شکل‌دهی وفقی و هوشمند پرتو در آرایه‌های میکروفونی Ad-hoc با استفاده از خوشه‌بندی و رتبه‌بندی میکروفون‌ها

سیّدحمید یزدانی و حمیدرضا ابوطالبی
دانشکدهٔ 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. شکل‌دهی پرتو یکی از روش‌های مطرح برای بهسازی گفتار است که با استفاده از آرایه‌های میکروفونی صورت می‌پذیرد. با توجه به محدودیت‌های موجود در نحوهٔ چینش میکروفون‌ها، پنجره‌ای به سوی بحث آرایه‌های میکروفونی با چیدمان نامنظم (Ad-hoc) گشوده شده است. با فرض عدم شناخت نسبت به مکان و چینش میکروفون‌ها و همچنین پراکنده‌بودن میکروفون‌ها در محیط، در این مقاله، نظریهٔ خوشه‌بندی میکروفون‌ها براساس انرژی دریافتی از منابع موجود در محیط مورد توجه قرار گرفته و روش جدیدی برای خوشه‌بندی پیشنهاد شده است. در ادامه، برای هر خوشه میکروفونی، دو نوع رتبه پیشنهاد شده که براساس میزان فاصله به منبع نوفه و گوینده می‌باشند. همچنین براساس نتیجهٔ خوشه‌بندی و رتبه‌بندی آنها، ساختار جدیدی برای شکل‌دهنده پرتو GSC (Generalized Sidelobe Canceller) ارائه شده است. برای حالتی که نوفه پخشنده نباشد، براساس انرژی منابع موجود در محیط روشی برای خوشه‌بندی پیشنهاد شده که می‌تواند برای سایر میدان‌های نوفه‌ای نیز به‌کار گرفته شود. با پیاده‌سازی و ارزیابی روش پیشنهادی دیده می‌شود که در بعضی از حالت‌ها، استفاده از یک خوشه در برابر به‌کارگیری تمام میکروفون‌ها نتیجهٔ بهتری را به‌دست می‌دهد که این خود حاکی از دستاوردی دیگر است. این دستاورد، کاهش تعداد میکروفون‌های مورد استفاده است که به نوبه خود، کاهش پیچیدگی و حجم محاسبات (در عین افزایش کیفیت خروجی) را به‌دنبال دارد.

واژگان کلیدی: بهسازی گفتار، شکل‌دهی پرتو، آرایه‌های میکروفونی Ad-hoc، خوشه‌بندی، شکل‌دهندهٔ پرتو GSC

۱- مقدّمه

با رشد و گسترش فناوری و پیشرفت در حوزه پردازش گفتار و زمینه‌های دیگری همچون پردازش تصویر، اهداف جدیدی در حوزه ارتباطات شکل گرفته است. در چند دهه اخیر، ارتباط بین انسان و رایانه یکی از مهم‌ترین زمینه‌های پژوهشی در حوزهٔ پردازش گفتار بوده است. با در نظر گرفتن گفتار به‌عنوان مهم‌ترین رابط کاربری با رایانه، دریافت گفتار با کیفیتی خوب بیش از پیش اهمیت می‌یابد. در بسیاری از شرایط محیطی طبیعی، سیگنال‌های صوتی توسط میکروفون‌هایی که از گوینده فاصلهٔ به‌نسبه زیادی دارند، ضبط می‌شوند. به‌دلیل فاصلهٔ زیاد گوینده و میکروفون‌ها و وجود عوامل مخربی همچون نوفه، پژواک، حرکت چندین گویندهٔ همزمان، تداخل و ...، سیگنال گفتار مطلوب توسط میکروفون‌ها با کیفیت پایینی ضبط می‌شود. کاهش کیفیت سیگنال دریافتی میکروفون‌ها، باعث افت عملکرد سامانه‌های پردازش گفتار می‌شود.

از دیرباز شیوه‌های متعددی برای بهسازی گفتار ارائه شده که یکی از روش‌های چندکانالهٔ مورد استفاده به این منظور، شکل‌دهی پرتو[1] بوده است. در شکل‌دهی پرتو با اعمال ضرایب (یا فیلترهایی) بر روی سیگنال‌های دریافتی از میکروفون‌های مختلف و ترکیب آنها با هم، به تعبیری سیگنال دریافتی از راستای سیگنال گفتارِ هدف تقویت و سیگنال‌های تداخل و نوفه دریافتی از سایر راستاها تضعیف می‌شود (ون وین و بوکلی، ۱۹۸۸).

[1] Beamforming

اگرچه در سامانه‌های کلاسیک آرایه میکروفونی، یک سری میکروفون در چیدمانی منظم (به‌عنوان مثال خطی، دایره‌ای یا نظایر آن) استفاده می‌شده، در چند سال اخیر، بحث به‌کارگیری مجموعه‌ای از میکروفون‌ها در یک چیدمان نامنظم مورد توجه قرار گرفته است (هیماوان و همکاران، ۲۰۱۱). این آرایه‌های میکروفونیِ با چیدمان نامنظم (یا به‌اصطلاح رایج، آرایه‌های میکروفونی Ad-hoc)، به‌طور معمول از میکروفون‌های موجود در سامانه‌های رایانه‌ای و ارتباطی قابل حمل (لپ‌تاپ‌ها، تلفن‌های همراه و ...) تشکیل می‌شود.

در آرایه‌های میکروفونی Ad-hoc، فرض می‌شود که مکان میکروفون‌ها برای شکل‌دهندهٔ پرتو مشخص نبوده و شکل‌دهی پرتو باید به‌صورت کور انجام شود؛ این بدان معنی است که الگوریتم‌های بهسازی، باید تنها با استفاده از سیگنال میکروفون‌ها (و نه اطلاعات هندسی محل آنها) پیاده شود.

با توجه به این که در اکثر روش‌های کلاسیک شکل‌دهی پرتو (با چیدمان میکروفونی منظم و مشخص)، فرض بر نزدیک‌بودن میکروفون‌ها به همدیگر بوده است، ماهیت آرایه‌های میکروفونی کلاسیک و روش‌های قابل پیاده‌سازی در این آرایه‌ها با آرایه‌های Ad-hoc متفاوت است. در آرایه‌های میکروفونی کلاسیک، چینش میکروفون‌ها به‌طور دقیقی چنان صورتی می‌پذیرد که بهسازی مؤثری برای آن شرایط محیطی خاص حاصل شود. در چینش‌های Ad-hoc، این موضوع کلیت نداشته و میکروفون‌ها به‌طور دلخواه و تصادفی در محیط پخش می‌شوند.

درحالی‌که در آرایه‌های کلاسیک، با افزایش تعداد میکروفون‌ها عملکرد، بهتر می‌شود (ون وین و بوکلی، ۱۹۸۸)، در آرایه‌های Ad-hoc این موضوع نیز عمومیت ندارد (هیماوان و همکاران، ۲۰۱۱)؛ زیرا در آرایه‌های کلاسیک میکروفون‌ها از لحاظ فیزیکی و مکانی و نوعی بسیار نزدیک به یکدیگر بوده و شرایط صوتی مشابهی دارند. این باعث می‌شود در تخمین پارامترها، داده‌های موجود بیشتر بوده و خطای کمتری در تخمین رخ دهد. در آرایه‌های میکروفونی Ad-hoc، ممکن است میکروفون‌ها فاصلهٔ زیادی از هم داشته باشند؛ همچنین ممکن است، نوع میکروفون‌ها نیز متفاوت باشد. در هر دو صورت، سیگنال دریافتی میکروفون‌ها شرایط به‌طور کامل متفاوتی داشته و نمی‌توان انتظار داشت با زیاد شدن تعداد میکروفون‌ها عملکرد سامانه لزوماً بهبود یابد (هیماوان و همکاران، ۲۰۱۱).

شکل‌دهی پرتو در آرایه‌های میکروفونی Ad-hoc (و در حالت جامع‌تر، در شبکه‌های حس‌گر صوتی Ad-hoc[1] (برتراند و همکاران، ۲۰۱۰ الف،ب، ج؛ روی و وترلی، ۲۰۰۹)، با دو رویکرد صورت می‌پذیرد: در رویکرد نخست، شکل هندسی آرایه تخمین زده شده و سپس از روش‌های کلاسیک شکل‌دهی پرتو استفاده می‌شود. در رویکرد دوم، روش‌های جدیدی برای شکل‌دهی پرتو بر پایه ساختار و شرایط خاص شبکه‌های میکروفونی Ad-hoc مطرح می‌شود.

به‌عنوان یک کار شاخص در حوزهٔ شکل‌دهی پرتو در آرایه Ad-hoc، هیماوان در سال ۲۰۱۰ در رسالهٔ دکترای خویش، روش خوشه‌بندی میکروفون‌ها را ارائه و به بررسی کاربرد این روش در بازشناسی گفتار پرداخت (هیماوان، ۲۰۱۰). در این روش، ابتدا میکروفون‌ها به خوشه‌هایی تفکیک می‌شوند. این امر به‌دلیل آن است که فاصلهٔ بین میکروفونی زیاد، مشکلاتی را در تخمین تأخیر سیگنال‌های رسیده ایجاد می‌کند. همچنین نظریه این است که خوشه‌ای انتخاب شود که نسبت سیگنال به نوفه بیشتری داشته باشد. در پژوهش‌های انجام‌شده توسط هیماوان، ابتدا دو روش برای خوشه‌بندی میکروفون‌ها معرفی شده است؛ سپس خوشه‌ها رتبه‌بندی شده و در انتها با دو روش (۱- نزدیک‌ترین خوشه، و ۲- ترکیب وزن‌دار از خوشه‌ها)، شکل‌دهی پرتو صورت گرفته است (هیماوان و همکاران، ۲۰۱۱).

در سال ۲۰۱۲، برتراند شکل‌دهنده پرتو LCMV[2] را در حالتی که گره‌ها (مجموعه چندین میکروفون نزدیک به هم با یک پردازنده مرکزی) در محیط پخش شده‌اند، بررسی کرد (برتراند و مونن، ۲۰۱۲).

کارهایی نظیر (سرینیواسان، ۲۰۱۱) و (یو و هنسن، ۲۰۱۰) را نیز که در آن‌ها فرض بر مشخص‌نبودن مکان میکروفون‌هاست می‌توان کارهایی در حوزهٔ آرایه‌های میکروفونی Ad-hoc در نظر گرفت. در (سرینیواسان، ۲۰۱۱) یک میکروفون در نزدیکی منبع نوفه با مکان نامشخص قرار داده شده و با ارسال طیف توان نوفه به میکروفون دیگر عمل کاهش نوفه صورت گرفته است. در (یو و هنسن، ۲۰۱۰) نیز نوفه موسیقی از سیگنال نوفه‌ای حذف می‌شود. نظریهٔ به‌کار 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 میکروفون‌ها نداشته و قابل پیاده‌سازی در آرایه‌های میکروفونی Ad-hoc است.

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

در سال ۲۰۱۳ نیز مارکوویچ گولان سناریوی چندین گره را در نظر گرفت و در این سناریو با تعریف یک تبدیل، شکل‌دهندهٔ پرتو GSC[1] 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 گرهٔ اختصاصی تعریف شده و سپس با درنظرگرفتن معیارهایی، سیگنال‌ها در شبکه به اشتراک گذاشته می‌شوند. در (مارکوویچ گولان و همکاران، ۲۰۱۳) نشان داده شده که روش پیشنهادی بدون نیاز به ارسال حجم عظیمی از سیگنال‌های میکروفونی به پردازندهٔ مرکزی، بسیار نزدیک به روش GSC متمرکز عمل می‌کند.

در اکثر موارد بالا، فرض این بوده که گره‌ها مشخص بوده و هدف این است که در هر گره، بهترین سیگنالی که می‌توان از ترکیب مجموعه سیگنال‌های همه گره‌ها به‌دست آورد، در اختیار داشته باشیم. به‌عبارت دیگر، در موارد بالا، هدف نهایی، یک خروجی بهینه نیست؛ بلکه هدف به‌دست آمدن خروجی مناسب در هر یک از گره‌هاست.

نکتهٔ قابل ذکر دیگر آن که در مقالات (برتراند و همکاران، ۲۰۱۲؛ یو و هنسن، ۲۰۱۰؛ سرینیواسان، ۲۰۱۱؛ مارکوویچ گولان و همکاران، ۲۰۱۳)، به اصطلاح آرایهٔ میکروفونی Ad-hoc اشاره‌ای نشده و موضوع در قالب یک شبکهٔ حس‌گر صوتی مورد بررسی قرار گرفته است.

در این مقاله، هدف، بهره‌گیری از ایدهٔ خوشه‌بندی میکروفون‌ها و همچنین استفادهٔ هوشمند از سیگنال‌های دریافتی به‌منظور بهبود عملکرد شکل‌دهی پرتو GSC است. در این پژوهش، با اعمال معیارهایی ساده و در عین حال قوی، همچون انرژی در فریم‌های اولیهٔ شروع گفتار، روشی را برای خوشه‌بندی میکروفون‌ها پیشنهاد کرده و با به‌کارگیری هوشمند سیگنال‌های دریافتی در شکل‌دهندهٔ پرتو GSC، عملکرد این شکل‌دهنده را در آرایه‌های میکروفونی Ad-hoc بهبود داده‌ایم.

ساختار مطالب این مقاله به‌شرح زیر است. در بخش دوم، فرض میدان نوفه‌ای پخشنده از مسأله خوشه‌بندی کنار گذاشته شده که به معنای عدم امکان استفاده از روش خوشه‌بندی ارائه‌شده در (هیماوان و همکاران، ۲۰۱۱) است. در این بخش، روش جدیدی برای خوشه‌بندی در شرایط حضور نوفهٔ جهتی پیشنهاد شده است. در بخش سوم، با توجه به شرایط نوفه‌ای جدید، به‌طور هدفمند از خوشه‌ها برای بهبود ساختار شکل‌دهندهٔ پرتو GSC استفاده شده است. در بخش چهارم نیز نتایج شبیه‌سازی و ارزیابی روش پیشنهادی آمده، و در بخش پنجم نتایج کلی و جمع‌بندی این پژوهش ذکر شده است.

۲- نوفه جهتی و خوشه‌بندی میکروفون‌ها

یکی از شرایطی که ممکن است در یک محیط برقرار باشد، حضور نوفهٔ مکانی است. نوفهٔ مکانی (نوفهٔ جهتی) همانند یک منبع صوت عمل می‌نماید که در محیط در جهت خاصی منتشر می‌شود. در این صورت، در سیگنال دریافتی در میکروفون‌ها، نوفه‌ای با همبستگی زیاد وجود خواهد داشت. در حالتی مشابه می‌توان به مسأله تداخل اشاره کرد. تداخل نیز یکی از عوامل مخرب سیگنال گفتار است که از دستگاه‌هایی مانند ضبط صوت، بلندگو و ... انتشار می‌یابد. همچنین وجود گوینده‌های دیگری (علاوه‌بر گوینده اصلی) به‌نوعی موجب ایجاد تداخل می‌شود که به این گوینده‌ها، گوینده رقیب نیز گفته می‌شود.

با توجه به این تعاریف، می‌توان گفت تداخل از جنس گفتار بوده و برخورد با این عامل از بقیهٔ عوامل مخرب نوعاً سخت‌تر است. توجیه این که تداخل و نوفهٔ جهتی در اینجا تحت یک عنوان مورد بررسی قرار می‌گیرد، این است که هر دو را می‌توان مانند یک منبع مکانی تولید صوت در نظر گرفت. از این دیدگاه، می‌توان برای بحث خوشه‌بندی میکروفون‌ها استفاده کرد.

نکتهٔ اساسی برای خوشه‌بندی میکروفون‌ها در شرایط نوفه‌ای مورد بحث حاضر، توجه به وجود چندین منبع در محیط است. در شرایطی که در این فصل در نظر گرفته شده، حداقل یک منبع نوفهٔ مکانی و یک منبع گفتار اصلی در محیط وجود دارد. با فرض اینکه محیط دارای انعکاس ناچیزی است، می‌توان از انرژی بازتاب‌ها صرف نظر کرد.

ورودی هر میکروفون یک نسخه فیلترشده از سیگنال منبع جمع شده با نوفه است:

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

در این رابطه، h_i پاسخ ضربه محیط بین منبع و میکروفون i-ام، $*$ نماد کانولوشن، $s_i(t)$ سیگنال گفتار تمیز و $v_i(t)$ نوفهٔ دریافتی در محل میکروفون i-ام است. در اینجا نوفه و سیگنال از لحاظ آماری مستقل فرض شده‌اند. در حوزه فرکانس، مدل سیگنال به‌صورت زیر نوشته می‌شود:

[1] Generalized Sidelobe Canceller

$$X_i(f) = H_i(f)S_i(f) + V_i(f) \quad (۲)$$

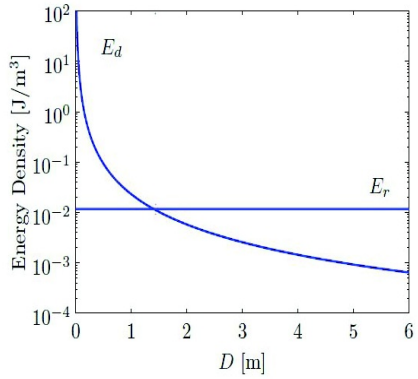
همان‌طورکه در روابط بالا مشخص است، سیگنال رسیده به میکروفون‌ها هم بر اثر نوفه و هم بر اثر پاسخ ضربهٔ محیط تغییر می‌کند. هدف تمامی الگوریتم‌های بهسازی گفتار، کم‌کردن اثر این عوامل تغییردهنده (مخرب) بر روی سیگنال تمیز است. انرژی سیگنال منتشرشده از هر منبع در مسیر مستقیم طبق رابطهٔ زیر تغییر می‌کند (هبتز، ۲۰۰۷):

$$E_d = \frac{QW_s}{4\pi cD^2} \quad (۳)$$

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

$$E_r = \frac{4W_s}{cR} \quad (۴)$$

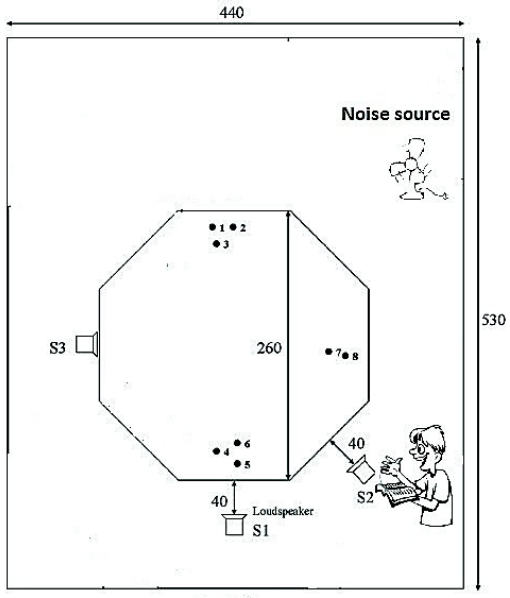
که در رابطهٔ بالا، R ثابت اتاق (وابسته به شرایط فیزیکی اتاق) است. همان‌طورکه از این رابطه مشخص است، انرژی بازتاب‌ها به فاصله بستگی نداشته و مقدار ثابتی است. در شکل (۱) چگالی انرژی مسیر مستقیم و بازتاب رسم شده است.



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

با فرض این‌که 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 خوشه‌بندی به‌نسبه مناسبی رسید. نظریهٔ اصلی این است که میکروفون‌هایی که به هم نزدیک باشند، انرژی مشابهی از هر منبع دریافت می‌کنند. هرچه میکروفون‌ها به هم نزدیک‌تر باشند، شباهت انرژی آنها برای منابع مختلف بیشتر است. بنابراین می‌توان از این نظریه برای به‌دست آوردن معیار جدیدی برای نزدیکی میکروفون‌ها استفاده کرد. درصورتی‌که تداخل و یا چندین گوینده در محیط حضور داشته باشند، این نظریه به شکل قوی‌تری قابل اعمال خواهد بود.

برای بیان بهتر نظریهٔ کار، سناریوی رسم‌شده در شکل (۲) را که سناریوی مورد استفاده برای ارزیابی نتایج این مقاله نیز می‌باشد؛ در نظر بگیرید. در این سناریو، ابعاد اتاق ۲.۷×۵.۳×۴.۴ (بر حسب متر) و فرکانس نمونه‌برداری نیز ۱۶۰۰۰ هرتز است. هشت میکروفون مورد استفاده از یک نوع بوده و تمام‌جهته[1] سیگنال را دریافت می‌کنند. نوفهٔ مکانی یک نوفهٔ سفید با میانگین صفر است. پاسخ ضربه اتاق با روش تصویر[2] (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 نوفه تنها و هم در حالت شروع گفتار، سطح انرژی مشابهی دارند. البته این موضوع را به‌طور دقیق‌تر می‌توان در فریم‌های ابتدایی گفتار مشاهده کرد؛ دلیل این امر این است که با گذشت زمان اندکی (حدود ۵۰ تا ۱۰۰

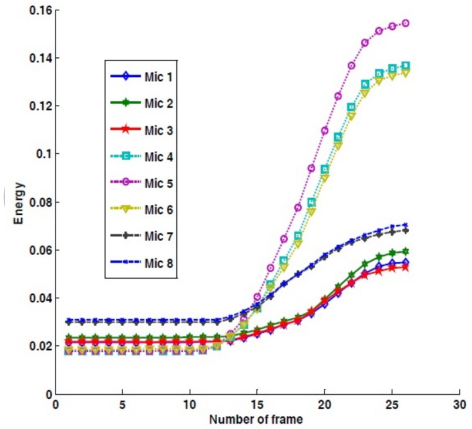
[1] Omni-directional
[2] Image method

میلی‌ثانیه)، بازتاب‌ها نیز به میکروفون‌ها رسیده و سطح انرژی را به‌طور مشخص نمی‌توان 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)، پایهٔ الگوریتم براساس قسمت سکوت سیگنال (زمانی که فقط منبع نوفه فعال است) می‌باشد. با توجه به مباحث بالا، الگوریتم زیر برای خوشه‌بندی میکروفون‌ها پیشنهاد می‌شود:

۱- اندازه‌گیری میانگین انرژی نوفه در هر میکروفون در قسمت‌های سکوت سیگنال

۲- خوشه‌بندی اولیه میکروفون‌ها با استفاده از الگوریتم k-means (مک کوین، ۱۹۶۷) براساس میانگین‌های محاسبه‌شده در مرحله قبل



(شکل- ۳): تغییرات میانگین انرژی در تمام میکروفون‌ها درلحظه شروع گفتار گویندهٔ نخست

۳- اندازه‌گیری میانگین انرژی سیگنال در فرکانس‌های پایین (با فرض طول تبدیل فوریه ۲۵۶ نقطه‌ای و فرکانس نمونه‌برداری ۱۶۰۰۰ هرتز، تا بین فرکانسی ۲۵، معادل با فرکانس حدود دو کیلوهرتز)

۴- خوشه‌بندی دوبارهٔ میکروفون‌ها در هر خوشهٔ نخستین و تشکیل خوشه‌های نهایی

در توضیح مرحلهٔ سوم الگوریتم بالا، ذکر این نکته لازم است که سیگنال گفتار، سیگنالی پایین‌گذر بوده و برای کاهش حجم محاسبات، می‌توان خوشه‌بندی 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. همان‌طورکه گفته شد، هر چه میکروفون‌ها به هم نزدیکتر باشند، میزان مشابهت انرژی دریافتی میکروفون‌ها از هر منبع افزایش می‌یابد. برای تعیین میزان نزدیکی خوشه‌ها به منبع نوفه و یا منبع گفتار از اختلاف میانگین انرژی‌های هر خوشه با یکدیگر استفاده می‌نماییم. البته لازم به ذکر است که برای سنجش نزدیکی منبع گفتار به خوشه‌ها، می‌توان از ویژگی‌هایی دیگری همچون زمان شروع سیگنال، TDOA و ... نیز استفاده کرد.

در جدول (۱)، الگوریتم خوشه‌بندی پیشنهادی با جزییات پردازشی بیان شده است.

در این پژوهش، بعد از خوشه‌بندی میکروفون‌ها، هر خوشه از دو روند مورد رتبه‌بندی قرار می‌گیرد. در رتبه‌بندی نخست، میزان نزدیکی خوشه‌ها به منبع گفتار با توجه به انرژی قاب‌های گفتاری رتبه‌دهی می‌شود. در مقالهٔ حاضر، به این رتبه، رتبهٔ گفتاری می‌گوییم. برای قوی‌ترشدن این رتبه‌بندی، علاوه‌بر انرژی، از اختلاف در لحظهٔ آغاز سیگنال نیز به‌عنوان معیار دوم استفاده می‌کنیم. 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 ۵) مقایسه می‌شود؛ زمانی که این نسبت از T_1 بیشتر باشد، از دو قاب قبل (همانند مرحله ۳-۲) الگوریتم پیشنهادی در جدول (۱)) با هم‌پوشانی بسیار زیاد (بالاتر از ۹۵ درصد) دوباره قاب‌بندی و با مقایسه با سطح آستانه T_2 (در این پژوهش برابر با ۳/۵) لحظه شروع سیگنال تخمین زده می‌شود. تفاوت این اقدام با آنچه در خوشه‌بندی صورت می‌گیرد، در این است که در فرایند رتبه‌بندی، معیار مورد استفاده، میانگین شماره فریم شروع گفتار بر روی تمام میکروفون‌های داخل هر خوشه است؛ هر خوشه‌ای که در آن، میانگین این شماره قاب‌ها کمتر باشد، به‌عنوان خوشهٔ نزدیک‌تر به منبع تلقی شده و رتبهٔ بهتری می‌گیرد. علت میانگین‌گیری نیز کم‌کردن اثر پخش‌شدگی میکروفون‌های یک خوشه در رتبه آن خوشه می‌باشد. لازم به ذکر است در این مقاله T_1 و T_2 به‌صورت دستی تنظیم شده که می‌توان در پژوهش‌های آینده، تعیین وقفی آنها را مورد بررسی قرار داد.

در رتبه‌بندی دوم، میزان نزدیکی خوشه‌ها به منبع نوفه با توجه به انرژی در قاب‌های نوفه‌ای بررسی و خوشه‌ها رتبه‌دهی می‌شوند. در این مقاله به این رتبه، رتبهٔ نوفه‌ای می‌گوییم. بنابراین پس از طی این فرایند، به هر خوشه دو عدد که نشان‌دهنده رتبهٔ گفتاری و رتبه نوفه‌ای آنهاست، اختصاص می‌یابد. در ادامه، خواهیم دید که این دو عدد

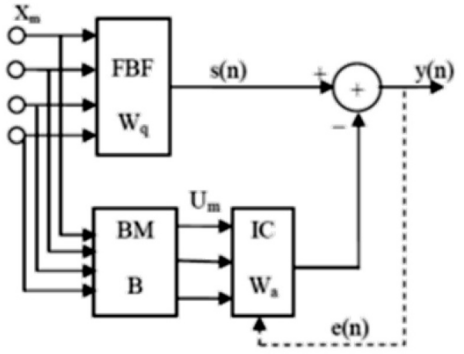
چگونه در اصلاح ساختار شکل‌دهندهٔ پرتو GSC مورد استفاده قرار می‌گیرد.

۳- معرفی ساختار GSC اصلاح‌شده برای شکل‌دهی پرتو

شکل‌دهندهٔ پرتو وفقی GSC دارای یک ساختار دوشاخه‌ای است. یکی از کاربردهای مهم این شکل‌دهندهٔ پرتو برای حذف نوفه (یا هر عامل مخربِ) مکانی و جهتی بوده و در حضور چنین میدان‌های نوفه‌ای، این شکل‌دهندهٔ پرتو عملکرد مناسبی نشان می‌دهد (گریفیت، ۱۹۸۲). با توجه به اینکه در این پژوهش، عامل مخرب سیگنال گفتار، تداخل و یا نوفهٔ جهتی فرض شده، به‌کارگیری شکل‌دهندهٔ پرتو GSC امری طبیعی به نظر می‌رسد.

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

همان‌طورکه شکل (۴) نشان می‌دهد، شکل‌دهندهٔ پرتو GSC، دارای دو شاخه است. در شاخهٔ بالایی یک شکل‌دهندهٔ ثابت پرتو (FBF[1]) قرار دارد که برای آن در اکثر موارد، از شکل‌دهندهٔ پرتو تأخیر و جمع[2] استفاده می‌شود. در شاخهٔ دوم نیز، با به‌کارگیری بلوک BM[3] سعی می‌شود سیگنال گفتار حذف‌شده و نمونه‌ای از نوفهٔ محیط ایجاد شود. این نمونه از نوفهٔ محیط، به‌عنوان نوفهٔ مرجع فرض شده و با ساختار فیلتر وفقی (IC[4])، نوفه در خروجی نهایی GSC حداقل می‌شود.



(شکل- ۴): ساختار کلی شکل‌دهندهٔ پرتو GSC

با توجه به توضیح بالا می‌توان دید که یکی از نقش‌های اصلی در این ساختار بر عهده فیلتر وفقی است. به‌عنوان نکته‌ای شناخته‌شده در رابطهٔ با فیلتر وفقی، می‌توان گفت هرچه قدر سیگنال ورودی فیلتر به نوفهٔ مرجع نزدیک‌تر (و با آن همبسته‌تر) باشد، عملکرد بهتری در کاهش نوفه دیده می‌شود. همان‌طورکه در ادامه توضیح خواهیم داد، این موضوع را می‌توان به‌عنوان مبنای پیشنهاد ساختار جدیدی برای GSC به‌کار گرفت.

(جدول- ۱): الگوریتم پیشنهادی برای خوشه‌بندی میکروفون‌ها

۱- فریم‌بندی سیگنال گفتار با هم‌پوشانی پنجاه درصد و در نظر گرفتن قسمت سکوت سیگنال (نوفهٔ خالص) برای خوشه‌بندی 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 i 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، به‌طور معمول بین‌های ۰ تا ۲۵ را شامل می‌شود).

(در این پژوهش، مقایسه زیر با حد آستانه $T_1 = 5$ انجام شده است.)

$$\frac{E_{ave,i,lowbins,x_i}}{E_{ave,i}} \geq T_1 \Rightarrow$$

: estimation of start frame in microphone i

۳-۲- افزایش میزان هم‌پوشانی قاب‌ها به ۹۵ درصد از دو قاب قبل از شماره به‌دست آمده برای میکروفونِ با تأخیر کمتر (حاصل از مرحله (۳-۱) یعنی فریم با شماره ($x_{i,\min} - 2$)) و قاب‌بندی مجدد و تکرار روند مشابه مرحله (۳-۱) با سطح آستانه کمتر (T_2) به‌منظور بالابردن دقت و ذخیرهٔ شمارهٔ قاب (N_i).

(این آستانه در اینجا برابر با $T_2 = 3.5$ درنظرگرفته شده است)

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

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

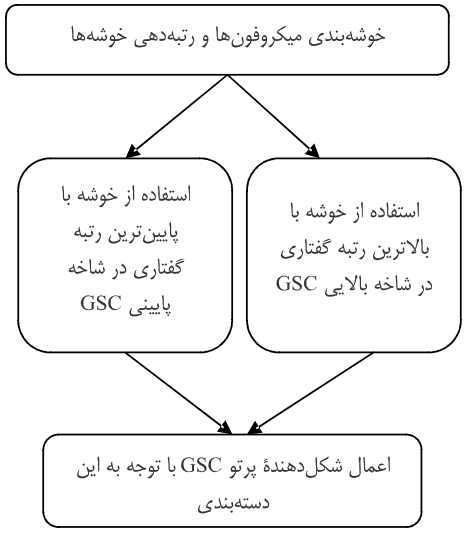
[1] Fixed BeamFormer
[2] Delay and Sum Beamformer
[3] Blocking Matrix
[4] Interference Canceller

با توجه به این که پژوهش حاضر، آرایه‌های میکروفونی Ad-hoc مورد بررسی و استفاده قرار گرفته و در این آرایه‌ها، میکروفون‌ها نوعاً در کل محیط پخش شده‌اند، بنابراین می‌توان انتظار داشت که بعضی از میکروفون‌ها به منبع نوفه و برخی از آنها به گویندهٔ اصلی نزدیک‌تر باشند. طبیعی است که میکروفون‌های دورتر از گویندهٔ اصلی، با توجه به رابطهٔ (۱) سیگنال گفتار را با تضعیف بیشتری دریافت می‌کنند.

یکی از مشکلاتی که در ساختار GSC اولیه (گریفیت، ۱۹۸۲) وجود دارد، نشت سیگنال گفتار به شاخه پایینی است که به نوبهٔ خود، باعث کاهش SNR[1] در خروجی نهایی می‌شود. هرچه نشت سیگنال گفتار به شاخهٔ پایینی GSC کمتر باشد، می‌توان گفت سیگنال خروجی بلوک حذف‌کنندهٔ گفتار (بلوک BM در شکل (۴))، دارای درصد بیشتری نوفه بوده و درنتیجه نزدیکی بیشتری با نوفهٔ مرجع خواهد داشت. بنابراین درصورتی‌که برای شاخهٔ پایینی GSC از میکروفون‌های با فاصله بیشتر نسبت به گوینده اصلی (به‌منظور وجود درصد کمتری از سیگنال گفتار) و در عین حال فاصله کمتر نسبت به منبع نوفه (به‌منظور وجود درصد بیشتری از نوفه) استفاده شود، عملکرد فیلتر وفقی می‌تواند بهبود یابد. همچنین استفاده از میکروفون‌های با نسبت سیگنال به نوفهٔ بالاتر در شاخه بالایی GSC می‌تواند عملکرد کلی این شکل‌دهنده پرتو را بهبود بخشد. این مزیت‌ها در آرایه‌های میکروفونی کلاسیک وجود ندارد؛ زیرا در آن آرایه‌های میکروفونی، اکثر میکروفون‌ها در فضای کوچکی از محیط قرار گرفته و SNR ورودی آنها به‌طور تقریبی برابر است.

بر اساس آنچه در بالا شرح داده شد، در این پژوهش، ساختاری اصلاح‌شده برای شکل‌دهنده پرتو GSC پیشنهاد می‌نماییم. نظریهٔ پیشنهادی آن است که به جای استفاده از تمام خوشه‌های میکروفونی در هر دو شاخهٔ بالا و پایین GSC، براساس رتبهٔ خوشه‌ها (رتبهٔ گفتاری و رتبهٔ نوفه‌ای) در هر شاخه تنها بخشی از خوشه‌ها به کار گرفته شوند. روند کلی به این صورت است که خوشه‌های با رتبهٔ بهتر گفتاری در شاخه بالایی و خوشه‌های با رتبهٔ بهتر نوفه‌ای در شاخهٔ پایینی استفاده شوند. البته به‌صورت دیگری نیز می‌توان عمل کرد و آن درنظرگرفتن تنها رتبهٔ گفتاری است؛ خوشه با بهترین رتبهٔ گفتاری در شاخه بالایی، و خوشه با بدترین رتبهٔ گفتاری در شاخهٔ پایین استفاده شود. در شکل (۵) روندکلی ساختار پیشنهادی جدید برای GSC رسم شده است.

[1] Signal to Noise Ratio



شکل ۵- روند کلی کارشکل‌دهنده پرتو پیشنهادی

۴- شبیه‌سازی و ارزیابی روش پیشنهادی

در این بخش، نتایج مربوط به پیاده‌سازی و ارزیابی روش‌های پیشنهادی برای خوشه‌بندی، رتبه‌بندی و شکل‌دهی پرتو وفقی GSC با ساختار جدید پیشنهادی آورده می‌شود.

ذکر این نکته ضروری است که چون هدف اصلی از ارزیابی‌ها، بررسی مزیت روش پیشنهادی بوده، در این مقاله نخست از GSC ساده (گریفیت، ۱۹۸۲) استفاده کرده‌ایم. در عین حال، برای توسعهٔ ارزیابی‌ها و مقایسهٔ کارایی روش خوشه‌بندی پیشنهادی در ترکیب با نسخه‌های جدیدتر سامانه GSC، از روش GSC همبسته (Correlative GSC) (تانسند، ۲۰۰۹) نیز بهره گرفته‌ایم. ایده و ویژگی اصلی این نسخهٔ جدید GSC، کاهش نشتی سیگنال در بلوک حذف‌کنندهٔ موجود در شاخه پایین است. این روش به‌دلیل ساختار و روش به کار رفته در آن نیازمند اطلاع نخستین از مکان میکروفون‌ها نیست. برای کارهای بعدی می‌توان نظریهٔ خوشه‌بندی را در ساختارهای پیچیده‌تر GSC نیز (به‌طور مثال، در ساختار ارائه شده در (مارکوویچ گولان و همکاران ۲۰۱۳؛ گنات و همکاران، ۲۰۰۱) بررسی کرد.

همان‌گونه که درقبل هم اشاره شد در شبیه‌سازی و ارزیابی، سناریوی رسم‌شده در شکل (۲) مورد نظر بوده است. در این سناریو، ابعاد اتاق ۷/۲×۳/۵×۴/۴ (بر حسب متر) فرض شده و هشت میکروفون یکسان به‌صورت تمام‌جهته سیگنال را دریافت می‌کنند. نوفهٔ مکانی یک نوفهٔ سفید با میانگین صفر است. به‌عنوان سیگنال منابع گفتار (گویندگان)

نیز از ۱۰ جمله نمونه سیگنال گفتاری از دادگان فارس‌دات (در فرکانس نمونه‌برداری 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[1] ارائه شده) استفاده کرده‌ایم. RT60 محیط (پارامتری برای نشان دادن میزان انعکاسی‌بودن 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، به‌دست 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 کرده‌ایم. بعد از قاب‌بندی و محاسبه میانگین انرژی در حوزهٔ 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 خوشه‌بندی روی آنها صورت گرفته است. در اینجا فرض شده که یک ثانیهٔ نخست سیگنال، سکوت است. این فرض در کلیت مسأله خللی وارد نمی‌کند.

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

همان‌طورکه از نتایج مندرج در جدول (۲) مشخص است، روش پیشنهادی در حالت گوینده سوم، دو خوشه (دوم و سوم) را به‌صورت یک خوشه ادغام کرده است. این مسأله را می‌توان چنین بیان کرد که در حالت گویندهٔ سوم، دو خوشه فاصله بسیار مشابهی نسبت به منبع نوفه و منبع گفتار دارند و بنابراین، روش خوشه‌بندی پیشنهادی نمی‌تواند بین این دو خوشه تفاوت چندانی قائل شود؛ اما در روش (هیماوان و همکاران، ۲۰۱۱) که نوفهٔ محیط پخشنده بوده و روش خوشه‌بندی براساس تابع هم‌دوسی این نوع از نوفه بنا نهاده شده، این مسأله وجود ندارد. با این وجود، توجه به این نکته اهمیت زیادی دارد که روش پیشنهادی در محیط نوفه جهتی پیشنهاد شده و برای بیشترکردن قدرت این روش و به‌وجودنیامدن 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 خوشه‌ها هم برای رتبهٔ گفتاری و هم برای رتبهٔ نوفه‌ای برای سه حالت مکانی منبع گفتار آورده شده است. لازم به ذکر است برای تعیین رتبه نوفه‌ای از انرژی و برای رتبهٔ گفتاری از میانگین شماره قاب آغاز سیگنال (به‌دست آمده از مقایسه سطح انرژی با یک آستانه طبق روش گفته‌شده در بخش‌های قبل) در هر خوشه استفاده شده است.

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

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

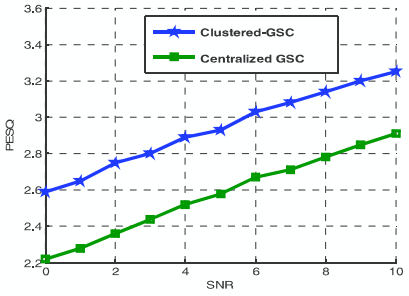
1 http://www.eric-lehmann.com (last accessed: 30 Nov. 2014)

(جدول- ۳): رتبه گفتاری و نوفه‌ای به‌دست‌آمده ازروش پیشنهادی برای سه حالت مکانی گوینده

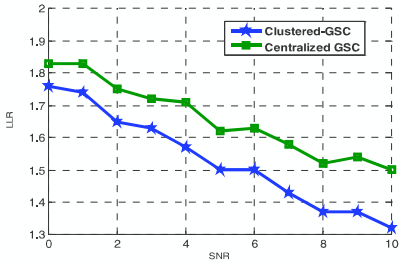
گوینده	خوشه نخست {1,2,3}		خوشه دوم {4,5,6}		خوشه سوم {7,8}	
	رتبه گفتاری	رتبه نوفه‌ای	رتبه گفتاری	رتبه نوفه‌ای	رتبه گفتاری	رتبه نوفه‌ای
اول	۳	۲	۱	۳	۲	۱
دوم	۳	۲	۱	۳	۲	۱
سوم	۱	۲	۲	۳	۳	۱

برای ارزیابی ساختار جدید GSC، عملکرد این روش در شرایط مختلف محیطی مورد بررسی قرار گرفته است. در ساختار پیشنهادی، از خوشه میکروفونی با رتبهٔ گفتاری نخست در شاخه بالایی GSC و از خوشهٔ میکروفونی با رتبهٔ گفتاری آخر در شاخهٔ پایینی GSC استفاده کرده‌ایم. در این ارزیابی، در مقادیر مختلف SNR ورودی، برای حالت گویندهٔ مستقر در مکان نخست، مقادیر PESQ[1] وLLR[2] محاسبه شده‌اند. به‌طور معمول معیار LLR برای سنجش اعوجاج استفاده شده و هر چه این مقدار کمتر باشد، کیفیت سیگنال از لحاظ اعوجاج بهتر است. معیار PESQ نیزکیفیت کلّی سیگنال گفتار را نشان می‌دهد.

شکل‌های (۶) و (۷) نتایج ارزیابی GSC جدید را نشان می‌دهد. در این شکل‌ها، نتایج برای GSC خوشه‌بندی‌شده و GSC متمرکز که در آن از تمامی میکروفون‌ها در دو شاخه بالایی و پایینی استفاده می‌شود، آورده شده است. در این آزمایش‌ها RT60 محیط برابر با ۱۵۰ میلی‌ثانیه فرض شده است. هر دو شکل، عملکرد برتر روش پیشنهادی را در مقایسه با ساختار GSC متمرکز نشان می‌دهد.

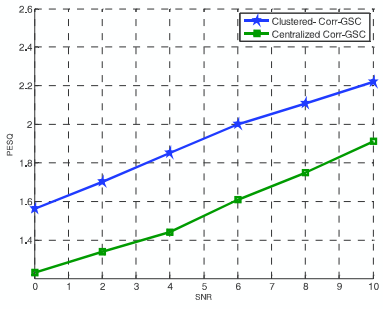


(شکل- ۶): نمودار تغییرات معیار کیفی PESQ بر حسب SNR ورودی برای دو روش GSC متمرکز و خوشه‌بندی پیشنهادشده (برای حالت گوینده نخست)

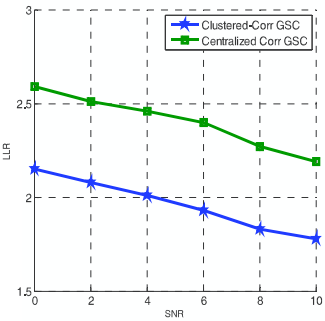


(شکل- ۷): نمودار تغییرات معیار کیفی LLR برحسب SNR ورودی برای دوروش GSC متمرکز و خوشه‌بندی پیشنهادشده (برای حالت گوینده نخست)

همان‌گونه که پیش از این اشاره شد، تأثیر روش خوشه‌بندی پیشنهادی در فرایند بهسازی گفتار، در نسخهٔ جدیدتری از ساختار GSC (یعنی Correlative GSC (تانسند، ۲۰۰۹)) نیز مورد ارزیابی قرار گرفته و نتایج مربوط به عملکرد این روش در دو حالت ۱) متمرکز، و ۲) با خوشه‌بندی پیشنهادی در شکل‌های (۸) و (۹) مقایسه شده است.



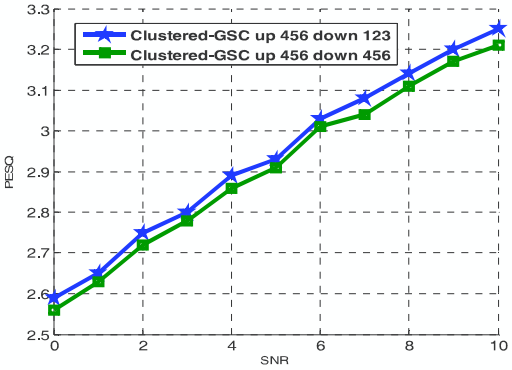
(شکل- ۸): نمودار تغییرات معیار کیفی PESQ بر حسب SNR ورودی برای دو روش Correlative GSC متمرکز و خوشه‌بندی پیشنهادشده (برای حالت گوینده نخست)



(شکل- ۹): نمودار تغییرات معیار کیفی LLR بر حسب SNR ورودی برای دو روش Correlative GSC متمرکز و خوشه‌بندی پیشنهادشده (برای حالت گوینده نخست)

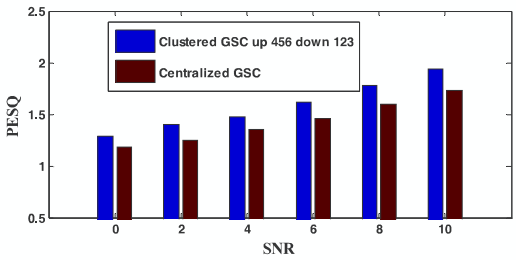
[1] Perceptual Evaluation of Speech Quality
[2] Log Likelihood Ratio

برای بررسی بیشتر اثر استفاده هوشمندانه از خوشه‌های میکروفونی در دو شاخه GSC، در آزمایشی دیگر، عملکرد ساختاری که در هر دو شاخه بالا و پایین آن، از همان خوشه با رتبهٔ گفتاری بهتر استفاده‌شده را بررسی می‌نماییم. شکل (۱۰)، خروجی این ساختار را با خروجی روش پیشنهادی که از خوشه با رتبهٔ گفتاری بالا در شاخهٔ بالا و از خوشه با رتبهٔ گفتاری پایین در شاخهٔ پایین استفاده شده براساس معیار PESQ مقایسه می‌نماید. همان‌طور که در این شکل مشخص است، نتیجه بهتر مربوط به حالتی است که خوشه میکروفونی با رتبهٔ گفتاری کمتر به شاخهٔ پایین GSC داده شود.

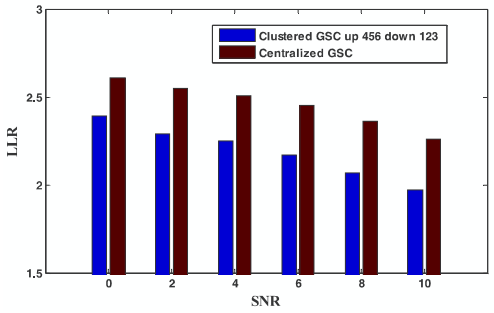


(شکل- ۱۰): مقایسه خروجی دو حالت برای GSC خوشه‌بندی شده. در حالت نخست روش پیشنهادی اعمال شده، ولی در حالت دوم در شاخه پایینی همان خوشهٔ شاخه بالا استفاده شده است.

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 خوشه‌بندی میکروفون‌ها در ساختار GSC، کاهش میزان نشت سیگنال گفتار در شاخه پایین GSC بود. همان‌طورکه در شکل (۱۲) مشخص است، میزان LLR خروجی که به‌نوعی بیان‌گر میزان اعوجاج می‌باشد، نسبت به GSC متمرکز کاهش یافته است.



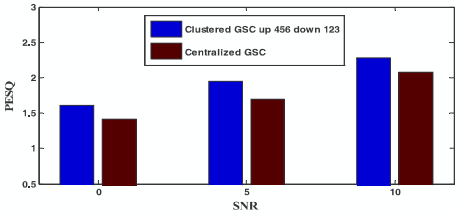
(شکل- ۱۱): نمودار تغییرات معیار کیفی PESQ برحسب SNR ورودی برای دوروش GSC متمرکز و خوشه‌بندی پیشنهادشده (برای گوینده دوم)



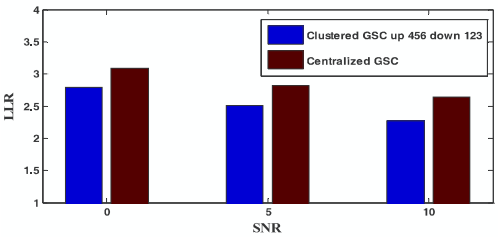
(شکل- ۱۲): نمودار تغییرات معیار کیفی LLR برحسب SNR ورودی برای دو روش GSC متمرکز و خوشه‌بندی پیشنهادشده (برای گوینده دوم)

در ارزیابی‌های بالا، RT60 محیط برابر با ۱۵۰ میلی‌ثانیه 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 میلی‌ثانیه باشد. با این فرض، دیگر نمی‌توان نوفه را به‌عنوان منبع جهتی در نظر گرفت و محیط به سمت میدان نوفه‌ای پخشنده میل می‌کند (هبتز، ۲۰۰۷). در چنین سناریویی و با درنظرگرفتن حالت مکانی گوینده نخست، با فرض اینکه خوشه‌بندی با اعمال روش‌هایی مانند روش ارائه‌شده در (هیماوان و همکاران، ۲۰۱۱) به‌صورت درست انجام شده باشد، روند ساختار GSC پیشنهادی را اعمال و کیفیت خروجی را در شکل‌های (۱۳) و (۱۴) با سامانهٔ GSC متمرکز مقایسه می‌کنیم. در محیط با انعکاس زیاد، ماهیت سیگنال‌های رسیده به میکروفون‌ها می‌توانند بسیار متفاوت باشد. این تفاوت ناشی از متفاوت بودن پاسخ ضربه محیط با توجه به رابطه (۱) است. این تفاوت برای میکروفون‌هایی که فاصلهٔ زیادی نسبت به یکدیگر دارند، بیشتر خواهد بود. با اعمال خوشه‌بندی و اختصاص خوشه‌های متفاوت حاوی

میکروفون‌های نزدیک به هم و درنتیجه میکروفون‌هایی با ماهیت انعکاسی مشابه به دو شاخه بالایی و پایینی GSC، احتمال نشت سیگنال گفتار در شاخهٔ پایین نسبت به ساختار GSC متمرکز کاهش می‌یابد. نتیجهٔ این مزیت در نمودار شکل (۱۳) به‌وضوح مشخص است.



(شکل- ۱۳): نمودار تغییرات معیار کیفی PESQ برحسب SNR ورودی برای دو روش GSC متمرکز و GSC پیشنهادی با فرض دانستن خوشه‌های درست برای گویندهٔ نخست در محیط با RT60=300ms



(شکل- ۱۴): نمودار تغییرات معیار کیفی LLR برحسب SNR ورودی برای دو روش GSC متمرکز و GSC پیشنهادی با فرض دانستن خوشه‌های درست برای گویندهٔ نخست در محیط با RT60=300ms

۵- نتیجه‌گیری

در این مقاله، با کنارگذاشتن فرض اساسی پخشنده بودن میدان نوفه‌ای در مقایسه با روش ارائه شده در (هیماوان و همکاران، ۲۰۱۱)، ابتدا سعی کردیم روش خوشه‌بندی مناسبی برای آرایه‌های میکروفونی Ad-hoc ارائه دهیم. بر همین اساس، یک روش خوشه‌بندی و یک روش رتبه‌بندی خوشه‌ها براساس انرژی و استفاده از چندین منبع پیشنهاد شد. در ادامه، این روش‌ها در چندین سناریو مورد ارزیابی قرار گرفته و نتایج مناسبی به‌دست آمد.

همچنین براساس خوشه‌بندی میکروفون‌ها، ساختار جدیدی برای شکل‌دهنده پرتو وفقی GSC پیشنهاد شد. در این ساختار جدید، به جای این که سیگنال تمام میکروفون‌ها در هر دو شاخه GSC به کار گرفته شود (GSC متمرکز)، در هر شاخه از این شکل‌دهندهٔ پرتو، تنها یک خوشه از میکروفون‌ها به‌کار گرفته می‌شود (GSC خوشه‌بندی شده).

برای تعیین این که کدام خوشه به کدام شاخه تعلق بگیرد، دو روش رتبه‌بندی تعریف شد. در نخستین روش، ملاک رتبه‌بندی میزان نزدیکی خوشه‌ها به گوینده، و در دومین روش، ملاک رتبه‌بندی میزان نزدیکی خوشه‌ها به منبع نوفه تعریف و نتایج حاصله به‌ترتیب رتبهٔ گفتاری و رتبهٔ نوفه‌ای هر خوشه نامیده شد. خوشهٔ با بالاترین رتبهٔ گفتاری به شاخهٔ بالایی GSC، و خوشهٔ با پایین‌ترین رتبهٔ گفتاری و یا در حالت دیگر، خوشهٔ با بالاترین رتبهٔ نوفه‌ای به شاخه پایین GSC داده شد. انگیزه و توجیه نظریهٔ حاضر این بود که در شاخه پایین، هدف ساختن نمونه‌ای مرجع از نوفهٔ محیط است و هرچه سیگنال گفتار تمیز در سیگنال ورودی این شاخه کمتر باشد، نشت کمتر سیگنال گفتار در شاخهٔ پایین را به‌دنبال خواهد داشت.

عملکرد روش پیشنهادی با معیارهای کیفی متنوعی 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 خوشه‌بندی‌شده کیفیت بهتری نسبت به خروجی ساختار GSC متمرکز دارد.

در شکل‌دهندهٔ پرتو GSC خوشه‌بندی‌شده، نحوهٔ انتخاب بهینه خوشه‌ها و در حالت کلی‌تر، میکروفون‌ها برای هر شاخه به‌صورت دقیق کاری سخت است که می‌تواند پیشنهادی برای ادامهٔ کار در این زمینه باشد.

۶- مراجع

Allen, J. B., and Berkley, D. A., 1979, Image method for efficiently simulating small room acoustics, J. Acoust. Soc. Amer., vol. 107, no. 4, pp. 943–950.

Bertrand A., and Moonen, M., 2010a, Distributed adaptive node-specific signal estimation in fully connected sensor networks - part I: sequential node updating, IEEE Trans. on Signal Processing, vol. 58, no. 10, pp. 5277-5291.

Bertrand, A., and Moonen, M., 2010b, Energy-based multi-speaker voice activity detection with an ad hoc microphone array, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), Dallas, Texas, USA, pp. 85-88.

Bertrand, A., Callebaut, J., and Moonen, M., 2010c, Adaptive distributed noise reduction for speech enhancement in wireless acoustic sensor networks, in Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC).

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.

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 دورهٔ کارشناسی را در سال ۱۳۹۰ در رشتهٔ مهندسی برق (مخابرات) در دانشگاه شیراز گذرانده و مدرک کارشناسی ارشد خود را در سال ۱۳۹۲ در رشتهٔ مهندسی برق (مخابرات- سیستم) از دانشگاه یزد اخذ کرد. زمینه‌های علمی مورد علاقهٔ وی پردازش آرایه‌ای و بهسازی گفتار می‌باشد.

نشانی رایانامهٔ ایشان عبارتست از:

seyedhamidyazdani@stu.yazd.ac.ir

حمیدرضا ابوطالبی دورهٔ کارشناسی و کارشناسی ارشد را به‌ترتیب در سال‌های ۱۳۷۴ و ۱۳۷۷ در رشتهٔ 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 رسالهٔ دکترای خویش، به‌مدت یک سال در دورهٔ فرصت مطالعاتی در دانشگاه واترلو کانادا به سر برد. دکتر ابوطالبی در سال ۱۳۸۲ به دانشکدهٔ مهندسی برق و کامپیوتر دانشگاه یزد پیوست و هم‌اکنون به‌عنوان دانشیار این دانشکده مشغول به فعالیت است. وی همچنین در سال ۹۰-۱۳۸۹ یک دوره فرصت مطالعاتی را در مرکز تحقیقاتی Idiap در سوئیس سپری کرد. زمینه‌های علمی مورد علاقهٔ وی پردازش آرایه‌ای سیگنال گفتار، بهسازی گفتار، مکان‌یابی گوینده، فیلترهای وفقی، و آنالیز زمان- فرکانس است.

نشانی رایانامهٔ ایشان عبارتست از:

habutalebi@yazd.ac.ir