خروجی صافی بول
الف	۱/۳۴	۱/۸۳	۲/۰۲
ب	۱/۷۴	۲/۱	۲/۴۱
ج	۲/۰۸	۲/۲۶	۲/۶۷

۳. نتایج

علامت‌های صوتی برای آزمایش روش پیشنهادی، از دادگان تی‌آی‌ام‌آی تی^۱ استخراج گردیده است و نوفه باد استفاده شده در روند پژوهش نیز مربوط به دادگان نوفه باد است که توسط دانشگاه فنی آخن^۲ ارائه گردیده است و از این دادگان جهت ارزیابی الگوریتم‌های کاهش نوفه استفاده می‌شود. نرخ نمونه‌برداری علامت‌های صوتی استفاده شده در روند آزمایش ۱۶۰۰۰ هرتز می‌باشد. برای انجام آزمون، ۶۱ علامت گفتار در سه وضعیت نوفه‌ای مختلف در نظر گرفته شده و در نهایت از نتایج به دست آمده در هر وضعیت میانگین گرفته شده است. در هر کدام از این حالت‌ها علامت نوفه‌ای دارای کیفیت متفاوتی می‌باشد. برای بررسی کیفیت علامت در حالت‌های مختلف معیار اس‌ان‌آر^۳ در نظر گرفته شده است. در این پژوهش، علامت گفتار همراه با نوفه باد با مقادیر اس‌ان‌آر، -۵- دسی‌بل (حالت الف)، ۰ دسی‌بل (حالت ب) و ۵ دسی‌بل (حالت ج) به روش پیشنهادی داده شده و خروجی روش در شرایط مختلف ارزیابی می‌شود. عملکرد روش پیشنهادی در غالب دو معیار اس‌ان‌آر و پی‌ای‌اس‌کیو^۴ در جداول ۲ و ۳ قابل مشاهده است. هم‌چنین در شکل ۳ اسپکتروگرام یک علامت در حالت‌های بدون نوفه، نوفه‌ای

^۱ TIMIT Acoustic-Phonetic Continuous Speech Corpus

^۲ Rheinisch-Westfälische Technische Hochschule Aachen

^۳ SNR

^۴ PESQ

جدول ۵ مقایسه‌ی نتایج افزایش اس‌ان‌آر.

روش	میزان افزایش معیار اس‌ان‌آر در		
	SNR=۵	SNR=۰	SNR=-۵
ام‌ام‌ئی‌اس	۱٫۶۷	۵٫۱	۷٫۳۴
لگاریتم ام‌ام‌اس‌ئی	۱٫۵۶	۵٫۵۷	۸٫۸
روش پیشنهادی	۵٫۱۵	۷٫۲۱	۹٫۸۳

با توجه به ویژگی‌های مختلف نوفه باد، از جمله ناپایداری زیاد آن، کاهش این نوع نوفه، نیازمند استفاده از روش‌های اختصاصی می‌باشد. در نتیجه همان‌طور که مشاهده می‌شود؛ روش پیشنهادی کاهش نوفه باد نسبت به روش‌های به‌سازی رایج، به دلیل در نظر گرفتن ناپایداری زیاد نوفه باد و کاهش آن، عملکرد بهتری را از خود نشان می‌دهد. همان‌طور که از نتایج به‌دست آمده مشخص است، روش مذکور عملکرد مناسبی در کاهش نوفه باد در علامت‌های گفتار را دارد. روش ارائه شده قادر است کیفیت علامت گفتار را از نظر معیار اس‌ان‌آر بیش از ۱۰ دسی‌بل بهبود بخشد و هم‌چنین از نظر کیفیت ادراکی این روش قادر است معیار پی‌ای‌اس‌کیو را حدود ۰٫۷ افزایش دهد. هم‌چنین این روش عملکرد مناسبی در کاهش میزان نوفه باد در علامت‌هایی با اس‌ان‌آر منفی (توان نوفه بیشتر از علامت) از خود نشان می‌دهد. با توجه به زمان محاسبات، می‌توان نتیجه گرفت که روش مذکور در علامت‌های ضبط شده کاربرد دارد و استفاده از این روش برای کاربردهایی از قبیل سمعک و یا پخش زنده مناسب نمی‌باشد.

۵. تقدیر و تشکر

این پژوهش با پشتیبانی دانشکده فنی و مهندسی رسانه دانشگاه صدا و سیما انجام شده است.

۶. تضاد منافع

هیچ‌گونه تضاد منافع در این مطالعه وجود ندارد.

فهرست منابع

- [1] H. Bai, F. Ge, Y. Yan, "DNN-based speech enhancement using soft audible noise masking

با اس‌ان‌آر برابر ۵- دسی‌بل، پس از الگوریتم اس‌اس‌ای و سپس گذر از صافی بول قابل مشاهده می‌باشد. با توجه به این که یکی از مرسوم‌ترین روش‌های حوزه‌ی به‌سازی گفتار، روش‌های بر پایه‌ی تخمین‌زن‌های میانگین حداقل مربعات ام‌ام‌اس‌ئی^۱ می‌باشد؛ برای مقایسه‌ی روش پیشنهادی با سایر روش‌ها، دو روش تخمین‌زن طیف ام‌ام‌اس‌ئی^۲ و لگاریتم ام‌ام‌اس‌ئی^۳ مورد بررسی قرار گرفته‌اند. نظر به کاربرد وسیع این روش‌ها و نتایج مناسب آن‌ها در حوزه‌ی به‌سازی گفتار، مقایسه‌ی این روش‌ها با روش پیشنهادی، محک مناسبی برای روش پیشنهادی می‌باشد. به صورت مشابه ۶۱ علامت گفتار تحت شرایط نوفه‌ای مختلف با استفاده از سه روش مذکور مورد به‌سازی قرار گرفتند و میانگین نتایج به‌دست آمده در جداول ۴ و ۵ نشان داده شده است.

۴. بحث و نتیجه‌گیری

در این پژوهش یک روش براساس الگوریتم‌های اس‌اس‌ای و کی-مینز و صافی (فیلتر) بول ارائه شد. یکی از مهم‌ترین تغییراتی که تأثیر فراوانی در اثربخشی روش پیشنهادی دارد؛ دی‌سی تی گرفتن از علامت گفتار است. هنگامی که از مقادیر علامت گفتار به‌صورت مستقیم در روش پیشنهادی استفاده شود؛ نتایج مناسبی در خروجی مشاهده نمی‌شود. به علاوه، از مقایسه‌ی نتایج روش پیشنهادی و نتایج روش‌های تخمین‌زن طیف ام‌ام‌اس‌ئی و لگاریتم ام‌ام‌اس‌ئی، واضح است که روش پیشنهادی نسبت به روش‌های دیگر، عملکرد بهتری را از خود نشان می‌دهد.

جدول ۴ مقایسه‌ی نتایج افزایش پی‌ای‌اس‌کیو.

روش	میزان افزایش معیار پی‌ای‌اس‌کیو در		
	SNR=۵	SNR=۰	SNR=-۵
ام‌ام‌ئی‌اس	۰٫۱۸	۰٫۲۹	۰٫۳۸
لگاریتم ام‌ام‌اس‌ئی	۰٫۱۷	۰٫۳۱	۰٫۴
روش پیشنهادی	۰٫۵۳	۰٫۶۴	۰٫۶۶

^۱ MMSE; Minimum Mean Square Estimators

^۲ MMES

^۳ logMMSE

- for wind noise reduction,” China Communications, vol. 15, no. 9, pp. 235-243, 2018.
- [2] K.T. Walker, M.A. Hedlin, “A review of wind-noise reduction methodologies,” Infrasound monitoring for atmospheric studies, pp. 141-182, 2010.
- [3] S. Zhao, M. Dabin, E. Cheng, X. Qiu, I. Burnett, J.C. Liu, “On the wind noise reduction mechanism of porous microphone windscreens,” The Journal of the Acoustical Society of America, vol. 142, no. 4, pp. 2454-2463, 2017.
- [4] S. Zhao, M. Dabin, E. Cheng, X. Qiu, I. Burnett, J.C. Liu, “Mitigating wind noise with a spherical microphone array,” The Journal of the Acoustical Society of America, vol. 144, no. 6, pp. 3211-3220, 2018.
- [5] C.M. Nelke, “Wind Noise Reduction: Signal Processing Concepts,” RWTH Aachen University, 2016.
- [6] G.W. Lee, K.M. Jeon, H.K. Kim, “U-Net-based single-channel wind noise reduction in outdoor environments,” 2020 IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2020.
- [7] D. Mirabilii, E.A. Habets, “Multi-channel wind noise reduction using the Corcos Model,” ICASSP, 2019.
- [8] A. Herzog, D. Mirabilii, E.A.P. Habets, “Direction preserving wind noise reduction of b-format signals,” 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 431-435, 2021.
- [9] S. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Transactions on acoustics, speech, and signal processing, vol. 27, no. 2, pp. 113-120, 1979.
- [10] O. Eldwaik, F. Li, “Microphone wind noise reduction using singular spectrum analysis techniques,” Proceedings of the Institute of Acoustics, vol. 39, no. 1, 2017.
- [11] H. Hassani, “Singular spectrum analysis: Methodology and comparison,” Journal of Data Science, vol. 5, 2007.

(Research Article)

Wind noise reduction in recorded speech using Singular Spectrum Analysis (SSA) dataset and machine learning method

M. Shafieian*, F. Ghiasvand

Faculty of Media Engineering, IRIB University

Received: 2021/09/26, Accepted: 2022/03/17

Abstract

In general, recording audio in different environments deals with different challenges. Wind noise in outdoors recording often leads to critical degradation to the speech signal. Because of wide band and nonstationary nature of wind noise, it is very difficult to remove it. Singular Spectrum Analysis (SSA) is a powerful method for time series analysis that is used in applications such as noise reduction, time series prediction and so on. The SSA selects and groups the principal components of the Eigen-space decomposed signals, based on their contributions; then the reconstructed main components are transferred to time domain. In our research, the SSA method combined with K-MEANS algorithm are used to reduce wind noise in speech signals. SNR and PESQ, are used to evaluate the proposed method's results. Speech signals from TIMIT dataset and wind noise from Rheinisch-Westfälische Technische Hochschule Aachen have been utilized in this research. The results show great increase in the quality of speech signal, in terms of perceptual quality and signal to noise ratio. The proposed algorithm shows the good performance compared with other reference algorithms.

Keywords: Singular Spectrum Analysis, Wind noise reduction, K-MEANS algorithm, Perceptual quality, Speech signal.

pp. 111-119 (In Persian)

* Corresponding author E-mail: masoume@gmail.com