



دانشگاه یزد
دانشکده مهندسی برق و کامپیوتر
گروه مهندسی مخابرات

پایان نامه
برای دریافت درجه کارشناسی ارشد
مهندسی برق (مخابرات)

استفاده از روش‌های جداسازی کور منابع در مکان‌یابی چند منبع گفتار

استاد راهنما: دکتر حمیدرضا ابوطالبی

استاد مشاور: دکتر مسعودرضا آقابزرگی

پژوهش و نگارش: مهدی آزادی حسین آباد

آذر ۱۳۹۱

تقدیم به

پدر و مادر همیشه خوبم

آنان که وجودم برایشان همزنج بود و وجودشان برایم همدم. مویشان سپیدی گرفت تا رویم سپید

بماند. آنان که فروغ نگاهشان، گرمی کلامشان و روشنی رویشان سرمایه جاودانی زندگیم است. آنان

که راستی قاسم در سلگتی قاتشان تجلی یافت. در برابر وجود کریشان زانوی ادب بر زمین می نهیم و با

دلی مملو از عشق و محبت بردستانشان بوسه می زنم.

تقدیر و شکر

سپاس فراوان نثار استادان بزرگوارمی که صداقتانه و صمیمانه در راه پرورش فرزندان این آب و خاک
قدم برمی دارند و در این راه از پیچ کوششی دریغ ندارند. سر تعظیم فرود می آورم بر این همه تواضع و
بزرگ منشی و به این همه مناعت طبعی که دارند و عاشقانه در راه اعتلای فرهنگ این مرز و بوم قدم
برمی دارند. از استاد راهنمایم آقای دکتر حمیدرضا ابوطالبی و همچنین استاد مشاورم آقای دکتر
مسعود رضا آقا بزرگی کمال شکر را به خاطر راهنمایی و مشاوره و گذاشتن وقت برای اینجانب در انجام بهتر
پروژه دارم. همچنین از آقای دکتر منصور نخکش و آقای دکتر وحید ابوطالبی به عنوان داوران داخلی
و خارجی کمال شکر را دارم.

چکیده

تخمین مقادیر TDOA (Time Differential Of Arrival) یکی از روش‌های پایه‌ای در مکان‌یابی منابع گفتار می‌باشد که به طور وسیع مورد بررسی قرار گرفته است. در دهه‌ی اخیر تحقیقات زیادی بر روی تخمین مقادیر TDOA با استفاده از فیلترهای بازترکیب مستخرج از جداسازی کور منابع، صورت گرفته است. در این پژوهش ضمن معرفی چندین روش از روش‌های مطرح‌شده در سال‌های اخیر، به بررسی و پیاده‌سازی روش SCT (State Coherence Transform) پرداخته و نقاط ضعف و قوت آن را بررسی می‌کنیم. در ادامه ضمن معرفی مشکل الیاسینگ فضایی که در صورت عدم رعایت نرخ نمونه‌برداری مکانی مناسب توسط میکروفون‌ها بوجود می‌آید، روش MSCT (Modified SCT) برای بهبود عملکرد مکان‌یابی در فواصل میکروفونی زیاد و کاهش اثر الیاسینگ فضایی پیشنهاد می‌شود که خطای مکان‌یابی در فاصله‌ی میکروفونی $l=1m$ را $1/74\%$ کاهش می‌دهد. با توجه به کاهش دقت عملکرد روش MSCT در شرایط نویزی، با وزن‌دهی روش MSCT که آن را WSCT (Weighted SCT) می‌نامیم، خطای مکان‌یابی در محیط‌های نویزی با $SNR=5dB$ را به میزان $89/82\%$ کاهش می‌دهیم.

در ادامه، با توجه به روند کاهشی دقت تخمین مقادیر TDOA با استفاده از روش‌های فوق در محیط‌های پژواکی، روش CC-BSS (Cross Correlation Blind Source Separation) با استفاده از همبستگی متقابل درایه‌های فیلتر بازترکیب مستخرج از BSS، معرفی می‌شود که در محیط با زمان انعکاس 700 میلی‌ثانیه خطای SCT را به میزان $21/34\%$ کاهش می‌دهد. سپس با پیاده‌سازی این روش در حوزه‌ی فرکانس و وزن‌دهی آن، روش GCC-BSS (Generalized CC-BSS) معرفی می‌شود که در آن از دو فیلتر گوسی و PHAT (PHase Transform) برای وزن‌دهی استفاده می‌کنیم که روش GCC-BSS-PHAT در محیط با زمان انعکاس 700 میلی‌ثانیه خطای مکان‌یابی با استفاده از روش CC-BSS را به میزان $66/41\%$ کاهش می‌دهد.

کلمات کلیدی: مکان‌یابی، TDOA، جداسازی کور منابع، الیاسینگ فضایی، فیلتر PHAT

فهرست مطالب

صفحه	عنوان
۱	فصل اول: مقدمه.....
۲	۱-۱- مقدمه
۲	۲-۱- معرفی روش جداسازی کور منابع (BSS)
۵	۱-۲-۱- مدل ترکیبی لحظه‌ای.....
۶	۲-۲-۱- BSS برای ترکیب‌های کانولوشنی
۷	۳-۲-۱- ارتباط بین مکان‌یابی و BSS
۷	۳-۱- مکان‌یابی با استفاده از روش‌های مبتنی بر BSS
۹	۴-۱- تقسیم‌بندی روش‌های مکان‌یابی مبتنی بر BSS
۹	۱-۴-۱- مکان‌یابی با استفاده از BSS-ADP
	۲-۴-۱- استفاده از روش اعمال فیلتر کور وقتی چند ورودی-چند خروجی (MIMO) در مکان‌یابی چند منبع
۱۲	۱۲
۱۳	۵-۱- ساختار پایان‌نامه
	فصل دوم: تبدیل حالت همدوسی تعمیم‌یافته (GSCT) برای تخمین TDOA منابع
۱۵	گفتار.....
۱۶	۱-۲- مقدمه
۱۷	۲-۲- BSS در حوزه‌ی فرکانس و تفسیر فیزیکی آن
۱۸	۱-۲-۲- حالت معین.....
۱۹	۲-۲-۲- حالت نامعین.....
۲۳	۳-۲- تبدیل حالت همدوسی تعمیم‌یافته (GSCT)
۲۵	۴-۲- مدل آماری برای حالت‌ها
۲۸	۵-۲- تخمین کرنل PDF تأخیر زمانی چندمتغیره.....
۲۸	۱-۵-۲- حالت تک‌مدی.....
۳۰	۲-۵-۲- حالت چندمدی.....

۳۱	۶-۲- نتایج بدست آمده از روش GSCT
	فصل سوم: ارائه‌ی روش‌های MSCT و WSCT برای مکان‌یابی در محیط‌های نویزی
۳۵
۳۶	۱-۳- مقدمه
۳۶	۲-۳- الیاسینگ فضایی
۳۸	۳-۳- معرفی روش MSCT برای حل مشکل الیاسینگ فضایی
۴۳	۴-۳- معرفی روش WSCT برای مکان‌یابی در محیط‌های نویزی
۴۴	۵-۳- نتایج شبیه‌سازی روش‌های MSCT و WSCT
۴۴	۱-۵-۳- تحلیل نتایج حاصل از پیاده‌سازی روش MSCT و WSCT در شرایط ایده‌آل ..
	۲-۵-۳- تحلیل نتایج حاصل از پیاده‌سازی روش MSCT و WSCT در شرایط غیرایده‌آل
۴۷
	فصل چهارم: ارائه‌ی روش‌های CC-BSS و GCC-BSS برای بهبود عملکرد مکان‌یابی در محیط‌های انعکاسی
۵۷
۵۸	۱-۴- مقدمه
۵۸	۲-۴- مدل‌های مختلف انتشار سیگنال برای تخمین TDOA
۵۹	۱-۲-۴- مدل انتشار ایده‌آل
۶۰	۲-۲-۴- مدل چندمسیری
۶۱	۳-۲-۴- مدل انعکاسی
۶۲	۳-۴- روش همبستگی متقابل برای تخمین مقادیر TDOA
۶۳	۴-۴- معرفی روش پیشنهادی CC-BSS
۶۶	۵-۴- معرفی روش پیشنهادی GCC-BSS
۶۸	۱-۵-۴- وزن‌دهی با فیلتر گوسی و معرفی روش GCC-BSS-Gaussian
۶۹	۲-۵-۴- وزن‌دهی با فیلتر PHAT و معرفی روش GCC-BSS-PHAT
۷۰	۶-۴- نتایج شبیه‌سازی روش‌های ارائه شده

۱-۶-۴- نتایج شبیه‌سازی روش‌های ارائه شده در حالت جداسازی و شرایط محیطی ایده‌آل	۷۰
۲-۶-۴- نتایج شبیه‌سازی روش‌های ارائه شده در حالت جداسازی و شرایط محیطی غیرایده‌آل	۷۲
فصل پنجم: جمع‌بندی و نتیجه‌گیری	۷۹
۱-۵- خلاصه‌ای از کارهای انجام شده	۸۰
۲-۵- نتیجه‌گیری	۸۱
۳-۵- دستاوردهای این پژوهش	۸۲
۴-۵- پیشنهادهایی برای ادامه‌ی کار	۸۳
منابع و مراجع	۸۵

فهرست اشکال

عنوان	صفحه
شکل (۱-۱): مدل عمومی ترکیب برای سیستم BSS [۲].....	۳
شکل (۲-۱): مدل عمومی بازترکیب برای سیستم BSS [۲].....	۵
شکل (۳-۱) نحوه‌ی عملکرد سیستم BSS [۳].....	۸
شکل (۵-۱): تخمین TDOA بر اساس مدل الف: SIMO و ب: MIMO بر پایه‌ی فیلترینگ و فقی کور [۵].....	۱۳
شکل (۱-۲): قطعه‌بندی بلوک داده‌ی سیگنال‌های ضبط‌شده [۷].....	۲۱
شکل (۲-۲) الف): تابع چگالی انباشته در حالت ۵ منبع و بلوک داده به اندازه‌ی ۸ فریم زمانی [۷].....	۲۱
شکل (۲-۲) ب): تابع چگالی انباشته در حالت ۵ منبع و بلوک داده به اندازه‌ی ۱ فریم زمانی [۷].....	۲۲
شکل (۲-۲) ج): تابع چگالی انباشته در حالت ۷ منبع و بلوک داده به اندازه‌ی ۸ فریم زمانی [۷].....	۲۲
شکل (۲-۲) د): تابع چگالی انباشته در حالت ۷ منبع و بلوک داده به اندازه‌ی ۱ فریم زمانی [۷].....	۲۲
شکل (۳-۲) الف): معیار شبیه‌نمایی تخمین‌زده‌شده طبق رابطه‌ی (۲-۱۴) برای فاصله‌ی میکروفونی $l=1\text{ m}$ [۷].....	۳۲
شکل (۳-۲) ب): معیار شبیه‌نمایی تخمین‌زده‌شده طبق رابطه‌ی (۲-۱۴) برای فاصله‌ی میکروفونی $l=0.1\text{ m}$ [۷].....	۳۳
شکل (۴-۲) الف): معیار شبیه‌نمایی تخمین‌زده‌شده طبق رابطه‌ی (۲-۲۹) برای فاصله‌ی میکروفونی $l=1\text{ m}$ [۷].....	۳۳
شکل (۴-۲) ب): معیار شبیه‌نمایی تخمین‌زده‌شده طبق رابطه‌ی (۲-۲۹) برای فاصله‌ی میکروفونی $l=0.1\text{ m}$ [۷].....	۳۳

- شکل (۲-۵): مقایسه‌ی تخمین دوبعدی تخمین مکان منابع برای ۳ ثانیه از فایل گفتار با فرکانس نمونه‌برداری ۱۶ کیلوهرتز (□ : مکان صحیح منابع و × : مکان‌های تخمین زده‌شده) [۷] ۳۴
- شکل (۳-۱): اختلاف فاز ایجادشده بین سیگنال‌های دریافتی از جفت میکروفون $(m+1, m)$ ناشی از منبع n م بر حسب فرکانس و فاصله‌ی میکروفونی $l=1\text{ cm}$ ۳۹
- شکل (۳-۲): فاز حالت‌های تعریف‌شده در روش SCT در دو حالت ایده‌آل (خط تیره) و حالت غیرایده‌آل (*) و فاصله‌ی میکروفونی $l=1\text{ cm}$ ۴۰
- شکل (۳-۳): منحنی فاز حالت‌های SCT بر حسب بازه‌های فرکانسی در حالت $l=10\text{ cm}$ و مقایسه‌ی مقدار تخمین زده‌شده‌ی TDOA (خط چین) با مقدار واقعی (خط تیره) ۴۱
- شکل (۳-۴): منحنی فاز حالت‌های SCT بر حسب بازه‌های فرکانسی در حالت $l=20\text{ cm}$ و مقایسه‌ی مقدار تخمین زده‌شده‌ی TDOA (خط چین) با مقدار واقعی (خط تیره) ۴۱
- شکل (۳-۵): فاز حالت‌های مربوط به منبع ۱ (ستاره)، منبع ۲ (مثلث) و منبع ۳ (دایره) بر حسب شماره‌ی بازه‌های فرکانسی و $l=10\text{ cm}$ ۴۵
- شکل (۳-۶): تابع SCT، MSCT و WSCT در شرایط سه منبع و $l=10\text{ cm}$ ۴۶
- شکل (۳-۷): فاز حالت‌های مربوط به منبع ۱ (ستاره) و منبع ۳ (دایره) بر حسب شماره‌ی بازه‌های فرکانسی و $l=1\text{ cm}$ ۴۸
- شکل (۳-۸): تابع SCT و MSCT در حالت دو منبع و $l=1\text{ cm}$ ۴۸
- شکل (۳-۹): فاز حالت‌های مربوط به منبع ۱ (ستاره) و منبع ۳ (دایره) بر حسب شماره‌ی بازه‌های فرکانسی و $l=10\text{ cm}$ ۴۹
- شکل (۳-۱۱): فاز حالت‌های مربوط به منبع ۱ (ستاره) و منبع ۳ (دایره) بر حسب شماره‌ی بازه‌های فرکانسی و $l=1\text{ m}$ ۵۰
- شکل (۳-۱۲): تابع SCT و MSCT در حالت دو منبع و $l=1\text{ m}$ ۵۱
- شکل (۴-۱): مدل انتشار چندمسیری [۲۱] ۶۰
- شکل (۴-۲): مدل انتشار چندمسیری [۲۱] ۶۱

شکل (۳-۴): معکوس فیلتر بازترکیب مستخرج از BSS، شکل بالایی دامنه و شکل پایینی توان را بر حسب dB نشان می‌دهد [۲۳].....	۶۴
شکل (۴-۴): دید نزدیک از ۱۰ میلی‌ثانیه از عکس فلیتر بازترکیب از منبع تا میکروفون نوعی. قسمت‌های مربوط مسیر مستقیم و بعضی قسمت‌های انعکاسی برجسته شده‌اند [۲۳].....	۶۵
شکل (۵-۴): توابع تخمین زده شده از روش‌های CC-BSS و GCC-BSS. محل ماکزیمم‌ها نشان‌دهنده‌ی مقادیر TDOA تخمینی هستند.....	۷۱
شکل (۶-۴): تابع همبستگی متقابل تخمین زده شده از روش CC-BSS. محل ماکزیمم‌ها نشان‌دهنده‌ی مقادیر TDOA تخمینی هستند.....	۷۳
شکل (۷-۴): مقادیر خطای روش‌های SCT و CC-PHAT به ازای زمان‌های انعکاس مختلف.....	۷۳
شکل (۸-۴): توابع تخمین زده شده از روش‌های CC-BSS و GCC-BSS در $T_{60}=200\text{ms}$. محل ماکزیمم‌ها نشان‌دهنده‌ی مقادیر TDOA تخمینی هستند.....	۷۴
شکل (۹-۴): توابع تخمین زده شده از روش‌های CC-BSS و GCC-BSS در $T_{60}=400\text{ms}$. محل ماکزیمم‌ها نشان‌دهنده‌ی مقادیر TDOA تخمینی هستند.....	۷۴
شکل (۱۰-۴): توابع تخمین زده شده از روش‌های CC-BSS و GCC-BSS در $T_{60}=600\text{ms}$. محل ماکزیمم‌ها نشان‌دهنده‌ی مقادیر TDOA تخمینی هستند.....	۷۵
شکل (۱۱-۴): توابع تخمین زده شده از روش‌های GCC-BSS-Gaussian و GCC-BSS-PHAT در $T_{60}=600\text{ms}$. محل ماکزیمم‌ها نشان‌دهنده‌ی مقادیر TDOA تخمینی هستند.....	۷۶
شکل (۱۲-۴): توابع تخمین زده شده از روش‌های CC-BSS-Gaussian و GCC-BSS-PHAT در $T_{60}=700\text{ms}$. محل ماکزیمم‌ها نشان‌دهنده‌ی مقادیر TDOA تخمینی هستند.....	۷۶
شکل (۱۳-۴): مقادیر خطای روش‌های SCT، GCC-PHAT و BSS-PHAT به ازای زمان‌های انعکاس مختلف.....	۷۷

فهرست جداول

صفحه	عنوان
۵۲	جدول (۱-۳) (الف): عملکرد روش SCT در محیط‌های با نویز مختلف
۵۲	جدول (۱-۳) (ب): عملکرد روش MSCT در محیط‌های با نویز مختلف
۵۳	جدول (۱-۳) (ج): عملکرد روش WSCT در محیط‌های با نویز مختلف
۵۴	جدول (۲-۳) (الف): عملکرد روش SCT در محیط‌های با پژواک مختلف
۵۴	جدول (۲-۳) (ب): عملکرد روش MSCT در محیط‌های با پژواک مختلف
۵۴	جدول (۲-۳) (ج): عملکرد روش WSCT در محیط‌های با پژواک مختلف
۵۵	جدول (۳-۳) (الف): عملکرد روش SCT در فواصل میکروفونی مختلف
۵۵	جدول (۳-۳) (ب): عملکرد روش MSCT در فواصل میکروفونی مختلف
۵۵	جدول (۳-۳) (ج): عملکرد روش WSCT در فواصل میکروفونی مختلف
۵۶	جدول (۴-۳): عملکرد روش WSCT با توابع وزن‌دهی مختلف

فصل اول:

مقدمه

۱-۱- مقدمه

مردم به طور عادی، با استفاده از گوش‌های خود، جهت صدای آمده از اطراف را تشخیص می‌دهند. ترکیب سیگنال‌های تقریباً متفاوت رسیده به گوش‌ها، ما را قادر به تشخیص حسی جهت صدای رسیده به گوش‌ها می‌سازد. به طور مشابه، با الهام از این قابلیت زیستی، می‌توان با استفاده از آرایه‌ای از میکروفون‌ها، یک سیستم مکان‌یابی صدا بنا نمود. اضافه بر این، با استفاده از همچون سیستمی از میکروفون‌ها می‌توان هر صدای منحصربه‌فرد را از مخلوط صداهای تولیدشده به صورت همزمان توسط چند منبع، استخراج نمود. چنین وضعیتی یادآور مسئله‌ی تلاش برای تمرکز روی صدای یک گوینده، در میان گفتگوهای ناهماهنگ و نویز موجود در محیط، است.

مکان‌یابی منبع، از مباحثی است که در زمینه‌های مختلفی همچون اتاق کنفرانس هوشمند، هدایت دوربین به سمت گوینده و... کاربرد دارد. جداسازی کور منابع (BSS^۱)، از حوزه‌هایی است که تحقیقات زیادی بر روی آن صورت گرفته است و هنوز راه‌حل عمومی برای آن ارائه نشده است. در این پایان‌نامه به تحقیق و بررسی در مورد کاربرد این روش برای حل مسئله‌ی مکان‌یابی چندگوینده پرداخته‌ایم.

۱-۲- معرفی روش جداسازی کور منابع (BSS)

جداسازی کور منابع (BSS) یکی از روش‌های نسبتاً جدید در پردازش سیگنال‌های دیجیتال است که در دو دهه اخیر بسیار مورد توجه قرار گرفته است [۱]. هدف اصلی این روش، جداسازی منابع ترکیب شده با کانال‌های نامعلوم، تنها با مشاهده‌ی سیگنال‌های ترکیب شده است. این تکنیک در بسیاری از زمینه‌های علمی همچون علوم زیستی، پردازش سیگنال‌های حیاتی، مخابرات دیجیتال و پردازش گفتار کاربرد دارد. کلمه‌ی «کور^۲» نشان دهنده‌ی آن است که جداسازی بدون استفاده از هرگونه اطلاعاتی درباره کانال‌های ترکیب کننده^۳ یا منابع صورت می‌گیرد.

^۱ Blind Source Separation

^۲ Blind

^۳ Mixing channels

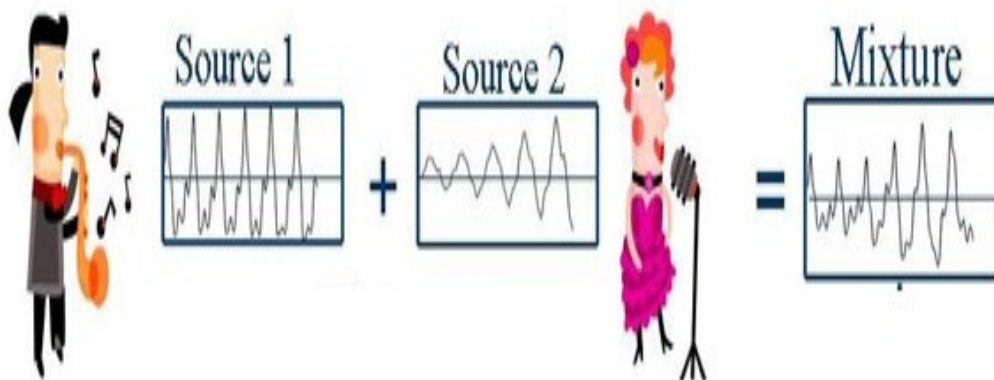
با در نظر گرفتن N منبع به صورت $s(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$ و M سیگنال ترکیب شده به صورت $x(t) = [x_1(t), x_2(t), \dots, x_N(t)]$ خواهیم داشت:

$$x(t) = H[s(t)] + e(t) \quad (1-1)$$

که $H[.]$ یک تابع انتقال عمومی^۱ و $e(t)$ بردار ستونی نشان‌دهنده‌ی نویز جمع‌شونده^۲ است. با فرض معکوس‌پذیری $H[.]$ هدف سیستم جداکننده‌ی منابع، تخمین تابع $W[.]$ ، برای بازسازی سیگنال‌های اصلی $s(t)$ به صورت زیر است:

$$Y(t) = W[x(t)] \quad (2-1)$$

بسیاری از روش‌های جداسازی منابع، روش‌های جداسازی کور هستند. کلمه‌ی کور به معنی تخمین تابع $W[.]$ بدون استفاده از اطلاعات اضافی و تنها با استفاده از سیگنال‌های ترکیب‌شده $x(t)$ است. در عمل هیچگاه نمی‌توان به صورت کاملاً کور به جداسازی دست یافت و ناچاریم حداقل از اطلاعات اولیه‌ای همچون اطلاعات آماری منابع استفاده کنیم. با این وجود روش BSS به دلیل عمومیت بیشتر و حساسیت کمتر به خطای مدل‌سازی سیستم ترکیب-بازترکیب و همچنین ویژگی‌های منابع از سایر روش‌های بهسازی گفتار^۳ متفاوت است. شکل‌های (۱-۱) و (۲-۱) به ترتیب نشان‌دهنده‌ی نحوه‌ی ترکیب و جداسازی دو سیگنال نوعی با استفاده از روش‌های جداسازی کور منابع را نشان می‌دهند.



شکل (۱-۱): مدل عمومی ترکیب برای سیستم BSS [۲]

^۱ Generic transfer function

^۲ Additive noise

^۳ Speech enhancement

پیچیدگی سیستم BSS تا حد زیادی به مدل مورد استفاده برای نشان دادن تابع انتقال عمومی وابسته است. در حالتی که منابع، صوتی باشند مدل‌های زیر بیشتر مورد توجه قرار می‌گیرند:

الف) مدل ترکیبی لحظه‌ای^۱

ب) مدل ترکیبی بدون انعکاس^۲

ج) مدل ترکیبی کانولوشنی^۳

ساده‌ترین حالت با توجه به مدل ترکیبی لحظه‌ای، زمانی است که سیگنال‌های مشاهده شده به صورت ترکیب خطی سیگنال‌های منابع باشند. مسائل موجود در این حالت تا حد زیادی به وسیله‌ی روش‌های آنالیز مؤلفه‌های مستقل (ICA^۴) حل شده است [۱]. هدف روش‌های ICA یافتن حضور داده‌ی غیرگوسی^۵ است پس مؤلفه‌های تخمین زده شده تا جای ممکن، به صورت آماری از یکدیگر مستقل خواهند بود.

در حالت دوم، BSS را می‌توان برای منابع ترکیب شده در شرایط بدون انعکاس بکار گرفت در این حالت، سیگنال‌های مشاهده شده به صورت ترکیب خطی مقیاس‌شده^۶ و تاخیر یافته‌ی سیگنال منابع اصلی هستند. بیشتر الگوریتم‌های ICA را می‌توان به این حالت تعمیم داد که مؤلفه‌ها با استفاده از مقادیر مختلط با یکدیگر ترکیب شده‌اند.

چالش برانگیزترین حالت، زمانی است که منابع توسط کانال‌های کانولوشنی با یکدیگر ترکیب شده‌اند. این وضعیت در کاربردهای صوتی که سیگنال‌ها عموماً در محیط انعکاسی ضبط شده‌اند، بسیار معمول است.

^۱ Instantaneous mixing model

^۲ Anechoic mixing model

^۳ Convolutional mixing model

^۴ Independent Component Analysis

^۵ Nongaussian

^۶ Scaled



شکل (۲-۱): مدل عمومی باز ترکیب برای سیستم BSS [۲]

۱-۲-۱- مدل ترکیبی لحظه‌ای

ساده‌ترین راه مدل‌سازی سیگنال مشاهده شده، در نظر گرفتن $x(t)$ به صورت ترکیب خطی لحظه‌ای از سیگنال‌های منبع، به صورت زیر است:

$$x(t) = Hs(t) \quad (۳-۱)$$

که $H(\cdot)$ نشان دهنده‌ی ماتریس ترکیب‌کننده است که برای سادگی تغییرناپذیر با زمان در نظر گرفته شده است. فرض کنید که تعداد سیگنال‌های مشاهده شده بیشتر از سیگنال‌های منبع باشند ($M \geq N$). در این حالت منابع اصلی را می‌توان با استفاده از تخمین شبه معکوس^۱ $H(\cdot)$ به صورت زیر به دست آورد:

$$W = H^+ = (H^T H)^{-1} H^T \quad (۴-۱)$$

که در حالتی که $M = N$ این مقدار برابر با معکوس $H(\cdot)$ است. بنابراین سیگنال‌های منبع به صورت زیر محاسبه می‌شوند:

$$y(t) = Wx(t) \quad (۵-۱)$$

اگر $N > M$ باشد، حالت نامعینی^۲ پیش می‌آید زیرا در این حالت ماتریس $H^T H$ دارای رتبه‌ی کامل^۳ نمی‌باشد در نتیجه معکوس کاذب W وجود نخواهد داشت. بنابراین تخمین سیگنال‌های خروجی با استفاده از باز ترکیب خطی بردار مشاهدات ممکن نمی‌باشد؛ با این وجود هنوز به صورت تئوری امکان تخمین ماتریس ترکیب‌کننده‌ی H وجود دارد.

^۱ Pseudo

^۲ Underdetermined

^۳ Full rank

۱-۲-۲- BSS برای ترکیب‌های کانولوشنی

مدل خطی برای جداسازی منابع صوتی در محیط‌های واقعی چندان مفید نیست. در حقیقت با مدل ترکیب خطی ساده به خوبی نمی‌توان سیستم ترکیب‌کننده را مدل کرد. در محیط داخل اتاق، میکروفون‌ها تعداد زیادی از سیگنال‌های مشابه سیگنال منبع اصلی و ناشی از انتشار در مسیرهای انعکاسی متفاوت را ضبط می‌کنند؛ بنابراین این سیگنال‌ها با تاخیرهای زمانی متفاوت به آرایه‌های میکروفونی می‌رسند. در نتیجه مدل بهتر برای مدل‌سازی سیگنال‌های ضبط شده با در نظر گرفتن جمع کانال‌های سیگنال منابع اصلی با پاسخ ضربه‌ی بین هر منبع تا میکروفون بدست می‌آید:

$$x(t) = h(t) * s(t) \quad (۶-۱)$$

که * نشان دهنده‌ی عملگر کانولوشن و $h(t)$ ماتریس پاسخ‌های ضربه‌ی محیط به صورت زیر می‌باشد:

$$h(t) = \begin{pmatrix} h_{11}(t) & \dots & h_{1N}(t) \\ \vdots & \ddots & \vdots \\ h_{M1}(t) & \dots & h_{MN}(t) \end{pmatrix} \quad (۷-۱)$$

اگر $M \geq N$ باشد، سیگنال‌های خروجی به صورت زیر است:

$$y(t) = w(t) * x(t) \approx s(t) \quad (۸-۱)$$

که $w(t)$ ماتریس دکانولوشن^۱ به صورت زیر است:

$$w(t) = \begin{pmatrix} w_{11}(t) & \dots & w_{1M}(t) \\ \vdots & \ddots & \vdots \\ w_{N1}(t) & \dots & w_{NM}(t) \end{pmatrix} \quad (۹-۱)$$

در تحقق عملی BSS کانولوشنی فیلترهای بازترکیب $w(t)$ به صورت گسسته و دارای طول محدود L می‌باشند که باید بر اساس زمان انعکاس اتاق و فرکانس نمونه برداری f_s سیگنال $x(t)$ انتخاب شود. تعداد عناصری که باید تخمین زده شوند، $M \times N \times L$ است؛ بنابراین

^۱ deconvolution

پیچیدگی متناسب با طول فیلترهای بازترکیب^۱ که برای جبران اثر فیلترهای ترکیب کننده استفاده می‌شوند، افزایش می‌یابد؛ بنابراین فرمول‌بندی الگوریتم BSS برای ترکیب کانولوشنی بسیار پیچیده‌تر از ترکیب لحظه‌ای است؛ زیرا تعداد زیادی پارامتر را باید باهم بهینه کرد. این پیچیدگی مانعی برای BSS کانولوشنی در حوزه‌ی زمان است؛ زیرا رفتار همگرایی آن با افزایش L ناپایدارتر می‌شود.

۱-۲-۳- ارتباط بین مکان‌یابی و BSS

جداسازی منابع صوتی یکی از موضوعات نسبتاً جدید در حوزه‌ی پردازش گفتار است، که به جداسازی همزمان چندین منبع صوتی به وسیله‌ی سیگنال‌های ضبط شده از یک یا چند میکروفون کمک می‌کند. در میان روش‌های متعدد جداسازی کور منابع، روش‌هایی که بر اساس آنالیز مؤلفه‌های مستقل (ICA) در حوزه‌ی فرکانس هستند، بیشتر مورد بررسی قرار گرفته‌اند. الگوریتم‌های جداسازی کور منابع (BSS) به صورت مجموعه‌ای از شکل‌دهنده‌های وفقی کور^۲ در نظر گرفته می‌شوند. گرچه این امر تحت شرایط واقعی کاملاً تحقق پذیر نیست، اما این بیان دیدی درباره‌ی قابلیت هدایت خودکار تکنیک‌های BSS می‌دهد. در واقع، در حالی که برای هدایت یک شکل‌دهنده‌ی پرتو نیاز به اطلاعات دقیق از مکان منبع داریم، BSS بدون استفاده از این اطلاعات اولیه امکان بازیابی سیگنال‌های منبع از مخلوط صداها را فراهم می‌سازد. از این قابلیت هدایت خودکار این‌طور استنباط می‌شود که فیلترهای تجزیه‌کننده اطلاعات مفیدی درباره‌ی مکان منبع در اختیار دارند؛ کافی است این اطلاعات از فیلترهای BSS استخراج شوند. در این فصل، روابط بین BSS و روش‌های شکل‌دهندگی پرتو مورد بررسی قرار می‌گیرند.

۱-۳- مکان‌یابی با استفاده از روش‌های مبتنی بر BSS

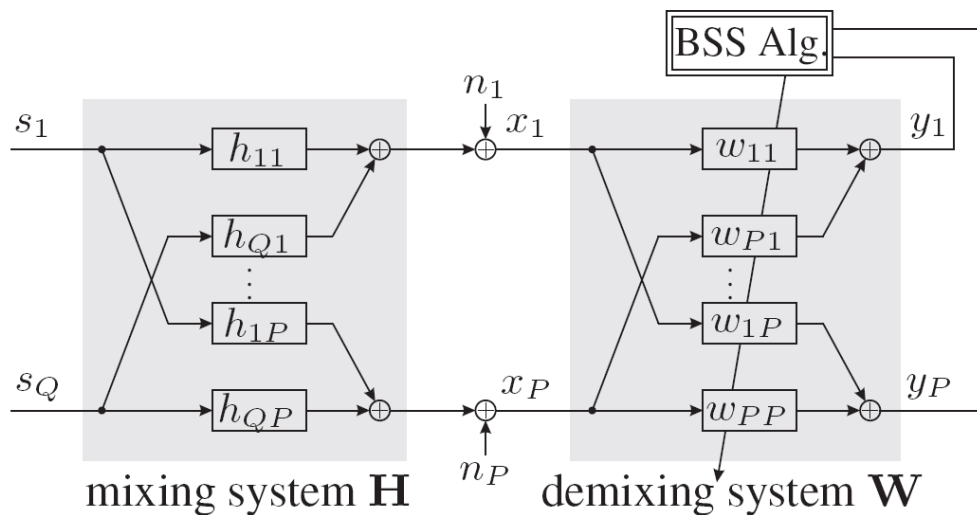
مکان‌یابی منابع صوتی از طریق تخمین مکان یک یا چند منبع صوتی با بهره‌گیری از تنوع فضایی^۳ ارائه شده توسط آرایه‌های میکروفونی صورت می‌گیرد. مکان‌یابی صحیح یک یا چند منبع

^۱ Demixing Filters

^۲ Blind adaptive beamformer

^۳ Spatial diversity

صوتی در کاربردهای زیادی همچون، هدایت خودکار^۱ شکل‌دهنده‌ی پرتو یا هدایت دوربین در راستای منبع صوتی، مورد استفاده قرار می‌گیرد. شکل (۳-۱) یک سیستم BSS عمومی را نشان می‌دهد. به دلیل وجود انعکاس در محیط آکوستیکی^۲، Q سیگنال منبع $s_q (q = 1 \dots Q)$ توسط سیستم ترکیب‌کننده‌ی چندورودی-چندخروجی (MIMO^۳) H ، که به صورت M اتصال فیلتر با پاسخ ضربه‌ی محدود (FIR^۴) h_{qp} ، بین q امین منبع و p امین میکروفون مدل شده است، فیلتر شده‌اند. تعداد P سیگنال ترکیب‌شده‌ی $x_p (p = 1 \dots P)$ ، به همراه نویز زمینه‌ای یا نویز میکروفون‌ها، n_p ، توسط آرایه‌های میکروفونی گرفته می‌شوند. سیگنال‌های منبع s_q ، به صورت دو به دو مستقل^۵ از یکدیگر فرض می‌شوند (یک فرض کلی برای سیگنال‌های صوتی). برای جداسازی سیگنال‌های منابع s_q بدون دسترسی به سیستم ترکیب‌کننده‌ی آکوستیکی H ، الگوریتم BSS سعی دارد تا با تنظیم وزن‌های سیستم بازترکیب W ، سیگنال‌های خروجی را به صورت آماری از یکدیگر مستقل بدست آورد؛ بنابراین L اتصال فیلترهای FIR جداکننده‌ی w_{pq} بین p امین میکروفون و q امین خروجی بدست می‌آید.



شکل (۳-۱) نحوه‌ی عملکرد سیستم BSS [۳]

^۱ Self steering

^۲ Acoustical

^۳ Multi-Input-Multi-Output

^۴ Finite Impulse Response

^۵ Mutually independent