

شکل دهی و فقی و هوشمند پرتو در آرایه های میکروفونی Ad-hoc با استفاده از خوشبندی و رتبه بندی میکروفون ها

سید حمید یزدانی و حمیدرضا ابوطالبی

دانشکده مهندسی برق و کامپیوتر، دانشگاه یزد، یزد، ایران

چکیده

با توجه به وجود عوامل بسیار در تخریب سیگنال گفتار، بهسازی این سیگنال اهمیت زیادی دارد. شکل دهی پرتو یکی از روش های مطرح برای بهسازی گفتار است که با استفاده از آرایه های میکروفونی صورت می پذیرد. با توجه به محدودیت های موجود در نحوه چیشنش میکروفون ها، پنجره ای به سوی بحث آرایه های میکروفونی با چیدمان نامنظم (Ad-hoc) گشوده شده است. با فرض عدم شناخت نسبت به مکان و چیشنش میکروفون ها و همچنین پراکنده بودن میکروفون ها در محیط، در این مقاله، نظریه خوشبندی میکروفون ها براساس انرژی دریافتی از منابع موجود در محیط مورد توجه قرار گرفته و روش جدیدی برای خوشبندی پیشنهاد شده است. در ادامه، برای هر خوشبندی میکروفونی، دو نوع رتبه پیشنهاد شده که براساس میزان فاصله به منبع نوفه و گوینده می باشند. همچنین براساس نتیجه خوشبندی و رتبه بندی آنها، ساختار جدیدی برای شکل دهنده پرتو GSC (Generalized Sidelobe Canceller) ارائه شده است. برای حالتی که نوفه پخششده نباشد، براساس انرژی منابع موجود در محیط روشی برای خوشبندی پیشنهاد شده که می تواند برای سایر میدان های نوفه ای نیز به کار گرفته شود. با پیاده سازی و ارزیابی روش پیشنهادی دیده می شود که در بعضی از حالت ها، استفاده از یک خوشه در برابر به کار گیری تمام میکروفون ها نتیجه بهتری را بدست می دهد که این خود حاکی از دستاوردهای دیگر است. این دستاوردها، کاهش تعداد میکروفون های مورد استفاده است که به نوبه خود، کاهش پیچیدگی و حجم محاسبات (در عین افزایش کیفیت خروجی) را به دنبال دارد.

وازگان کلیدی: بهسازی گفتار، شکل دهی پرتو، آرایه های میکروفونی Ad-hoc، خوشبندی، شکل دهنده پرتو GSC

۱- مقدمه

میکروفون ها با کیفیت پایینی ضبط می شود. کاهش کیفیت سیگنال دریافتی میکروفون ها، باعث افت عملکرد سامانه های پردازش گفتار می شود.

از دیربارز شیوه های متعددی برای بهسازی گفتار ارائه شده که یکی از روش های چند کاناله مورد استفاده به این منظور، شکل دهی پرتو^۱ بوده است. در شکل دهی پرتو با اعمال ضرایب (یا فیلترهایی) بر روی سیگنال های دریافتی از میکروفون های مختلف و ترکیب آنها با هم، به تعییری سیگنال دریافتی از راستای سیگنال گفتار هدف تقویت و سیگنال های تداخل و نوفه دریافتی از سایر راستاها تضعیف می شود (ون وین و بوکلی، ۱۹۸۸).

با رشد و گسترش فناوری و پیشرفت در حوزه پردازش گفتار و زمینه های دیگری همچون پردازش تصویر، اهداف جدیدی در حوزه ارتباطات شکل گرفته است. در جند دهه اخیر، ارتباط بین انسان و رایانه یکی از مهم ترین زمینه های پژوهشی در حوزه پردازش گفتار بوده است. با در نظر گرفتن گفتار به عنوان مهم ترین رابط کاربری با رایانه، دریافت گفتار با کیفیتی خوب بیش از پیش اهمیت می باید. در بسیاری از شرایط محیطی طبیعی، سیگنال های صوتی توسط میکروفون هایی که از گوینده فاصله ب نسبه زیادی دارند، ضبط می شوند. بدلیل فاصله زیاد گوینده و میکروفون ها و وجود عوامل محرکی همچون نوفه، پژواک، حرکت چندین گوینده هم زمان، تداخل و ... سیگنال گفتار مطلوب توسط

^۱ Beamforming

شکل دهی پرتو در آرایه های میکروفونی Ad-hoc (و در
حالت جامع تر، در شبکه های حسگر صوتی Ad-hoc^۱
(برتراند و همکاران، ۲۰۱۰ الف، ب، ج؛ روی و وترلی، ۲۰۰۹)،
با دو رویکرد صورت می پذیرد: در رویکرد نخست، شکل
هندسی آرایه تخمين زده شده و سپس از روش های کلاسیک
شکل دهی پرتو استفاده می شود. در رویکرد دوم، روش های
جدیدی برای شکل دهی پرتو بر پایه ساختار و شرایط خاص
شبکه های میکروفونی Ad-hoc مطرح می شود.

به عنوان یک کار شاخص در حوزه شکل دهی پرتو در
آرایه های Ad-hoc، هیماوان در سال ۲۰۱۰ در رساله دکترای
خویش، روش خوش بندی میکروفون ها را ارائه و به بررسی
کاربرد این روش در بازناسانی گفتار پرداخت (هیماوان،
۲۰۱۰). در این روش، ابتدا میکروفون ها به خوش بندی
تفکیک می شوند. این امر به دلیل آن است که فاصله بین
میکروفونی زیاد، مشکلاتی را در تخمين تأخیر سیگنال های
رسیده ایجاد می کند. همچنان نظریه این است که خوش بندی
انتخاب شود که نسبت سیگنال به نویه بیشتری داشته باشد.
در پژوهش های انجام شده توسط هیماوان، ابتدا دو روش برای
خوش بندی میکروفون ها معرفی شده است؛ سپس خوش بندی
رتیبه بندی شده و در انتها با دو روش (۱- نزدیک ترین خوش،
و ۲- ترکیب وزن دار از خوش ها)، شکل دهی پرتو صورت
گرفته است (هیماوان و همکاران، ۲۰۱۱).

در سال ۲۰۱۲، برتراند شکل دهنده پرتو LCMV^۲ را
در حالتی که گره ها (مجموعه چندین میکروفون نزدیک به
هم با یک پردازنده مرکزی) در محیط پخش شده اند، بررسی
کرد (برتراند و منون، ۲۰۱۲).

کارهایی نظیر (سرینیوسان، ۲۰۱۱) و (یو و هنسن،
۲۰۱۰) را نیز که در آن ها فرض بر مشخص نبودن مکان
میکروفون هاست می توان کارهایی در حوزه آرایه های
میکروفونی می کرد، در نظر گرفت. در (سرینیوسان،
۲۰۱۱) یک میکروفون در نزدیکی منبع نویه با مکان نامشخص قرار
داده شده و با ارسال طیف توان نویه به میکروفون دیگر عمل
کاهش نویه صورت گرفته است. در (یو و هنسن، ۲۰۱۰) نیز
نویه موسیقی از سیگنال نویه ای حذف می شود. نظریه به کار
گرفته شده در (یو و هنسن، ۲۰۱۰) بر این اساس است که
میزان نوسان طیف گفتار و موسیقی متفاوت است. این معیار
وابستگی به مکان میکروفون ها نداشته و قابل پیاده سازی در
آرایه های میکروفونی Ad-hoc است.

اگرچه در سامانه های کلاسیک آرایه میکروفونی، یک
سری میکروفون در چیدمانی منظم (به عنوان مثال خطی،
دایره ای یا نظایر آن) استفاده می شده، در چند سال اخیر،
بحث به کارگیری مجموعه ای از میکروفون ها در یک چیدمان
نامنظم مورد توجه قرار گرفته است (هیماوان و همکاران،
۲۰۱۱). این آرایه های میکروفونی با چیدمان نامنظم (یا
به اصطلاح رایج، آرایه های میکروفونی Ad-hoc)، به طور
معمول از میکروفون های موجود در سامانه های رایانه ای و
ارتیاطی قبل حمل (لپتاپ ها، تلفن های همراه و ...) تشکیل
می شود.

در آرایه های میکروفونی Ad-hoc، فرض می شود که
مکان میکروفون ها برای شکل دهنده پرتو مشخص نبوده و
شکل دهی پرتو باید به صورت کور انجام شود؛ این بدان معنی
است که الگوریتم های به سازی، باید تنها با استفاده از سیگنال
میکروفون ها (و نه اطلاعات هندسی محل آنها) پیاده شود.
با توجه به این که در اکثر روش های کلاسیک
شکل دهی پرتو (با چیدمان میکروفونی منظم و مشخص)،
فرض بر نزدیک بودن میکروفون ها به هم دیگر بوده است،
ماهیت آرایه های میکروفونی کلاسیک و روش های قابل
پیاده سازی در این آرایه ها با آرایه های Ad-hoc متفاوت است.
در آرایه های میکروفونی کلاسیک، چینش میکروفون ها به طور
دقیقی چنان صورتی می پذیرد که به سازی مؤثری برای آن
شرایط محیطی خاص حاصل شود. در چینش های Ad-hoc،
این موضوع کلیت نداشته و میکروفون ها به طور دلخواه و
تصادفی در محیط پخش می شوند.

در حالی که در آرایه های کلاسیک، با افزایش تعداد
میکروفون ها عملکرد، بهتر می شود (ون وین و بوکلی،
۱۹۸۸)، در آرایه های Ad-hoc این موضوع نیز عمومیت ندارد
(هیماوان و همکاران، ۲۰۱۱)؛ زیرا در آرایه های کلاسیک
میکروفون ها از لحاظ فیزیکی و مکانی و نوعی بسیار نزدیک
به یکدیگر بوده و شرایط صوتی مشابهی دارند. این باعث
می شود در تخمين پارامترها، داده های موجود بیشتر بوده و
خطای کمتری در تخمين رخ دهد. در آرایه های میکروفونی
Ad-hoc، ممکن است میکروفون ها فاصله زیادی از هم داشته
باشند؛ همچنان ممکن است، نوع میکروفون ها نیز متفاوت
باشد. در هر دو صورت، سیگنال دریافتی میکروفون ها شرایط
به طور کامل متفاوتی داشته و نمی توان انتظار داشت با زیاد
شدن تعداد میکروفون ها عملکرد سامانه لزوماً بهبود یابد
(هیماوان و همکاران، ۲۰۱۱).

¹ Ad-hoc acoustic sensor networks
² Linearly Constrained Minimum Variance





بخش چهارم نیز نتایج شبیه سازی و ارزیابی روش پیشنهادی آمده، و در بخش پنجم نتایج کلی و جمع بندی این پژوهش ذکر شده است.

۲- نوافه جهتی و خوشبندی میکروفون ها

یکی از شرایطی که ممکن است در یک محیط برقرار باشد، حضور نوافه مکانی است. نوافه مکانی (نوافه جهتی) همانند یک منبع صوت عمل می نماید که در محیط در جهت خاصی منتشر می شود. در این صورت، در سیگنال دریافتی در میکروفون ها، نوافه ای با همبستگی زیاد وجود خواهد داشت. در حالت مشابه می توان به مسأله تداخل اشاره کرد. تداخل نیز یکی از عوامل مخرب سیگنال گفتار است که از دستگاه های مانند ضبط صوت، بلندگو و ... انتشار می یابد. همچنین وجود گوینده های دیگری (علاوه بر گوینده اصلی) بهنوعی موجب ایجاد تداخل می شود که به این گوینده ها، گوینده رقیب نیز گفته می شود.

با توجه به این تعاریف، می توان گفت تداخل از جنس گفتار بوده و برخورد با این عامل از بقیه عوامل مخرب نوعاً سخت تر است. توجیه این که تداخل و نوافه جهتی در اینجا تحت یک عنوان مورد بررسی قرار می گیرد، این است که هر دو را می توان مانند یک منبع مکانی تولید صوت در نظر گرفت. از این دیدگاه، می توان برای بحث خوشبندی میکروفون ها استفاده کرد.

نکته اساسی برای خوشبندی میکروفون ها در شرایط نوافه ای مورد بحث حاضر، توجه به وجود گندین منبع در محیط است. در شرایطی که در این فصل در نظر گرفته شده، حداقل یک منبع نوافه مکانی و یک منبع گفتار اصلی در محیط وجود دارد. با فرض اینکه محیط دارای انعکاس ناچیزی است، می توان از انرژی بازتاب ها صرف نظر کرد.

ورودی هر میکروفون یک نسخه فیلتر شده از سیگنال منبع جمع شده با نوافه است:

$$x_i(t) = h_i * s_i(t) + v_i(t) \quad (1)$$

در این رابطه، h_i پاسخ ضربه محیط بین منبع و میکروفون i -ام، $*$ نماد کالولوشن، $s_i(t)$ سیگنال گفتار تمیز و (t) نوافه دریافتی در محل میکروفون i -ام است. در اینجا نوافه و سیگنال از لحاظ آماری مستقل فرض شده اند. در حوزه فرکانس، مدل سیگنال به صورت زیر نوشته می شود:

در سال ۲۰۱۳ نیز مارکوویچ گولان سناریوی چندین گره را در نظر گرفت و در این سناریو با تعریف یک تبدیل، شکل دهنده پرتو GSC^۱ متمرکز را به چندین GSC در هر گره تفکیک کرد (مارکوویچ گولان و همکاران، ۲۰۱۳). در این روش، فرض شده است که چندین گوینده در محیط وجود دارد. برای هر گوینده یک گره اختصاصی تعریف شده و سپس با درنظر گرفتن معیارهایی، سیگنال ها در شبکه به اشتراک گذاشته می شوند. در (مارکوویچ گولان و همکاران، ۲۰۱۳) نشان داده شده که روش پیشنهادی بدون نیاز به ارسال حجم عظیمی از سیگنال های میکروفونی به پردازنده مرکزی، بسیار نزدیک به روش GSC متمرکز عمل می کند.

در اکثر موارد بالا، فرض این بوده که گره ها مشخص بوده و هدف این است که در هر گره، بهترین سیگنالی که می توان از ترکیب مجموعه سیگنال های همه گره ها به دست آورده، در اختیار داشته باشیم، به عبارت دیگر، در موارد بالا هدف نهایی، یک خروجی بهینه نیست؛ بلکه هدف به دست آمدن خروجی مناسب در هر یک از گره هاست.

نکته قابل ذکر دیگر آن که در مقالات (برتراند و همکاران، ۲۰۱۲؛ یو و هنسن، ۲۰۱۰؛ سرینیوسان، ۲۰۱۱؛ مارکوویچ گولان و همکاران، ۲۰۱۳)، به اصطلاح آرایه میکروفونی Ad-hoc اشاره ای نشده و موضوع در قالب یک شبکه حسگر صوتی مورد بررسی قرار گرفته است.

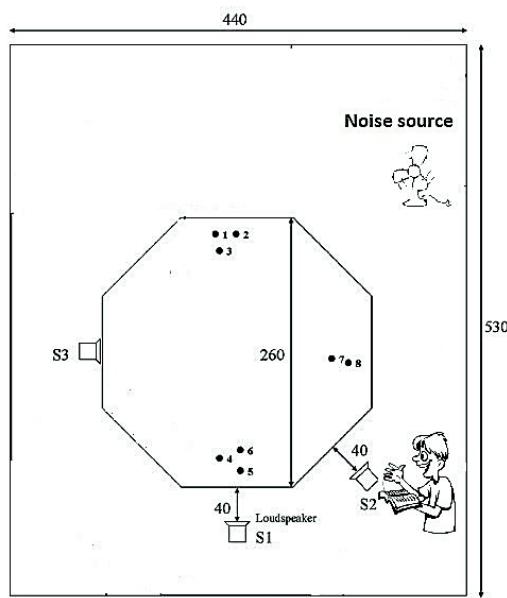
در این مقاله، هدف، بهره گیری از ایده خوشبندی میکروفون ها و همچنین استفاده هوشمند از سیگنال های دریافتی به منظور بهبود عملکرد شکل دهنده پرتو GSC است. در این پژوهش، با اعمال معیارهایی ساده و در عین حال قوی، همچون انرژی در فریم های اولیه شروع گفتار، روشی را برای خوشبندی میکروفون ها پیشنهاد کرده و با به کار گیری هوشمند سیگنال های دریافتی در شکل دهنده پرتو GSC، عملکرد این شکل دهنده را در آرایه های میکروفونی Ad-hoc بهبود داده ایم.

ساختم مطالب این مقاله به شرح زیر است. در بخش دوم، فرض میدان نوافه ای پخش شده از مسأله خوشبندی کnar گذاشته شده که به معنای عدم امکان استفاده از روش خوشبندی ارائه شده در (هیماون و همکاران، ۲۰۱۱) است. در این بخش، روش جدیدی برای خوشبندی در شرایط حضور نوافه جهتی پیشنهاد شده است. در بخش سوم، با توجه به شرایط نوافه ای جدید، به طور هدفمند از خوشبندی برای بهبود ساختار شکل دهنده پرتو GSC استفاده شده است. در

^۱ Generalized Side lobe Canceller

باشند، شباهت انرژی آنها برای منابع مختلف بیشتر است. بنابراین می‌توان از این نظریه برای بهدست آوردن معیار جدیدی برای نزدیکی میکروفون‌ها استفاده کرد. در صورتی که تداخل و یا چندین گوینده در محیط حضور داشته باشند، این نظریه به شکل قوی‌تری قابل اعمال خواهد بود.

برای بیان بهتر نظریه کار، سناریوی رسم شده در شکل (۲) را که سناریوی مورد استفاده برای ارزیابی نتایج این مقاله نیز می‌باشد؛ در نظر بگیرید. در این سناریو، ابعاد اتاق نیز می‌باشد؛ در حسب متر) $4.4 \times 5.3 \times 2.7$ (بر حسب متر) و فرکانس نمونه‌برداری نیز 16000 هرتز است. هشت میکروفون مورد استفاده از یک نوع بوده و تمام جهته^۱ سیگنال را دریافت می‌کنند. نویفه مکانی یک نویفه سفید با میانگین صفر است. پاسخ ضربه اتاق با روش تصویر^۲ (آلن و برکلی، ۱۹۷۹) شبیه‌سازی و بهدست آمده است.



(شکل-۲): سناریوی مورد استفاده در این پژوهش

برای نمایش بهتر مبنای ایده مورد بحث، تغییرات میانگین انرژی میکروفون‌های مختلف در فریم‌های متوالی، در شکل (۳) رسم شده است. در این شکل مشاهده می‌شود میکروفون‌هایی که از لحاظ مکانی نزدیک به یکدیگر هستند، هم در حالت منبع نویفه تنها و هم در حالت شروع گفتار، سطح انرژی مشابهی دارند. البته این موضوع را به طور دقیقتر می‌توان در فریم‌های ابتدایی گفتار مشاهده کرد؛ دلیل این امر این است که با گذشت زمان اندکی (حدود ۵۰ تا ۱۰۰

^۱ Omni-directional

^۲ Image method

$$X_i(f) = H_i(f)S_i(f) + V_i(f) \quad (2)$$

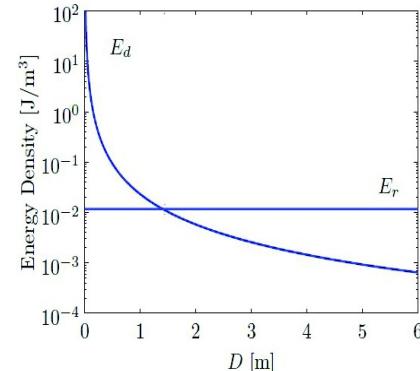
همان‌طور که در روابط بالا مشخص است، سیگنال رسیده به میکروفون‌ها هم بر اثر نویفه و هم بر اثر پاسخ ضربه محیط تغییر می‌کند. هدف تمامی الگوریتم‌های بهسازی گفتار، کم کردن اثر این عوامل تغییردهنده (مخرب) بر روی سیگنال تمیز است. انرژی سیگنال منتشرشده از هر منبع در مسیر مستقیم طبق رابطه زیر تغییر می‌کند (هبتز، ۲۰۰۷) :

$$E_d = \frac{QW_s}{4\pi c D^2} \quad (3)$$

که در این رابطه W_s توان منبع به وات، Q یک ثابت برای منبع (وابسته به جهت الگوی تشعشعی موج)، E_d انرژی مؤلفه مسیر مستقیم، c سرعت صوت و D فاصله منبع تا میکروفون است. همان‌طور که از این رابطه مشخص است، با زیادشدن فاصله منبع تا میکروفون انرژی کاوش می‌یابد؛ همچنین، در مورد انرژی بازتاب‌ها (مؤلفه‌های دریافتی از غیر مسیر مستقیم) رابطه زیر برقرار است (هبتز، ۲۰۰۷) :

$$E_r = \frac{4W_s}{cR} \quad (4)$$

که در رابطه بالا، R ثابت اتاق (وابسته به شرایط فیزیکی اتاق) است. همان‌طور که از این رابطه مشخص است، انرژی بازتاب‌ها به فاصله بستگی نداشته و مقدار ثابتی است. در شکل (۱) چگالی انرژی مسیر مستقیم و بازتاب رسم شده است.



(شکل-۱): چگالی انرژی مسیر مستقیم و بازتاب‌ها بر حسب فاصله (گرفته شده از (هبتز، ۲۰۰۷))

با فرض این که در محیط مورد بحث، حداقل دو منبع صوتی وجود داشته باشد، می‌توان با ترکیب نتایج مربوط به میزان انرژی دریافتی از هریک از منابع، به خوشبندی به‌نسبه مناسبی رسید. نظریه اصلی این است که میکروفون‌هایی که به هم نزدیک باشند، انرژی مشابهی از هر منبع دریافت می‌کنند. هرچه میکروفون‌ها به هم نزدیک‌تر



میکروفون ها به هم نزدیکتر باشند، میزان مشابهت انرژی دریافتی میکروفون ها از هر منبع افزایش می یابد. برای تعیین میزان نزدیکی خوشه ها به منبع نوشه و یا منبع گفتار از اختلاف میانگین انرژی های هر خوشه با یکدیگر استفاده می نماییم. البته لازم به ذکر است که برای سنجش نزدیکی منبع گفتار به خوشه ها، می توان از ویژگی هایی دیگری همچون زمان شروع سیگنال، TDOA و ... نیز استفاده کرد. در جدول (۱)، الگوریتم خوشبندی پیشنهادی با جزییات پردازشی بیان شده است.

در این پژوهش، بعد از خوشبندی میکروفون ها، هر خوشه از دو روند مورد رتبه بندی قرار می گیرد. در رتبه بندی نخست، میزان نزدیکی خوشه ها به منبع گفتار با توجه به انرژی قاب های گفتاری رتبه دهی می شود. در مقاله حاضر، به این رتبه، رتبه گفتاری می گوییم. برای قوی تر شدن این رتبه بندی، علاوه بر انرژی، از اختلاف در لحظه آغاز سیگنال نیز به عنوان معیار دوم استفاده می کنیم. برای این کار، مقدار انرژی هر قاب نسبت به مقدار میانگین انرژی از نخست تا آن قاب با یک حد آستانه T_1 (در این پژوهش برابر با 5) مقایسه می شود؛ زمانی که این نسبت از T_1 بیشتر باشد، از دو قاب قبل (همانند مرحله (۲-۳) الگوریتم پیشنهادی در جدول (۱)) با هم پوشانی بسیار زیاد (بالاتر از 95 درصد) دوباره قاب بندی و با مقایسه با سطح آستانه T_2 (در این پژوهش برابر با $3/5$) لحظه شروع سیگنال تخمین زده می شود. تفاوت این اقدام با آنچه در خوشبندی صورت می گیرد، در این است که در فرایند رتبه بندی، معیار مورد استفاده، میانگین شماره قاب ها کمتر باشد، به عنوان خوشة نزدیکتر به منبع تلقی شده و رتبه بهتری می گیرد. علت میانگین گیری نیز کم کردن اثر پخش شدگی میکروفون های یک خوشه در رتبه آن خوشه می باشد. لازم به ذکر است در این مقاله T_1 و T_2 به صورت دستی تنظیم شده که می توان در پژوهش های آینده، تعیین وقتي آنها را مورد بررسی قرار داد.

در رتبه بندی دوم، میزان نزدیکی خوشه ها به منبع نوشه با توجه به انرژی در قاب های نوشه ای بررسی و خوشه ها رتبه دهی می شوند. در این مقاله به این رتبه، رتبه نوشه ای می گوییم. بنابراین پس از طی این فرایند، به هر خوشه دو عدد که نشان دهنده رتبه گفتاری و رتبه نوشه ای آنهاست، اختصاص می یابد. در ادامه، خواهیم دید که این دو عدد

میلی ثانیه، بازتاب ها نیز به میکروفون ها رسیده و سطح انرژی را به طور مشخص نمی توان وابسته به فاصله تا گوینده دانست. لازم به ذکر است در رسم این شکل (۳)، گوینده در مکان نخست (S1) قرار داشته است.

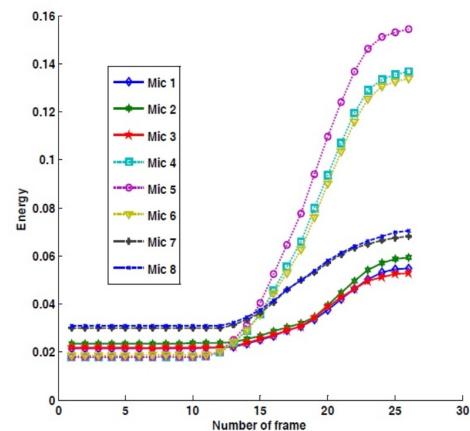
با توجه به شکل (۳)، پایه الگوریتم براساس قسمت سکوت سیگنال (زمانی که فقط منبع نوشه فعال است) می باشد. با توجه به مباحثه بالا، الگوریتم زیر برای خوشبندی میکروفون ها پیشنهاد می شود:

۱- اندازه گیری میانگین انرژی نوشه در هر میکروفون در

قسمت های سکوت سیگنال

۲- خوشبندی اولیه میکروفون ها با استفاده از الگوریتم k-means (مک کوین، ۱۹۶۷) براساس

میانگین های محاسبه شده در مرحله قبل



(شکل - ۳): تغییرات میانگین انرژی در تمام میکروفون ها در لحظه شروع گفتار گوینده نخست

۳- اندازه گیری میانگین انرژی سیگنال در فرکانس های پایین (با فرض طول تبدیل فوریه ۲۵۶ نقطه ای و فرکانس نمونه برداری 16000 هرتز، تا بین فرکانسی 25 ، معادل با فرکانس حدود دو کیلوهرتز)

۴- خوشبندی دوباره میکروفون ها در هر خوشه نخستین و تشکیل خوشه های نهایی در توضیح مرحله سوم الگوریتم بالا، ذکر این نکته لازم است که سیگنال گفتار، سیگنالی پایین گذر بوده و برای کاهش حجم محاسبات، می توان خوشبندی را براساس اطلاعات باندهای پایین سیگنال به انجام رساند. در پژوهش حاضر، محاسبه میانگین انرژی در باند فرکانسی صفر تا دو کیلوهرتز صورت گرفته است. همان طور که گفته شد، هر چه

سال ۱۳۹۴ شماره ۳ پیاپی ۲۵

با توجه به توضیح بالا می‌توان دید که یکی از نقش‌های اصلی در این ساختار بر عهده فیلتر وفقی است. به عنوان نکته‌ای شناخته شده در رابطه با فیلتر وفقی، می‌توان گفت هرچه قدر سیگنال ورودی فیلتر به نووفه مرجع نزدیک‌تر (و با آن همبسته‌تر) باشد، عملکرد بهتری در کاهش نووفه دیده می‌شود. همان‌طور که در ادامه توضیح خواهیم داد، این موضوع را می‌توان به عنوان مبنای پیشنهاد ساختار جدیدی برای GSC به کار گرفت.

(جدول-۱): الگوریتم پیشنهادی برای خوشبندی میکروفون‌ها

۱- فریم‌بندی سیگنال گفتار با همپوشانی پنجاه درصد و در نظر گرفتن قسمت سکوت سیگنال (نووفه خالص) برای خوشبندی اولیه ($T_{silence}$)
۲- اندازه‌گیری میانگین انرژی در $T_{silence}$ برای فریمهای متوالی در تمام میکروفون‌ها ($E_{ave,i}$)
۳- تخمین لحظات نخستین شروع گفتار در هر میکروفون. این کار در دو مرحله انجام می‌شود:
۳-۱- تخمین نخستین شماره قاب شروع سیگنال در میکروفون با مقایسه انرژی میانگین قاب‌های متوالی در بین‌های فرکانسی دارای حضور غالب گفتار با $E_{ave,i}$ (با فرض طول قاب ۲۵۶ و فرکانس نمونه‌برداری ۱۶ کیلوهرتز، محدوده فرکانسی دارای گفتار غالب، به طور معمول بین‌های ۰ تا ۲۵۶ شامل می‌شود). (در این پژوهش، مقایسه زیر با حد آستانه ۵ T_1 انجام شده است).

$$\frac{E_{ave,i,lowbins,x_i}}{E_{ave,i}} \geq T_1 \Rightarrow \\ : \text{estimation of start frame in microphone } i$$

۴- افزایش میزان همپوشانی قاب‌ها به ۹۵ درصد از دو قاب قبل از شماره به دست آمده برای میکروفون با تأخیر کمتر (حاصل از مرحله ۳-۱) یعنی فریم با شماره $(-2 - x_{i,\min})$ و قاب‌بندی مجدد و تکرار روند مشابه مرحله ۳-۱ با سطح آستانه کمتر (T_2) به منظور بالابردن دقت و ذخیره شماره قاب (N_i). (این آستانه در اینجا برابر با $T_2 = 3.5$ در نظر گرفته شده است)

$$\frac{E_{ave,i,x_i,lowbins,newframes}}{E_{ave,i}} \geq T_2 \Rightarrow N_i = x_i$$

۵- اعمال الگوریتم دسته‌بندی K-means روی مقادیر $E_{ave,i}$ و تعیین خوشبندی‌های نخستین و سپس استفاده مجدد از الگوریتم K-means در هر خوشبندی اولیه با استفاده از مقادیر شماره قاب شروع سیگنال در هر میکروفون (N_i).

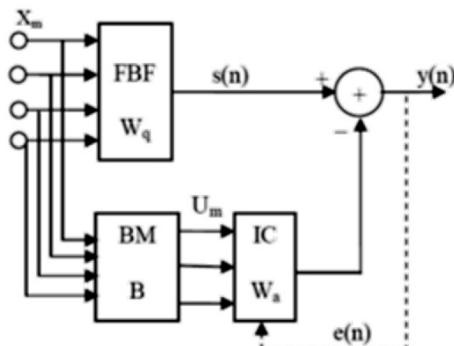
چگونه در اصلاح ساختار شکل دهنده پرتو GSC مورد استفاده قرار می‌گیرد.

۳- معرفی ساختار GSC اصلاح شده برای شکل دهنده پرتو

شکل دهنده پرتو وفقی GSC دارای یک ساختار دو شاخه‌ای است. یکی از کاربردهای مهم این شکل دهنده پرتو برای حذف نووفه (یا هر عامل مخرب) مکانی و جهتی بوده و در حضور چنین میدان‌های نووفه‌ای، این شکل دهنده پرتو عملکرد مناسبی نشان می‌دهد (گریفیت، ۱۹۸۲). با توجه به اینکه در این پژوهش، عامل مخرب سیگنال گفتار، تداخل و یا نووفه جهتی فرض شده، به کارگیری شکل دهنده پرتو GSC امری طبیعی به نظر می‌رسد.

برای توجیه کامل‌تر دلیل استفاده از این شکل دهنده پرتو در بحث حاضر، ابتدا نگاهی عمیق‌تر به این شکل دهنده پرتو داشته و از منظر آرایه میکروفونی Ad-hoc آن را مورد بررسی قرار می‌دهیم.

همان‌طور که شکل (۴) نشان می‌دهد، شکل دهنده پرتو GSC، دارای دو شاخه است. در شاخه بالایی یک شکل دهنده ثابت پرتو (FBF) ^۱ قرار دارد که برای آن در اکثر موارد، از شکل دهنده پرتو تأخیر و جمع ^۲ استفاده می‌شود. در شاخه دوم نیز، با به کارگیری بلوک BM ^۳ سعی می‌شود سیگنال گفتار حذف شده و نمونه‌ای از نووفه محیط ایجاد شود. این نمونه از نووفه محیط، به عنوان نووفه مرجع فرض شده و با ساختار فیلتر وفقی (IC) ^۴، نووفه در خروجی نهایی GSC حداقل می‌شود.



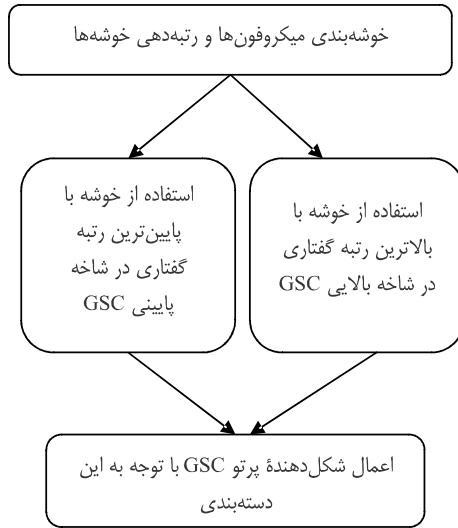
(شکل-۴): ساختار کلی شکل دهنده پرتو GSC

^۱ Fixed BeamFormer

^۲ Delay and Sum Beamformer

^۳ Blocking Matrix

^۴ Interference Canceller



شکل ۵- روند کلی کار شکل دهنده پرتو پیشنهادی

۴- شبیه سازی و ارزیابی روش پیشنهادی

در این بخش، نتایج مربوط به پیاده سازی و ارزیابی روش های پیشنهادی برای خوشبندی، رتبه بندی و شکل دهنده پرتو و فوئی GSC با ساختار جدید پیشنهادی آورده می شود.

ذکر این نکته ضروری است که چون هدف اصلی از ارزیابی ها، بررسی مزیت روش پیشنهادی بوده، در این مقاله نخست از GSC ساده (گریفیت، ۱۹۸۲) استفاده کرده ایم. در عین حال، برای توسعه ارزیابی ها و مقایسه کارایی روش خوشبندی پیشنهادی در ترکیب با نسخه های جدیدتر سامانه GSC، از روش GSC همبسته (Correlative GSC) (بررسی ۲۰۰۹)، نیز بهره گرفته ایم. ایده و ویژگی اصلی این نسخه جدید GSC، کاهش نشتی سیگنال در بلوك حذف کننده موجود در شاخه پایین است. این روش به دلیل ساختار و روش به کار رفته در آن نیازمند اطلاع نخستین از مکان میکروفون ها نیست. برای کارهای بعدی می توان نظریه خوشبندی را در ساختارهای پیچیده تر GSC نیز (به طور مثال، در ساختار ارائه شده در (مارکو ویج گولان و همکاران ۲۰۱۳)، گنات و همکاران، ۲۰۰۱) بررسی کرد.

همان گونه که در قبل هم اشاره شد در شبیه سازی و ارزیابی، سناریوی رسم شده در شکل (۲) مورد نظر بوده است. در این سناریو، بعد اتفاق ۷/۲×۵/۴ (بر حسب متر) فرض شده و هشت میکروفون یکسان به صورت تمام جهتی سیگنال را دریافت می کنند. نویفه مکانی یک نویفه سفید با میانگین صفر است. به عنوان سیگنال منابع گفتار (گویندگان)

با توجه به این که پژوهش حاضر، آرایه های میکروفونی Ad-hoc مورد بررسی و استفاده قرار گرفته و در این آرایه ها، میکروفون ها نوعاً در کل محیط پخش شده اند، بنابراین می توان انتظار داشت که بعضی از میکروفون ها به منبع نویفه و برخی از آنها به گویندۀ اصلی نزدیک تر باشند. طبیعی است که میکروفون های دورتر از گویندۀ اصلی، با توجه به رابطه (۱) سیگنال گفتار را با تضعیف بیشتری دریافت می کنند.

یکی از مشکلاتی که در ساختار GSC اولیه (گریفیت، ۱۹۸۲) وجود دارد، نشت سیگنال گفتار به شاخه پایینی است که به نوبه خود، باعث کاهش SNR^۱ در خروجی نهایی GSC می شود. هر چه نشت سیگنال گفتار به شاخه پایینی کمتر باشد، می توان گفت سیگنال خروجی بلوك حذف کننده گفتار (بلوك BM در شکل (۴))، دارای درصد بیشتری نویفه بوده و در نتیجه نزدیکی بیشتری با نویفه مرجع خواهد داشت. بنابراین در صورتی که برای شاخه پایینی GSC از میکروفون های با فاصله بیشتر نسبت به گویندۀ اصلی (به منظور وجود درصد کمتر از سیگنال گفتار) و در عین حال فاصله کمتر نسبت به منبع نویفه (به منظور وجود درصد بیشتری از نویفه) استفاده شود، عملکرد فیلتر و فوئی می تواند بهبود یابد. همچنین استفاده از میکروفون های با نسبت سیگنال به نویفه بالاتر در شاخه بالایی GSC می تواند عملکرد کلی این شکل دهنده پرتو را بهبود بخشد. این مزیت ها در آرایه های میکروفونی کلاسیک وجود ندارد؛ زیرا در آن آرایه های میکروفونی، اکثر میکروفون ها در فضای کوچکی از محیط قرار گرفته و در نتیجه SNR و رودی آنها به طور تقریبی برابر است.

بر اساس آنچه در بالا شرح داده شد، در این پژوهش، ساختاری اصلاح شده برای شکل دهنده پرتو GSC پیشنهاد می نماییم. نظریه پیشنهادی آن است که به جای استفاده از تمام خوشبندی های میکروفونی در هر دو شاخه بالا و پایین GSC، براساس رتبه خوشبندی ها (رتبه گفتاری و رتبه نویفه ای) در هر شاخه تنها بخشی از خوشبندی ها به کار گرفته شوند. روند کلی به این صورت است که خوشبندی های با رتبه بهتر گفتاری در شاخه بالایی و خوشبندی های با رتبه نویفه ای در شاخه پایینی استفاده شوند. البته به صورت دیگری نیز می توان عمل کرد و آن در نظر گرفتن تنها رتبه گفتاری است، خوشبندی با بهترین رتبه گفتاری در شاخه بالایی، و خوشبندی با بدترین رتبه گفتاری در شاخه پایین استفاده شود. در شکل (۵) روند کلی ساختار پیشنهادی جدید برای GSC رسم شده است.

^۱ Signal to Noise Ratio

و بنابراین، روش خوشبندی پیشنهادی نمی‌تواند بین این دو خوشبندی چندانی قائل شود؛ اما در روش (هیماون و همکاران، ۲۰۱۱) که نوفة محیط پخشندۀ بوده و روش خوشبندی براساس تابع همدوسی این نوع از نوفة بنا نهاده شده، این مسئله وجود ندارد. با این وجود، توجه به این نکته اهمیت زیادی دارد که روش پیشنهادی در محیط نوّفه جهتی پیشنهاد شده و برای بیشتر کردن قدرت این روش و به وجودنیامدن چنین مشکلاتی، به تعداد منبع صوتی بیشتری نیاز است. بنابراین در محیط‌های واقعی که نوعاً چندین گوینده در آنها وجود دارد، احتمال رخداد این مشکل بسیار کم است.

در جدول (۳) نتایج مربوط به رتبه‌بندی خوشبندی هم برای رتبه گفتاری و هم برای رتبه نوّفه‌ای برای سه مکانی منبع گفتار آورده شده است. لازم به ذکر است برای تعیین رتبه نوّفه‌ای از انرژی و برای رتبه گفتاری از میانگین شماره قاب آغاز سیگنال (به دست آمده از مقایسه سطح انرژی با یک آستانه طبق روش گفته شده در بخش‌های قبل) در هر خوشبندی استفاده شده است.

(جدول-۲): نتیجه اعمال روش خوشبندی پیشنهادی برای سه حالت مکانی گوینده در مقایسه با نتیجه روش خوشبندی در (هیماون و همکاران، ۲۰۱۱)

خوشبندی سوم	خوشبندی دوم	خوشبندی اول	شماره گوینده
7,8	1,2,3	4,5,6	گوینده نخست روش پیشنهادی
7,8	1,2,3	4,5,6	گوینده نخست روش (هیماون و همکاران، ۲۰۱۱)
7,8	1,2,3	4,5,6	گوینده دوم روش پیشنهادی
7,8	1,2,3	4,5,6	گوینده دوم روش (هیماون و همکاران، ۲۰۱۱)
-	1,2,3,7,8	4,5,6	گوینده سوم روش پیشنهادی
7,8	1,2,3	4,5,6	گوینده سوم روش (هیماون و همکاران، ۲۰۱۱)

نیز از ۱۰ جمله نمونه سیگنال گفتاری از دادگان فارس دات (در فرکانس نمونه برداری ۱۶ کیلوهرتز) بهره گرفته شده است. برای ساختن سیگنال‌های دریافتی میکروفون‌ها، از مدل‌سازی پاسخ ضربه محیط توسط روش تصویر (آلن و برکلی، ۱۹۷۹) (و شبیه‌سازی آن در محیط Matlab که در اینترنت^۱ ارائه شده) استفاده کردند. محیط (پارامتری برای نشان دادن میزان انعکاسی بودن محیط) نیز برابر با ۱۵۰ میلی‌ثانیه در نظر گرفته شده است. سیگنال رسیده از هر منبع به هر میکروفون، از کانولوشن سیگنال منبع در پاسخ ضربه محیط بین منبع و میکروفون مربوطه، به دست آمده است. در ارزیابی میزان بهسازی گفتار، سیگنال ورودی با SNR بین صفر تا ده دسیبل مورد بررسی قرار گرفته است.

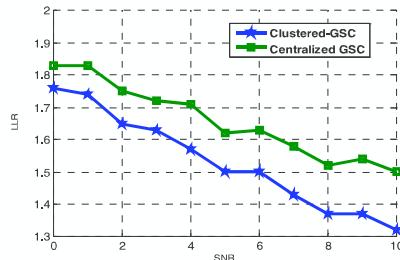
سیگنال رسیده به میکروفون‌ها ابتدا فریم‌بندی می‌شود. در این پژوهش، قاب‌ها با طول ۲۵۶ و با هم‌پوشانی پنجاه درصد در نظر گرفته شده و از پنجره همینگ استفاده کردند. بعد از قاب‌بندی و محاسبه میانگین انرژی در حوزه فرکانس، روی دنباله قاب‌ها نیز میانگین‌گیری می‌شود. بعد از چهار الی پنج قاب، نتایج به حالت به‌طور تقریبی ایستان و ثابتی (با نوسان بسیار ناچیز) نزدیک می‌شود. در این پژوهش، متوسط‌گیری روی پنج فریم انجام شده و سپس خوشبندی روی آنها صورت گرفته است. در اینجا فرض شده که یک ثالثیه نخست سیگنال، سکوت است. این فرض در کلیت مسئله خللی وارد نمی‌کند.

براساس روش خوشبندی (طرح‌شده در بخش ۲)، با محاسبه انرژی میانگین، میکروفون‌ها به خوشبندی‌هایی تقسیم می‌شوند که در هر خوشبندی، انرژی‌ها به هم نزدیک هستند. با شروع سیگنال گفتار، خوشبندی مجدد صورت گرفته و آن دسته از میکروفون‌هایی که مقادیر انرژی متوسطشان نزدیک به هم باقی می‌ماند، به عنوان خوشبندی‌هایی در نظر گرفته می‌شوند. در جدول (۲) نتایج مربوط به پیاده‌سازی الگوریتم خوشبندی در سناریوی شکل (۲) و برای سه حالت مکانی منبع گفتار (یا به تعبیری، سه گوینده) آورده شده است.

همان‌طور که از نتایج مندرج در جدول (۲) مشخص است، روش پیشنهادی در حالت گوینده سوم، دو خوشبندی (دوم و سوم) را به صورت یک خوشبندی ادغام کرده است. این مسئله را می‌توان چنین بیان کرد که در حالت گوینده سوم، دو خوشبندی فاصله بسیار مشابهی نسبت به منبع نوّفه و منبع گفتار دارند.

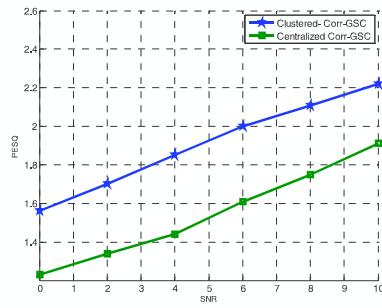
۱ <http://www.eric-lehmann.com> (last accessed: 30 Nov. 2014)



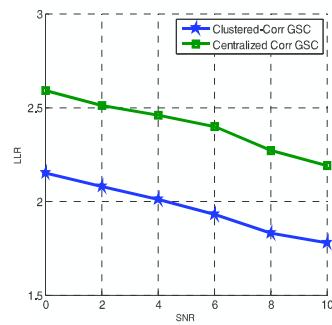


(شکل - ۷): نمودار تغییرات معیار کیفی LLR بر حسب SNR ورودی برای دو روش GSC متتمرکز و خوبه بندی پیشنهاد شده (برای حالت گوینده نخست)

همان گونه که پیش از این اشاره شد، تأثیر روش خوبه بندی پیشنهادی در فرایند بهسازی گفتار، در نسخه Correlative GSC (یعنی GSC جدیدتری از ساختار GSC) (تائیسنده، ۲۰۰۹) نیز مورد ارزیابی قرار گرفته و نتایج مربوط (شکل ۱) با عملکرد این روش در دو حالت (۱) متتمرکز، و (۲) خوبه بندی پیشنهادی در شکل های (۸) و (۹) مقایسه شده است.



(شکل - ۸): نمودار تغییرات معیار کیفی PESQ بر حسب SNR ورودی برای دو روش Correlative GSC متتمرکز و خوبه بندی پیشنهاد شده (برای حالت گوینده نخست)



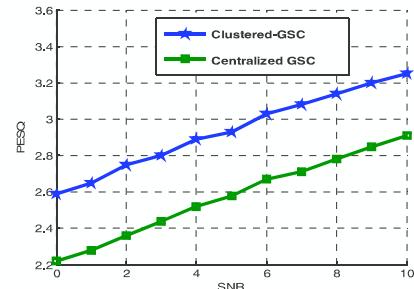
(شکل - ۹): نمودار تغییرات معیار کیفی LLR بر حسب SNR ورودی برای دو روش Correlative GSC متتمرکز و خوبه بندی پیشنهاد شده (برای حالت گوینده نخست)

(جدول - ۳): رتبه گفتاری و نووفای به دست آمده از روش پیشنهادی برای سه حالت مکانی گوینده

گوینده	خوشه نخست {1,2,3}		خوشه دوم {4,5,6}		خوشه سوم {7,8}	
	رتبه گفتاری	رتبه نووفای	رتبه گفتاری	رتبه نووفای	رتبه گفتاری	رتبه نووفای
اول	۲	۳	۱	۲	۳	۱
دوم	۳	۱	۲	۳	۲	۱
سوم	۱	۲	۳	۲	۱	۳

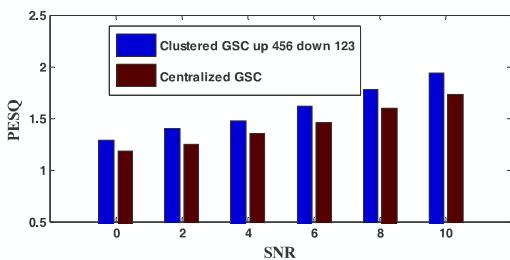
برای ارزیابی ساختار جدید GSC، عملکرد این روش در شرایط مختلف محیطی مورد بررسی قرار گرفته است. در ساختار پیشنهادی، از خوشه میکروفونی با رتبه گفتاری نخست در شاخه بالایی GSC و از خوشه میکروفونی با رتبه گفتاری آخر در شاخه پایینی GSC استفاده کرده ایم. در این ارزیابی، در مقادیر مختلف SNR ورودی، برای حالت گوینده مستقر در مکان نخست، مقادیر^۱ PESQ و^۲ LLR محاسبه شده اند. به طور معمول معیار LLR برای سنجش اعوجاج استفاده شده و هر چه این مقدار کمتر باشد، کیفیت سیگنال از لحاظ اعوجاج بهتر است. معیار PESQ نیز کیفیت کلی سیگنال گفتار را نشان می دهد.

شکل های (۶) و (۷) نتایج ارزیابی GSC جدید را نشان می دهد. در این شکل ها، نتایج برای GSC خوبه بندی شده و GSC متتمرکز که در آن از تمامی میکروفون ها در دو شاخه بالایی و پایینی استفاده می شود، آورده شده است. در این آزمایش ها RT60 محیط برابر با ۱۵۰ میلی ثانیه فرض شده است. هر دو شکل، عملکرد برتر روش پیشنهادی را در مقایسه با ساختار GSC متتمرکز نشان می دهد.

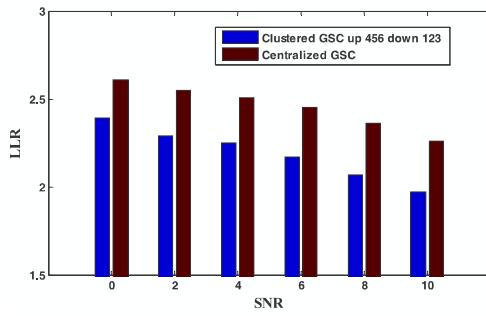


(شکل - ۶): نمودار تغییرات معیار کیفی PESQ بر حسب SNR ورودی برای دو روش GSC متتمرکز و خوبه بندی پیشنهاد شده (برای حالت گوینده نخست)

^۱ Perceptual Evaluation of Speech Quality
^۲ Log Likelihood Ratio



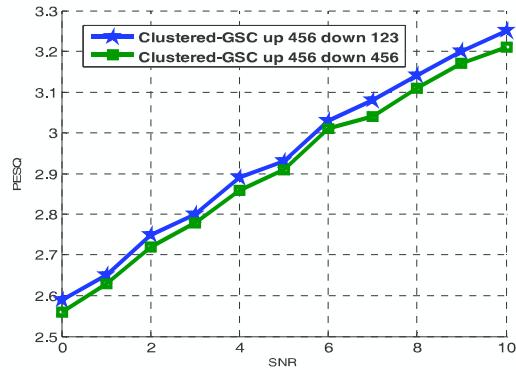
(شکل-۱۱): نمودار تغییرات معیار کیفی PESQ بر حسب SNR ورودی برای دو روش GSC متمرکز و خوشبندی پیشنهادشده (برای گوینده دوم)



(شکل-۱۲): نمودار تغییرات معیار کیفی LLR بر حسب SNR ورودی برای دو روش GSC متمرکز و خوشبندی پیشنهادشده (برای گوینده دوم)

در ارزیابی‌های بالا، RT60 محیط برابر با ۱۵۰ میلی ثانیه و بنابراین میزان انعکاس محیط ناچیز بوده است. برای بررسی عملکرد روش پیشنهادی در شرایط محیطی با انعکاس (شبیه محیط‌های معمول اداری و دفتری)، در ادامه فرض می‌کنیم که مقدار RT60 محیط برابر با ۳۰۰ میلی ثانیه باشد. با این فرض، دیگر نمی‌توان نوفه را به عنوان منبع جهتی در نظر گرفت و محیط به سمت میدان نوفه‌ای پخششده میل می‌کند (هبتز، ۲۰۰۷). در چنین سناریویی و با درنظر گرفتن حالت مکانی گوینده نخست، با فرض اینکه خوشبندی با اعمال روش‌هایی مانند روش ارائه شده در (هیماون و همکاران، ۲۰۱۱) به صورت درست انجام شده باشد، روند ساختار GSC پیشنهادی را اعمال و کیفیت خروجی را در شکل‌های (۱۳) و (۱۴) با سامانه GSC متمرکز مقایسه می‌کنیم. در محیط با انعکاس زیاد، ماهیت سیگنال‌های رسیده به میکروفون‌ها می‌تواند بسیار متفاوت باشد. این تفاوت ناشی از متفاوت بودن پاسخ ضربه محیط با توجه به رابطه (۱) است. این تفاوت برای میکروفون‌هایی که فاصله زیادی نسبت به یکدیگر دارند، بیشتر خواهد بود. با اعمال خوشبندی و اختصاص خوشبندی‌های متفاوت حاوی

برای بررسی بیشتر اثر استفاده هوشمندانه از خوشبندی میکروفونی در دو شاخه GSC در آزمایشی دیگر، عملکرد ساختاری که در هر دو شاخه بالا و پایین آن، از همان خوشبندی با رتبه گفتاری بالا در شاخه بالا و از خوشبندی با رتبه گفتاری پایین در شاخه پایین استفاده شده براساس معیار PESQ مقایسه می‌نماید. همان‌طور که در این شکل مشخص است، نتیجه بهتر مربوط به حالتی است که خوشبندی میکروفونی با رتبه گفتاری کمتر به شاخه پایین GSC داده شود.



(شکل-۱۰): مقایسه خروجی دو حالت برای GSC خوشبندی شده. در حالت نخست روش پیشنهادی اعمال شده، ولی در حالت دوم در شاخه پایینی همان خوشبندی شاخه بالا استفاده شده است.

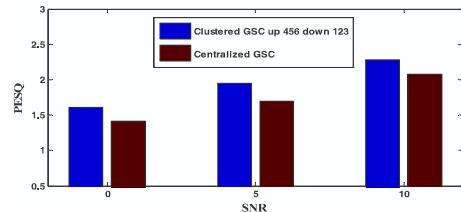
در شکل‌های (۱۱) و (۱۲) نمودار نتایج مربوط به اعمال الگوریتم پیشنهادی برای گوینده دوم نمایش داده شده است. در شکل (۱۱) میزان PESQ خروجی‌ها برای الگوریتم پیشنهادی و الگوریتم کلاسیک GSC و در شکل (۱۲) نیز میزان LLR خروجی‌ها آورده شده است. همان‌طور که از نمودارها مشخص است، روش پیشنهادی عملکرد بهتری داشته است. یکی از اهداف خوشبندی میکروفون‌ها در ساختار GSC، کاهش میزان نشت سیگنال گفتار در شاخه پایین GSC بود. همان‌طور که در شکل (۱۲) مشخص است، میزان LLR خروجی که به نوعی بیان‌گر میزان اعوجاج می‌باشد، نسبت به GSC متمرکز کاهش یافته است.

برای تعیین این که کدام خوشه به کدام شاخه تعلق بگیرد، دو روش رتبه بندی تعریف شد. در نخستین روش، ملاک رتبه بندی میزان نزدیکی خوشه ها به گوینده، و در دومین روش، ملاک رتبه بندی میزان نزدیکی خوشه ها به منبع نویه ای تعریف و نتایج حاصله به ترتیب رتبه گفتاری و رتبه نویه ای هر خوشه نامیده شد. خوشه با بالاترین رتبه گفتاری به شاخه بالایی GSC، و خوشه با پایین ترین رتبه گفتاری و یا در حالت دیگر، خوشه با بالاترین رتبه نویه ای به شاخه پایین GSC داده شد. انگیزه و توجیه نظریه حاضر این بود که در شاخه پایین، هدف ساختن نمونه ای مرجع از نویه محیط است و هرچه سیگنال گفتار تمیز در سیگنال ورودی این شاخه کمتر باشد، نشت کمتر سیگنال گفتار در شاخه پایین را به دنبال خواهد داشت.

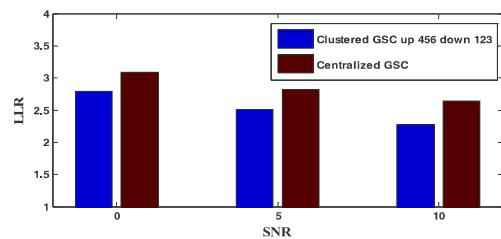
عملکرد روش پیشنهادی با معیارهای کیفی متعددی بررسی و ارزیابی شد. در بررسی عملکرد روش پیشنهادی از دو روش GSC کلاسیک و روش GSC همبسته (تانسند، ۲۰۰۹) استفاده شد. نشان داده شد که خروجی ساختار GSC خوشه بندی شده کیفیت بهتری نسبت به خروجی ساختار GSC متمرکز دارد.

در شکل دهنده پرتو GSC خوشه بندی شده، نحوه انتخاب بهینه خوشه ها و در حالت کلی تر، میکروفون ها برای هر شاخه به صورت دقیق کاری سخت است که می تواند پیشنهادی برای ادامه کار در این زمینه باشد.

میکروفون های نزدیک به هم و درنتیجه میکروفون های با ماهیت انعکاسی مشابه به دو شاخه بالایی و پایینی GSC. احتمال نشت سیگنال گفتار در شاخه پایین نسبت به ساختار GSC متمرکز کاهش می پاید. نتیجه این مزبت در نمودار شکل (۱۳) بهوضوح مشخص است.



(شکل-۱۳): نمودار تغییرات معیار کیفی PESQ بر حسب ورودی برای دو روش GSC متمرکز و GSC پیشنهادی با فرض دانستن خوشه های درست برای گوینده نخست در محیط با RT60=300ms



(شکل-۱۴): نمودار تغییرات معیار کیفی LLR بر حسب ورودی برای دو روش GSC متمرکز و GSC پیشنهادی با فرض دانستن خوشه های درست برای گوینده نخست در محیط با RT60=300ms

۵- نتیجه گیری

در این مقاله، با کنار گذاشتن فرض اساسی پخششنه بودن میدان نویه ای در مقایسه با روش ارائه شده در (هیماوان و همکاران، ۲۰۱۱)، ابتدا سعی کردیم روش خوشه بندی مناسبی برای آرایه های میکروفونی Ad-hoc ارائه دهیم. بر همین اساس، یک روش خوشه بندی و یک روش رتبه بندی خوشه ها بر اساس انرژی و استفاده از چندین منبع پیشنهادی شد. در ادامه، این روش ها در چندین ستاریو مورد ارزیابی قرار گرفته و نتایج مناسبی به دست آمد.

همچنین بر اساس خوشه بندی میکروفون ها، ساختار جدیدی برای شکل دهنده پرتو و فقی GSC پیشنهاد شد. در این ساختار جدید، به جای این که سیگنال تمام میکروفون ها در هر دو شاخه GSC به کار گرفته شود (GSC متمرکز)، در هر شاخه از این شکل دهنده پرتو، تنها یک خوشه از میکروفون ها به کار گرفته می شود (GSC خوشه بندی شده).

Van Veen, B. D., and Buckley, K. M., 1988, Beamforming: A versatile approach to spatial filtering, IEEE Trans. Acoust., Speech, Signal Process. Mag., vol. 5, no. 2, pp. 4–24.



سید حمید یزدانی دوره کارشناسی را در سال ۱۳۹۰ در رشته مهندسی برق (مخابرات) در دانشگاه شیراز گذرانده و مدرک کارشناسی ارشد خود را در سال ۱۳۹۲ در رشته مهندسی برق (مخابرات-سیستم) از دانشگاه یزد اخذ کرد. زمینه‌های علمی مورد علاقه وی پردازش آرایه‌ای و بهسازی گفتار می‌باشد.

نشانی رایانماء ایشان عبارتست از:
seyedhamidyazdani@stu.yazd.ac.ir



همیدرضا ابوطالبی دوره کارشناسی و کارشناسی ارشد را به ترتیب در سال‌های ۱۳۷۴ و ۱۳۷۷ در رشته مهندسی برق (مخابرات) در دانشگاه صنعتی شریف گذرانده و مدرک دکترای خود را در سال ۱۳۸۲ در رشته مهندسی برق (مخابرات) از دانشگاه صنعتی امیرکبیر اخذ کرد. ایشان در جریان رساله دکترای خویش، به‌مدت یک سال در دوره فرست مطالعاتی در دانشگاه واترلو کانادا به سر برده. دکتر ابوطالبی در سال ۱۳۸۲ به دانشکده مهندسی برق و کامپیوتر دانشگاه یزد پیوست و هم‌اکنون به عنوان دانشیار این دانشکده مشغول به فعالیت است. وی همچنین در سال ۱۳۸۹-۹۰ یک دوره فرست مطالعاتی را در مرکز تحقیقاتی Idiap در سوئیس سپری کرد. زمینه‌های علمی مورد علاقه وی پردازش آرایه‌ای سیگنال گفتار، بهسازی گفتار، مکان‌بایی گوینده، فیلترهای ورقی، و آنالیز زمان-فرکانس است.

نشانی رایانماء ایشان عبارتست از:

habutalebi@yazd.ac.ir

Bertrand, A., and Moonen, M., 2012, Distributed node-specific LCMV beamforming in wireless sensor networks, IEEE Trans. Audio, Speech, Lang. Process., vol. 60, no. 1, pp. 233–246, pp. 785–797.

Gannot, S., Burshtein, D., and Weinstein, E., 2001, Signal enhancement using beamforming and nonstationarity with applications to speech, IEEE Trans. on Signal Processing, vol. 49, pp. 1614–1626.

Griffiths, L. J., and Jim, C. W., 1982, An alternative approach to linearly constrained adaptive beamforming, IEEE Trans. Antenna and Propagation, vol. 30, pp. 27–34.

Habets, E. A. P., 2007, Single- and multi-Microphone speech dereverberation using spectral enhancement, Ph.D. Thesis, Dept. of Electrical Engineering, University of Eindhoven.

Himawan, I., 2010, Speech recognition using Ad-hoc microphone array, Ph.D. Thesis, Queensland University of technology, Brisbane, Queensland.

Himawan, I., Sridharan, S., and McCowan, I., 2011, Clustered blind beamforming from Ad-hoc microphone arrays, IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 4, pp. 661–676.

Markovich-Golan, S., Gannot, S., and Cohen, I., 2013, Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks, IEEE Trans. Audio, Speech and Language Processing, vol. 21, pp. 343–356.

McQueen, J. B., 1967, Some methods for classification and analysis of multivariate observations, in Proc. 5th Berkeley Symp. Math. Statist. Probability, vol. 1, pp. 281–297.

Roy, O., and Vetterli, M., 2009, Rate-constrained collaborative noise reduction for wireless hearing aids, IEEE Trans. on Signal Processing, vol. 57, no. 2, pp. 645–6579.

Srinivasan, S., 2011, Using a remote wireless microphone for speech enhancement in nonstationary noise, Tech. Report, Digital Signal Processing Group, Philips Research, Netherlands.

Townsend, P., 2009, Enhancements to the generalized sidelobe canceller for audio beamforming in an immersive environment, MSc Thesis, University of Kentucky.

Yu, T., Hansen, J. H. L., 2010, Automatic beamforming for blind extraction of speech from music environment using variance of spectral fluxinspired criterion, IEEE Journal of selected topics in signal process., vol. 4, no. 5.

