

مکان‌یابی منابع چندگانه صوتی در محیط انعکاسی به کمک BSS و استفاده از ویژگی‌های سیگنال گفتار برای رفع ابهام جایگشت عمومی

وحید خان آقا و محمد حسین کهایبی
دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران

چکیده:

در این مقاله یک چارچوب الگوریتمی جدید برای مکان‌یابی منابع چندگانه صوتی در محیط انعکاسی معرفی می‌شود. مبنای مکان‌یابی بر اساس آمیزش چندین تخمین TDOA هر یک از منابع است که این تخمین‌ها با استفاده از الگوریتم جداسازی کور منابع (BSS) در حوزه زمان به دست می‌آیند. به این منظور یک الگوریتم BSS حوزه زمان جدید پیشنهاد شده که نسبت به روش مرجع کیفیت جداسازی و شناسایی کانال را بهبود داده و بار محاسباتی آن نیز در شرایطی کاهش یافته است. سپس برای رفع ابهام جایگشت عمومی که در ذات الگوریتم‌های BSS وجود دارد، پیشنهاد شده که از ویژگی‌های وابسته به گوینده سیگنال گفتار استفاده شود. در برابر معیار همبستگی مورد استفاده در مقاله مرجع، نتایج شبیه‌سازی توانایی خوب این ویژگی‌ها را در رفع ابهام جایگشت نشان می‌دهد.

واژگان کلیدی: تخمین TDOA، بهینه‌سازی PSO، جداسازی کور منابع، رفع ابهام جایگشت عمومی BSS.

۱- مقدمه

مکان‌یابی منابع صوتی، نقش مهمی در بسیاری از سیستم‌های پردازش صوت مثل سیستم‌های مراقبتی، ویدئو کنفرانس‌ها و رابط‌های انسان و ماشین دارد. مکان‌یابی یک منبع صوتی در محیط بدون انعکاس، به سادگی انجام پذیر است. اما در محیط واقعی پهنای باند زیاد سیگنال گفتار، تعدد منابع و تعدد پرتوهای انعکاسی ناشی از هر منبع، مکان‌یابی را بسیار دشوار می‌سازد.

یکی از محبوب‌ترین روش‌های مکان‌یابی غیرفعال استفاده از تخمین تفاضل زمان رسیدن (TDOA) سیگنال هدف به عناصر آرایه حس‌گری می‌باشد. در این روش در دو مرحله ابتدا چندین تخمین TDOA با استفاده از چند آرایه سنسوری مستقل استخراج می‌شود، سپس این تخمین‌ها در فضای دوبعدی یا سه‌بعدی به صورت هندسی ترکیب می‌شوند تا موقعیت فضایی منبع معلوم شود. اگرچه تخمین TDOA مربوط به یک منبع در فضای آزاد بدون مشکل انجام می‌شود (Di Claudio, et al., 1999) ولی در حضور منابع چندگانه و یا در محیط‌های انعکاسی، روش‌های سنتی تخمین TDOA

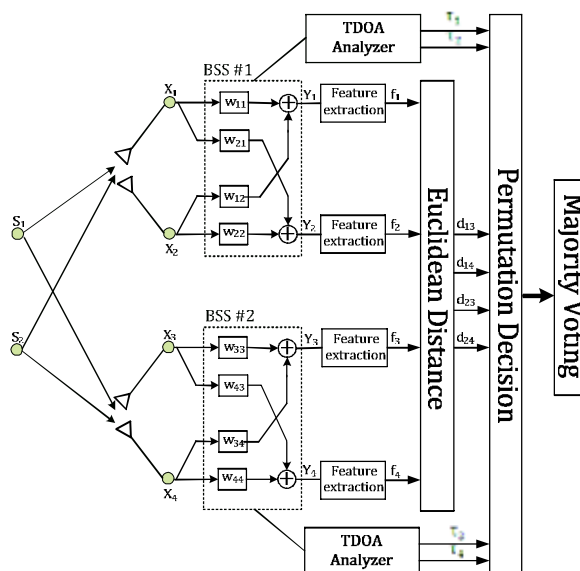
کارآمدی لازم را نداشته‌اند (Chen, et al., 2005). برای رفع این مشکل دسته‌ای از محققان با معرفی رابطه میان مسئله جداسازی کور منابع (BSS) با مسئله شناسایی کور سیستم چند - ورودی - چند - خروجی (MIMO) از الگوریتم‌های جداسازی منابع در محیط انعکاسی برای استخراج تخمین TDOA استفاده کرده‌اند (Buchner, et al., 2005. Buchner, et al., 2006) که جزئیات آن در بخش ۲ تشریح خواهد شد. به این ترتیب تخمین TDOA مقاومی به عنوان محصول جانبی الگوریتم BSS به دست خواهد آمد که مکان‌یابی نهایی با ترکیب چندین تخمین TDOA میسر خواهد شد.

(شکل ۱) بلوک دیاگرام کلی مکان‌یابی به کمک BSS را برای دو منبع نشان می‌دهد که در آن با اجرای BSS روی دو آرایه‌ی میکروفونی دوتایی، به ازای هر آرایه، دو تخمین TDOA و دو سیگنال جدا شده به دست می‌آید (از این پس این برای ایجاد تمایز میان دو الگوریتم BSS که روی دو آرایه‌ی متفاوت اجرا می‌شوند، از نمادگذاری BSS1 و BSS2 استفاده می‌کنیم).

BSS2 متناظر هستند (اگرچه که سیستم شنوایی انسان غلبه منبع صوتی مشابهی را در آنها ادراک کند).

در این مقاله پیشنهاد شده تا از ویژگی‌های وابسته به گوینده سیگنال گفتار برای رفع ابهام جایگشت عمومی استفاده شود. به این صورت که تخمین‌های TDOA ای به منبع یکسانی نسبت داده شوند که سیگنال جداشده متناظرشان بردارهای ویژگی مشابهی داشته باشند. نتایج شبیه‌سازی نشان می‌دهد که در شرایط واقعی برخلاف معیار همبستگی که به طور کامل ناتوان از رفع ابهام جایگشت می‌باشد، ویژگی‌های مورد استفاده به خوبی از پس این مشکل برمی‌آیند.

هم‌چنین یک الگوریتم BSS حوزه زمان ابتکاری معرفی شده که با تمرکز بر شناسایی کانال، با استخراج موقعیت نسبی همبستگی‌های فضایی در سیستم جداساز ایده‌آل را بازسازی می‌کند. این الگوریتم بر تحلیل پیش‌گویی خطی و الگوریتم بهینه‌سازی هوشمند PSO¹ استوار است و به میزان قابل توجهی کیفیت شناسایی کانال و جداسازی را بهبود می‌دهد. در این مقاله ابتدا در بخش دو به چگونگی انجام تخمین TDOA به کمک BSS می‌پردازیم. سپس جزئیات روش جداسازی پیشنهادی در بخش سه بیان می‌شود و منطق تصمیم‌گیری پیشنهادی برای رفع ابهام جایگشت در بخش چهارم تشریح خواهد شد. در نهایت پس از ارائه نتایج شبیه‌سازی در بخش پنج، جمع‌بندی پایانی در بخش ششم ارائه خواهد شد.



(شکل ۱) نمایش ساختار فیلترهای مخلوط کننده و جداساز. پاسخ ضربه میان منابع صوتی تا میکروفون‌ها به صورت h_{ij} و پاسخ ضربه مربوط به فیلترهای جداساز با w'_{ij} مشخص شده‌اند.

به دلیل ابهام جایگشت عمومی ذاتی الگوریتم‌های BSS، فعلاً تعیین این که کدام تخمین TDOA حاصل از BSS1 مربوط به کدام تخمین TDOA حاصل از BSS2 می‌باشد، ناممکن است. تنها داشته ما برای حل این ابهام، خروجی‌های جداشده Y_i می‌باشند؛ در واقع هر تخمین τ_i متعلق به منبع صوتی‌ای می‌باشد که در Y_i به دست آمده است. بنابراین اگر خروجی Y_i از BSS1 و خروجی Y_j از BSS2 متعلق به منبع گفتار مشابهی باشند، τ_i و τ_j هم مربوط به همان منبع خواهند بود و می‌توانند برای استخراج محل آن منبع در فضای دوبعدی مورد استفاده قرار بگیرند.

مشکل آن جاست که الگوریتم‌های جداسازی حوزه زمان، توانایی جداسازی ایده‌آل را ندارند. بلکه در عمل اگرچه در هر خروجی صدای غالب، قابل تشخیص خواهد بود، ولی به طور عمومی صدای منبع دیگر هم (هرچند ضعیف‌تر) قابل تشخیص خواهد بود. حتی ممکن است در عین عملکرد خوب سیستم جداساز، به دلیل تفاوت شدت بیان گفتار توسط اشخاص مختلف، یکی از منابع در هر دو خروجی غالب باشد. نتایج شبیه‌سازی نشان می‌دهند که تحلیل‌های ریاضی ابتدایی، مانند تابع همبستگی مورد استفاده در (Knapp and Carter, 1976)، توانایی رفع ابهام جایگشت را در شرایط واقعی ندارند؛ به این معنا که تابع همبستگی نمی‌تواند مشخص کند که کدام دو خروجی از BSS1 و

۲- تخمین TDOA به کمک BSS حوزه زمان

مشهورترین روش تخمین TDOA استفاده از تابع همبستگی متقابل تعمیم‌یافته یا Generalized Cross Correlation (GCC) می‌باشد (Knapp and Carter, 1976). در این روش از موقعیت قله‌های تبدیل فوریه معکوس چگالی طیفی توان متقابل سیگنال‌های دریافتی روی حس‌گرها، برای استخراج تخمین TDOA استفاده می‌شود. در حضور منابع چندگانه و یا در محیط دارای پژواک تابع GCC قله‌های چندگانه‌ای پیدا می‌کند. چنانچه در محیط انعکاسی واقعی به صورت هم‌زمان چند منبع حضور داشته باشند، تعدد قله‌ها در عمل تخمین TDOA را غیر ممکن می‌سازد (Chen, et al., 2005).

¹ Particle Swarm Optimization

با جاگذاری (۱) و (۲) در (۳) و (۴) تخمین TDOA با استفاده از پاسخ ضربه‌های سیستم جداساز ایده‌آل، قابل محاسبه خواهد بود:

$$\tau_1 = \operatorname{argmax}_n |w_{22}(n)| - \operatorname{argmax}_n |w_{21}(n)| \quad (5)$$

$$\tau_2 = \operatorname{argmax}_n |w_{11}(n)| - \operatorname{argmax}_n |w_{12}(n)| \quad (6)$$

بنابراین با فرض برقراری روابط (۱) و (۲) در نتیجه هم‌گرایی الگوریتم BSS حوزه زمان، می‌توان از فیلترهای سیستم جداساز برای انجام تخمین TDOA و درنهایت مکان‌یابی منابع چندگانه صوتی در محیط انعکاسی استفاده کرد (Buchner, et al., 2006). بنیان الگوریتم BSS پیشنهادی نیز بر تمرکز مستقیم بر برقراری روابط فوق استوار است که جزئیات آن در ادامه تشریح خواهد شد.

۳- آموزش انتخابی فیلترهای جداساز

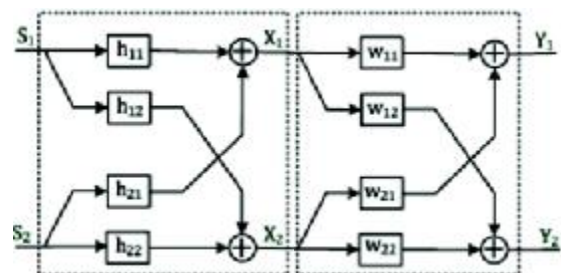
در این بخش به تشریح جزئیات روش BSS حوزه زمان پیشنهادی می‌پردازیم. در روش پیشنهادی در دوگام، ابتدا زیر مجموعه‌ای از ضرایب فیلترهای جداساز شکل ۱ که متناظر با بیشترین همبستگی‌های فضایی هستند انتخاب می‌شوند و در گام دوم با استفاده از الگوریتم PSO مقدار بهینه این ضرایب معین می‌شود.

۳-۱- شناسایی ساختار همبستگی‌های فضایی

منظور از همبستگی فضایی، همبستگی میان نمونه‌های سیگنال ضبط شده توسط یک میکروفون با نمونه‌های سیگنال میکروفون دیگر می‌باشد که می‌توان آن را به صورت $R_{x_1 x_2}(n, m) = \langle x_1(n) x_2(n - m) \rangle$ بیان نمود. با در نظر داشتن (۱) و (۲) و نیز با توجه به این حقیقت که فیلترهای سیستم مخلوط‌کننده، عامل اصلی همبستگی‌های فضایی هستند، می‌توان با شناسایی ساختار این همبستگی‌های فضایی به فیلترهای جداساز ایده‌آل روابط (۱) و (۲) دست یافت. در واقع می‌توان انتظار داشت که موقعیت‌های بیشینه‌گی وابستگی‌های فضایی با موقعیت‌های بیشینه‌گی فیلترهای مخلوط‌کننده و در نتیجه فیلترهای جداساز متناظر باشند.

برای شناسایی موقعیت‌های بیشینه‌گی همبستگی‌های فضایی، ماتریس همبستگی بلوکی مخلوط‌ها (سیگنال‌های

به‌طور اصولی مدل تحلیلی این روش و سایر روش‌های سنتی مبتنی بر مدل انتشار امواج آزاد توانایی دریافت ماهیت چندمسیرگی و چندگانگی منابع را ندارد (Chen, et al., 2005). برای رفع دو مشکل فوق، دسته‌ای از محققان روش به‌طور کامل متفاوتی را با تمرکز مستقیم بر استخراج پاسخ ضربه میان منابع و حس‌گرها در پیش گرفتند (Buchner, et al., 2005) برای توضیح این روش، (شکل ۲) را در نظر بگیرید که دیگرام ساختاری سیستم مخلوط‌کننده را همراه با سیستم جداساز نشان می‌دهد. تمام فیلترهای این مدل تحلیلی FIR در نظر گرفته شده‌اند.



(شکل ۲) نمایش ساختار فیلترهای مخلوط‌کننده و جداساز. پاسخ ضربه میان منابع صوتی تا میکروفون‌ها به صورت h_{ij} و پاسخ ضربه مربوط به فیلترهای جداساز با w_{ij} مشخص شده‌اند.

در نقطه بهینه جداسازی انتظار داریم که در هر خروجی Y_i اثر تنها یکی از منابع اصلی S_i موجود باشد. به‌عنوان مثال فرض کنید بخواهیم اثر منبع صدای S_1 را در خروجی Y_2 حذف کنیم. در (شکل ۲) مشاهده می‌شود که این منبع از دو مسیر $h_{11} \otimes w_{12}$ و $h_{12} \otimes w_{22}$ به خروجی Y_2 می‌رسد (علامت \otimes نشان‌دهنده کانولوشن خطی است). برای حذف اثر S_1 کافی است این دو کانال قرینه یکدیگر باشند. از آن‌جا که در حالت کلی h_{11} و h_{12} مستقل از یکدیگر و متفاوت هستند، تنها راه حل کلی برای چنین قرینه‌گی عبارتست از (Buchner, et al., 2005):

$$w_{12} = -\alpha_1 h_{12} \quad w_{22} = \alpha_1 h_{11} \quad (1)$$

به روش مشابه برای حذف اثر S_2 را در خروجی Y_1 خواهیم داشت:

$$w_{11} = \alpha_2 h_{22} \quad w_{21} = -\alpha_2 h_{21} \quad (2)$$

از طرفی مطابق تعریف، تفاضل زمان رسیدن (TDOA) هر یک از منابع به دو میکروفون به‌صورت زیر محاسبه می‌شود:

$$\tau_1 = \operatorname{argmax}_n |h_{11}(n)| - \operatorname{argmax}_n |h_{12}(n)| \quad (3)$$

$$\tau_2 = \operatorname{argmax}_n |h_{22}(n)| - \operatorname{argmax}_n |h_{21}(n)| \quad (4)$$

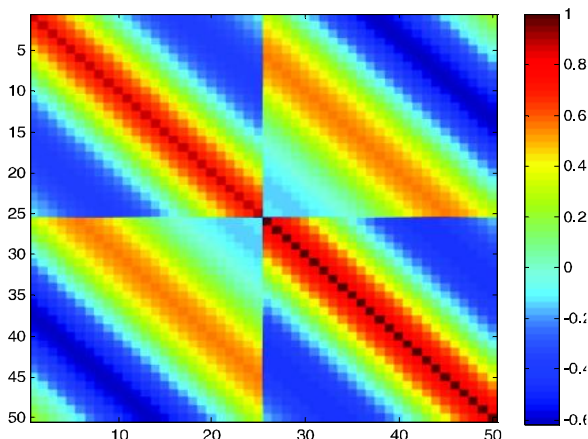
میکروفون‌ها) تشکیل داده می‌شود. این ماتریس همبستگی با استفاده از فرمولاسیون ماتریسی بلوکی با ساختار ویژه سیلوستری که در (Buchner, et al., 2006) معرفی شده تشکیل می‌شود.

با توجه به ساختار ویژه ماتریس همبستگی می‌توان به صورت شهودی به اطلاعات خوبی در مورد سیستم مخلوط‌کننده دست یافت. برای توضیح بهتر این موضوع، ماتریس همبستگی سیگنال‌های مخلوط را در حالتی که سیستم مخلوط‌کننده در ساده‌ترین شکل خود تنها با یک تضعیف و تأخیر به حس‌گرها رسیده‌اند (یعنی حالتی که فقط یکی از ضرایب فیلترهای FIR مخلوط‌کننده h_{ij} غیرصفر می‌باشد) در (شکل ۳) آورده‌ایم. در این شکل، چهار ناحیه بلوکی قابل تمیز هستند؛ دو بلوک قطری، خودبستگی هر یک از مخلوط‌های مشاهده شده در میکروفون‌ها و بلوک‌های غیرقطری، همبستگی متقابل مخلوط‌ها را نشان می‌دهند. همچنین هر یک از خطوط اریب مشاهده شده در این شکل، نتیجه ساختار سیلوستر مورد استفاده در (Kokkinakis and Nandi, 2006) بوده و هر کدام متناظر با یکی از ضرایب فیلترهای FIR است. لازم به توضیح است که تنها نیمه پایین مثلثی هر بلوک در رابطه با فیلتر متناظر، قابل مطالعه است.

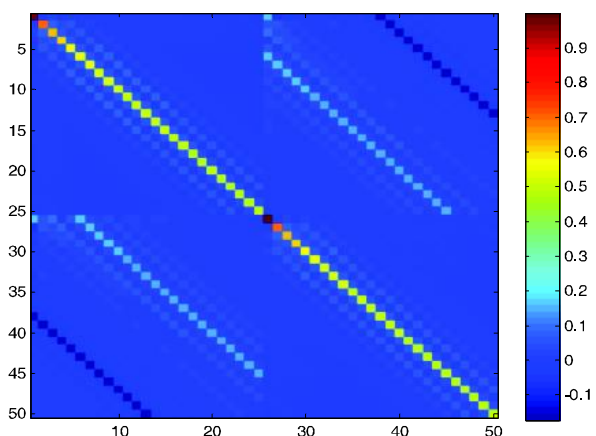
در بلوک‌های غیرقطری (شکل ۳) دو ناحیه از همبستگی به صورت خطوط مورب دارای شدت بیشتر، به رنگ‌های آبی پررنگ و نارنجی قابل تمیز هستند. در عمل مرکز این خطوط مورب، به طور دقیق متناظر با ضریب غیر صفر فیلتر مخلوط می‌باشد؛ ولی مشکل این‌جاست که به خاطر پهن بودن این خطوط، تعیین دقیق تأخیر متناظر با هر خط امکان‌پذیر نیست. به‌ویژه در مخلوط واقعی در محیط انعکاسی با افزایش تعداد ضرایب غیر صفر فیلترهای مخلوط‌کننده، تعداد این خطوط بیشتر و بیشتر می‌شود و تداخل دامنه آن‌ها تعیین موقعیت واقعی خطوط را ناممکن می‌سازد.

دلیل پهن بودن این خطوط، ساختار زمانی سیگنال گفتار است که در آن هر نمونه سیگنال گفتار، همبستگی زیادی با نمونه‌های قبلی و بعدی خود دارد. برای رفع این مشکل از یک طبقه پیش‌پردازش برای از بین بردن این ساختار زمانی استفاده می‌کنیم. در عمل می‌توان با استفاده از مانده‌های فیلتر پیش‌بینی خطی سیگنال گفتار، نسخه‌ای از سیگنال‌های مخلوط را به دست آورد که خودبستگی‌های زمانی‌شان را از دست داده‌اند (سفید شده‌اند) ولی

همبستگی‌های متقابل فضایی‌شان را حفظ کرده‌اند. این کار به‌شرطی امکان‌پذیر است که سفیدسازی در بازه‌های زمانی کوچک‌تر از پنج میلی‌ثانیه انجام شود تا وابستگی‌های فضایی مخلوط‌ها از بین نرود (Jan and Flanagan, 1998). (شکل ۴) نتیجه اعمال این طبقه پیش‌پردازش را نشان می‌دهد.



(شکل ۳) نمایش ماتریس همبستگی یک مخلوط ساده. خطوط اریب متناظر با همبستگی‌های فضایی در بلوک‌های غیرقطری قابل مشاهده، اما غیرمتمم‌گرااند.



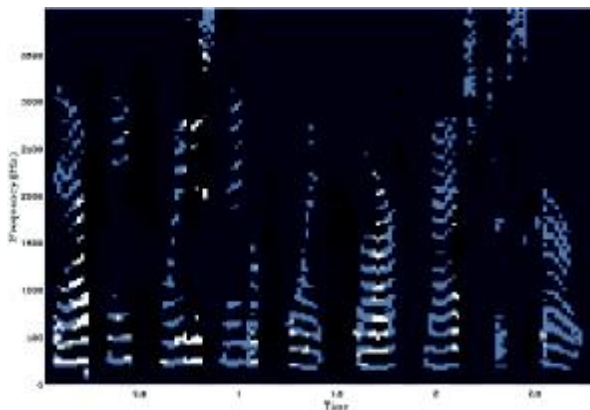
(شکل ۴) نمایش ماتریس همبستگی مانده‌های تحلیل پیش‌گویی خطی یک مخلوط ساده. خطوط اریب متناظر با ضرایب غیرصفر فیلترهای مخلوط‌کننده، به‌طور کامل قابل تمیز هستند.

مشاهده می‌شود که این بار در بلوک‌های غیرقطری دو خط باریک به‌طور کامل متمایز به‌وجود آمده که به‌طور دقیق متناظر با ضرایب غیرصفر فیلتر مخلوط‌کننده هستند. البته همچنان نقاط و خطوط اضافی با شدت کم‌تر در این ماتریس وجود دارند که با انجام عملیات آستانه‌ای ستونی و اعمال فیلتر میانه‌ای (median) دوبعدی در امتداد خطوط اریب در ناحیه پایین مثلثی هر بلوک، قابل حذف هستند. (شکل ۵) نتیجه نهایی را نشان می‌دهد که با ضرب درایه به درایه در ماتریس همبستگی اولیه، میزان شدت اولیه هر خط بازبایی

هم‌گرا می‌شود. بنابراین مهم‌ترین موضوع در پیاده‌سازی PSO اختیار کردن یک تابع شایستگی خوش‌رفتار است که یک بیشینه (یا کمینه) عمومی در نقطهٔ بهینهٔ جداسازی داشته باشد. در عین حال باید بار محاسباتی این تابع به قدر کافی پایین باشد تا در هر مرتبه تکرار الگوریتم در زمان کوتاهی برای همهٔ N_p ذره قابل محاسبه‌پذیر باشد.

در این مقاله با استفاده از خاصیت ناهم‌پوشانی نمایش زمان-فرکانس (TF) سیگنال‌های گفتار گویندگان متفاوت، تابع شایستگی مشخص‌کنندهٔ میزان کمی جداسازی را تعریف می‌کنیم. دلیل این ناهم‌پوشانی، تُنک بودن سیگنال گفتار در فضای تبدیل TF است (O'Grady, et al., 2005). منظور از تنک بودن آن است که در یک فضای تبدیل، سیگنال بیش از آن‌چه از واریانس آن انتظار می‌رود، مقادیر نزدیک به صفر اختیار کند (O'Grady, et al., 2005).

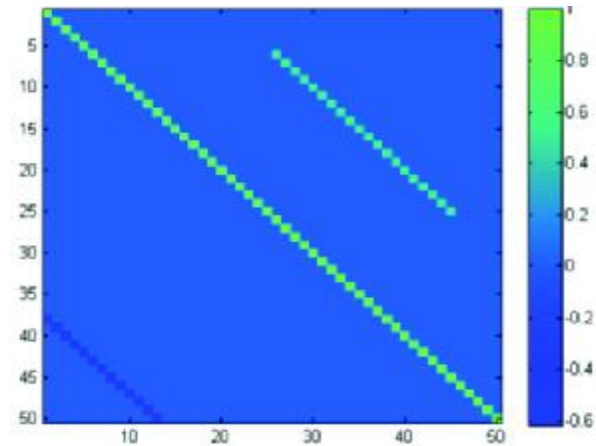
(شکل ۶) نمایش طیف TF دو سیگنال تمیز گفتار را نشان می‌دهد و (شکل ۷) نمایش TF را برای مخلوط‌های همین سیگنال‌ها نشان می‌دهد. مشاهده می‌شود که تعداد نقاط هم‌پوشانی دو سیگنال تمیز (شکل ۶) بسیار اندک می‌باشد ولی در دو سیگنال مخلوط (شکل ۷) مقدار هم‌پوشانی بسیار زیاد شده و به‌طور تقریبی هم‌پوشانی به تمامی انجام شده است. بنابراین به‌نظر می‌رسد کمینه‌کردن این هم‌پوشانی (تعداد نقاط سفید) به‌عنوان تابع شایستگی، معیار خوبی از میزان جداسازی حاصل شده باشد.



(شکل ۶) نمایش TF توأم دو سیگنال تمیز گفتار. نقاط سیاه نقاط شدت منبع شماره ۱، نقاط طوسی روشن نقاط شدت منبع شماره ۲ و نقاط سفید نقاطی را نشان می‌دهند که هر دو منبع با شدت زیادی حاضر هستند. مشاهده می‌شود که تعداد نقاط هم‌پوشانی بسیار اندک می‌باشد.

مشکل این معیار آن است که ممکن است آموزش ضرایب به سمتی هدایت شود که فیلترهای جداساز درگیر در محاسبهٔ هر خروجی در (شکل ۲)، بازه‌های فرکانسی

شده است. به این ترتیب می‌توان با مشاهدهٔ موقعیت خطوط اریب نشان‌دهندهٔ همبستگی فضایی، موقعیت‌های احتمالی بیشینگی فیلترهای جداساز را معلوم نمود. پس از شناسایی این ضرایب، از الگوریتم PSO برای تعیین بهترین مقدار ممکن آن‌ها استفاده می‌شود.



(شکل ۵) نمایش ماتریس همبستگی بعد از انجام عملیات آستانه‌ای ستونی و اعمال فیلتر دوبعدی median در ناحیهٔ پایین‌مثلثی هر بلوک.

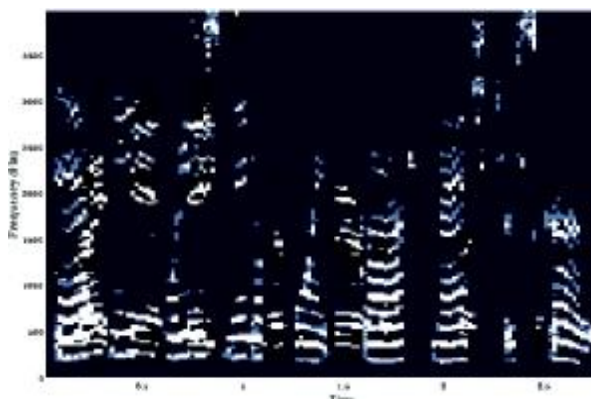
۳-۱- الگوریتم یادگیری PSO

PSO یک الگوریتم تکاملی آماری برای حل مسئله در فضای پیچیدهٔ پارامتری چندبعدی است که ریشه در مطالعات زیستی روی حرکت جمعی پرندگان دارد. شیوهٔ کار به این صورت است که یک تابع هزینه (شایستگی) بر روی یک فضای پارامتری چندبعدی تعریف می‌شود. جمعیتی از N_p ذره. به‌صورت اتفاقی در فضای N_w بعدی به‌صورت $p_i = [p_{iq}, \dots, p_{iN_w}]$ مقداردهی اولیه می‌شوند. این موقعیت هر ذره را در فضای پارامتری مشخص می‌کند. این ذرات به‌کمک هوش جمعی در فضای پارامتری حرکت داده می‌شوند (جستجو می‌کنند) تا نقطهٔ بهینه‌گی تابع شایستگی پیدا شود (Schutte and Groenwold, 2005). در طول جستجو، به کمک تابع شایستگی میزان شایستگی موقعیت فعلی هر ذره سنجیده می‌شود و سابقهٔ بهترین موقعیتی که هر ذره تجربه کرده $(pbest_i, i = 1, \dots, N_p)$ و بهترین موقعیتی که تمامی ذرات تجربه کرده‌اند $(gbest)$ نگهداری می‌شود و با استفاده از این سوابق، جمعیت ذرات به سوی نقطهٔ بهینه‌گی تابع هزینه هدایت می‌شوند (Das and Abraham, 2006).

در یک فضای پارامتری چندبعدی یک تابع شایستگی خوش‌رفتار، این الگوریتم به‌طور معمول با تعداد اندکی تکرار

تمایزی را عبور دهند. به عنوان مثال ممکن است Y_1 به یک سیگنال بالاگذر و سیگنال Y_2 به یک سیگنال پایین گذر تبدیل شده باشند و به این صورت شرط ناهمپوشانی طیف زمان فرکانس برقرار شده باشد؛ درحالی که جداسازی واقعی صورت نگرفته باشد. چون همان طور که در (شکل ۶) دیده می شود، دو سیگنال متمایز گفتار در اکثر بازه های فرکانسی به صورت هم زمان حضور دارند. در نتیجه انتخاب گری فرکانسی فیلترهای جداساز ممکن است باعث شود که بدون حصول جداسازی واقعی، هم پوشانی خروجی های Y_1 و Y_2 کمینه شود.

برای رفع این مشکل، تابع شایستگی به این صورت اصلاح می شود که به جای معیار کمینه شدن هم پوشانی، معیار معادل بیشینه شدن ناهمپوشانی در پنجره های زمان-فرکانس کوچک تر، و تنها در پنجره هایی که هر دو سیگنال خروجی در بیش از نیمی از نقاط پنجره، شدتی بیشتر از سطح نویز، محاسبه شود. یعنی هر پنجره در صورتی در محاسبه تابع شایستگی کل وارد شود که هر دو منبع خروجی در آن پنجره حاضر باشند و حداقل هم پوشانی را نیز داشته باشند. سطح نویز نمایش TF با توجه به نرمالیزاسیون اندازه سیگنال های میکروفون ها برابر -35dB در نظر گرفته شده است.



(شکل ۷) نمایش TF توأم دو سیگنال مخلوط. به طور تقریبی تعداد نقاط هم پوشانی (سفید) به شدت افزایش یافته و تعداد نقاطی که تنها یکی از منابع شدت دارند (سیاه یا طوسی روشن) به شدت کاهش یافته است.

در ابتدا به دلیل هم پوشانی به طور تقریبی کامل توزیع TF دو مخلوط، تعداد پنجره های معتبر که هر دو منبع در آن حاضرند، زیاد است درحالی که مقدار ناهم پوشانی آن ها بسیار کم است. به تدریج الگوریتم PSO مقدار ناهم پوشانی ها را بیشتر می کند و در عین حال ممکن است تعدادی از پنجره ها که اتفاقاً فقط یک مخلوط در آن ها حاضر بوده، کاهش یابد. به هر حال الگوریتم به بیشینه کردن ناهم پوشانی در باقی مانده پنجره ها ادامه می دهد.

طول پنجره انتخاب شده در محور فرکانس بسته به تفکیک پذیری فرکانسی فیلترهای جداساز انجام می شود. هر چه طول پاسخ ضربه فیلترهای جداساز بزرگ تر انتخاب شود، انتخاب گری فرکانسی آن ها بیشتر می شود. اگر این انتخاب گری کم تر از طول فرکانسی پنجره باشد، احتمال ضعیفی وجود دارد که مشکل جداسازی فرکانسی به جای جداسازی واقعی در هر پنجره تکرار شود. اگر چه شبیه سازی ها انتخاب طول فرکانسی پنجره را چندان تأثیرگذار نشان ندادند. طول زمانی پنجره، باید به اندازه کافی بزرگ انتخاب شود تا تعداد پنجره های محتوی فعالیت هم زمان هر دو منبع گفتار زیاد شود. زیرا تابع شایستگی تنها در پنجره هایی حساب می شود که هر دو منبع فعال باشند. بنابراین حداقل طول انتخاب 0.5 ثانیه (مدت زمان ادای یک کلمه کوتاه) انتخاب می شود.

۴- رفع ابهام جایگشت به کمک ویژگی های سیگنال گفتار

در این مقاله از ویژگی های ویژه گوینده سیگنال گفتار استفاده می شود. این ویژگی ها با توجه به مدل ریاضی تولید سیگنال گفتار در دو دسته ویژگی های وابسته به منبع تحریک تارهای صوتی و ویژگی های وابسته به فیلتر لوله صوتی و لب ها در روش های شناسایی هویت گوینده بررسی شده اند و به طور عمومی با تقریب سیستم شنوایی انسان، افزونگی سیگنال حوزه زمان را به بردارهای فشرده محتوی اطلاعات مربوط به گوینده کاهش می دهند و در نتیجه تقریب خوبی از ادراک شنوایی انسان هستند.

۴-۱- ویژگی های وابسته به منبع

ویژگی های وابسته به منبع، ویژگی هایی هستند که با مدل سازی موج صدای اولیه که وارد حنجره می شود، ارتباط دارند و بنابراین بیشتر با مشخصات تارهای صوتی ارتباط دارند. جدا از ویژگی های وابسته به زبان شناسی که اندازه گیری آن ها دشوار و نیازمند داده های آموزشی طولانی می باشد، مشهورترین ویژگی وابسته به منبع، فرکانس بنیادی سیگنال گفتار (F_0) می باشد. در (Wildermoth, 2001) نشان داده شده که این ویژگی تنها در حروف صدادار (بازه های پررنگ بودن سیگنال گفتار) قابل مشاهده است و به علاوه تمایز معناداری برای اشخاص موجود در طبقه جنسیتی مشابه (به خصوص برای زنان) ایجاد نمی کند. بنابراین برای عمل بازشناسی گوینده، کاربرد چندانی ندارد. البته می توان به جای محاسبه این فرکانس

موفق‌ترین بردارهای ویژگی در سیستم‌های شناسایی گوینده بوده‌اند (Buchner, et al., 2006). تفاوت این دو در نوع تقریبات آن‌ها از سیستم شنوایی انسان می‌باشد. این دو اگرچه در ذیل ویژگی‌های وابسته به سیستم دسته‌بندی می‌شوند، ولی در واقع علاوه بر اطلاعات فیلتر لوله صوتی، حامل اطلاعات دینامیکی منبع نیز هستند.

۴-۳- منطق تصمیم‌گیری

در کاربرد ویژگی‌های سیگنال گفتار برای رفع ابهام جایگشت و تعیین خروجی‌های متناظر BSS1 و BSS2، داده آموزشی در اختیار نداریم. بنابراین نمی‌توانیم مانند روش‌های بازشناسی گوینده از توسعه مدل‌های آماری برای شناسایی استفاده کنیم. در عوض به علت برقراری هم‌زمانی میان خروجی‌های دو واحد الگوریتم BSS نشان داده‌شده در (شکل ۱)، می‌توانیم به سادگی از اطلاعات مربوط به تغییرات دینامیکی سیگنال گفتار استفاده کنیم. به این دلیل، در این پروژه بردار ویژگی‌های سیگنال گفتار، فریم به فریم محاسبه می‌شود و خروجی‌هایی از دو الگوریتم BSS را متناظر در نظر می‌گیریم که بردارهای ویژگی آن‌ها کم‌ترین فاصله اقلیدسی را داشته باشند. به این شکل، فریم‌به‌فریم تصمیم‌گیری انجام می‌شود. پس از انجام این تصمیم‌گیری برای تمامی فریم‌های یک سیگنال، از معیار رأی اکثریت برای اجماع استفاده می‌شود و جایگشتی انتخاب می‌شود که رأی اکثریت را به دست آورده باشد. برای این منظور دو متغیر تصمیم‌گیری به صورت زیر تعریف می‌شوند:

$$\text{decision_variable}_1 = d_{13} + d_{24} \quad (7)$$

$$\text{decision_variable}_2 = d_{14} + d_{23} \quad (8)$$

که d_{ij} نشان‌دهنده فاصله اقلیدسی بین بردار ویژگی خروجی‌های Y_1 و Y_j در شکل ۱ می‌باشد. تصمیم‌گیری برای رفع ابهام جایگشت برای هر فریم به صورت زیر انجام می‌شود:

اگر $\text{decision_variable}_1 < \text{decision_variable}_2$

زوج تخمین‌های (τ_1, τ_3) و (τ_2, τ_4) برای مکان‌یابی استفاده شوند.

اگر $\text{decision_variable}_1 > \text{decision_variable}_2$

زوج تخمین‌های (τ_1, τ_4) و (τ_2, τ_3) برای مکان‌یابی استفاده شوند.

به صورت میانگین تخمین به دست آمده از تعداد زیادی از فریم‌های سیگنال، منحنی تغییرات این تخمین در طول زمان بیان یک جمله را به عنوان یک ویژگی متمایز کننده در نظر گرفت. در (Weddin, 2005) نشان داده شده که حتی در بیان جمله یکسان توسط گوینده‌های متفاوت، منحنی تغییرات فرکانس بنیادی چه از نظر مقدار و چه از نظر مدت زمان ادای جملات متفاوت است.

با در نظر گرفتن مشکلات فوق، برای در نظر گرفتن این منحنی تغییرات، در این مقاله با استفاده از روش پیش‌نهادشده در (Deshmukh, et al., 2005) از معیار به عنوان "معیار مجموع" استفاده می‌شود که معیاری است کمی که تمامی قطعات سیگنال را از نظر پریودی بودن، میزان انرژی پریودیک، غیرپریودیک بودن و میزان انرژی غیرپریودیک، دسته‌بندی می‌کند. این معیار تابع معروف تابع تفاضل اندازه متوسط (AMDF) را در زیرباندهای شنیداری سیگنال گفتار محاسبه کرده و با مطالعه توزیع چاله‌های AMDF در روی محور زمان معیاری کمی در مورد میزان تناوبی بودن هر فریم ارائه می‌کند. جزییات نحوه محاسبه در (Deshmukh, et al., 2005) بیان شده است.

نکته مهم آن است که برای استفاده از معیار مجموع در رفع ابهام جایگشت، تمایز بین گفتارها به دلیل وقوع دو رخداد اتفاق می‌افتد: اول تمایز میان فرکانس‌های بنیادی گویندگان هنگام وقوع هم‌زمان فریم‌های صدادار این گویندگان و دومی هم‌زمان شدن یک فریم پریودیک از یک گوینده با فریم غیرپریودیک از گوینده دیگر. بنابراین میزان موفقیت در تمیزگذاری بین دو گوینده، با تعدد دفعات وقوع دو رخداد فوق متناسب است. در مخلوط گفتار دو گوینده، احتمال وقوع یکی از حالات فوق بسیار زیاد است. اما وقتی تعداد گوینده‌ها بیشتر شود انتظار داریم که احتمال نقض این دو حالت زیادتر شود.

۴-۲- ویژگی‌های وابسته به فیلتر

ویژگی‌های وابسته به فیلتر، ویژگی‌هایی هستند که اطلاعات مربوط به فیلتر محفظه صوتی را شامل می‌شوند. سازه‌ها به عنوان قلّه‌های پوش طیفی سیگنال گفتار، اطلاعاتی در مورد شکل و اندازه لوله صوتی را حمل می‌کنند و تنها به مشخصات فیزیولوژیک گوینده وابسته هستند. ضرایب MFCC و PLPC از دیگر بردارهای ویژگی مورد استفاده هستند که با تقریب دادگان فیلترهای شنیداری سیستم شنوایی انسان، افزونگی سیگنال گفتار را کاهش می‌دهند و

۵-۱- کارآیی الگوریتم جداسازی پیشنهادی

برای بررسی کارآیی الگوریتم پیشنهادی در بخش ۳، از آن جا که تابع شایستگی پیشنهاد شده وابسته به توزیع زمان فرکانس منابع گفتار می باشد، لازم است که حساسیت آن نسبت به سیگنال های گفتار متفاوت سنجیده شود. بنابراین دویست آزمایش با استفاده از مخلوط های گفتار دو گوینده که به طور اتفاقی از پایگاه داده انتخاب شده اند، انجام شد. در این دویست اجرا ماتریس پاسخ ضربه میان دو منبع تا دو میکروفون به صورت زیر قرار داده شد.

$$\begin{bmatrix} 1 - .7z^{-25} + .3z^{-30} - .1z^{-66} & .5z^{-23} - .3z^{-54} + .3z^{-66} - .1z^{-88} \\ .6z^{-18} - .04z^{-28} + .3z^{-66} & 1 - .6z^{-20} + .4z^{-88} + .1z^{-98} \end{bmatrix}$$

نتایج بهره سیگنال به اعوجاج به دست آمده برای این توابع ضربه در (جدول ۱) آورده شده است. منظور از SIR1 و SIR2 به ترتیب بهره سیگنال به اعوجاج به دست آمده در دو خروجی الگوریتم می باشد.

(جدول ۱) مقایسه نتایج بهره سیگنال به اعوجاج.

Average(dB)	SIR2 (dB)	SIR1(dB)	
۱۲/۴۱	۱۳/۵۵	۱۱/۳۳	روش پیشنهادی
۱۰/۷۸	۱۱/۳۵	۱۰/۲۱	SOS [4]

با مطالعه (جدول ۱) مشاهده می شود که کیفیت جداسازی منابع به وضوح نسبت به (Buchner, et al., 2006) برتری دارد. هم چنین برای نشان دادن توانایی الگوریتم پیشنهادی در معکوس سازی کانال، تابع کارایی مربع خطای نرمالیزه مربوط به شناسایی کانال در همان دویست اجرای مختلف محاسبه شد. تعریف ریاضی این تابع به صورت زیر است (Tugnait, 1997).

$$NMSE_{ij} = \frac{0.5 \times \sum_{t=1}^L \left[\sum_{\tau=1}^L (h_{ij}(\tau) - w_{ij}(\tau))^2 \right]}{\sum_{\tau=1}^L (w_{ij}(\tau))^2} \quad (9)$$

$$ONMSE = (4)^{-1} \sum_{i=1}^2 \sum_{j=1}^2 NMSE_{ij} \quad (10)$$

که در آن L طول پاسخ ضربه فیلترهای سیستم مخلوط (h_{ij}) و جداساز (w_{ij}) می باشد. نتایج حاصله در (جدول ۲) خلاصه شده است. در این جدول منظور از NMSE1 و NMSE2 مربع خطای نرمالیزه شناسایی کانال برای فیلترهای محاسبه کننده هر یک از دو خروجی می باشد.

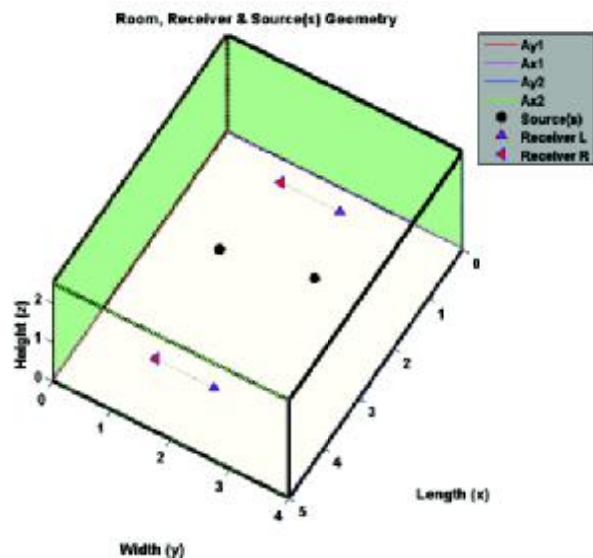
(جدول ۲) مقایسه نتایج بهره سیگنال به اعوجاج.

Average	NMSE1(dB)	NMSE2(dB)	
-۵/۲	-۵/۵	-۵	روش پیشنهادی
-۰/۲۵	-۰/۳	-۰/۲	SOS [4]

همین منطق تصمیم گیری به سادگی برای مکان یابی هم زمان سه منبع صوتی توسط دو آرایه سه میکروفونی قابل تعمیم است؛ در این مورد برای رفع ابهام جایگشت به هنگام حضور سه منبع صوتی تعداد جایگشت های ممکن، ۶ حالت خواهد بود و بنابراین ۶ متغیر تصمیم گیری به روش مشابه روال فوق تعریف خواهد شد که در این مورد تنها به ذکر نتایج شبیه سازی در فصل پنجم بسنده خواهیم کرد.

۵- نتایج شبیه سازی

در این مقاله از یک پایگاه داده گفتار با فرکانس ۱۶ کیلوهرتز استفاده شده که طول هر تکه سیگنال گفتار آن سه ثانیه می باشد. به منظور شبیه سازی واقعی تر محیط آکوستیکی و به منظور دستیابی به پاسخ ضربه هایی که محتوی راهنماهای فضایی معنادار در رابطه با موقعیت منابع هستند، از جعبه ابزار نرم افزاری RoomSim (Campbell, et al., 2005) استفاده شده که با استفاده از مدل منابع مجازی، شرایط آکوستیکی تا حد ممکن واقعی یک اتاق را شبیه سازی می کند. به این معنا که با توجه به چیدمان هندسی انتخاب شده پاسخ ضربه های مسیر بین هر منبع تا هر میکروفون را به گونه ای محاسبه می کند که بازتاب دهنده این چیدمان هندسی-آکوستیکی باشند. (شکل ۸) چیدمان منابع و دو آرایه دومیکروفونی مورد استفاده را برای آزمایش کیفیت رفع ابهام جایگشت (بخش ۵-۲) نشان می دهد. انتخاب فاصله یک متری میکروفون های هر آرایه، به این دلیل است که تفکیک پذیری درجه ای به کم تر از یک درجه برسد.



(شکل ۸) چیدمان میکروفون ها و منابع در بخش ۵-۲.

غلبه دارد ولی منبع دیگر هنوز حضور قابل درکی دارد. نسبت سیگنال به اعوجاج ناشی از فیلترهای دو سیستم جداساز به دست آمده، در (جدول ۳) نشان داده شده است.

۲۵۰ ترکیب دوبه‌دو از سی صدای موجود در پایگاه داده که متعلق به پانزده مرد و پانزده زن هستند برای سنجش توانایی ویژگی‌ها در رفع ابهام جایگشت استفاده شد. طول فریم‌های پردازشی سی میلی‌ثانیه است که معادل با بازه ایستانی سیگنال گفتار است و تعداد فریم‌های مورد استفاده در هر مرتبه مقایسه ۱۹۵ عدد می‌باشد. توانایی ویژگی‌های فهرست شده در زیر مورد بررسی قرار گرفته است:

- فرکانس‌های فورمانت: فورمانت‌های F_2, F_3, F_4 و F_5 مورد استفاده قرار گرفتند. طول تابع تمام قطب $p=14$ می‌باشد.
- ضرایب PLPC: از درجه ۱۳.
- ضرایب MFCC: از درجه ۱۳.
- معیار مجموع: ۶۰ فیلتر گاما - تون به طول ۵۱۲.

(جدول ۳) نسبت سیگنال به اعوجاج به دست آمده از سیستم جداساز.

سیستم جداساز شماره ۱ (BSS1)	
نسبت توان پاسخ‌ضربه سیگنال به اعوجاج ۱	۱۰/۰۴ دسی‌بل
نسبت توان پاسخ‌ضربه سیگنال به اعوجاج ۲	۸/۸۳ دسی‌بل
سیستم جداساز شماره ۲ (BSS2)	
نسبت توان پاسخ‌ضربه سیگنال به اعوجاج ۱	۷/۴۱ دسی‌بل
نسبت توان پاسخ‌ضربه سیگنال به اعوجاج ۲	۱۳/۱۲ دسی‌بل

هم‌چنین برای مقایسه با (Buchner, et al., 2006) که از معیار همبستگی برای رفع ابهام جایگشت استفاده نموده است، برای هر فریم به روش معرفی شده در (Buchner, et al., 2006) همبستگی متقابل خروجی‌ها را محاسبه و تصمیم در مورد چگونگی ارتباط خروجی‌ها اتخاذ می‌گردد. در نهایت با توجه به معیار رأی اکثریت برای انجام تصمیم‌گیری، تعداد تصمیم‌های درست در هر مرتبه اجرا شمرده شده و میانگین تصمیم‌های درست در (جدول ۴) نشان داده شده است. هم‌چنین، در این جدول میزان انحراف از معیار متغیر تصمیم‌گیری برای نشان دادن میزان پایداری هر ویژگی نشان داده شده است. به علاوه، تعداد تصمیم‌های به‌طور کامل اشتباه (رأی کم‌تر از پنجاه درصد) نیز در این جدول نشان داده شده است.

با مطالعه (جدول ۴) مشاهده می‌شود که بهترین نتایج چه از نظر تعداد رأی صحیح و چه از نظر میزان پایداری، متعلق به ویژگی معیار مجموع می‌باشد. این نتیجه مورد

با مطالعه مقادیر گزارش شده در (جدول ۲) ملاحظه می‌شود که الگوریتم مرجع بسیار از معکوس‌سازی کانال (و در واقع شناسایی آن) دور افتاده است. این درحالی است که روش پیشنهادی تا حد قابل قبولی به معکوس‌سازی نزدیک شده است. دلیل بهبود کیفیت شناسایی کانال آن است که الگوریتم پیشنهادی توانایی حذف همبستگی‌های فضایی در بلوک‌های قطری ماتریس همبستگی بلوکی در (شکل ۴) را نیز دارد. الگوریتم مرجع (Buchner, et al., 2006) به دلیل دوری جستن از حذف ساختار زمانی سیگنال گفتار، تابع هزینه را به صورت بلوکی قطری می‌کند و توانایی حذف خودبستگی‌های فضایی را ندارد. این درحالی است که الگوریتم پیشنهادی توانایی حذف این خودبستگی‌های فضایی را نیز دارد. گذشته از بهبودهای کیفی منعکس شده در (جداول ۱ و ۲) لازم است به مزیت‌های اجرایی الگوریتم جداسازی ارایه شده اشاره کنیم. مشکل عمومی الگوریتم‌های حوزه زمان مانند (Buchner, et al., 2006) آن است که پیچیدگی محاسباتی آن‌ها در صورت انتخاب طول بزرگ‌تر برای فیلترهای جداساز، به صورت نمایی رشد می‌کند. بار محاسباتی الگوریتم پیشنهادی وابستگی نمایی به طول فیلتر جداساز ندارد و تنها به صورت خطی با تعداد ضرایب انتخاب شده برای آموزش رشد می‌کند. از طرفی الگوریتم‌های افقی مانند (Buchner, et al., 2006) حساسیت شدیدی به پارامتر یادگیری واقعی دارند به گونه‌ای که عملکرد ایده‌آل آن‌ها تنها در صورت تنظیم مناسب این پارامتر متصور است. این درحالی است که الگوریتم پیشنهادی فاقد پارامتری است که نیاز به تنظیم دقیق داشته باشد.

۵-۲- کیفیت رفع ابهام جایگشت

برای مقایسه توانایی ویژگی‌های معرفی شده در (بخش ۴) در شرایط مساوی (از نظر سیستم مخلوط‌کننده و جداساز اعمال شده) ابتدا با استفاده از دو گوینده اتفاقی انتخاب شده از پایگاه داده، مخلوطی را با استفاده از چیدمان (شکل ۸) ایجاد کرده و با اجرای الگوریتم BSS1 و BSS2 از (شکل ۱)، پاسخ ضربه‌ی فیلترهای جداساز متناظر این دو را به دست آوردیم. سپس همین پاسخ ضربه فیلترهای جداساز به دست آمده برای BSS1 و BSS2، به مخلوط‌هایی که با اعمال سیستم مخلوط‌کننده مشابه به صداهای مختلف تولید شده‌اند، اعمال می‌شود و خروجی‌ها از نظر رفع ابهام جایگشت بررسی می‌شوند.

دو سیستم جداساز کیفیت جداسازی متوسطی اعمال می‌کنند؛ به این معنا که در هر خروجی یکی از منابع گفتار

انتظار می‌باشد؛ چون این ویژگی، تنها ویژگی است که به‌صراحت تغییرات دینامیکی سیگنال گفتار را استفاده می‌کند و با استفاده از دادگان فیلترهای شنیداری، همان‌طور که انسان میزان ادراک سیگنال اعوجاج‌کننده را کاهش می‌دهد، اثر سیگنال اعوجاج را کاهش می‌دهد. به این معنا که اعوجاج موجود تنها در معدود زیرباندهای شنیداری که غالب است، باعث اشتباه در سنجش چاله‌ها می‌شود و مشخصات سیگنال اصلی که در بیش‌تر زیرباندها غالب هستند، در مجموع غالب می‌شوند.

(جدول ۴) مقایسه توانایی ویژگی‌های مختلف در رفع ابهام جایگشت در حالت حضور دو گوینده.

ویژگی	رای درست	انحراف معیار	تصمیم اشتباه	مدت محاسبه (ثانیه)
هم‌بستگی [۴]	٪۵۹	٪۱۰/۴	۲۶	۰/۰۰۵۱
MFCC	٪۸۰/۶	٪۶/۲	۰	۰/۰۰۵۵
PLPC	٪۷۹/۱	٪۶	۰	۰/۰۰۵۸
فورمانتها	٪۶۳/۴۱	٪۶/۲۵	۴	۰/۰۳۲
معیار مجموع	٪۸۷/۲	٪۴/۴۷	۰	۴/۹۵

ویژگی دیگری که نتایج قابل مقایسه‌ای نسبت به معیار مجموع به‌دست آورده، ضرایب MFCC می‌باشد. که اتفاقاً این ویژگی نیز از دادگان فیلترهای شنیداری استفاده می‌کند و تغییرات دینامیکی منبع تحریک گفتار را حمل می‌کند. دلیل فاصله اندک کارایی آن با معیار مجموع را می‌توان در وابستگی هم‌زمان آن به مشخصات فیلتر لوله صوتی نسبت داد. به‌نظر می‌رسد تأثیر چهار فیلتر FIR مربوط به سیستم‌های مخلوط و جداساز روی هر یک از خروجی‌ها به‌اندازه‌ای بوده است که پاسخ فرکانسی آن‌ها بر پوش طیفی فیلتر لوله صوتی غالب شده و مشخصات سیستمی را بدون استفاده کرده است. این موضوع به‌خصوص در مورد نتایج ویژگی فرکانس‌های فورمانت قابل مشاهده است که تنها بازتاب‌دهنده موقعیت قله‌های پوش طیفی سیگنال گفتار هستند که نتیجه پاسخ فرکانسی لوله صوتی می‌باشند.

در این میان نتیجه حاصل از اعمال معیار هم‌بستگی، که در (Buchner, et al., 2006) استفاده شده، به‌وضوح از سایر معیارها ضعیف‌تر است و تعدد تصمیمات به‌طور کامل اشتباه آن استفاده از آن را به‌تمامی ناممکن می‌سازد. شبیه‌سازی‌ها نشان می‌دهند که معیار هم‌بستگی در مخلوط‌هایی اشتباه نمی‌کند که منابع حاضر در حوزه زمان

ناهم‌پوشان باشند و به عبارت دیگر بازه‌های شدت یکی بر بازه‌های خاموشی دیگری منطبق باشد. هرچه هم‌پوشانی روی محور زمان بیش‌تر باشد، تعداد اشتباه هم اضافه می‌شود. البته بدیهی است که چنان‌چه الگوریتم BSS در شرایط ایده‌آل به‌سمت جداسازی خیلی خوب برود، این معیار کافی خواهد بود. اما با ملاحظه گزارش‌های متعدد در مورد الگوریتم‌های BSS معلوم می‌شود که مقادیر متداول بهبود SIR در محیط‌های عملیاتی، در حدود ۹ تا ۱۰ دسی‌بل می‌باشد که مشابه بهره SIR استفاده در سناریوی آزمایشی می‌باشد که خود نتیجه اجرای شبیه‌سازی در یک محیط تا حد امکان واقعی می‌باشد. تلاش این مقاله رفع مشکل جداسازی غیرایده‌آل برای رفع ابهام بوده است و سناریوی فوق به این منظور تدارک دیده شده است تا عملکردها مستقل از چیدمان منابع و میکروفون‌ها سنجیده شوند. با این وجود بررسی تأثیر چیدمان‌های متفاوت بر عملکرد کلی سیستم جداساز و مکان‌یابی، موضوع مناسبی برای تحقیقات آینده می‌باشد.

در آزمایش دیگری برای بررسی کارایی ویژگی‌ها برای رفع ابهام جایگشت برای مخلوط سه گوینده، سناریوی مشابهی با استفاده از سه منبع گفتار و دو آزایی سه‌تایی ترتیب داده‌شد که نتایج تعداد تصمیمات درست در رفع ابهام جایگشت در (جدول ۵) آورده شده است. تعداد ترکیبات گفتار مورد استفاده سیصد عدد بوده است.

(جدول ۵) مقایسه توانایی ویژگی‌های مختلف در رفع ابهام جایگشت در حالت حضور سه گوینده.

ویژگی	رای درست	انحراف از معیار	تعداد تصمیم اشتباه
هم‌بستگی [۴]	٪۴۹	٪۷/۱	۱۳۰
ضرایب MFCC	٪۷۲/۹۳	٪۷/۷۳	۱
ضرایب PLPC	٪۶۷/۵۲	٪۶/۴۸	۲
فورمانتها	٪۳۳/۵	٪۵/۴	۲۹۵
معیار مجموع	٪۵۰/۷	٪۷/۱۶	۱۳۰

افت مشاهده‌شده در عملکرد معیار مجموع، مطابق توضیحات ارائه‌شده در بخش ۲-۴ توجیه می‌شود. در واقع شرایط تمیزپذیری فریم‌ها آن قدر کم اتفاق می‌افتد که در عمل کارایی این معیار برای رفع ابهام به‌شدت کاهش می‌یابد. همین‌طور با افزایش تعداد فیلترهای مؤثر در محاسبه هر خروجی، پوش طیفی اولیه سیگنال گفتار آن قدر تغییر می‌کند که همان‌طور که در نتایج مربوط به فورمانتها

سیگنال‌های گفتار، به رفع این ابهام در آرایه‌های مختلف پرداختیم و توانایی هر یک از ویژگی‌ها را در دو دسته ویژگی‌های وابسته به منبع تحریک و ویژگی‌های وابسته به فیلتر لوله صوتی بررسی شدند؛ و نشان داده شد که ضرایب PLPC با اعمال پیش‌پردازش‌های مشابه سیستم شنوایی، اطلاعات بی‌اهمیت از نظر شنیداری را دور می‌ریزد و حتی در حالت مخلوط سه تایی هم نتایج قابل قبولی به دست می‌دهد. هم‌چنین ضرایب MFCC با اعمال پیش‌پردازش‌های مشابه و با حفظ هم‌زمان ویژگی‌های وابسته به منبع تحریک و فیلتر لوله صوتی، بهترین نتایج را در رفع ابهام جایگشت حاصل کردند. تمام ویژگی‌های فوق بر معیار همبستگی (Buchner, et al., 2006) که به دلیل تعدد تصمیمات اشتباه، به‌طور کامل غیر قابل استفاده بود، برتری داشتند.

۷- تشکر و قدردانی

این پژوهش با حمایت مالی مرکز تحقیقات مخابرات ایران انجام شده است. مؤلفان مراتب قدردانی خود را از این حمایت اعلام می‌دارند.

۸- مراجع

Di Claudio, E.D., Parisi, R., Orlandi, G., 1999. Multi source localization in reverberant environments, INFOCOM Dpt, University of Rome "La Sapienza".

Chen, J., Hill, M., Benesty, J., 2005. Performance of GCC- and AMDF-Based Time-Delay Estimation in Practical Reverberant Environments, EURASIP Journal on Applied Signal Processing 2005:1, 25-36.

Buchner, H., Aichner, R., Kellermann, W., 2005. Relation between blind system identification and convolutive blind source separation, in Proc. Joint Workshop on Hands-Free Communication and Microphone Arrays.

Buchner, H., Aichner, R., Stenglein, J., Teutsch, H., Kellermann, W., 2006. Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering, Springer Handbook on Speech Processing and Speech Communication.

Knapp, C.H., Carter, G.C., 1976. The generalized correlation method for estimation of time delay, IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 320-327.

Kokkinakis, K., Nandi, A.K., 2006. Multichannel blind deconvolution for source separation in convolutive mixtures of speech, IEEE Trans. Audio, Speech, Lang. Process., Vol. 14, No. 1, pp. 200-212.

مشاهده می‌شود، به‌هیچ‌وجه برای رفع ابهام جایگشت مؤثر نیست. مطمئن‌ترین تصمیم‌گیری توسط ضرایب MFCC و PLPC به‌دست آمده که دلیل آن عدم وجود محدودیت‌های فوق برای این ویژگی‌ها می‌باشد. هم‌چنین چنان‌چه بار محاسباتی پارامتر تأثیرگذاری باشد، با مطالعه مقادیر گزارش شده در (جدول ۴) مشاهده می‌شود که برخلاف بار محاسباتی بسیار زیاد معیار مجموع، ضرایب MFCC و PLPCC بار محاسباتی مشابهی با معیار همبستگی داشته و از این نظر محدودیتی ندارند.

۶- بحث و نتیجه‌گیری

در این پژوهش یک چارچوب الگوریتمی جدید برای مکان‌یابی منابع چندگانه صوتی به کمک BSS حوزه زمان معرفی شد. ابتدا یک الگوریتم BSS حوزه زمان جدید معرفی شد که در آن بر آموزش انتخابی ضرایبی از فیلترهای جداساز که تأثیر بیشتری در جداسازی دارند، تمرکز می‌شود. میزان تأثیرگذاری هر یک از ضرایب با مشاهده ماتریس همبستگی فضایی نسخه سفیدشده مخلوط‌ها مشخص شد و آموزش انتخابی توسط الگوریتم هوش مصنوعی PSO انجام شد. مهم‌ترین مرحله اجرایی الگوریتم، تعریف تابع شایستگی خوش‌رفتار برای کمی‌سازی میزان جداسازی حاصل‌شده در هر گام از اجرای الگوریتم، بود.

این چارچوب از طرفی با تمرکز بر شناسایی کانال، با اطمینان و تنها در یک مرتبه اجرا تخمین‌های TDOA را به‌دست می‌دهد. از نظر کیفیت جداسازی در شرایط شبیه‌سازی مشابه، الگوریتم پیشنهادی سه دسی‌بل بهبود نسبت به روش (Buchner, et al., 2006) نشان داد. هم‌چنین به دلیل توانایی الگوریتم پیشنهادی در رفع خودبستگی‌های فضایی (در بلوک‌های قطری ماتریس همبستگی فضایی) کیفیت شناسایی کانال نیز بهبود قابل ملاحظه‌ای داشت. در این مورد حدود پنج دسی‌بل بهبود در خطای نرمالیزه شناسایی کانال مشاهده شد. مزیت اجرایی روش پیشنهادی استقلال بار محاسباتی از طول پاسخ‌ضربه فیلترهای جداساز (برخلاف سایر روش‌های حوزه زمان که به دلیل محاسبات پیچیده ماتریسی، بار محاسباتی‌شان به صورت نمایی با افزایش طول فیلتر جداساز رشد می‌کند) و بی‌نیازی آن از تنظیم پارامترت دقیق مورد نیاز (Buchner, et al., 2006) می‌باشد.

در ادامه پس از تشریح مشکل ابهام جایگشت عمومی BSS حوزه زمان در مکان‌یابی، با استفاده از ویژگی‌های

وی می‌باشد. در این زمینه، جداسازی کور از مخلوط‌های کانولوتیو، بخش‌بندی خودکار آوایی (phonetic segmentation) و مدل‌های غیرخطی تولید سیگنال صحبت از جمله علاقه‌مندی‌های وی می‌باشد.

نشانی رایانامک ایشان عبارتست از:

vahidkh62@gmail.com



محمد حسین کهائی در سال ۱۳۶۴

مدرک کارشناسی خود را در رشته مهندسی

برق، مخابرات از دانشگاه صنعتی اصفهان

اخذ و از سال ۱۳۶۵ تا ۱۳۶۹ در مرکز

تحقیقات مخابرات ایران مشغول به کار شد. سپس با ادامه

تحصیل در دانشگاه ریوکیو، ژاپن، مدرک کارشناسی ارشد

خود را در گرایش پردازش وفقی سیگنال‌ها در سال ۱۳۷۳

اخذ و به دنبال آن در سال ۱۳۷۷ مدرک دکترای تخصصی

خود را در زمینه پردازش آماری سیگنال‌ها از دانشگاه

صنعتی کوئینزلند، استرالیا دریافت نمود. دکتر کهائی از سال

۱۳۷۷ به‌عنوان عضو هیئت علمی دانشکده مهندسی برق

دانشگاه علم و صنعت ایران در گروه مخابرات مشغول به

فعالیت می‌باشد. زمینه‌های تحقیقاتی ایشان پردازش آرایه‌ای

سیگنال‌ها در کاربردهای جداسازی کور سیگنال‌ها، کنترل

فعال نویز، جهت یابی، محل یابی، ره‌گیری اهداف و مخابرات

سیستم می‌باشد.

نشانی رایانامک ایشان عبارتست از:

kahaei@iust.ac.ir

Jan, E.-E., Flanagan, J., 1998. Sound Source Localization in Reverberant Environments using an Outlier Elimination Algorithm, In Proc. of the International Conference on Spoken Language Processing, p1321-1324.

Schutte, J.F., Groenwold, A., 2005. A Study of Global Optimization Using Particle Swarms, Journal of Global Optimization 31: pp 93-108.

Das, S., Abraham, A., 2006. Synergy of Particle Swarm Optimization with Evolutionary Algorithms for Intelligent Search and Optimization, In IEEE International Congress on Evolutionary Computation, Vol. 1, 84-88.

Wildermoth, B.R., 2001. Text-Independent Speaker Recognition using Source Based Features, Master's thesis, Griffith University, Australia.

Weddin, M.E.M., 2005. Speaker Identification for Hearing Instruments, Master's Thesis, Denmark's Technical University.

O'Grady, P.D., Pearlmutter, B.A., Rickard, S.T., 2005. Survey of sparse and non-sparse methods in source separation, IJIST, vol. 15, pp. 18-33.

Campbell, D., Palomäki, K., Brown, G., 2005. A matlab simulation of 'shoebox' room acoustics for use in research and teaching. Computing and Information Systems Journal, 9(3):48-51.

Deshmukh, O., Espy-Wilson, C., Salomon, A., Singh, J., 2005. Use of Temporal Information: Detection of the Periodicity, Aperiodicity and Pitch in Speech, IEEE Trans. on Speech and Audio Proc., vol.13, pp. 776 - 786.

Tugnait, J.K., 1997. Identification and Deconvolution of Multichannel Linear Non-Gaussian Processes Using Higher Order Statistics and Inverse Filter Criteria, IEEE TRANS. ON SIGNAL PROCESSING, VOL.45, NO.3.



وحيد خان آقا در سال ۱۳۸۵ مدرک

کارشناسی خود را در رشته مهندسی برق،

الکترونیک از دانشگاه علم و صنعت ایران

اخذ نمود و سپس مطالعات خود را در

مقطع کارشناسی ارشد در گرایش مخابرات سیستم در

همین دانشگاه ادامه داد. در سال ۱۳۸۹ تحصیل در مقطع

دکترای تخصصی در رشته انفورماتیک در موسسه ملی

تحقیقات انفورماتیک فرانسه، دانشگاه بوردو، آغاز

نمود. پردازش سیگنال صحبت زمینه عمومی تحقیقات