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

وحید خان آقا و محمد حسین کهایی
دانشکدهٔ مهندسی کامپیوتر، دانشگاه علم و صنعت ایران

چکیده:

در این مقاله یک چارچوب الگوریتمی جدید برای مکان‌یابی منابع چندگانهٔ صوتی در محیط انعکاسی معرّفی 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 حوزهٔ زمان به‌دست می‌آیند. به این منظور یک الگوریتم BSS حوزهٔ 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 گویندهٔ سیگنال گفتار استفاده شود. در برابر معیار همبستگی مورد استفاده در مقالهٔ مرجع، نتایج شبیه‌سازی توانایی خوب این ویژگی‌ها را در رفع ابهام جایگشت نشان می‌دهد.

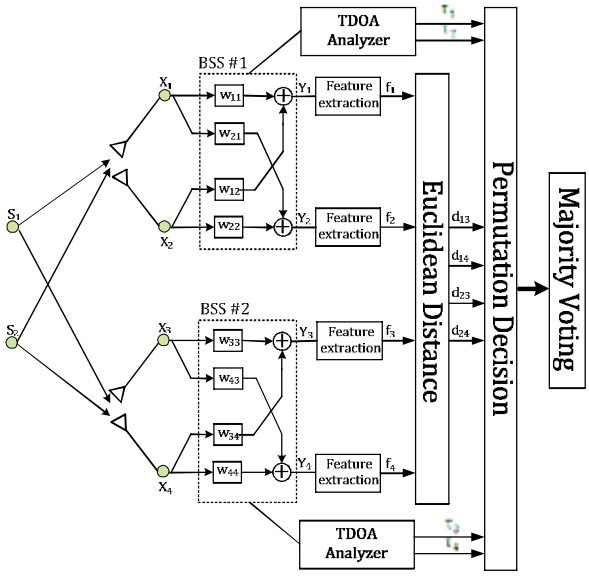
واژگان کلیدی: تخمین TDOA ، بهینه‌سازی PSO، جداسازی کور منابع، رفع ابهام جایگشت عمومی BSS.

۱- مقدمه

مکان‌یابی منابع صوتی، نقش مهمّی 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، به‌سادگی انجام‌پذیر است. امّا 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 آرایهٔ حس‌گری می‌باشد. در این روش در دو مرحله ابتدا چندین تخمین TDOA با استفاده از چند آرایهٔ سنسوری مستقل استخراج می‌شود، سپس این تخمین‌ها در فضای دوبعدی یا سه‌بعدی به‌صورت 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 سنّتی تخمین TDOA کارآمدی لازم را نداشته‌اند *(Chen, et al., 2005)*. برای رفع این مشکل دسته‌ای از محقّقان با معرّفی رابطهٔ میان مسئلهٔ جداسازی کور منابع (BSS) با مسئلهٔ 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 به‌عنوان محصول جانبی الگوریتم BSS به‌دست خواهد آمد که مکان‌یابی نهایی با ترکیب چندین تخمین TDOA میسر خواهد شد.

(شکل ۱) بلوک دیاگرام کلیِ مکان‌یابی به‌کمک BSS را برای دو منبع نشان می‌دهد که در آن با اجرای BSS روی دو آرایه‌ی میکروفونی دوتایی، به‌ازای هر آرایه، دو تخمین TDOA و دو سیگنالِ جداشده به‌دست 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).



(شکل ۱) نمایش ساختار فیلترهای مخلوط کننده و جداساز. پاسخ ضربهٔ میان منابع صوتی تا میکروفون‌ها به‌صورت h_{ij} و پاسخ ضربهٔ مربوط به فیلترهای جداساز با w_{ij} مشخّص شده‌اند.

به‌دلیل ابهام جایگشت عمومی ذاتی الگوریتم‌های BSS، فعلاً تعیینِ این‌که کدام تخمین TDOA حاصل از BSS1 مربوط به کدام تخمین TDOA حاصل از BSS2 می‌باشد، ناممکن است. تنها داشتهٔ ما برای حلّ این ابهام، خروجی‌های جداشدهٔ Yi می‌باشند؛ درواقع هر تخمین τ_i متعلّق به منبع صوتی‌ای می‌باشد که در Yi به‌دست آمده است. بنابراین اگر خروجی Yi از BSS1 و خروجی Yj از BSS2 متعلّق به منبع گفتار مشابهی باشند، τ_i و τ_j هم مربوط به همان منبع خواهند بود و می‌توانند برای استخراج محلّ آن منبع در فضای دوبعدی مورد استفاده قرار بگیرند.

مشکل آن‌جاست که الگوریتم‌های جداسازی حوزهٔ زمان، توانایی جداسازی ایده‌آل را ندارند. بلکه درعمل اگرچه در هر خروجی صدای غالب، قابل تشخیص خواهد بود، ولی به‌طور عمومی صدای منبع دیگر هم (هرچند ضعیف‌تر) قابل تشخیص خواهد بود. حتّی ممکن است در عین عملکرد خوب سیستم جداساز، به‌دلیل تفاوتِ شدّتِ بیانِ گفتار توسط اشخاص مختلف، یکی از منابع در هر دو خروجی غالب باشد. نتایج شبیه‌سازی نشان می‌دهند که تحلیل‌های ریاضی ابتدایی، مانند تابع همبستگیِ مورد استفاده در *(Knapp and Carter, 1976)*، توانایی رفع ابهام جایگشت را در شرایط واقعی ندارند؛ به این معنا که تابع همبستگی نمی‌تواند مشخّص کند که کدام دو خروجی از BSS1 و BSS2 متناظرِ هستند (اگرچه که سیستم شنوایی انسان غلبهٔ منبعِ صوتیِ مشابهی را در آن‌ها ادراک کند).

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

هم‌چنین یک الگوریتم BSS حوزهٔ زمان ابتکاری معرّفی شده که با تمرکز بر شناساییِ کانال، با استخراجِ موقعیتِ نسبی همبستگی‌های فضایی در سیستمِ جداساز ایده‌آل را بازسازی می‌کند. این الگوریتم بر تحلیل پیش‌گویی خطّی و الگوریتم بهینه‌سازی هوشمندِ PSO[1] 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 می پردازیم. سپس جزییات روش جداسازی پیشنهادی در بخش سه بیان می‌شود و منطق تصمیم‌گیری پیشنهادی برای رفع ابهام جایگشت در بخش چهارم تشریح خواهد شد. درنهایت پس از ارائهٔ نتایج شبیه‌سازی در بخش پنج، جمع‌بندی پایانی در بخش ششم ارایه خواهد شد.

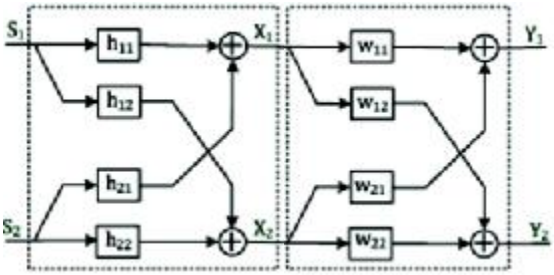
۲- تخمین TDOA به کمک BSS حوزهٔ زمان

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 قلّه‌های تبدیل فوریه معکوسِ چگالی طیفی توان متقابل سیگنال‌های دریافتیِ روی حس‌گرها، برای استخراج تخمین TDOA استفاده می‌شود. در حضور منابع چندگانه و یا در محیط دارای پژواک تابع GCC قلّه‌های چندگانه‌ای پیدا می‌کند. چنان‌چه در محیط انعکاسی واقعی به‌صورت هم‌زمان چند منبع حضور داشته باشند، تعدد قلّه‌ها درعمل تخمین TDOA را غیر ممکن می‌سازد *(Chen, et al., 2005)*.

[1] Particle Swarm Optimization

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 محقّقان روش به‌طور کامل متفاوتی را با تمرکز مستقیم بر استخراج پاسخ ضربه میان منابع و حس‌گرها در پیش‌گرفتند.*(Buchner, et al., 2005)* برای توضیح این روش، (شکل ۲) را در نظر بگیرید که دیاگرام ساختاری سیستم مخلوط کننده را همراه با سیستم جداساز نشان می‌دهد. تمام فیلترهای این مدل تحلیلی FIR درنظرگرفته‌شده‌اند.



(شکل ۲) نمایش ساختار فیلترهای مخلوط‌کننده و جداساز. پاسخ ضربهٔ میان منابع صوتی تا میکروفون‌ها به‌صورت h_{ij} و پاسخ ضربهٔ مربوط به فیلترهای جداساز با w_{ij} مشخص‌شده‌اند.

در نقطهٔ بهینهٔ جداسازی انتظار داریم که در هر خروجی Yi اثر تنها یکی از منابع اصلی Si موجود باشد. به‌عنوان مثال فرض کنید بخواهیم اثر منبع صدای S1 را در خروجی Y_2 حذف کنیم. در (شکل ۲) مشاهده می‌شود که این منبع از دو مسیر $w_{12} \otimes h_{11}$ و $h_{12} \otimes w_{22}$ به خروجی Y_2 می‌رسد (علامت $\otimes$ نشان‌دهندهٔ کانولوشن خطی است). برای حذف اثر S_1 کافی است این دو کانال قرینهٔ یکدیگر باشند. از آن‌جا که در حالت کلّی h_{11} و h_{12} مستقل از یکدیگر و متفاوت هستند، تنها راه حلِّ کلی برای چنین قرینه‌گی عبارتست از*(Buchner, et al., 2005)*:

$$w_{12} = -\alpha_1 h_{12} \qquad w_{22} = \alpha_1 h_{11} \qquad (۱)$$

به روش مشابه برای حذف اثر S2 را در خروجی Y_1 خواهیم داشت:

$$w_{11} = \alpha_2 h_{22} \qquad w_{21} = -\alpha_2 h_{21} \qquad (۲)$$

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

$$\tau_1 = \text{argmax}_n|h_{11}(n)| - \text{argmax}_n|h_{12}(n)| \qquad (۳)$$

$$\tau_2 = \text{argmax}_n|h_{22}(n)| - \text{argmax}_n|h_{21}(n)| \qquad (۴)$$

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

$$\tau_1 = \text{argmax}_n|w_{22}(n)| - \text{argmax}_n|w_{21}(n)| \qquad (۵)$$

$$\tau_2 = \text{argmax}_n|w_{11}(n)| - \text{argmax}_n|w_{12}(n)| \qquad (۶)$$

بنابراین با فرض برقراری روابط (۱) و (۲) درنتیجه هم‌گرایی الگوریتم BSS حوزهٔ زمان، می‌توان از فیلترهای سیستم جداساز برای انجام تخمین TDOA و درنهایت مکان‌یابی منابع چندگانهٔ 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 جزییات آن در ادامه تشریح خواهد شد.

۳- آموزش انتخابی فیلترهای جداساز

در این بخش به تشریح جزییات روش BSS حوزهٔ 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 بهینهٔ 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 به‌صورت $R_{x_1x_2}(n,m) = \langle x_1(n)x_2(n-m)\rangle$ بیان نمود. با در نظر داشتن (۱) و (۲) و نیز با توجّه به این حقیقت‌که فیلترهای سیستم مخلوط کننده، عامل اصلی همبستگی‌های فضایی هستند، می‌توان با شناسایی ساختارِ این همبستگی‌های فضایی به فیلترهای جداساز ایده‌آل روابط (۱) و (۲) دست یافت. درواقع می‌توان انتظار داشت که موقعیت‌های بیشینه‌گیِ وابستگی‌های فضایی با موقعیت‌های بیشینه‌گی فیلترهای مخلوط‌کننده و درنتیجه فیلترهای جداساز متناظر باشند.

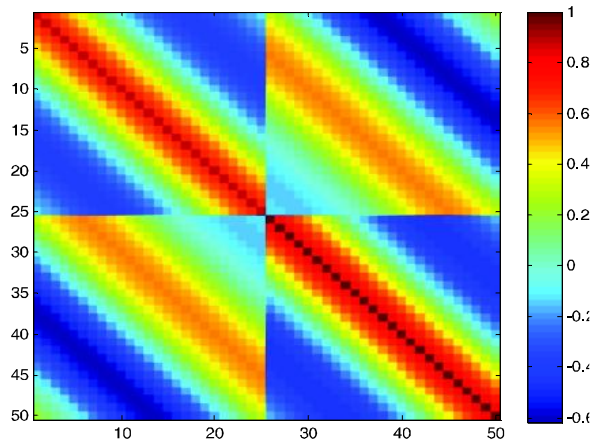
برای شناسایی موقعیت‌های بیشینه‌گیِ همبستگی‌های فضایی، ماتریس همبستگیِ بلوکیِ مخلوط‌ها (سیگنال‌های

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

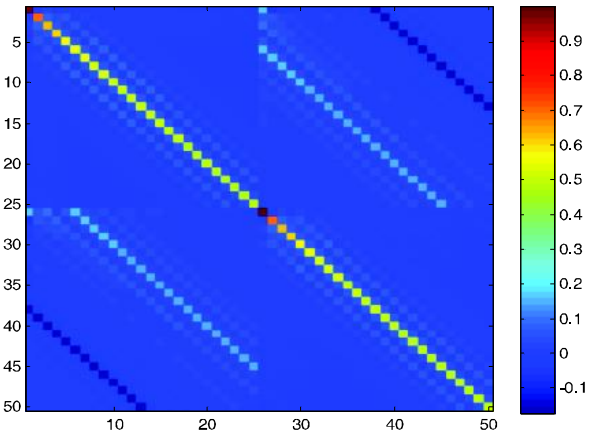
با توجّه به ساختار ویژهٔ ماتریس همبستگی می‌توان به صورت شهودی به اطّلاعات خوبی در مورد سیستم مخلوط‌کننده دست یافت. برای توضیح بهتر این موضوع، ماتریس همبستگی سیگنال‌های مخلوط را در حالتی‌که سیستم مخلوط‌کننده در ساده‌ترین شکل خود تنها با یک تضعیف و تأخیر به حس‌گرها رسیده‌اند (یعنی حالتی‌که فقط یکی از ضرایب فیلترهای FIR مخلوط‌کننده h_{ij} غیرصفر می‌باشد) در (شکل ۳) آورده‌ایم. در این شکل، چهار ناحیه بلوکی قابل تمیز هستند؛ دو بلوک قطری، خودبستگیِ هریک از مخلوط‌های مشاهده شده در میکروفون‌ها و بلوک‌های غیرقطری، همبستگیِ متقابلِ 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 نیمهٔ پایین‌مثلثیِ هر بلوک در رابطه با فیلتر متناظر، قابل مطالعه است.

در بلوک‌های غیرقطریِ (شکل ۳) دو ناحیه از همبستگی به‌صورت خطوط موربِ دارای شدّت بیشتر، به رنگ‌های آبی‌پررنگ و نارنجی قابل تمیز هستند. درعمل مرکز این خطوط مورب، به‌طور دقیق متناظر با ضریب غیر صفر فیلتر مخلوط می‌باشد؛ ولی مشکل این‌جاست که به خاطر پهن‌بودنِ این خطوط، تعیین دقیق تأخیر متناظر با هر خط امکان‌پذیر نیست. به‌ویژه در مخلوطِ واقعی در محیط انعکاسی با افزایش تعداد ضرایب غیر صفر فیلترهای مخلوط‌کننده، تعداد این خطوط بیش‌تر و بیش‌تر می‌شود و تداخل دامنهٔ آن‌ها تعیین موقعیت واقعی خطوط را ناممکن می‌سازد.

دلیل پهن‌بودن 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 طبقهٔ پیش‌پردازش برای از بین بردن این ساختار زمانی استفاده می‌کنیم. درعمل می‌توان با استفاده از مانده‌های فیلتر پیش‌بینی خطی سیگنالِ گفتار، نسخه‌ای از سیگنال‌های مخلوط را به‌دست آورد که خودبستگی‌های زمانی‌شان را از دست داده‌اند (سفید شده‌اند) ولی همبستگی‌های متقابلِ فضایی‌شان را حفظ کرده‌اند. این کار به‌شرطی امکان‌پذیر است که سفیدسازی در بازه‌های زمانیِ کوچک‌تر از پنج میلی‌ثانیه انجام شود تا وابستگی‌های فضایی مخلوط‌ها از بین نرود *(Jan and Flanagan, 1998)*. (شکل ۴) نتیجهٔ اعمال این طبقه پیش‌پردازش را نشان می‌دهد.



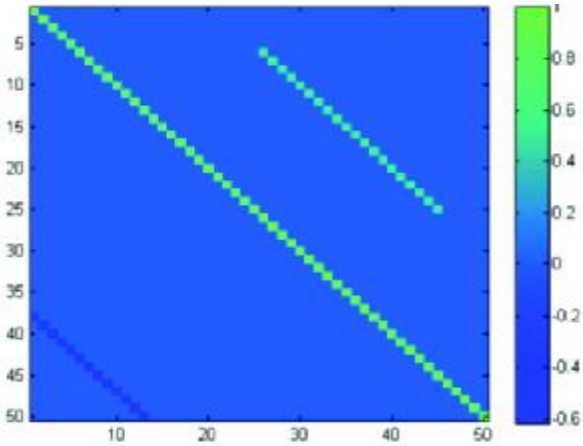
(شکل۳) نمایش ماتریس همبستگی یک مخلوط ساده. خطوط اریبِ متناظر با همبستگی‌های فضایی در بلوک‌های غیرقطری قابل مشاهده، امّا غیرمتمرکزاند.



(شکل۴) نمایش ماتریس همبستگی مانده‌های تحلیل پیش‌گویی خطّی یک مخلوط ساده . خطوط اریبِ متناظر با ضرایبِ 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 شدّت کم‌تر در این ماتریس وجود دارند که با انجام عملیات آستانه‌ایِ ستونی و اعمال فیلتر میانه‌ای (median) دوبعدی در امتداد خطوط اریب در ناحیهٔ پایین‌مثلثی هر بلوک، قابل حذف هستند. (شکل ۵) نتیجهٔ نهایی را نشان می‌دهد که با ضرب درایه به درایه در ماتریس همبستگی اولیّه، میزان شدّت اوّلیهٔ هر خط بازیابی

شده است. به این ترتیب می‌توان با مشاهدهٔ موقعیت خطوط اریبِ 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.



(شکل۵) نمایش ماتریس همبستگی بعد از انجام عملیات آستانه‌ای ستونی و اعمال فیلتر دوبعدی median در ناحیهٔ پایین‌مثلثی هر بلوک.

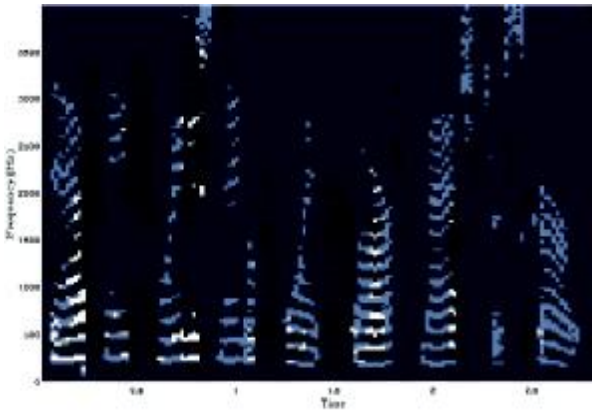
۳-۱- الگوریتم یادگیری PSO

PSO یک الگوریتم تکاملیِ 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) برروی یک فضای پارامتریِ چندبعدی تعریف می‌شود. جمعیتی از N_p ذره. به‌صورت اتّفاقی در فضای N_w بعدی به‌صورت $p_i = [p_{i_q}, \dots, p_{i_{N_w}}]$ مقداردهی اولیه می‌شوند. p_i موقعیت هر ذره را در فضای پارامتری مشخص می‌کند. این ذرّات 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 هرینه هدایت می‌شوند *(Das and Abraham, 2006)*.

در یک فضای پارامتری چندبعدی یک تابع شایستگی خوش‌رفتار، این الگوریتم به‌طور معمول با تعداد اندکی تکرار هم‌گرا می‌شود. بنابراین مهم‌ترین موضوع در پیاده‌سازی PSO اختیار کردن یک تابع شایستگی خوش‌رفتار است که یک بیشینه (یا کمینهٔ) 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 کمّیِ جداسازی را تعریف می‌کنیم. دلیل این ناهمپوشانی، تُنُک بودنِ 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 واریانسِ آن انتظار می‌رود، مقادیر نزدیک به صفر اختیار کند*(O'Grady, et al., 2005)*.

(شکل۶) 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 انجام‌شده است. بنابراین به‌نظر می‌رسد کمینه‌کردن این هم‌پوشانی (تعداد نقاط سفید) به‌عنوان تابع شایستگی، معیار خوبی از میزان جداسازیِ حاصل‌شده باشد.

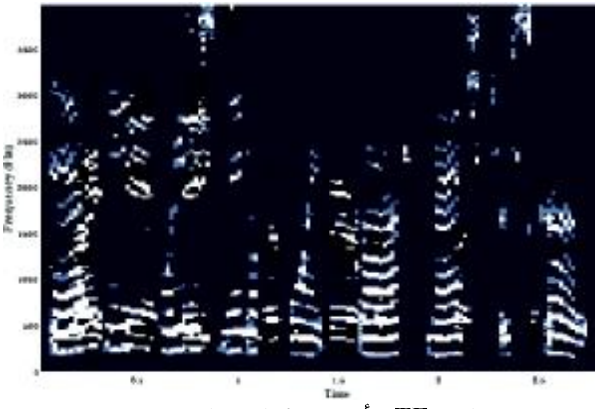


(شکل ۶) نمایش TF توأم دو سیگنال تمیز گفتار. نقاط سیاه نقاطِ شدّتِ منبع شمارهٔ ۱، نقاطِ طوسی روشن نقاط شدّت منبع شمارهٔ ۲ و نقاط سفید نقاطی را نشان می‌دهند که هر دو منبع با شدّت زیادی حاضر هستند. مشاهده می‌شود که تعداد نقاط هم‌پوشانی بسیار اندک می‌باشد.

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

متمایزی را عبور دهند. به‌عنوان مثال ممکن است Y_1 به یک سیگنال بالاگذر و سیگنال Y_2 به یک سیگنال پایین‌گذر تبدیل شده باشند و به این صورت شرط ناهمپوشانی طیف زمان‌فرکانس برقرار شده باشد؛ درحالی‌که جداسازی واقعی صورت نگرفته باشد. چون همان‌طور که در (شکل ۶) دیده می‌شود، دو سیگنالِ متمایز گفتار در اکثر بازه‌های فرکانسی به‌صورت هم‌زمان حضور دارند. درنتیجه انتخاب‌گری فرکانسی فیلترهای جداساز ممکن است باعث شود که بدون حصول جداسازی واقعی، هم‌پوشانی خروجی‌های Y_1 و Y_2 کمینه شود.

برای رفع این مشکل، تابع شایستگی به این صورت اصلاح می‌شود که به جای معیار کمینه‌شدنِ هم‌پوشانی، معیار معادلِ بیشینه‌شدنِ ناهمپوشانی در پنجره‌های زمان-فرکانسِ کوچک‌تر، و تنها در پنجره‌هایی که هر دو سیگنالِ خروجی در بیش از نیمی از نقاط پنجره، شدتی بیشتر از سطح نویز، محاسبه‌شود. یعنی هر پنجره در صورتی در محاسبهٔ تابع شایستگی کل وارد شود که هر دو منبع خروجی در آن پنجره حاضر باشند و حداقلِ هم‌پوشانی را نیز داشته باشند. سطح نویز نمایش TF با توجّه به نرمالیزاسیون اندازهٔ سیگنال‌های میکروفون‌ها برابر 35dB- درنظرگرفته‌شده است.



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

در ابتدا به دلیل هم‌پوشانی به‌طور تقریبی کامل توزیع TF دو مخلوط، تعداد پنجره‌های معتبر که هر دو منبع در آن حاضرند، زیاد است درحالی‌که مقدار ناهم‌پوشانیِ آن‌ها بسیار کم است. به‌تدریج الگوریتم PSO مقدار ناهم‌پوشانی‌ها را بیشتر می‌کند و در عین حال ممکن است تعدادی از پنجره‌ها که اتفاقاً فقط یک مخلوط در آن‌ها حاضر بوده، کاهش یابد. به هر حال الگوریتم به بیشینه‌کردنِ ناهم‌پوشانی در باقی‌ماندهٔ پنجره‌ها ادامه می‌دهد.

طول پنجرهٔ انتخاب‌شده در محور فرکانس بسته به تفکیک‌پذیری فرکانسی فیلترهای جداساز انجام می‌شود. هر چه طول پاسخ ضربهٔ فیلترهای جداساز بزرگ‌تر انتخاب شود، انتخاب‌گری فرکانسی آن‌ها بیشتر می‌شود. اگر این انتخاب‌گری کم‌تر از طول فرکانسی پنجره باشد، احتمال ضعیفی وجود دارد که مشکل جداسازی فرکانسی به‌جای جداسازی واقعی در هر پنجره تکرار شود. اگرچه شبیه‌سازی‌ها انتخاب طولِ فرکانسی پنجره را چندان تأثیرگذار نشان ندادند. طول زمانی پنجره، باید به اندازهٔ کافی بزرگ انتخاب شود تا تعداد پنجره‌های محتوی فعالیتِ هم‌زمانِ هر دو منبع گفتار زیاد شود. زیرا تابع شایستگی تنها در پنجره‌هایی حساب می‌شود که هر دو منبع فعّال باشند. بنابراین حدّاقل طول انتخاب ۰/۵ ثانیه (مدت زمان ادای یک کلمهٔ کوتاه) انتخاب می‌شود.

۴-رفع ابهام جایگشت به کمک ویژگی‌های سیگنال گفتار

در این مقاله از ویژگی‌های ویژهٔ گویندهٔ سیگنال گفتار استفاده می‌شود. این ویژگی‌ها با توجّه به مدل ریاضی تولید سیگنال گفتار در دو دستهٔ ویژگی‌های وابسته به منبع تحریک تارهای صوتی و ویژگی‌های وابسته به فیلتر لوله صوتی و لب‌ها در روش‌های شناسایی هویت گوینده بررسی شده‌اند و به‌طور عمومی با تقریب سیستم شنوایی انسان، افزونگی سیگنال حوزهٔ زمان را به بردارهای فشردهٔ محتوی اطّلاعات مربوط به گوینده کاهش می‌دهند و درنتیجه تقریب خوبی از ادراک شنوایی انسان هستند.

۴-۱- ویژگی‌های وابسته به منبع

ویژگی‌های وابسته به منبع، ویژگی‌هایی هستند که با مدل‌سازی موج صدای اولیه که وارد حنجره می‌شود، ارتباط دارند و بنابراین بیشتر با مشخّصات تارهای صوتی ارتباط دارند. جدا از ویژگی‌های وابسته به زبان‌شناسی که اندازه‌گیری آن‌ها دشوار و نیازمندِ داده‌های آموزشی طولانی می‌باشد، مشهورترین ویژگی وابسته به منبع، فرکانس بنیادی سیگنال گفتار (F_0) می‌باشد. در *(Wildermoth, 2001)* نشان‌داده‌شده که این ویژگی تنها در حروف صدادار (بازه‌های پریودیک بودن سیگنال گفتار) قابل مشاهده است و به‌علاوه تمایز معناداری برای اشخاص موجود در طبقهٔ جنسیتی مشابه (به‌خصوص برای زنان) ایجاد نمی‌کند. بنابراین برای عمل بازشناسی گوینده، کاربرد چندانی ندارد. البته می‌توان به جای محاسبهٔ این فرکانس

به‌صورت میانگینِ تخمینِ به‌دست‌آمده از تعداد زیادی از فریم‌های سیگنال، منحنی تغییراتِ این تخمین در طولِ زمانِ بیان یک جمله را به‌عنوان یک ویژگی متمایز کننده در نظر گرفت. در *(Weddin, 2005)* نشان داده‌شده که حتّی در بیان جملهٔ یکسان توسط گوینده‌های متفاوت، منحنی تغییرات فرکانس بنیادی چه از نظر مقدار و چه از نظر مدت زمان ادای جملات متفاوت است.

با درنظر گرفتن مشکلاتِ فوق، برای در نظر گرفتنِ این منحنی تغییرات، در این مقاله با استفاده از روشِ پیش‌نهادشده در *(Deshmukh, et al., 2005)* از معیار به‌عنوان "معیار مجموع" استفاده می‌شود که معیاری است کمی که تمامی قطعاتِ سیگنال را از نظر پریودیک‌بودن، میزانِ انرژی پریودیک، غیرپریودیک بودن و میزانِ انرژی غیرپریودیک، دسته‌بندی می‌کند. این معیار تابع معروفِ تابع تفاضل اندازهٔ 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 بودنِ هر فریم ارایه می‌کند. جزییات نحوهٔ محاسبه در *(Deshmukh, et al., 2005)* بیان شده است.

نکتهٔ مهم آن است که برای استفاده از معیارمجموع در رفع ابهام جایگشت، تمایز بین گفتارها به‌دلیل وقوع دو رخداد اتفاق می‌افتد: اوّل تمایزِ میانِ فرکانس‌های بنیادی گویندگان هنگامِ وقوعِ هم‌زمانِ فریم‌های صدادار این گویندگان و دومی هم‌زمان‌شدن یک فریم پریودیک از یک گوینده با فریم غیرپریودیک از گویندهٔ دیگر. بنابراین میزانِ موفقیت در تمیزگذاری بین دو گوینده، با تعدد دفعات وقوع دو رخدادِ فوق متناسب است. در مخلوط گفتار دو گوینده، احتمال وقوع یکی از حالات فوق بسیار زیاد است. امّا 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 اطّلاعات مربوط به فیلتر محفظهٔ صوتی را شامل می‌شوند. سازه‌ها به‌عنوان قلّه‌های پوش طیفی سیگنال گفتار، اطّلاعاتی در مورد شکل و اندازهٔ لولهٔ 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 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 ذیلِ ویژگی‌های وابسته به سیستم دسته‌بندی می‌شوند، ولی در واقع علاوه‌بر اطّلاعات فیلترِ لولهٔ صوتی، حامل اطّلاعات دینامیکی منبع نیز هستند.

۴-۳- منطق تصمیم‌گیری

در کاربرد ویژگی‌های سیگنال گفتار برای رفع ابهام جایگشت و تعیین خروجی‌های متناظرِ BSS1 و BSS2، دادهٔ آموزشی در اختیار نداریم. بنابراین نمی‌توانیم مانند روش‌های بازشناسی گوینده از توسعهٔ مدل‌های آماری برای شناسایی استفاده کنیم. در عوض به‌علت برقراری هم‌زمانی میانِ خروجی‌های دو واحد الگوریتمِ BSS نشان داده‌شده در (شکل ۱)، می‌توانیم به‌سادگی از اطّلاعات 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 فاصلهٔ 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 به‌دست آورده باشد. برای این منظور دو متغیر تصمیم‌گیری به‌صورت زیر تعریف می‌شوند:

$$\text{decision\_variable}_1 = d_{13} + d_{24} \quad (۷)$$

$$\text{decision\_variable}_2 = d_{14} + d_{23} \quad (۸)$$

که d_{ij} نشان‌دهندهٔ فاصلهٔ اقلیدسی بین بردار ویژگی خروجی‌های Y_i و Y_j در شکل ۱ می‌باشد. تصمیم‌گیری برای رفع ابهام جایگشت برای هر فریم به‌صورت زیر انجام می‌شود:

اگر $\text{decision\_varable}_1 < \text{decision\_varable}_2$

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

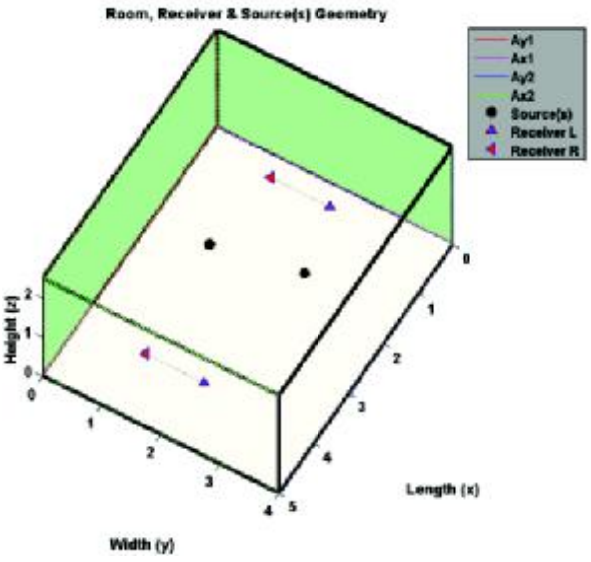
اگر $\text{decision\_varable}_1 > \text{decision\_varable}_2$

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

همین منطق تصمیم‌گیری به‌سادگی برای مکان‌یابی هم‌زمانِ سه منبع صوتی توسط دو آرایهٔ سه‌میکروفونی قابل تعمیم است؛ در این مورد برای رفع ابهام جایگشت به هنگام حضور سه منبع صوتی تعداد جایگشت‌های ممکن، ۶ حالت خواهد بود و بنابراین ۶ متغیرِ تصمیم‌گیری به روش مشابه روال فوق تعریف خواهد شد که در این مورد تنها به ذکر نتایج شبیه‌سازی در فصل پنجم بسنده خواهیم کرد.

۵- نتایج شبیه‌سازی

در این مقاله از یک پایگاه دادهٔ گفتار با فرکانس ۱۶ کیلوهرتز استفاده شده که طول هر تکهٔ سیگنال گفتار آن سه ثانیه می‌باشد. به‌منظور شبیه‌سازی واقعی‌ترِ محیط آکوستیکی و به‌منظور دست‌یابی به پاسخ ضربه‌هایی که محتوی راهنماهای فضایی معنادار در رابطه با موقعیت منابع هستند، از جعبه‌ابزار نرم‌افزاری *RoomSim (Campbell, et al., 2005)* استفاده شده که با استفاده از مدلِ منابعِ مجازی، شرایط آکوستیکی تا حدّ ممکن واقعی یک اتاق را شبیه‌سازی می‌کند. به این معنا که با توجّه به چیدمان هندسی انتخاب شده پاسخ ضربه‌های مسیر بین هر منبع تا هر میکروفون را به‌گونه‌ای محاسبه می‌کند که بازتاب‌دهندهٔ این چیدمان هندسی-آکوستیکی باشند. (شکل ۸) چیدمان منابع و دو آرایهٔ دومیکروفونیِ مورد استفاده را برای آزمایش کیفیت رفع ابهام جایگشت (بخش ۵-۲) نشان می‌دهد. انتخاب فاصلهٔ یک‌متریِ میکروفون‌های هر آرایه، به این دلیل است که تفکیک‌پذیری درجه‌ای به کم‌تر از یک درجه برسد.



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

۵-۱- کارآیی الگوریتم جداسازی پیشنهادی

برای بررسی کارآیی الگوریتم پیشنهادی در بخش ۳، از آن‌جا که تابع شایستگی پیشنهادشده وابسته به توزیع زمان‌فرکانسِ منابع گفتار می‌باشد، لازم است که حساسیتِ آن نسبت به سیگنال‌های گفتار متفاوت سنجیده شود. بنابراین دویست آزمایش با استفاده از مخلوط‌های گفتار دو گوینده که به‌طور اتفاقی از پایگاه داده انتخاب شده‌اند، انجام شد. در این دویست اجرا ماتریس پاسخ‌ضربهٔ میان دو منبع تا دو میکروفون به‌صورت زیر قرار داده‌شد.

$$\begin{bmatrix} 1-.7z^{-25}+.3z^{-30}-.1z^{-66} & .5z^{-23}-.3z^{-54}+.3z^{-66}-.1z^{-88} \\ .6z^{-18}-0.4z^{-28}+.3z^{-66} & 1-.6z^{-20}+.4z^{-88}+.1z^{-98} \end{bmatrix}$$

نتایجِ بهرهٔ سیگنال به اعوجاج به‌دست آمده برای این توابع ضربه در (جدول ۱) آورده شده است. منظور از SIR1 و SIR2 به‌ترتیب بهرهٔ سیگنال به اعوجاج به‌دست‌آمده در دو خروجی الگوریتم می‌باشد.

(جدول ۱) مقایسهٔ نتایج بهرهٔ سیگنال به اعوجاج.

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

با مطالعهٔ (جدول ۱) مشاهده می‌شود که کیفیت جداسازی منابع به‌وضوح نسبت به *(Buchner, et al., 2006)* برتری دارد. هم‌چنین برای نشان دادن توانایی الگوریتم پیشنهادی در معکوس‌سازی کانال، تابع کارایی مربع خطای نرمالیزهٔ مربوط به شناسایی کانال در همان دویست اجرای مختلف محاسبه شد. تعریف ریاضی این تابع به‌صورت زیر است *(Tugnait, 1997)*:

$$\mathrm{NMSE}_{ij} = \frac{0.5\times\sum_{l=1}^{L}\left[\sum_{\tau=1}^{L}\left(h_{ij}^{l}(\tau)-w_{ij}(\tau)\right)^2\right]}{\sum_{\tau=1}^{L}\left(w_{ij}(\tau)\right)^2} \qquad (۹)$$

$$\mathrm{ONMSE} = (4)^{-1}\sum_{i=1}^{2}\sum_{j=1}^{2}\mathrm{NMSE}_{ij} \qquad (۱۰)$$

که در آن L طول پاسخ‌ضربهٔ فیلترهای سیستم مخلوط (h_{ij}) و جداساز (w_{ij}) می‌باشد. نتایج حاصله در (جدول ۲) خلاصه‌شده است. در این جدول منظور از NMSE1 و NMSE2 مربع خطای نرمالیزهٔ شناسایی کانال برای فیلترهای محاسبه‌کنندهٔ هریک از دو خروجی می‌باشد.

(جدول ۲) مقایسهٔ نتایج بهرهٔ سیگنال به اعوجاج.

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

با مطالعهٔ مقادیر گزارش‌شده 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 حدّ 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)* به‌دلیل دوری جستن از حذف ساختار زمانی سیگنال گفتار، تابع هزینه را به‌صورت 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 مزیّت‌های اجرایی الگوریتم جداسازی ارایه شده اشاره کنیم. مشکل عمومی الگوریتم‌های حوزهٔ زمان مانند *(Buchner, et al., 2006)* آن است که پیچیدگی محاسباتی آن‌ها در صورت انتخاب طول بزرگ‌تر برای فیلترهای جداساز، به‌صورت نمایی رشد می‌کند. بار محاسباتی الگوریتم پیشنهادی وابستگی نمایی به‌طول فیلتر جداساز ندارد و تنها به‌صورت خطّی با تعداد ضرایب انتخاب شده برای آموزش رشد می‌کند. از طرفی الگوریتم‌های وفقی مانند *(Buchner, et al., 2006)* حساسیت شدیدی به پارامتر یادگیری وفقی دارند به‌گونه‌ای که عملکرد ایده‌آل آن‌ها تنها در صورت تنظیم مناسب این پارامتر متصوّر است. این درحالی است که الگوریتم پیشنهادی فاقد پارامتری است که نیاز به تنظیم دقیق داشته باشد.

۵-۲- کیفیت رفع ابهام جایگشت

برای مقایسهٔ توانایی ویژگی‌های معرّفی شده در (بخش ۴) در شرایط مساوی (از نظر سیستمِ مخلوط‌کننده و جداساز اعمال شده) ابتدا با استفاده از دو گویندهٔ اتفاقیِ انتخاب‌شده 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 جداسازِ متناظر این دو را به‌دست آوردیم. سپس همین پاسخ‌ضربهٔ فیلترهای جداساز به‌دست آمده برای BSS1 و BSS2، به مخلوط‌هایی که با اعمال سیستم مخلوط‌کنندهٔ 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 غلبه دارد ولی منبع دیگر هنوز حضور قابل درکی دارد. نسبتِ سیگنال به اعوجاجِ ناشی از فیلترهای دو سیستم جداسازِ به‌دست‌آمده، در (جدول ۳) نشان‌داده‌شده است.

۲۵۰ ترکیبِ 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 بازهٔ ایستانی سیگنال گفتار است و تعداد فریم‌های مورداستفاده در هر مرتبهٔ مقایسه ۱۹۵ عدد می‌باشد. توانایی ویژگی‌های فهرست‌شده 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 درجهٔ ۱۳.
- ضرایب MFCC : از درجهٔ ۱۳.
- معیار مجموع : ۶۰ فیلتر گاما - تون به طول ۵۱۲.

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

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

هم‌چنین برای مقایسه با *(Buchner, et al., 2006)* که از معیار همبستگی برای رفع ابهامِ جایگشت استفاده نموده است، برای هر فریم به روش معرّفی‌شده در *(Buchner, et al., 2006)* هم‌بستگی متقابل خروجی‌ها را محاسبه و تصمیم در مورد چگونگی ارتباط خروجی‌ها اتّخاذ می‌گردد. درنهایت با توجّه به معیار رأی اکثریت برای انجام تصمیم‌گیری، تعداد تصمیم‌های درست در هر مرتبهٔ اجرا شمرده شده و میانگین تصمیم‌های درست در (جدول ۴) نشان‌داده‌شده است. هم‌چنین، در این جدول میزان انحراف از معیارِ متغیرِ تصمیم‌گیری برای نشان‌دادن میزانِ پایداری هر ویژگی نشان‌داده‌شده است. به‌علاوه، تعداد تصمیم‌های به‌طور کامل اشتباه (رأی کم‌تر از پنجاه درصد) نیز در این جدول نشان‌داده‌شده است.

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

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) مقایسهٔ توانایی ویژگی‌های مختلف در رفع ابهام جایگشت در حالت حضور دو گوینده.

ویژگی	رأی درست	انحراف‌معیار	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 تحریکِ گفتار را حمل می‌کند. دلیل فاصلهٔ اندک کارآیی آن با معیار مجموع را می‌توان در وابستگی هم‌زمانِ آن به مشخصات فیلتر لولهٔ صوتی نسبت داد. به‌نظر می‌رسد تأثیر چهار فیلتر FIR مربوط به سیستم‌های مخلوط و جداساز روی هریک از خروجی‌ها به‌اندازه‌ای بوده است که پاسخ فرکانسی آن‌ها بر پوش طیفی فیلتر لولهٔ صوتی غالب شده و مشخّصات سیستمی را بدون استفاده کرده است. این موضوع به‌خصوص در مورد نتایج ویژگی فرکانس‌های فورمانت قابل مشاهده است که تنها بازتاب‌دهندهٔ موقعیتِ قلّه‌های پوش طیفی سیگنال گفتار هستند که نتیجهٔ پاسخ فرکانسی لولهٔ صوتی می‌باشند.

در این میان نتیجهٔ حاصل از اعمالِ معیارِ هم‌بستگی، که در *(Buchner, et al., 2006)* استفاده‌شده، 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 حوزهٔ زمان ناهم‌پوشان باشند و به عبارت دیگر بازه‌های شدّت 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 خیلی‌خوب برود، این معیار کافی خواهد بود. امّا با ملاحظهٔ گزارش‌های متعدد در مورد الگوریتم‌های BSS معلوم می‌شود که مقادیر متداولِ بهبودِ SIR در محیط‌های عملیاتی، در حدود ۹ تا ۱۰ دسی‌بل می‌باشد که مشابهِ بهرهٔ SIR استفاده در سناریوی آزمایشی می‌باشد که خود نتیجهٔ اجرای شبیه‌سازی در یک محیطِ تا حدّ 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 کلّی سیستم جداساز و مکان‌یابی، موضوع مناسبی برای تحقیقات آینده می‌باشد.

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

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

ویژگی	رأی درست	انحراف‌ازمعیار	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 ارایه‌شده در بخش ۲-۴ توجیه می‌شود. در واقع شرایط تمیزپذیریِ فریم‌ها آن‌قدر کم اتفاق می‌افتد که درعمل کارآیی این معیار برای رفع ابهام به‌شدّت کاهش می‌یابد. همین‌طور با افزایش تعداد فیلترهای مؤثر در محاسبهٔ هر خروجی، پوش طیفی اولیهٔ سیگنالِ گفتار آن‌قدر تغییر می‌کند که همان‌طور که در نتایج مربوط به فورمانت‌ها

مشاهده می‌شود، به‌هیچ‌وجه برای رفع ابهام جایگشت مؤثّر 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 مطالعهٔ مقادیر گزارش شده در (جدول ۴) مشاهده می‌شود که برخلاف بار محاسباتی بسیار زیاد معیار مجموع، ضرایب MFCC و PLPCC بارمحاسباتی مشابهی با معیارهمبستگی داشته و از این نظر محدودیتی ندارند.

۶- بحث و نتیجه‌گیری

در این پژوهش یک چارچوب الگوریتمی جدید برای مکان‌یابی منابع چندگانهٔ صوتی به کمک BSS حوزهٔ زمان معرفی شد. ابتدا یک الگوریتم BSS حوزهٔ 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 مشاهدهٔ ماتریس همبستگی فضایی نسخهٔ سفیدشدهٔ مخلوط‌ها مشخص شد و آموزش انتخابی توسط الگوریتم هوش مصنوعی PSO انجام شد. مهم‌ترین مرحلهٔ اجرایی الگوریتم، تعریف تابع شایستگی خوش‌رفتار برای کمّی‌سازی 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)* نشان‌داد. هم‌چنین به‌دلیل 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 نرمالیزهٔ شناسایی کانال مشاهده‌شد. مزیتِ اجراییِ روشِ پیشنهادی استقلال بار محاسباتی از طول پاسخ‌ضربهٔ فیلترهای جداساز (برخلاف سایر روش‌های حوزهٔ زمان که به‌دلیل محاسبات پیچیدهٔ ماتریسی، بار محاسباتی‌شان به صورت نمایی با افزایش طول فیلتر جداساز رشد می‌کند) و بی‌نیازی آن از تنظیم پارامتر وفقی مورد نیازِ *(Buchner, et al., 2006)* می‌باشد.

در ادامه پس از تشریح مشکل ابهام جایگشت عمومی BSS حوزهٔ زمان در مکان‌یابی، با استفاده از ویژگی‌های سیگنال‌های گفتار، به رفع این ابهام در آرایه‌های مختلف پرداختیم و توانایی هر یک از ویژگی‌ها را در دو دستهٔ ویژگی‌های وابسته به منبع تحریک و ویژگی‌های وابسته به فیلتر لولهٔ صوتی بررسی شدند؛ و نشان داده شد که ضرایب PLPC با اعمال پیش‌پردازش‌های مشابه سیستم شنوایی، اطّلاعات بی‌اهمیت از نظر شنیداری را دور می‌ریزد و حتّی در حالت مخلوط سه تایی هم نتایج قابل قبولی به‌دست می‌دهد. هم‌چنین ضرایب MFCC با اعمال پیش‌پردازش‌های مشابه و با حفظ هم‌زمان ویژگی‌های وابسته به منبع تحریک و فیلتر لولهٔ صوتی، بهترین نتایج را در رفع ابهام جایگشت حاصل کردند. تمام ویژگی‌های فوق بر معیار همبستگی *(Buchner, et al., 2006)* که به‌دلیل 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. مؤلّفان 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.

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, Griffth University, Australia.

Weddin, M.E.M., 2005. Speaker Identification for Hearing Instruments, Master's Thesis, Denmarks 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.

وحید خان‌آقا در سال ۱۳۸۵ مدرک کارشناسی خود را در رشتهٔ 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 تخصّصی در رشتهٔ انفورماتیک در موسسهٔ ملّی تحقیقات انفورماتیک فرانسه، دانشگاه بوردو، آغاز نمود. پردازش سیگنال صحبت زمینهٔ عمومی تحقیقات وی می‌باشد. در این زمینه، جداسازی کور از مخلوط‌های کانولوتیو، بخش‌بندی خودکار آوایی (phonetic segmentation) و مدل‌های غیرخطّی 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 رشتهٔ 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 ادامهٔ تحصیل در دانشگاه ریوکیو، ژاپن، مدرک کارشناسی ارشد خود را در گرایش پردازش وفقی سیگنال‌ها در سال ۱۳۷۳ اخذ و به دنبال آن در سال۱۳۷۷ مدرک دکترای تخصّصی خود را در زمینهٔ پردازش آماری سیگنال‌ها از دانشگاه صنعتی کوئینزلند، استرالیا دریافت نمود. دکتر کهائی از سال ۱۳۷۷ به‌عنوان عضو هیئت علمی دانشکدهٔ مهندسی برق دانشگاه علم و صنعت ایران در گروه مخابرات مشغول به فعّالیت می‌باشد. زمینه‌های تحقیقاتی ایشان پردازش آرایه‌ای سیگنال‌ها در کاربردهای جداسازی کور سیگنال‌ها، کنترل فعّال نویز، جهت یابی، محل‌یابی، ره‌گیری اهداف و مخابرات سیستم می‌باشد.

نشانی رایانامک ایشان عبارتست از:

kahaei@iust.ac.ir