کاربرد الگوریتم جداسازی کور منابع در جداسازی مخلوط سیگنال‌های گفتار و موسیقی

آرش پیشرویان، مسعودرضا آقا بزرگی صحاف و حمیدرضا ابوطالبی
گروه تحقیقاتی پردازش سیگنال، دانشکدهٔ مهندسی برق و کامپیوتر، دانشگاه یزد

چکیده

در این مقاله کاربرد یکی از الگوریتم‌های تجزیه به مؤلفه‌های مستقل (ICA) در جداسازی ترکیب‌های لحظه‌ای سیگنال‌های گفتار و موسیقی مورد بحث قرار می‌گیرد. الگوریتم جداساز در حوزهٔ زمان بوده و برای کمینه‌سازی اطلاعات متقابل به تخمین تابع رتبه نیاز دارد. برای تخمین تابع رتبه نمونه‌هایی از سیگنال‌های ترکیب شده، خروجی الگوریتم جداساز باید انتخاب گردد که اثر منابع اولیه در آن وجود داشته باشد. از آن‌جا که سیگنال‌های گفتار و موسیقی در بازه‌هایی از زمان می‌توانند دارای فواصل سکوت باشند، انتخاب قاب از سیگنال‌های ترکیب‌شده با اثر هر دو منبع اولیه در مسألهٔ ما اهمیت پیدا می‌کند. روش پیشنهادی ما استفاده از پارامتر تفاضل توابع رتبه است که با انتخاب یک حدّ آستانه می‌توان قاب بهینه را انتخاب نمود. کارایی الگوریتم جداساز با دو تخمین‌زن ترکیب‌های گوسی و حدّاقل میانگین مربع خطا مقایسه شده است. نتایج شبیه‌سازی نشان می‌دهد اوّل این‌که روش پیشنهادی عملکرد خوبی را در حذف بازه‌های سکوت از خود نشان می‌دهد و دوم این‌که الگوریتم جداساز با تخمین‌زن ترکیب‌های گوسی می‌تواند به تفکیک بهتر سیگنال‌ها و زمان پردازش کمتر الگوریتم جداساز منجر شود.

واژگان کلیدی: سیگنال‌های گفتار و موسیقی، الگوریتم جداساز، تفاضل توابع رتبه، تخمین‌زن تابع رتبه.

۱- مقدمه

تاکنون سیستم‌ها و الگوریتم‌های مختلفی توسط پژوهش‌گران در راستای جداسازی منابع صوتی[1] و در کاربردهای مختلف ارائه شده است *(Vincent, et al., 2005)*، *(Didiot, et al., 2010)*. این روش‌ها به دو دستهٔ الگوریتم‌های تک‌کاناله و الگوریتم‌های چندکاناله تقسیم‌بندی می‌شوند. در الگوریتم‌های تک‌کاناله، تنها یک سیگنال ترکیب‌شدهٔ خروجی جهت پردازش موجود بوده و تفکیک سیگنال‌ها که به‌طور معمول بر جداسازی یک منبع سیگنال تمرکز دارد از روی آن صورت می‌پذیرد. این‌گونه روش‌ها به‌طور معمول بر پایهٔ ویژگی‌ها و فرضیاتی است که در طبیعت سیگنال‌های منابع وجود دارد و با استفاده از این ویژگی‌ها و شرایط خاص آماری سیگنال‌ها، الگوریتم تفکیک کننده قابل اجرا می‌باشد *(Panagiotakis and Tziritas, 2005)*.

از مهم‌ترین روش‌های چند کاناله می‌توان به تکنیک جداسازی کور منابع (BSS[2]) اشاره کرد. جداسازی کور منابع یکی از مسائل اساسی و رو به گسترش در حوزهٔ پردازش سیگنال است که به مسألهٔ جداسازی و بازآوری سیگنال‌های نامعلومی که فقط مشاهداتی از ترکیب‌های آن‌ها (به واسطهٔ ترکیب‌کننده‌ای نامعلوم) توسط چند گیرنده دریافت شده است، می‌پردازد. از ویژگی‌های منحصر به فرد این روش ماهیت کوربودن الگوریتم می‌باشد که بر فرض‌های بسیار ضعیف بر روی سیگنال‌ها و سیستم ترکیب‌کننده استوار است.

یکی از فرض‌های ضعیفی که به‌طور معمول برای حل مسألهٔ BSS در نظر گرفته می‌شود، استقلال آماری منابع اولیه است که منجر به تکنیک تجزیه به مؤلّفه‌های مستقل (ICA[3]) می‌شود. هدف الگوریتم‌های ICA جداسازی اجزای مستقل از روی سیگنال‌های مشاهده شده است.

[1] Audio Source Separation

[2] Blind Source Separation
[3] Independent Component Analysis

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

درحقیقت با توجّه به فرض استقلال منابع اولیه، خروجی‌های تولید شده توسط الگوریتم ICA به‌نحوی است که تا حدّ امکان مستقل از یکدیگر باشند (*Choi, et al., 2005*)، (*Naik and Kumar, 2011*).

معیارهای مختلف استقلال در حل مسألهٔ ICA منجر به ایجاد الگوریتم‌های مختلفی می‌شود که این الگوریتم‌ها بسته به شرایط آن‌ها در کاربردهای گوناگون مورد استفاده قرار می‌گیرند که از جمله می‌توان به به‌کارگیری ICA جهت جداسازی سیگنال‌های گفتار و موسیقی اشاره نمود (*Pishravian, et al., 2008*). هدف این مقاله این است که به‌کمک یکی از الگوریتم‌های ICA (*Babaie-Zade and Jutten, 2005*) سیگنال‌های ترکیب‌شدهٔ لحظه‌ای گفتار و موسیقی را با ارایهٔ پیشنهادی جهت انتخاب قاب‌های مناسب برای اعمال به الگوریتم از یکدیگر تفکیک کند.

معیار در نظر گرفته شده برای سنجش استقلال در الگوریتم جداساز مورد بررسی، اطّلاعات متقابل[1] بین خروجی‌های الگوریتم می‌باشد. از این‌رو ابتدا در بخش دو صورت کلی مسأله و معیار استقلال نشان داده شده است. سپس در بخش سه الگوریتم جداساز و نحوهٔ عملکرد آن توضیح داده شده است. کمینه‌سازی اطّلاعات متقابل در الگوریتم جداساز، منجر به تخمین تابع رتبه می‌شود. برای ارزیابی کارایی الگوریتم جداساز در بخش چهار دو روش برای تخمین تابع رتبه توضیح شده است. همان‌طور که در بخش چهار مشاهده خواهید نمود است تخمین تابع رتبه از روی تعدادی از نمونه‌های سیگنال‌های ترکیب شده منابع اولیه صورت می‌پذیرد. برای پیاده‌سازی این الگوریتم در مسألهٔ مورد نظر از آن‌جا که منابع اولیهٔ گفتار و موسیقی می‌توانند در فواصل زمانی نامشخصی دارای سکوت باشند؛ از این‌رو ابتدا باید قابی از سیگنال‌های ترکیب شده را که در آن اثر هر دو منبع اوّلیه وجود داشته باشد، انتخاب کنیم. روش پیشنهادی استفاده از پارامتر تفاضل توابع رتبه (SFD[2]) است که در بخش پنج توضیح داده شده است. درانتها و در بخش شش نتایج شبیه‌سازی عملکرد الگوریتم جداساز و مقایسهٔ دو تخمین‌زن تابع رتبه و در بخش هفت نتیجه‌گیری ارائه شده است.

۲- مدل‌سازی مسأله و معیار استقلال

برای مسئله مدل BSS در نظر گرفته می‌شود. یعنی فرض می‌شود M منبع سیگنال $\mathbf{s}(t)=(s_1(t),...,s_M(t))^T$ وجود دارد که به‌طور مستقیم قابل مشاهده نیستند و هیچ‌گونه اطّلاعاتی از آن‌ها در دسترس نیست. سیگنال‌های منابع در یک کانال لحظه‌ای با هم ترکیب شده و توسط N گیرنده دریافت می‌شوند. بنابراین هر گیرنده یک ترکیب متفاوت از منابع اولیه را ضبط می‌کند و یک مجموعه از N سیگنال مشاهده شدهٔ $\mathbf{x}(t)=(x_1(t),...,x_N(t))^T$ که ترکیبی از سیگنال‌های منابع اولیه هستند، در دسترس است. با این فرضیات مدل فرآیند ترکیب لحظه‌ای به‌صورت زیر خواهد بود:

$$\mathbf{x}(t)=\mathbf{a}_1.s_1(t)+\square+\mathbf{a}_M.s_M(t) \quad (۱)$$

که به‌صورت برداری زیر قابل نمایش است:

$$\mathbf{x}(t)=\mathbf{A}\mathbf{s}(t) \quad (۲)$$

که در آن $\mathbf{x}(t)\in\Re^{N\cdot 1}$ بردار مشاهدات، $\mathbf{s}(t)\in\Re^{M\cdot 1}$ بردار منابع اولیه و $\mathbf{A}=[\mathbf{a}_1,...,\mathbf{a}_M]\in\Re^{N\cdot M}$ ماتریس ترکیب کننده نامعلوم می‌باشند.

در این‌شرایط الگوریتم BSS ماتریس جداساز $\mathbf{B}$ را طوری به‌دست می‌آورد که مؤلّفه‌های بردار خروجی $\mathbf{y}(t)=(y_1(t),...,y_M(t))^T$ تخمینی از منابع اولیهٔ $\mathbf{s}(t)$ باشند:

$$\mathbf{y}(t)=\mathbf{B}\mathbf{x}(t) \quad (۳)$$

که در آن $\mathbf{B}\in\Re^{M\cdot N}$ ماتریس جداساز نام دارد. درواقع اگر رابطهٔ $\mathbf{B}=\mathbf{A}^{-1}$ برقرار باشد منابع اولیه به‌طور دقیق بازیابی شده‌اند.

فرض استقلال منابع اوّلیه در حل مسألهٔ BSS منجر به تکنیک ICA می‌شود. هدف الگوریتم‌های ICA جداسازی اجزای مستقل از روی سیگنال‌های مشاهده شده است. در حقیقت با توجه به فرض استقلال منابع اولیه، خروجی‌های تولید شده توسط الگوریتم ICA به‌نحوی است که تا حدّ امکان مستقل از یکدیگر باشند. در مسئلهٔ مورد نظر با توجّه به استقلال منابع اولیهٔ گفتار و موسیقی، برای جداسازی سیگنال‌های ترکیب شدهٔ آن‌ها می‌توان از یک الگوریتم ICA 2 · 2 استفاده کرد.

در ICA برای اندازه‌گیری استقلال، به یک معیار نیاز است. تاکنون معیارهای مختلفی برای این منظور معرّفی شده‌اند (*Hyvärinen and Oja, 2000*)، (*Pedersen, et al., 2007*). یکی از این معیارها اطلاعات متقابل است که در الگوریتم‌های گوناگونی مورد کاربرد قرار گرفته (*Karvanen and Koivunen, 2002*) و روش‌هایی نیز برای تخمین آن معرفی شده است (Pham, 2003).

در الگوریتم مورد بررسی در این مقاله برای اندازه‌گیری استقلال بین متغیرهای تصادفی $y_1,y_2...,y_N$ از اطلاعات متقابل بین آن‌ها استفاده می‌شود:

[1] Mutual Information
[2] Score Function Difference

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

$$I(\mathbf{y}) = \int_{\mathbf{y}} P_{\mathbf{y}}(\mathbf{y}) \operatorname{Ln} \frac{P_{\mathbf{y}}(\mathbf{y})}{\prod_i P_{y_i}(y_i)} d\mathbf{y} \qquad (۴)$$

که در آن $p_{\mathbf{y}}$ و p_{y_i} به‌ترتیب توابع چگالی احتمال[1] $\mathbf{y}$ و y_i می‌باشند. رابطهٔ ۴ نشان می‌دهد که $I(\mathbf{y})$ همیشه غیر منفی می‌باشد و اگر و فقط اگر $P_{\mathbf{y}}(\mathbf{y}) = \prod_{i=1}^{N} p_{y_i}(y_i)$ (یا متغیرهای $y_1,...,y_N$ مستقل باشند) صفر خواهد بود. بنابراین برای جداسازی سیگنال‌ها باید اطلاعات متقابل بین خروجی‌های الگوریتم مینیمم شود.

۳- الگوریتم جداساز

در این بخش، الگوریتم جداساز که از آن برای تفکیک سیگنال‌های گفتار و موسیقی استفاده شده است، توضیح داده می‌شود. ابتدا برخی از تعاریف و مفاهیم پایه را که در ادامه به آن‌ها نیاز خواهیم داشت، بیان می‌کنیم *(Babaie-Zadeh, et al., 2001)*.

تابع رتبه: تابع رتبهٔ متغیر تصادفی y به‌صورت زیر تعریف می‌شود:

$$\psi_y(y) = -\frac{d}{dy} \ln p_y(y) = -\frac{p'_y(y)}{p_y(y)} \qquad (۵)$$

که در آن p_y تابع چگالی متغیر تصادفی y می‌باشد.

برای بردار تصادفی $\mathbf{y} = (y_1,...,y_N)^T$ دو نوع تابع رتبه تعریف می‌شود:

تابع رتبهٔ حاشیه‌ای (MSF[2]): تابع رتبهٔ حاشیه‌ای بردار تصادفی $\mathbf{y}$ به‌صورت بردار متشکل از توابع رتبه مؤلّفه‌های آن تعریف می‌شود:

$$\psi_{\mathbf{y}}(\mathbf{y}) = (\psi_1(y_1), \dots, \psi_N(y_N))^T \qquad (۶)$$

که در آن $\psi_i(y_i)$ تابع رتبه مؤلّفهٔ iاُم بردار تصادفی $\mathbf{y}$ می‌باشد.

تابع رتبهٔ توأم (JSF[3]): تابع رتبهٔ توأم بردار تصادفی y به‌صورت بردار $\varphi_{\mathbf{y}}(\mathbf{y}) = (\varphi_1(\mathbf{y}),...,\varphi_N(\mathbf{y}))^T$ تعریف می‌شود، که در آن:

$$\varphi_i(\mathbf{y}) = -\frac{\partial}{\partial y_i} \ln p_{\mathbf{y}}(\mathbf{y}) = -\frac{\frac{\partial}{\partial y_i} p_{\mathbf{y}}(\mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})} \qquad (۷)$$

تفاضل توابع رتبهٔ (SFD): که به‌صورت زیر تعریف می‌شود:

$$\beta_{\mathbf{y}}(\mathbf{y}) = \psi_{\mathbf{y}}(\mathbf{y}) - \varphi_{\mathbf{y}}(\mathbf{y}) \qquad (۸)$$

برخی از خواص مورد استفاده توابع فوق نیز از قرار زیر می‌باشند:

خاصیت ۱: مؤلّفه‌های بردار تصادفی $\mathbf{y} = (y_1,...,y_N)^T$ مستقل هستند اگر و تنها اگر $\beta_{\mathbf{y}}(\mathbf{y}) = 0$، و یا به‌عبارت دیگر $\psi_{\mathbf{y}}(\mathbf{y}) = \varphi_{\mathbf{y}}(\mathbf{y})$.

خاصیت ۲: فرض کنیم f یک تابع مشتق‌پذیر باشد که در رابطهٔ $\lim_{y\to\pm\infty} f(y)p_y(y) = 0$ صدق کند، آن‌گاه داریم:

$$E\{f(y)\psi_y(y)\} = E\{f'(y)\} \qquad (۹)$$

خاصیت ۳: برای بردار تصادفی کرانه‌دار $\mathbf{y}$ رابطهٔ $E\{\varphi_{\mathbf{y}}(\mathbf{y})\mathbf{y}^T\} = \mathbf{I}$ برقرار است که در آن $\mathbf{I}$ ماتریس واحد می‌باشد.

دیفرانسیل اطّلاعات متقابل: اگر $\mathbf{y}$ یک بردار تصادفی و Δ یک بردار تصادفی کوچک باشد آن‌گاه *(Babaie-Zade, et al., 2004)*:

$$I(\mathbf{y}+\Delta) - I(\mathbf{y}) = E\{\Delta^T \beta_{\mathbf{y}}(\mathbf{y})\} + o(\Delta) \qquad (۱۰)$$

که در آن $\beta_{\mathbf{y}}$ تفاضل توابع رتبهٔ بردار $\mathbf{y}$ و $o(\Delta)$ ترم‌های مرتبهٔ بالا بر حسب Δ می‌باشد.

الگوریتم جداساز از روش گرادیان طبیعی برای کمینه‌سازی اطّلاعات متقابل استفاده می‌کند. با توجّه به قضیهٔ دیفرانسیل اطّلاعات متقابل، گرادیان طبیعی اطّلاعات متقابل نسبت به ماتریس جداساز به‌صورت زیر به‌دست می‌آید (Babaie-Zade and Jutten, 2005):

$$\nabla_{\mathbf{B}} I = \frac{\partial I(\mathbf{y})}{\partial \mathbf{B}} \mathbf{B}^T = E\{\beta_{\mathbf{y}}(\mathbf{y})\mathbf{y}^T\} \qquad (۱۱)$$

رابطهٔ بالا را می‌توان با توجّه به خاصیت سه به‌صورت زیر نوشت:

$$\nabla_{\mathbf{B}} I = E\{\psi_{\mathbf{y}}(\mathbf{y})\mathbf{y}^T\} - \mathbf{I} \qquad (۱۲)$$

درنهایت الگوریتم جداساز ماتریس $\mathbf{B}$ را به‌صورت زیر به‌دست می‌آورد:

$$\mathbf{B} \longleftarrow (\mathbf{I} - \propto \nabla_{\mathbf{B}} I)\mathbf{B} \qquad (۱۳)$$

که در آن $\propto$ پارامتر گام می‌باشد.

همان‌طور که روابط ۱۲ و ۱۳ نشان می‌دهند الگوریتم جداساز به تخمین تابع رتبه نیاز دارد. درحقیقت در هر مرحله از تکرار الگوریتم جداساز، ابتدا تابع رتبهٔ حاشیه‌ای خروجی‌ها تخمین زده شده و سپس ماتریس $\mathbf{B}$ به‌صورت رابطهٔ ۱۳ به‌روز می‌شود. روند فوق آن‌قدر تکرار می‌شود تا الگوریتم هم‌گرا شده و خروجی‌های مستقلی تولید کند.

در (شکل ۱) ساختار الگوریتم جداساز نشان داده شده است. توجّه شود که برای همگرایی الگوریتم جداساز و

[1] Probability Density Function
[2] Marginal Score Function
[3] Joint Score Function

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

غلبه بر ابهام دامنه، در هر تکرار الگوریتم سیگنال‌های خروجی نرمالیزه می‌شوند.

- $\mathbf{B} = \mathbf{I}$
- حلقه
  - ۱) $\mathbf{y} = \mathbf{Bx}$
  - ۲) تخمین تابع رتبهٔ حاشیه‌ای $\psi_{\mathbf{y}}(\mathbf{y})$
  - ۳) $\nabla_{\mathbf{B}} I = \mathrm{E}\{\psi_{\mathbf{y}}(\mathbf{y})\mathbf{y}^T\} - \mathbf{I}$
  - ۴) نرمالیزه کردن
- تکرار تا همگرایی الگوریتم

(شکل ۱): ساختار الگوریتم

۴- تخمین تابع رتبه

از آن‌جا که دقّت و سرعت الگوریتم جداساز به تخمین MSF بستگی دارد، در این بخش دو الگوریتم تخمین‌زن تابع رتبه برای این‌منظور توضیح داده می‌شود. (در بخش شش، نحوهٔ عملکرد دو تخمین‌زن در کاربرد مسألهٔ مورد نظر مقایسه می‌شوند و نتایج شبیه‌سازی الگوریتم جداساز در تفکیک سیگنال‌های گفتار و موسیقی ارائه می‌شوند.)

۴-۱- تخمین‌زن کرنل‌های گوسی

همان‌طور که رابطهٔ (۵) نشان می‌دهد، تابع رتبهٔ یک متغیر تصادفی را می‌توان از روی تابع چگالی احتمال آن به‌دست آورد و در نتیجه تخمین تابع رتبه مستلزم تخمین تابع چگالی یک متغیر تصادفی خواهد بود. یکی از روش‌هایی که برای تخمین تابع چگالی به‌کار برده می‌شود، استفاده از مدل ترکیب‌کننده‌های گوسی[1] است (*McLachlan and Peel, 2000*) که در این بخش نحوهٔ تخمین تابع رتبه با استفاده از آن نشان داده می‌شود (*Vlassis and Motomura, 2001*، *Kokkinakis and Nandi, 2007*).

در مدل ترکیب‌کننده گوسی، تابع چگالی با مجموع m کرنل گوسی و به فرم زیر مدل می‌شود:

$$f(y) = \sum_{k=1}^{m} \pi_k g(y, \propto_k, \sigma_k^2) \qquad (۱۴)$$

که در آن $g(.)$ کرنل گوسی با مرکز $\propto_k$ و واریانس σ_k^2 به‌صورت زیر می‌باشد:

$$g(y, \propto_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{(y-\propto_k)^2}{2\sigma_k^2}\right] \qquad (۱۵)$$

تحت این شرایط تابع رتبهٔ متغیر تصادفی y به‌شکل زیر خواهد بود:

$$\psi_y(y) = -\frac{f'(y)}{f(y)} = \sum_{k=1}^{m} P(k|y)\frac{(y-\propto_k)}{\sigma_k^2} \qquad (۱۶)$$

که در آن $P(k|y) = \pi_k g(y, \propto_k, \sigma_k^2)$ $f(y)$ می‌باشد.

برای تخمین سریع و دقیق تابع رتبه، تعداد زیادی از کرنل‌ها که مرکز آن‌ها به‌طور مساوی بر روی مقادیر y قرار گرفته و واریانس آن‌ها ثابت σ^2 است در نظر گرفته می‌شوند. مرکز کرنل اوّل $\propto_1 \le \min(y)$ و مرکز کرنل kام $\propto_k = \propto_1 + (k-1)\delta$ در نظر گرفته شده که $\delta = (\propto_m - \propto_1)/(m-1)$ فاصلهٔ بین مرکزها می‌باشد. در این حالت تابع رتبه به‌صورت زیر به‌دست خواهد آمد:

$$\psi(y) = \frac{1}{\sigma^2}\left[y - \sum_{k=1}^{m} p(k|y)\propto_k\right] \qquad (۱۷)$$

که در آن $\sum_{k=1}^{m} P(k|y)\propto_k \equiv h(y)/f(y)$ برابر است با:

$$h(y) = \sum_{k=1}^{m} \pi_k \propto_k \exp\left[-0.5(y-\propto_k)^2/\sigma^2\right]$$
$$f(y) = \sum_{k=1}^{m} \pi_k \exp\left[-0.5(y-\propto_k)^2/\sigma^2\right] \qquad (۱۸)$$

روشی که در (*Vlassis and Motomura, 2001*) برای محاسبهٔ سریع معادلهٔ بالا ارائه شده است، بدین صورت است که به‌جای این‌که مقادیر رابطهٔ بالا را در نقاط y_i به‌دست آوریم، می‌توانیم $h(y)$ و $f(y)$ را ابتدا فقط در نقاط $\propto_k$ به‌دست آورده و سپس با درون‌یابی[2] مقادیر توابع بالا را در نقاط y_i به‌دست آوریم. مزیت انجام این کار کاهش پیچیدگی و افزایش سرعت محاسبات است، زیرا محاسبهٔ توابع $h(y)$ و $f(y)$ در نقاط $\propto_k$ با تکنیک FFT[3] به‌صورت زیر قابل پیاده‌سازی می‌باشد:

$$h(\propto_l) = \sum_{k=1}^{m} \pi_k \propto_k \exp\left[-0.5(l-k)^2\delta^2/\sigma^2\right]$$
$$f(\propto_l) = \sum_{k=1}^{m} \pi_k \exp\left[-0.5(l-k)^2\delta^2/\sigma^2\right] \qquad (۱۹)$$

هر دو رابطهٔ بالا بصورت کانولوشن گسستهٔ $h_l = \sum_{k=1}^{m} a_k g_{l-k}$ بوده که به‌آسانی با الگوریتم FFT به سرعت قابل حل می‌باشند.

۴-۲- تخمین‌زن کمینهٔ میانگین مربع خطا

خاصیت دو برای طراحی یک تخمین‌زن تابع رتبه با معیار کمینهٔ میانگین مربع خطا (MMSE[4]) بکار برده می‌شود (*Babaie-Zade and Jutten, 2005*). فرض کنیم می‌خواهیم

[1] Gaussian Mixture Model

[2] Interpolation

[3] Fast Fourier Transform

[4] Minimum Mean Square Error

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

$\psi_y(y)$ را با تابع پارامتری $f(y;\mathbf{w})$ (که در آن $\mathbf{w}=(w_1,....,w_L)^T$ بردار پارامتر است) تخمین بزنیم؛ با توجه به خاصیت دو نتیجه می‌شود:

$$\arg\min_{\mathbf{w}} \mathrm{E}\{(\psi_y(y)-f(y;\mathbf{w}))^2\} = \arg\min_{\mathbf{w}} \{\mathrm{E}\{\mathrm{f}^2(y;\mathbf{w})\}-2\mathrm{E}\{\frac{\partial f}{\partial y}(y;\mathbf{w})\}\} \quad (۲۰)$$

برای تخمین تابع رتبه بر حسب ترکیب خطّی توابع $k_1(y),...,k_L(y)$ داریم:

$$\hat{\psi_y}(y)=\sum_i^L w_i k_i(y)=\mathbf{k}^T(y)\mathbf{w} \quad (۲۱)$$

که در آن $\mathbf{k}(y)=(k_1(y),....,k_L(y))^T$. از آن‌جا که معیار تخمین، MMSE می‌باشد ضرایب $\mathbf{w}$ باید به‌نحوی تعیین شوند که خطای $\varepsilon=\mathrm{E}\{[\psi_y(y)-\hat{\psi_y}(y)]^2\}$ حدّاقل شود. هم‌چنین از قانون تعامد می‌توان $\mathrm{E}\{\mathbf{k}(y)(\psi_y(y)-\hat{\psi_y}(y))\}=0$ را نتیجه گرفت. این رابطه با توجّه به خاصیت دو به‌صورت زیر خلاصه می‌شود:

$$\mathrm{E}\{\mathbf{k}(y)\mathbf{k}^T(y)\}\mathbf{w}=\mathrm{E}\{\mathbf{k}'(y)\} \quad (۲۲)$$

با در نظر گرفتن $\mathbf{k}(y)=(1,\ y,\ y^2,y^3)^T$ ضرایب $\mathbf{w}$ به‌راحتی تعیین می‌شوند.

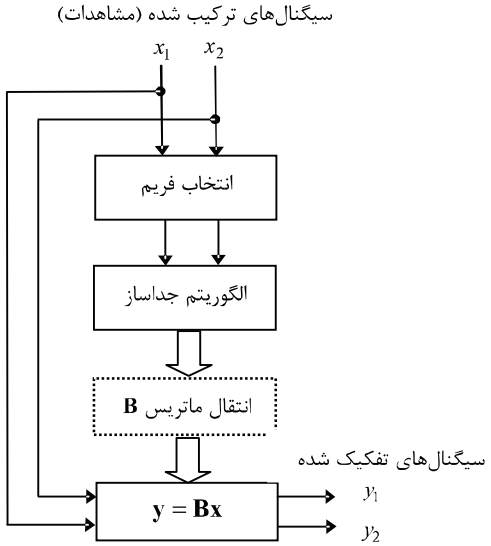
۵- تشخیص فواصل سکوت

همان‌طور که در بخش سه دیده شد، در هر تکرار الگوریتم جداساز تابع رتبهٔ حاشیه‌ای باید تخمین زده شود. روش‌های ارائه شده برای تخمین تابع رتبه نیز به تعدادی نمونه از سیگنال‌های مشاهده شدهٔ گفتار و موسیقی نیاز دارند. توجّه شود که در انتخاب نمونه‌ها و اعمال آن به الگوریتم جداساز باید اثر هر دو منبع اولیه وجود داشته باشد و هر دو سیگنال گفتار و موسیقی در قاب انتخاب شده وجود داشته باشد؛ چون همان‌طور که در بخش دو مقاله ذکر شده است، یکی از فرضیات مسئلهٔ جداسازی کور منابع این است که هر گیرنده باید شامل ترکیب متفاوتی از منابع اوّلیه باشد. حال اگر در مسئلهٔ مورد نظر در فاصلهٔ زمانی خاصّی یکی از منابع اوّلیه سکوت داشته باشد در این‌صورت هر دو گیرنده دارای یکی از منابع اولیه خواهد بود که این در جداسازی منابع مشکل ایجاد می‌کند. لذا انتخاب قاب مناسب با توجّه به نواحی سکوت دارای اهمّیت می باشد.

(شکل ۲) بلوک دیاگرام فرآیند جداسازی را نشان می‌دهد. ابتدا قابی از سیگنال‌های ترکیب شدهٔ گفتار و موسیقی انتخاب می‌شود که در آن اثر هر دو منبع اولیه وجود داشته باشد. سپس این قاب به‌عنوان ورودی الگوریتم جداساز اعمال می‌گردد. درنهایت بعد از هم‌گرایی الگوریتم جداساز، ماتریس به‌دست‌آمدهٔ $\mathbf{B}$ به کلّ سیگنال‌های مشاهده شدهٔ گفتار و موسیقی اعمال شده تا منابع اوّلیه به‌دست آورده شوند.

روشی که در این مقاله برای انتخاب قاب و حذف فواصل سکوت پیشنهاد شده است، به‌کارگیری پارامتر SFD می‌باشد. همان‌طور که در بخش سه گفته شد، دو متغیر تصادفی y_1 و y_2 مستقل هستند، اگر و تنها اگر $\beta_y(\mathbf{y})=0$. به‌عبارت دیگر اگر دو سیگنال، مستقل از یک‌دیگر باشند، مقدار SFD تخمین‌زده شدهٔ آن‌ها باید نزدیک صفر باشد (خاصیت یک). از این ویژگی می‌توان برای حذف فواصل سکوت در مسألهٔ مورد نظر استفاده کرد، به این‌صورت که اگر در یک بازهٔ زمانی خاص، یکی از منابع اولیهٔ گفتار یا موسیقی دارای سکوت باشند، در این‌صورت سیگنال‌های مشاهده شده، فقط شامل یکی از منابع اوّلیه خواهند بود. حال اگر پارامتر SFD را در این بازهٔ زمانی به‌دست آوریم، مقدار آن باید زیادتر از SFD فریمی باشد که شامل هر دو منبع اوّلیه است.

در این‌صورت برای تشخیص فواصل سکوت ابتدا سیگنال‌های مشاهده شده را به قاب‌هایی تقسیم‌بندی می‌کنیم. سپس پارامتر SFD را برای هر فریم به‌طور جداگانه به‌دست می‌آوریم. درنهایت با انتخاب یک حدّ آستانهٔ مناسب می‌توان قاب‌هایی را که در آن‌ها یکی از منابع اوّلیه دارای فواصل سکوت هستند و فقط سیگنال منبع دوم وجود دارد، شناسایی و حذف کرد.



(شکل ۲): فرآیند جداسازی سیگنال‌های گفتار و موسیقی

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

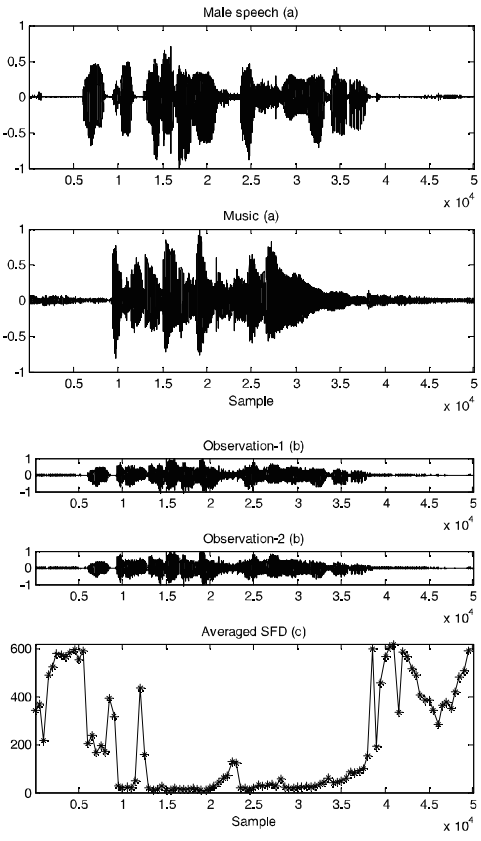
همان‌طور که دیده شد برای حذف فواصل سکوت باید تفاضل توابع رتبه را در هر قاب به‌دست آوریم. یک روش برای تخمین SFD استفاده از تعریف آن است. از آن‌جا که $\beta_{\mathbf{y}}(\mathbf{y}) = \psi_{\mathbf{y}}(\mathbf{y}) - \varphi_{\mathbf{y}}(\mathbf{y})$، بنابراین ابتدا می‌توان JSF و MSF را تخمین‌زده و سپس با تفاضل آن‌ها SFD را به‌دست آورد. در بخش ۴-۲ نحوهٔ تخمین MSF با معیار MMSE توضیح داده شده است. برای تخمین JSF با معیار MMSE نیز می‌توان از خاصیت زیر توابع رتبه استفاده کرد *(Babaie-Zadeh, et al., 2001)*:

فرض کنیم بخواهیم $\varphi_i(\mathbf{y})$ را به‌صورت ترکیب خطّی توابع چندمتغیرهٔ $\{k_1(\mathbf{y}),...,k_L(\mathbf{y})\}$، تخمین بزنیم، به‌عبارت دیگر $\hat{\varphi}_i(\mathbf{y}) = \mathbf{k}^T(\mathbf{y})\mathbf{w}$، که در آن $\mathbf{k}(\mathbf{y}) = (k_1(\mathbf{y}),....,k_L(\mathbf{y}))^T$ و $\mathbf{w} = (w_1,....,w_L)^T$، در این‌صورت:

$$\mathrm{E}\{\mathbf{k}(\mathbf{y})\mathbf{k}^T(\mathbf{y})\}\mathbf{w} = \mathrm{E}\{\frac{\partial \mathbf{k}}{\partial y_i}(\mathbf{y})\} \qquad (۲۳)$$

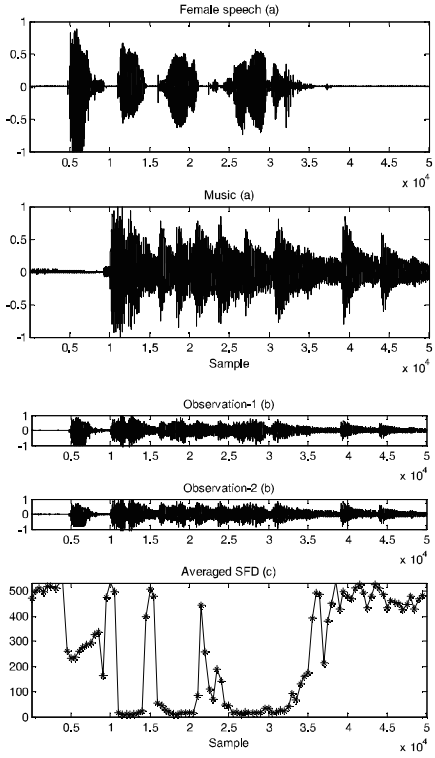
با در نظر گرفتن تابع چند متغیرهٔ $\mathbf{k}(\mathbf{y})$ به‌صورت زیر به‌راحتی می‌توان ضرایب $\mathbf{w} = (w_1, w_2)^T$ را به‌دست آورد.

$$\mathbf{k}(\mathbf{y}) = (1, y_1, y_1^2, y_1^3, y_2, y_2^2, y_2^3)^T \qquad (۲۴)$$



(شکل ۳): a)سیگنال‌های اولیه موسیقی و گویندهٔ مرد، b)سیگنال‌های ترکیب شدهٔ آن‌ها، c)مقدار میانگین SFD برای هر قاب

با انجام دو آزمایش صحت روش پیشنهاد شده را نشان می‌دهیم. برای انجام آزمایش‌ها از دو سیگنال گویندهٔ (مرد و زن) از دادگان FARSDAT و دو سیگنال موسیقی که همگی دارای فرکانس نمونه‌برداری ۱۶KHZ هستند، استفاده شده است. ابتدا منابع اولیه را با یک ماتریس مخلوط‌کننده با هم ترکیب کرده و سپس سیگنال‌های مشاهده شده را به قاب‌های پانصد نمونه‌ای تقسیم می‌کنیم. سپس مقدار میانگین پارامتر SFD را برای هر قاب به‌طور مستقل تخمین می‌زنیم. شکل‌های ۳ و ۴ نتایج شبیه‌سازی را نشان می‌دهند. همان‌طور که دیده می‌شود در قاب‌هایی که فقط اثر یکی از منابع اوّلیه وجود دارد (سیگنال دوم فاصلهٔ سکوت دارد) مقدار میانگین SFD زیاد و در قاب‌های دیگر مقدار آن کم‌تر می‌باشد.



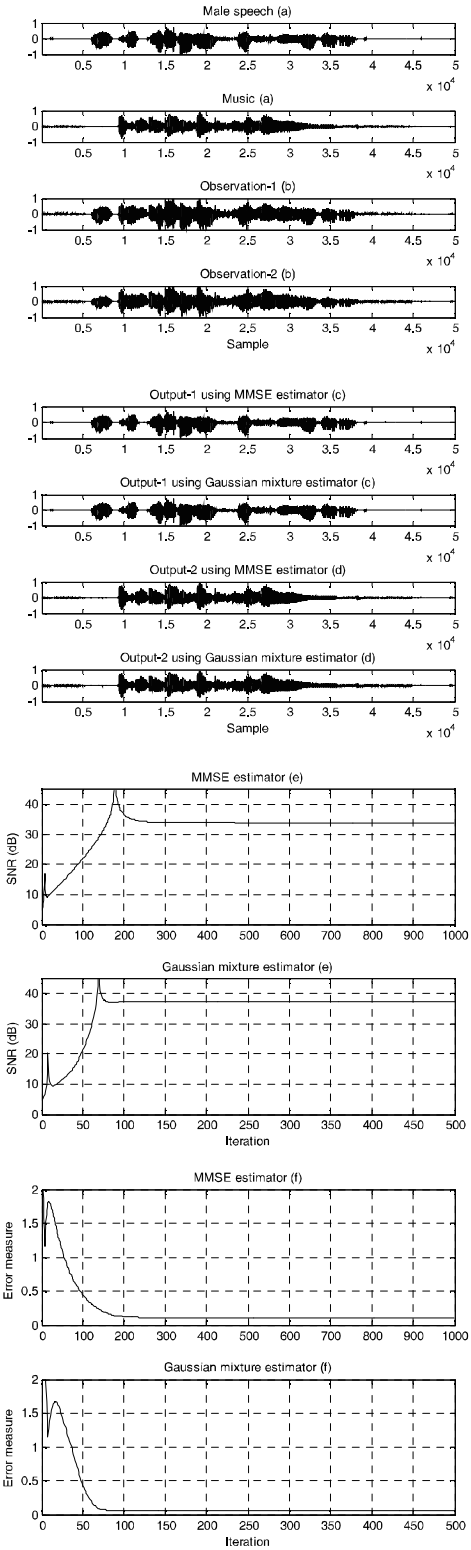
(شکل ۴): a)سیگنال‌های اوّلیه موسیقی و گویندهٔ زن، b)سیگنال‌های ترکیب شدهٔ آن‌ها، c)مقدار میانگین SFD برای هر قاب

۶- نتایج شبیه‌سازی

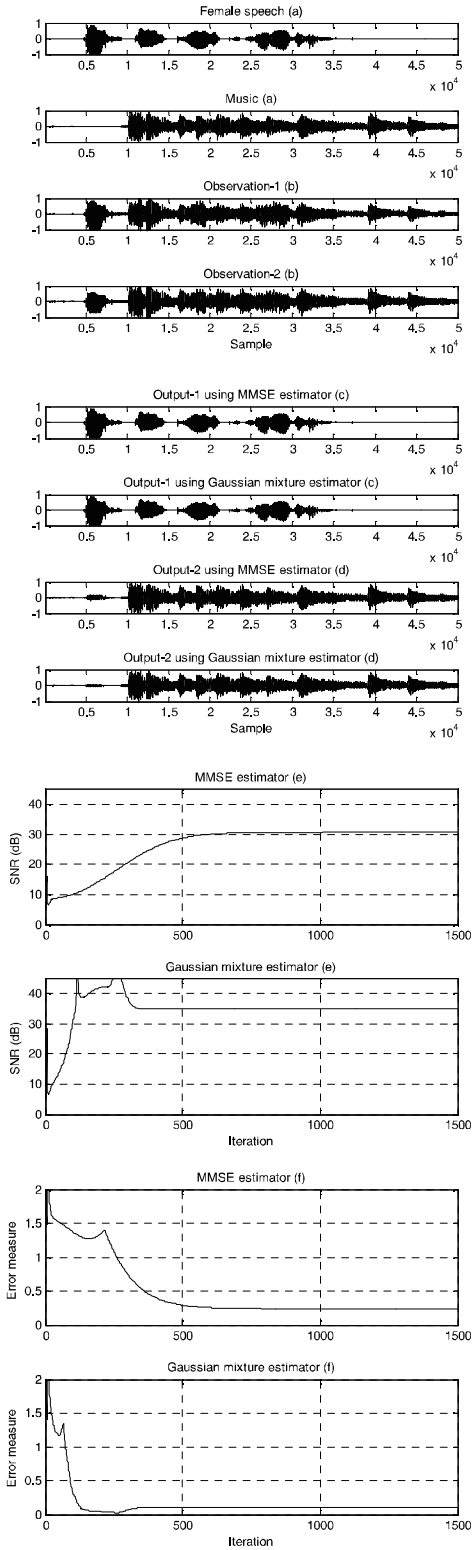
در این بخش نتایج عملکرد الگوریتم جداساز را بر روی ترکیب‌های گفتار و موسیقی (بخش ۵) نشان می‌دهیم.

هم‌چنین کارایی دو تخمین‌زن کمینهٔ میانگین مربع خطا و کرنل‌های گوسی در الگوریتم جداساز با هم مقایسه می‌شوند. در تمامی آزمایش‌های این بخش، پارامتر $\alpha = 0.1$ در نظر گرفته شده است.

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]



(شکل ۵): a) منابع اوّلیه موسیقی و گویندهٔ مرد، b) مشاهدات، c و d) مقایسهٔ سیگنال‌های خروجی الگوریتم با دو تخمین‌زن، e و f) مقایسهٔ مقدار SNR و معیار خطای خروجی‌های الگوریتم با دو تخمین‌زن



(شکل ۶): a) منابع اوّلیه موسیقی و گویندهٔ زن، b) مشاهدات، c و d) مقایسهٔ سیگنال‌های خروجی الگوریتم با دو تخمین‌زن، e و f) مقایسهٔ مقدار SNR و معیار خطای خروجی‌های الگوریتم با دو تخمین‌زن

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

برای سنجش کیفیت سیگنال‌های تفکیک شده در خروجی‌های الگوریتم، از دو معیار SNR و معیار خطای E استفاده شده است. با فرض عدم جایگشت[1] در خروجی‌های الگوریتم، مقدار SNR بر حسب dB در خروجی iاُم به‌صورت زیر محاسبه می‌شود:

$$\mathrm{SNR}_i = 10\log_{10}\frac{\mathrm{E}\{y_i^2\}}{\mathrm{E}\{y_i^2\big|_{s_i=0}\}} \qquad (۲۵)$$

که در آن $y_i|_{s_i=0}$ سیگنال خروجی iاُم است، زمانی که منبع iاُم برابر صفر در نظر گرفته شود. با توجّه به رابطهٔ بالا می‌توان نتیجه گرفت که هرچه مقدار SNR_i بیشتر باشد، کیفیت جداسازی سیگنال‌ها بهتر و اثر منابع دیگر در خروجی iاُم کم‌تر خواهد بود. از این‌رو رابطهٔ بالا می‌تواند معیار خوبی برای سنجش کیفیت جداسازی در خروجی‌های الگوریتم باشد. درنهایت مقادیر SNR کل از میانگیری روی مقادیر SNR_i و به‌صورت $\mathrm{SNR} = (\mathrm{SNR}_1 + \mathrm{SNR}_2)/2$ به‌دست می‌آید.

هم‌چنین عملکرد الگوریتم با معیار خطا E که به‌صورت زیر تعریف می‌شود *(Amari, et al., 1996)* قابل ارزیابی می‌باشد:

$$E = \sum_{i=1}^{d}\left(\sum_{j=1}^{d}\frac{|p_{ij}|}{\max_k|p_{ik}|}-1\right)+ \sum_{j=1}^{d}\left(\sum_{i=1}^{d}\frac{|p_{ij}|}{\max_k|p_{kj}|}-1\right) \qquad (۲۶)$$

که در آن p_{ij} عناصر ماتریس $\mathbf{P} = \mathbf{BA}$ می‌باشد. در حالتی که جداسازی سیگنال‌ها به‌طور کامل و دقیق صورت گیرد، ماتریس جداساز $\mathbf{B}$ برابر معکوس ماتریس $\mathbf{A}$ بوده و معیار خطای E صفر خواهد بود. درحالتی‌که تفکیک سیگنال‌ها به‌طور نسبی کامل صورت گیرد، ماتریس $\mathbf{P}$ به‌طور تقریبی به‌صورت یک ماتریس قطری بوده و معیار خطای E نزدیک صفر خواهد بود. در حقیقت رابطهٔ بالا نشان می‌دهد، تا چه اندازه ماتریس جداساز $\mathbf{B}$ به مقدار ایده‌آل آن یعنی $\mathbf{A}^{-1}$ نزدیک می‌باشد. نتایج آزمایش‌ها در شکل‌های ۵ و ۶ نشان داده شده است. همانطور که نتایج شبیه‌سازی نشان می‌دهند الگوریتم جداساز با تخمین‌زن کرنل‌های گوسی عملکرد بهتری را در تفکیک سیگنال‌های ترکیب شدهٔ گفتار و موسیقی از خود نشان می‌دهد؛ و موجب SNR بیش‌تر (کاهش معیار خطای E) می‌شود. هم‌چنین مشاهده می‌شود که الگوریتم جداساز با تخمین‌زن کرنل‌های گوسی سریع‌تر، هم‌گرا شده و درنتیجه زمان پردازش الگوریتم کاهش می‌یابد.

هم‌چنین مقایسهٔ زمان اجرای دو تخمین‌زن در MATLAB نشان می‌دهد که زمان پردازش الگوریتم جداساز با تخمین‌زن کرنل های گوسی و تخمین‌زن حداقل میانگین مربع خطا برای یک‌صد تکرار به‌ترتیب برابر ۱/۷۳ و ۱/۰۶ ثانیه می‌باشد.

۶- نتیجه‌گیری

در این مقاله ضمن بررسی یکی از الگوریتم‌های جداسازی کور منابع، کاربرد آن در جداسازی مخلوط سیگنال‌های گفتار و موسیقی نشان داده شد.

با توجّه به آن‌که عملکرد دقیق و زمان پردازش الگوریتم جداساز به‌دقّت و سرعت تخمین تابع رتبه وابسته است، دو الگوریتم کمینهٔ میانگین مربع خطا و کرنل‌های گوسی برای تخمین تابع رتبه توضیح داده شد. از طرفی تخمین تابع رتبه از روی نمونه‌های سیگنال‌های ترکیب شده صورت می‌پذیرد که در این نمونه‌ها باید اثر هر دو منبع اوّلیه وجود داشته باشد؛ لذا برای پیاده‌سازی الگوریتم جداساز در مسألهٔ مورد نظر مقاله، انتخاب قاب مناسب که در آن ترکیب سیگنال‌های هر دو منبع اولیه گفتار و موسیقی وجود داشته باشند، اهمیت داشت. روش پیشنهادی برای انتخاب قاب بهینه استفاده از پارامتر میانگین SFD بود که با انتخاب یک حدّ آستانه می‌توان قاب‌هایی که در آن‌ها یکی از منابع اوّلیه دارای فاصلهٔ زمانی سکوت می‌باشند، حذف نمود. نتایج شبیه‌سازی‌های انجام شده بر روی ترکیب‌های گفتار و موسیقی صحّت روش پیشنهادی را نشان دادند.

هم‌چنین بر اساس نتایج شبیه‌سازی، نشان داده شد که تخمین‌زن کرنل‌های گوسی می‌توانند عملکرد بهتری نسبت به تخمین‌زن کمینهٔ میانگین مربع خطا در الگوریتم جداساز از خود نشان دهد و موجب SNR بیشتر (کاهش خطا) سیگنال‌های خروجی الگوریتم و کاهش زمان همگرایی الگوریتم جداساز شود.

لازم به ذکر است روش جداسازی مخلوط گفتار-موسیقی این مقاله برای محیط‌های با کانال مخلوط کننده کانولوتیو مؤثّر نخواهد بود و مستلزم تعمیم الگوریتم در مخلوط‌های کانولوتیو و یا ترکیب آن با روش‌های حوزهٔ فرکانس می‌باشد.

[1] Permutation

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]

۷- تقدیر و تشکّر

بدین‌وسیله از حمایت و پشتیبانی مالی مرکز تحقیقات مخابرات ایران تقدیر و تشکّر می‌گردد.

۸- منابع

Amari, S., Cichocki, A., Yang, H.H., 1996. A new learning algorithm for blind signal separation, in Advances in Neural Information Processing Systems, MIT Press, Vol. 8, 757-763.

Babaie-Zade, M., Jutten, C., 2005. A general approach for mutual information minimization and its application to blind source separation, Signal Processing, Vol. 85, No. 5, 975-995.

Babaie-Zadeh, M., Jutten, C., Nayebi, K., 2001. Separating convolutive mixtures by mutual information minimization, In Proceedings of IWANN'2001, Granada, Spain, 834-842.

Babaie-Zade, M., Jutten, C., Nayebi, K., 2004. Differential of mutual information function, IEEE Signal Processing Letters, Vol. 11, No. 1, 48-51.

Choi, S., Cichocki, A., Park, H. M. Lee, S.Y., 2005. Blind source separation and independent component analysis: a review, Neural Information Processing - Letters and Reviews, Vol. 6, No. 1, 1-57.

Didiot, E., Illina, I., Fohr, D., Mella O., 2010. A wavelet-based parameterization for speech/music discrimination", Computer Speech & Language, Vol. 24, No. 2, 341-357.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications, Neural Networks, Vol. 13, No. 4, 411-430.

Karvanen, J., Koivunen, V. (2002). Blind separation methods based on pearson system and its extensions, Signal Processing, Vol. 82, No. 4, 663–673.

Kokkinakis, K., Nandi, A.k., 2007. Generalized gamma density-based score functions for fast and flexible ICA, Signal Processing, Vol. 87, No.5, 1156-1162.

McLachlan, G. J., Peel, D., 2000. Finite Mixture Models, Wiley, New York.

Naik, G.R., Kumar, D. 2011. An Overview of Independent Component Analysis and Its Applications, Informatica, Vol. 35, No. 1, 63-81.

Panagiotakis, C., Tziritas, G., 2005. A speech/music discriminator based on RMS and zero-crossings, IEEE Transactions on Multimedia, Vol. 7, No. 1, 155-166.

Pedersen, M. S., Larsen, J., Kjems, U., Parra, L. C., 2007. A survey of convolutive blind source separation methods, Springer Handbook on Speech Processing and Speech Communication, Springer, 1-34.

Pham, D. T., 2003. Fast algorithm for estimating mutual information, entropies and score functions, Proceedings of ICA2003 Conference, Nara, Japan, 17-22.

Pishravian, A., Aghabozorgi Sahaf, M. R., Abutalebi, H. R., 2008. Application of efficient score function estimation in blind speech-music separation, In Proceedings of the 9th International Conference on Signal Processing (ICSP08), Beijing, China, (Oct. 26-29, 2008), 618-621.

Vincent, E., Jafari, M. G., Abdallah, S. A., Plumbley, M. D., Davies, M. E., 2005. Blind Audio Source Separation, Queen Mary University of London.

Vlassis, N., Motomura, Y., 2001. Efficient source adaptivity in independent component analysis, IEEE Transactions on Neural Networks, Vol. 12, No. 3, 559-566.



آرش پیشرویان دورهٔ کارشناسی خود را سال ۱۳۸۴ رشتهٔ مهندسی برق (الکترونیک) گذرانده و مدرک کارشناسی ارشد خود را در سال ۱۳۸۷ در رشتهٔ مهندسی برق (مخابرات) از دانشگاه یزد اخذ نمود. زمینه‌های علمی مورد علاقهٔ وی پردازش سیگنال، به‌ویژه پردازش سیگنال گفتار می‌باشد.

نشانی رایانامک ایشان عبارت است از:

ar_pishravian@yahoo.com

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]



مسعودرضا آقابزرگی دورهٔ کارشناسی خود را در سال ۱۳۷۲ در رشتهٔ مهندسی برق (الکترونیک) در دانشگاه صنعتی اصفهان گذرانده و مدارک کارشناسی ارشد و دکترای خود را به‌ترتیب در سال‌های ۱۳۷۵ و ۱۳۸۲ در رشتهٔ مهندسی برق (مخابرات) از دانشگاه صنعتی اصفهان اخذ نمود. ایشان از سال ۱۳۸۲ بعنوان عضو هیئت علمی در دانشکده مهندسی برق و کامپیوتر دانشگاه یزد مشغول به فعالیت می‌باشد. زمینه‌های علمی مورد علاقهٔ وی پردازش سیگنال، جداسازی کور منابع، آنالیز زمان- فرکانس و تئوری اطّلاعات می‌باشد.
نشانی رایانامک ایشان عبارت است از:
aghabozorgi@yazduni.ac.ir



حمیدرضا ابوطالبی دورهٔ کارشناسی و کارشناسی ارشد خود 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 رشتهٔ مهندسی برق (مخابرات) از دانشگاه صنعتی امیرکبیر اخذ نمود. وی در جریان رساله دکترای خویش، به‌مدّت یک سال در دورهٔ فرصت مطالعاتی در دانشگاه واترلو کانادا به سر برد. ایشان در سال ۱۳۸۲ به دانشکدهٔ مهندسی برق و کامپیوتر دانشگاه یزد پیوست و هم‌اکنون به‌عنوان دانشیار این دانشکدهٔ مشغول به فعّالیت می‌باشد. وی هم‌چنین در سال ۹۰-۱۳۸۹ یک دوره فرصت مطالعاتی را در مرکز تحقیقاتی Idiap در سوئیس سپری نمود. زمینه‌های علمی مورد علاقهٔ وی پردازش آرایه‌ای سیگنال گفتار، بهسازی گفتار، مکان‌یابی گوینده، فیلترهای وفقی و آنالیز زمان- فرکانس می‌باشد.
نشانی رایانامک ایشان عبارت است از:
habutalebi@yazduni.ac.ir

[Downloaded from jsdp.rcisp.ac.ir on 2026-11-02]