عنوان پروپوزال به فارسی:

جداسازی سیگنال های صوتی با الهام از عملکرد سیستم شنوایی و روانشناسی هیئت نگر (گشتالت)

عنوان پروپوزال به انگلیسی:

Audio Signal Separation with Inspire of Auditory System Function and Gestalt Theory

بیان مسئله

در این تحقیق، عملیات جداسازی سیگنال های صوتی با در نظر گرفتن محدوده شنوایی و عملکرد آن مبتنی بر روانشناسی هیئت نگر (گشتالت)، صورت خواهد گرفت. برای این منظور، در ابتدا یک پایگاه داده از سیگنال های صوتی دریافت می شود که این سیگنال ها به صورت مخلوط می باشند. در هر کدام از این سیگنال های مخلوط، یک فیلتر بانک اجرا می شود و سپس با استفاده از تبدیل فوریه، ویژگی هایی از سیگنال مشخص می گردد. در ادامه، سیگنال های مخلوط فیلتر شده، به زیر باندهایی تبدیل می شوند که در هر یک از این زیرباندها، روش ICA جداگانه اعمال می شود تا به یک سری سیگنال های جداگانه تقسیم گردد. در ادامه برای گروه بندی سیگنال های جدا شده و شناسایی نوع ارتباط سیگنال ها با هم، از اصول گشتالت استفاده می گردد. تعریف چهار مفهوم اساسی این تحقیق که حوه اصلی را پیش می برد، امری لام است که در ذیل صورت گرفته است:

- ✓ جداسازی سیگنال صوتی: جداسازی سیگنال های صوتی یکی از مشکلات اساسی پردازش سیگنال دیجیتال در سیستم های مختلف اطلاعات، پیش بینی و کنترل است. برای تشخیص نقاط تغییر سیگنال می توان روش های زیادی را اعمال کرد. هدف از جداسازی سیگنال های صوتی، به دست آوردن یک سری ویژگی ها ا سیگنال مرجع است که در فرایند پردازشی به کار می آید.
- ✓ تعریف BSS: به مشکلی اشاره دارد که منابع و روش اختلاط ناشناخته است و فقط سیگنال های مخلوط برای فرایند جداسازی بیشتر در دسترس هستند. در چندین موقعیت، بازیابی همه منابع منفرد از سیگنال مخلوط یا حداقل تفکیک یک منبع خاص مطلوب است. در شرایط آزمایشگاهی، اکثر الگوریتم ها در جایی که

گروه نامیرا - www.nmze.com

سیگنال های ورودی هستند، بسیار خوب کار می کنند. منبع موجود در سیگنال های مخلوط، روش اختلاط و غیره برای فرآیند جداسازی کاملا شناخته شده است.

- ✓ تعریف ICA: یک روش یادگیری ماشین برای جدا کردن منابع مستقل از یک سیگنال مخلوط است. برخلاف تجزیه و تحلیل مولفه های اصلی که بیشترین واریانس نقاط داده را متمرکز می کند، تجزیه و تحلیل مولفه های مستقل بر استقلال، یعنی مولفه های مستقل تمرکز دارند.
- ✓ قضیه کوکتل پارتی: اثر کوکتل پارتی، پدیده توانایی مغز در تمرکز توجه شنیداری فرد به یک محرک خاص در حالی که طیف وسیعی از محرک های دیگر را فیلتر می کند، مانند زمانی که یک طرف کوکتل پارتی می تواند بر روی یک مکالمه در یک اتاق پر سر و صدا تمرکز کند، می باشد.

اهداف تحقیق

- ✓ ارائه یک روش بهینه و جدید جهت جداسازی سیگنال های صوتی با رویکرد ترکیبی BSS و ICA با در نظر گرفتن نظریه گشتالت
- ✓ در نظر گرفتن عملیات پیش پردازش، استخراج ویژگی و طبقه بندی برای تخمین و جداسازی سیگنال های صوتی با حداقل خطا و بهترین دقت نسبت به روش های پیشین

فرضیات تحقیق

- ✓ با در نظر گرفتن نظریه گشتالت و عملکرد سیستم شنوایی، می توان یک رویکرد بهینه جداسازی سیگنال های صوتی ارائه کرد.
- ✓ رویکرد ترکیبی BSS و ICA می تواند جداسازی سیگنال ها را انجام بدهد.
- ✓ دقت رویکرد پیشنهادی نسبت به روش های پیشین بالاتر و بهینه است و نقاط ضعف روش های پیشین را برطرف و پوشش می دهد و نقاط قوت ان را را نیز پشتیبانی می نماید.

www.nn٤e.com – گروه نامیرا

سوالات تحقیق

- ✓ چگونه می توان از نظریه گشتالت و در نظر گرفتن عملکرد سیستم شنوایی، یک رویکرد بهینه جداسازی سیگنال های صوتی ارائه نمود؟
- ✓ آیا رویکرد ترکیبی BSS و ICA می تواند جداسازی سیگنال ها را انجام بدهد؟
- ✓ دقت رویکرد پیشنهادی تا چه اندازه است و چه شکاف های تحقیقاتی را در مسائل و پیشینه تحقیق، حل می نماید؟

مروری بر ادبیات و پیشینه تحقیق

افزایش بیش از حد داده های چندرسانه ای از طریق اینترنت تغییر بزرگی به سمت خدمات آنلاین ایجاد کرده است. در بیشتر برنامه های چندرسانه ای، اطلاعات صوتی بخش مهمی است. رایج ترین و محبوب ترین نمونه اطلاعات آنلاین، موسیقی است [۱].تجزیه و تحلیل صوتی، تجزیه و تحلیل فیلم و درک مطلب را می توان با جداسازی و طبقه بندی یک جریان صوتی بر اساس محتوای آن انجام داد [۲]. برای این منظور، روشی کارآمد و دقیق لازم است که بخش هایی از یک جریان صوتی را جداسازی می کند. تکنیکی که در آن یک جریان صوتی به مناطق همگن (مشابه) تقسیم می شود ، جداسازی سیگنال صوتی نامیده می شود [۱]. سیگنال صوتی در توالی های کوتاه نمونه برداری می شوند که در چندین کلاس طبقه بندی می شوند. سپس هر توالی بسته به کلاس متعلق به آن، قابل تجزیه و تحلیل است. ظهور چندرسانه ای و فناوری شبکه منجر به ظهور داده های دیجیتال می شود و این باعث افزایش علاقه به بازیابی اطلاعات مبتنی بر محتوای چندرسانه ای می شود. برای تجزیه و تحلیل و درک یک سیگنال صوتی، گام اساسی برای تمایز یک سیگنال صوتی بر اساس محتوای آن است. طبقه بندی صدا و جداسازی یک مسئله تشخیص الگو است. این مسئله شامل دو مرحله اصلی است: استخراج ویژگی ها و سپس طبقه بندی بر اساس این ویژگی ها یا اطلاعات آماری استخراج شده [۳]. کاربردهای تجزیه و تحلیل محتوای صوتی را می توان در دو بخش طبقه بندی کرد. یک بخش، تفکیک جریان صوتی به مناطق همگن و بخش دیگر، تفکیک جریان گفتار به بخش ها، از بلندگوهای مختلف است.

گروه نامیرا - www.nmze.com

جداسازی سیگنال های صوتی و طبقه بندی آن ها کاربردهای بسیاری دارد. طبقه بندی و بازیابی صوتی مبتنی بر محتوا بیشتر در صنعت سرگرمی، مدیریت بایگانی صوتی، استفاده از موسیقی به صورت تجاری، نظارت و موارد دیگر استفاده می شود. امروزه، در شبکه جهانی وب، میلیون ها پایگاه داده وجود دارند. برای جستجوی صوتی و فهرست بندی این داده های سیگنال، از جداسازی صوتی و طبقه بندی آن ها استفاده می شود. الگوریتم های جداسازی سیگنال های صوتی را می توان به سه دسته کلی تقسیم کرد. در دسته اول، طبقه بندی ها طراحی می شوند[۴]. ویژگی ها در دامنه زمانی و دامنه فرکانس استخراج می شوند. سپس از طبقه بندیگر برای تفکیک سیگنال های صوتی بر اساس محتوای آن استفاده می شود. دسته دوم جداسازی سیگنال های صوتی ویژگی هایی را در آمار استخراج می کند که توسط طبقه بندی کننده برای تبعیض استفاده می شود. به این نوع ویژگی ها، ویژگی های مبتنی بر احتمال خلفی[۱] گفته می شود. مقادیر زیادی از داده های آموزش توسط طبقه بندیگر برای ارائه نتایج دقیق مورد نیاز است. دسته سوم الگوریتم جداسازی سیگنال صوتی بر تنظیم طبقه بندی های موثر تاکید دارد .طبقه بندی کننده های استفاده شده در این گروه عبارتند از: معیار اطلاعات بیزی[۲]، نسبت احتمال گاوسی[۳] و طبقه بندی کننده مدل پنهان مارکوف[۴]. این طبقه بندی ها همچنین هنگامی که داده های آموزشی بزرگی ارائه می شود نتایج خوبی می دهند [۴].

طرح طبقه بندی ترکیبی به منظور طبقه بندی فایل های سیگنال صوتی به انواع داده های اصلی ارائه شده است. قبل از طبقه بندی یک مرحله از پیش طبقه بندی انجام می شود که هر قاب پنجره ای از سیگنال صوتی را به طور جداگانه تجزیه و تحلیل می کند. سپس مرحله استخراج ویژگی انجام می شود که از آن یک بردار ویژگی عادی به دست می آید. پس از استخراج ویژگی از روش طبقه بندی ترکیبی استفاده می شود. مرحله اول سیگنال / فریم های صوتی را با استفاده از یک روش، طبقه بندی می کند. از آن جا که فریم های سکوت بیشتر در سیگنال گفتار وجود دارند، بنابراین بخش گفتار بر اساس طبقه بندی مبتنی بر قاعده به بخش های سکوت و گفتار خالص طبقه بندی می شود. سرانجام، ازیک طبقه بندی دیگر برای تمایز بیشتر بخش های غیر حرفه ای در بخش موسیقی صدا و محیط استفاده می شود. این طرح ترکیبی برای دستیابی به دقت طبقه بندی بالا استفاده می شود و می تواند برای برنامه های مختلف در زمان واقعی چندرسانه ای مورد استفاده قرار گیرد. در صورت کلی جداسازی سیگنال های صوتی، جریان

[1] posterior probability
[2] Bayesian information criterion
[3] Gaussian likelihood ratio
[4] Hidden Markov model (HMM)

صوتی به عنوان ورودی گرفته می شود، سپس به ۸۰۰۰ کیلوهرتز کاهش می یابد، مرحله پیش طبقه بندی بر روی این جریان صوتی اعمال می شود، دارای سرعت عبور صفر، انرژی کوتاه مدت، شار طیف، ضرایب غشایی با فرکانس Mel و تجزیه و تحلیل تناوبی است. در انتها نیز یک روش طبقه بندی ترکیبی استفاده می شود. با استفاده از ویژگی های عبور صفر، انرژی کوتاه مدت، شار طیف، و ضرایب سپسترال[1] با فرکانس Mel، سیگنال صوتی به بخش گفتاری و غیر واژگان طبقه بندی می شود. از ویژگی های شار طیف، آنالیز تناوبی و ضرایب سپسترال با فرکانس Mel استفاده می شود و بخش های غیر متناسب با استفاده از یک ساختار بهینه به موسیقی و صدای محیط طبقه بندی می شوند. از طبقه بندیگر مبتنی بر قاعده برای تبعیض بخش های سکوت و گفتار استفاده می شود. تجزیه و تحلیل گفتار در فایل های صوتی یک مشکل پیچیده است و سیستم های عملکردی بهبود یافته مورد نیاز است. در بسیاری از برنامه های پردازش صوتی، جداسازی سیگنال های صوتی نقش مهمی در مرحله پیش پردازش بازی می کند. همچنین تاثیر بسزایی در عملکرد تشخیص گفتار دارد. به همین دلیل یک الگوریتم طبقه بندی و جداسازی سیگنال های صوتی سریع و بهینه شده در این تحقیق ارائه می شود که می تواند برای برنامه های بلادرنگ چندرسانه ای استفاده شود. ورودی صوتی به چهار نوع اصلی صوتی طبقه بندی شده و تقسیم می شود: گفتار خالص، موسیقی، صدای محیط و سکوت. در این تحقیق الگوریتمی ارائه شده است که به داده های آموزشی کمتری نیاز دارد و از آن به دقت بالایی می توان رسید، یعنی، میزان سوء طبقه بندی حداقل است.

جداسازی منبع کور یا BSS[2] اولین بار در سال ۱۹۸۲ با بحث ساده بین برنارد آنس[3]، جانی هررو[4] و کریستین جوتن[5] با ژان پیر رول[6]، یک متخصص علوم اعصاب، در مورد رمزگشایی حرکت در مهره ها مورد بررسی قرار گرفت. حرکت مفصل به دلیل انقباض عضلات صورت می گیرد. هر فیبر عضلانی توسط یک موتور نورون توسط مغز کنترل می شود. علاوه بر این، بر روی هر فیبر، میزان انقباض عضله اندازه گیری می شود و اطلاعات مشابه توسط دو نوع انتهای حسی، که در تاندون قرار دارند، به سیستم عصبی مرکزی منتقل می شود و انتهای اولیه و فرعی نامیده می شود. نتایج با میانگین پاسخ های حسی واحد حاصل از تعداد زیادی از الیاف، مربوط به تکرار همان حرکت اجباری حاصل می شوند. با کمال تعجب، در حالی که می توان تصور کرد که هر نوع پایان فقط یک نوع اطلاعات

[1] cepstral
[2] Blind Source Separation (BSS)
[3] Bernard Ans
[4] JeannyHérault
[5] Christian Jutten
[6] Jean-Pierre Roll

را اعم از کشش یا سرعت منتقل می کند، اطلاعات حسی منتقل شده توسط انتها، ترکیبی از اطلاعات کششی و سرعت است. در جداسازی منبع کور یا BSS، مشاهدات متعدد توسط آرایه ای از حسگرها انجام می شود تا به منظور بازیابی ترکیب اولیه سیگنال های منبع، پردازش شوند. اصطلاح ناب به این واقعیت اطلاق می شود که هیچ اطلاعات خاصی در مورد روند اختلاط یا سیگنال های منبع موجود وجود ندارد. درک جداسازی از منبع کور مربوط به تجزیه و تحلیل مؤلفه های مستقل یا ICA[1] است. با این حال، ICA را می توان به عنوان ابزاری با هدف کلی جایگزین تجزیه و تحلیل مولفه های اصلی یا PCA[2] دانست که به معنای کاربرد آن برای طیف گسترده ای از مشکلات است. برخی از حوزه های برنامه جداسازی منبع کور عبارتند از: تجزیه و تحلیل سیگنال زیست پزشکی، پردازش داده های ژئوفیزیکی، داده کاوی، ارتباطات بیسیم و پردازش آرایه حسگر.

کاری که این تحقیق سعی در ارائه آن دارد، به کارگیری عملکرد سیستم شنوایی با کمک روان شناسی هئیت نگر یا گشتالت[3] است که به روش هوشمند مبتنی بر الگوریتم ها، عملیات جداسازی سیگنال های صوتی را انجام می دهد. در واقع قرار است از جداسازی منبع کور یا BSS[4] مبتنی بر الگوریتم تحلیل مولفه های اصلی یا ICA[5] انجام شود. اما تفسیر مراحل کلی کار و روش های موجود در آن نیز حادز اهمیت است. به صورت کلی، رویکردی که این تحقیق ارائه می دهد، به صورت ذیل است:

مرحله پیش پردازش: سیگنال صوتی به عنوان ورودی وارد سیستم می شود. جدا کردن منبع یا بخش های مورد نظر در چارچوب تجزیه و تحلیل مؤلفه مستقل به عنوان جداسازی منبع کور شناخته می شود [۷-۵]. منبع کور به طور کلی روشی است که برای جدا کردن سیگنال مختلط به منابع مستقل استفاده می شود (وقتی فرآیند اختلاط مشخص نیست) [۸]. اکثر تکنیک های جداسازی منبع کور از آمار مرتبه بالاتر استفاده می کنند. برای آمار مرتبه بالاتر این الگوریتم ها نیاز به محاسبات تکراری دارند [۹]. از روش ملجدی[6] و شوستر برای جداسازی سیگنال ها بر اساس آمار مرتبه دوم (همبستگی) استفاده می شود. این امر به آمار مرتبه بالاتر و محاسبات تکراری نیاز ندارد. ساختار زمانی سیگنال ها تجزیه و تحلیل می شود و جداسازی بر این اساس انجام می شود. سیگنال مختلط ابتدا با استفاده از تبدیل

[1] Independent Component Analysis (ICA)
[2] Principle Component Analysis (PCA)
[3] gestalt theory
[4] Blind Source Separating (BSS)
[5] Independent Component Analysis (ICA)
[6] Molgedey

فوریه در فواصل زمانی کوتاه به دامنه فرکانس زمان که به آن اسپکتروگرام سیگنال[1] نیز می گویند، تبدیل می شود. به منظور جلوگیری از اختلاط طیف سنج ها، هر طیف سنج به طور جداگانه مورد بررسی قرار می گیرد. همبستگی در تمام این فواصل کوتاه انجام می شود. مرحله چرخش[2] انجام می شود. بخش چرخشی به سیگنال های منبع متعامد به یک مختصات مشاهده ای، گفته می شود. مشاهده در واقع یک پیش بینی سیگنال های منبع در جهت معینی است و مشاهدات اصلی متعامد نیستند. این مشاهدات با استفاده از کلام مرتب شده اند، به گونه ای که به یکدیگر متصل می شوند. ابهام در چرخش، حتی پس از مشخص شدن، هم چنان باقی می ماند. بنابراین می توان چرخش صحیح را با از بین بردن تمام مشاهدات خارج از حالت موربی موجود در ماتریس همبستگی مورد بررسی قرار داد. مورب مصنوعی همزمان[3] [۱۱، ۱۰] در تاخیرها، چندین بار اعمال می شود. مرحله بازسازی در طیف سنج در هر سیگنال به طور مجزا انجام می شود. سپس تمام اجزای فرکانس تجزیه شده ترکیب می شوند. در پایان مرحله جابجایی برای یافتن رابطه بین سیگنال های جدا شده انجام می شود و تصمیم با استفاده از طبقه بندیگر گرفته می شود.

مرحله استخراج ویژگی[4]: فرآیند تبدیل یک سیگنال صوتی به دنباله ای از بردارهای ویژگی، فرآیند استخراج ویژگی نامیده می شود. بردارهای این ویژگی اطلاعات مشخصه زمانی و همچنین طیفی را در مورد سیگنال صوتی دارند. بردارهای ویژگی بر اساس پنجره محاسبه می شوند. انتخاب ویژگی، تاثیر زیادی در عملکرد سیستم های جداسازی سیگنال های صوتی دارد. سه نوع از ویژگی ها در این کار پیشنهادی محاسبه می شود: ضرایب سرعتی فرکانس Mel[5]، ویژگی های دامنه زمانی و دامنه فرکانس. برای شکل گیری یک بردار ویژگی ها، این ویژگی های عادی با هم ترکیب می شوند. در ابتدا جریان صوتی با سرعت نمونه برداری از ۸ کیلوهرتز به ۱۶ کیلوهرتز، به یک تکه بیت تبدیل می شود. مرحله استخراج ویژگی بر روی سیگنال های جدا شده به دست آمده پس از مرحله پیش طبقه بندی انجام می شود. این سیگنال های جدا شده به فریم های غیرمستقیم تقسیم می شوند. این فریم ها به عنوان واحد طبقه بندی استفاده می شوند. بر اساس طبقه بندی، نتایج جداسازی سیگنال صوتی، انجام می شود. همان طور که توسط [۱۱] پیشنهاد شده است، از مرتبه دوازدهم ضرایب سپسترال با فرکانس Mel استفاده می شود. ویژگی های

[1] Signal Spectrogram
[2] sphering
[3] Simultaneous diagonalization
[4] Feature Extraction
[5] Mel-frequency cepstral coefficients (MFCCs)

دامنه زمانی شامل نرخ عبور صفر[1]، انرژی کوتاه مدت[2] و تجزیه و تحلیل تناوبی[3] است. ویژگی فرکانس دامنه، شار طیف است. نرخ عبور صفر، اندازه گیری از تغییرات سیگنال است که از مثبت به منفی یا برعکس رخ می دهد. تعریف کلی میزان نرخ عبور صفر در یک قاب است. نرخ عبور صفر از زبان گفتار و موسیقی را به طور موثر متمایز می کند، زیرا گفتار حاوی مناطق ساکت تر در مقایسه با موسیقی است، بنابراین میزان عبور صفر برای گفتار از موسیقی بیشتر است. انرژی کوتاه مدت، اندازه گیری از کل انرژی یک قاب است. انرژی کوتاه مدت یک ویژگی مفید برای تمایز بخش های گفتار و موسیقی است. اندازه گیری انرژی کوتاه مدت برای سیگنال های گفتاری تغییرات زیادی در مقایسه با سیگنال های موسیقی دارد [۱۲]، زیرا ویژگی های فرکانس صدای انسان با دستگاه موسیقی بسیار متفاوت است. شار طیف، اندازه گیری از طیف توان متغیر سیگنال صوتی است. با محاسبه فاصله بین قاب فعلی و قاب قبلی محاسبه می شود. دقیقا با محاسبه فاصله اقلیدسی بین دو طیف نرمال، می توان شار طیف را به دست آورد. شار طیف در تمایز گفتار، موسیقی و صدای محیط کمک می کند. سیگنال های گفتاری در مقایسه با موسیقی دارای تغییرات طیفی بالاتری هستند. با این حال، برای صدای محیط، تنوع طیفی بالاتر از موسیقی و گفتار است [۱۲]. ضرایب سپسترال با فرکانس Mel یا MFCC، اندازه گیری لگاریتمی طیف Mel بزرگی است که توسط فیلتر باند انتقال مثلثی محاسبه می شود. این مقادیر با استفاده از تبدیل کسینوس گسسته انجام می شوند. MFCC یک اجرای واقعی با ارزش بسیار زیاد سپستروم پیچیده است. به همین دلیل است که با گرفتن تبدیل فوریه سریع، محاسبه می شود. تجزیه و تحلیل دوره ای را می توان با برآورد تناوبی هر قاب محاسبه کرد و تناوبی با همبستگی به دست آورد. تجزیه و تحلیل دوره ای برای موسیقی بالاتر از صدای محیط است، زیرا سیگنال های موسیقی در مقایسه با سیگنال های صدا در محیط، دارای ماهیت بیشتری هستند [۱۲].

روانشناسی گشتالت با هدف درک نظریه های موجود در توانایی دستیابی و حفظ درک معنادار در دنیایی آشکار است. اصل مهم گشتالت این است که ادراک کل جهانی را با تمایلات خود سازماندهی تشکیل می دهد. اصل اساسی درک گشتالت، قانون دسته بندی با نام مستعار Prägnanz است، از جمله هشت قانون گروه بندی شامل نزدیکی، شباهت، بستن، تقارن، سرنوشت مشترک، استمرار، ژست خوب و تجربه گذشته. در این تحقیق، از قوانین مشابهت و سرنوشت مشترک استفاده می شود تا به ترتیب عدم قطعیت شایستگی مکانی و زمانی را در جداسازی سیگنال های

[1] zero-crossing rate
[2] short-time energy
[3] periodicity analysis

صوتی با رویکرد ترکیبی BSS و ICA بتوان محاسبه نمود. قانون تشابه ادعا می کند که اگر عناصر مشابه یکدیگر باشند، عناصر تمایل به درک در یک گروه دارند. اصل سرنوشت رایج بیان می کند که اگر عناصر در یک سمت حرکت کنند، عناصر تمایل دارند به عنوان گروه درک شوند. اگر عناصر در همان جهت و سرعت حرکت کنند، آن ها حتی در مسافت های بزرگ به عنوان یک گروه درک می شوند. این دو قانون برای تخمین عدم قطعیت فضای مکانی و شفافیت زمانی به شرح زیر استفاده می شوند: اول، یک سیگنال صوتی که بیشتر به مرکزیت در یک سیگنال شباهت دارد، احتمالا یک سیگنال برجسته است. دوم، یک سیگنال صوتی به همان شیوه حرکت با مرکز حساسیت به احتمال زیاد یک سیگنال برجسته است. این مفهوم، اصل تئوری گشتالت را در حوزه این تحقیق، مورد نظر قرار می دهد [۱۳]. واضح و مبرهن است که تمامی حالات ممکن مبتنی بر عملکرد سیستم شنوایی، انجام می شود. همین طور از تئوری گشتالت در فاز استخراج ویژگی ها و همین طور طبقه بندی نهایی نیز استفاده می شود.

در زمینه جداسازی سیگنال های صوتی، مطالعات مختلفی صورت گرفته است. یکی از این مطالعات که در [۱۴] ارائه شده است، به جداسازی و طبقه بندی سیگنال ها با استفاده از اصول و عملکرد شنوایی انسان می پردازد که از ماشین بردار پشتیبان و شبکه های عصبی[1] برای جداسازی و طبقه بندی بهره می برد. بخش گفتار بر اساس طبقه بندی مبتنی بر قاعده، به بخش های سکوت و گفتار خالص طبقه بندی می شود. این تحقیق از حداقل داده ها برای طبقه بندی آموزش استفاده می کند.این رویکرد از روش های گروهی[2] برای به حداقل رساندن میزان طبقه بندی نادرست استفاده می کند و تقریبا ۹۸٪ قطعه های دقیق سیگنالی به دست می آورد. به صورت کلی در این مقاله، یک الگوریتم سریع و کارآمد طراحی شده است که می تواند با برنامه های چندرسانه ای در زمان واقعی مورد استفاده قرار گیرد.

انتخاب داده اقتباس برای جداسازی سیگنال های صوتی مبتنی بر یادگیری عمیق در [۱۵] با کاربرد در کانال پخش چند ژانری در وب سایت های موزیک ارائه شده است. روش جدید به مدل های گفتاری/غیر گفتاری مبتنی بر GMM متکی است که به صورت گفتاری و گفتاری آموزش داده می شود. از اطلاعات طولانی مدت برای انتخاب داده های آموزشی قابل اعتماد برای مدل های گفتاری/غیر گفتاری از گفتارهای موجود استفاده می شود. نتایج تجربی نشان می دهد که این روش انتخاب داده ها یک الگوریتم جداسازی سیگنال های صوتی قدرتمند به نوبه خود است. همچنین مشاهده شده است که شبکه های عصبی عمیق آموزش داده شده با استفاده از داده های انتخاب

[1] Neural Networks (NNs)
[2] Ensemble Methods

شده توسط این روش نسبت به داده های آموزش دیده با داده های انتخاب شده با دو روش مقایسه ای، برتر هستند. علاوه بر این، می توان با ترکیب روش جداسازی سیگنال های صوتی مبتنی بر یادگیری عمیق با روش انتخاب داده ها، عملکرد بهتری کسب کرد.

در [۱۶] تکنیک های ساده برای جداسازی سیگنال های صوتی در دو یا سه کلاس شرح داده می شود. این روش ها به دامنه، تحلیل طیفی و مدل های کلاسیک پنهان مارکوف متکی هستند. با توجه به محدودیت های این رویکردها، یک فرایند جداسازی سیگنال های صوتی دو سطح در این مقاله پیشنهاد شده است. جداسازی سیگنال های صوتی با محاسبه چندین ویژگی برای هر دنباله صوتی انجام می شود. این ویژگی ها یا در یک قطعه صوتی کامل یا بر روی یک قاب (مجموعه ای از نمونه ها) که زیر مجموعه بخش صوتی است، محاسبه می شود. روش پیشنهادی برای ریزنشت بخشیدن به داده های صوتی شامل ترکیبی از طبقه بندی کننده K-Means در سطح قطعه و یک سیستم مدل مارکوف پنهان چند بُعدی با استفاده از تجزیه قاب سیگنال است. طبقه بندی اول با استفاده از طبقه بندی K-Means و ویژگی های مبتنی بر قطعه به دست می آید. سپس نتیجه نهایی از استفاده از مدل های چند بُعدی پنهان مارکوف و ویژگی های مبتنی بر فریم شامل نتایج موقتی حاصل می شود. مدل های مارکوف پنهان چند بُعدی، افزونه ای از مدل های کلاسیک پنهان مارکوف اختصاص داده شده به داده های چند جزئی است. روش این تحقیق در زمینه تجزیه و تحلیل سیگنال های صوتی فوتبال نشان داده شده است.

در [۱۷] یک رویکرد جدید به مسئله جداسازی سیگنال های صوتی ارائه شده است که یک مرحله قبلی غیرقابل اجتناب برای نمایه سازی صوتی است. از اطلاعات متقابل برای ارزیابی صحت جداسازی سیگنال های صوتی استفاده می شود، به عنوان تابعی که توسط یک الگوریتم تبرید شبیه شبیه سازی یا SA[1] به صورت بیشینه یا حداکثر می شود. این تحقیق یک اپراتور جهش[2] جدید برای الگوریتم تبرید شبیه شبیه سازی معرفی نموده است به نام اپراتور متوالی Bits Mutation یا جهش بیت ها که عملکرد الگوریتم تبرید شبیه شبیه سازی را در این مسئله بهبود می بخشد. همچنین از اصطلاح عامل فشرده سازی استفاده می شود که به الگوریتم تبرید شبیه شبیه سازی اجازه می دهد تا در یک فضای جستجوی کاهش یافته فعالیت کند. الگوریتم ارائه شده در جداسازی سیگنال های صوتیاز یک

[1] Simulated Annealing (SA)
[2] Mutation

پایگاه داده صوتی واقعی مورد آزمایش قرار گرفته است و با چندین الگوریتم موجود برای جداسازی سیگنال های صوتی مقایسه شده است و در مشکلات آزمون در نظر گرفته شده، نتایج بسیار خوبی را کسب می کند.

رویکرد ارائه شده در [۱۸] از تئوری مجموعه فازی[1] برای ارزیابی جداسازی سیگنال های صوتی با وضوح و دقت بالا استفاده می کند. این روش براساس مقایسه مرزهای یافت شده به صورت خودکار با گراند تراث[2] است. از این رو، این روش دقیق تر و قادر به درک مسئله ارزیابی به روشی شبیه به ارزیابی انجام شده توسط یک انسان است که از عملکرد سیست شنواییانسان، استفاده می کند. روش های سنتی اغلب در درجه بندی الگوریتم های جداسازی سیگنال های صوتی، به ویژه آن هایی که از کیفیت های نسبتا مشابهی برخوردار هستند، ناکام می باشند. این تحقیق یک تابع عضویت[3] فازی را تعریف می کند که میزان اندازه گیری بخش های به دست آمده توسط یک روش خودکار را با نتایج یک جداسازی سیگنال های صوتی صحیح اندازه گیری می کند. برای شناسایی یک جفت از بخش های معادل، یک تابع هم ترازی فازی تنظیم می شود که جفت قطعه های به دست آمده توسط یک جداسازی خودکار را با بخش های مربوطه از یک جداسازی سیگنال های صوتی صحیح نشان می دهد.

در [۱۹] به جداسازی سیگنال های صوتی از یک خواننده معروف اپرای چین به نام Kungu پرداخته شده است. این رویکرد از ویژگی های نرخ عبور صفر استفاده می کند و نتایج آن نشان می دهد که دقت این روش تا حدود ۹۱,۶۱٪ می باشد. به کارگیری ساختار توالی-توالی اتوانکورد[4] به عنوان یک روش بدون نظارت[5] در [۲۰] به عنوان یک روش جداسازی سیگنال ارائه شده است که هر متن یا جمله در یک سیگنال را به یک بردار تبدیل می نماید، از این رو، به نام روش Audio Word2vec مطرح می باشد.

امکان بومی سازی اشیای بصری که با یک منبع صوتی در ارتباط هستند و در عین حال برای جدا کردن سیگنال صوتی، سنگ بنای برنامه های پردازش سیگنال صوتی و تصویری است. با این حال، روش های موجود عمدتا بر بومی سازی فقط اشیای بصری، بدون توانایی جداسازی صوتی متمرکز است. علاوه بر این، این روش ها اغلب به مراحل پیش پردازش دشوار برای جداسازی قاب های ویدئویی در مناطق معنایی یا نظارت های اضافی برای هدایت بومی سازی آن ها متکی هستند. در [۲۱]، مسئله محلی سازی منبع بصری و جداسازی صوتی به روشی بدون نظارت را

[1] Fuzzy
[2] Ground Truth
[3] Membership Function (MF)
[4] sequence-to-sequence auto encoder (SA)
[5] Unsupervised Learning

برطرف کرده است و از انجام کلیه مراحل پیش پردازش یا پردازش بعد از آن جلوگیری نموده است. برای این منظور، یک روش تجزیه ساختار ماتریس جدید طراحی شده است که ماتریس داده های هر روش را به عنوان یک ترکیب از سه اصطلاح تجزیه می کند: ۱) یک ماتریس با درجه پایین که اطلاعات پیش زمینه را ضبط می کند، ۲) یک ماتریس پراکنده که اجزای همبسته را در بین دو روش به دست می آورد و از این رو پرده برداری از منبع صدا در حالت بصری و صدای همراه در مد صوتی انجام می شود و ۳) یک ماتریس پراکنده سوم برای اجزای نامربوط مانند حواس پرتی اشیا در حالت تصویری و صدای بی ربط در مد صوتی را محاسبه می کند. عمومیت روش ارائه شده با استفاده از آن بر روی سه برنامه نشان داده شده است، یعنی: ۱) بومی سازی بصری یک منبع صدا، ۲) جداسازی صوتی با کمک نمایش بصری و ۳) تشخیص سخنگوی فعال. نتایج تجربی نشان دهنده اثربخشی روش پیشنهادی در این حوزه کاربردی است.

طی سالیان متمادی تعدادی روش ایجاد شده است تا ایستگاه پخش بتواند صدا را در ۷ کلاس تقسیم و طبقه بندی کند که عبارتند از: گفتار مرد، گفتار زن، گفتار با نویز، گفتار با موسیقی، نویز ، موسیقی و سکوت. برخی از روش های اصلی شامل ماشین بردار پشتیبان، GMM و مدل مخفی مارکُف که در بین روش های پیشین همین بخش نیز مورد بررسی قرار گرفت، در بین سایر موارد به شدت مورد استفاده و کاربردی است، اما صحت کلی این روش های موجود رضایت بخش نیست. در [۲۱]، یک روش شبکه عصبی کانولوشن تطبیقی[1] ارائه می شود که سیگنال های صوتی را مستقیما بر اساس نقاط نمونه صوتی طبقه بندی می کند. به صورت کلی می توان گفت، این اولین کاری است که می تواند جداسازی سیگنال و طبقه بندی را در حوزه پخش توسط شبکه عصبی عمیق کانولوشن تطبیقی انجام دهد. هر لایه جمع کننده معادل استخراج ویژگی سیگنال صوتی است. در این تحقیق از روش شبکه عصبی عمیق کانولوشن تطبیقی با افزایش میزان محاسبه استفاده شده است تا مناسب ترین ویژگی ها استخراج شود. به عبارت دیگر، شبکه عصبی عمیق کانولوشن تطبیقی روند انتخاب ویژگی را در مقایسه با شبکه عصبی کانولوشن افزایش می دهد. تعداد زیادی از آزمایشات انجام گرفته است و با سایر رویکردها مقایسه شده و نتایج نشان می دهد که تکنیک شبکه عصبی عمیق کانولوشن تطبیقی کاهش خطای قابل توجهی را به دست آورده است.

[1] Adaptive Convolution Neural Network (ACNN)

ضرورت انجام تحقیق

ضرورت اصلی این تحقیق از جایی نشات می گیرد که اولا بتوان ضعف روش های پیشین را تا حدی بهبود و پوشش داد و دوما سیستمی باشد که از نظریه گشتالت که مبتنی بر روان شناسی است، پیروی کند. یکی از ضعف های مهم، عدم ارائه یک ساختار جداسازی سیگنال های صوتی با در نظر گرفتن عملکرد شنوایی انسان می باشد و در آن از مراحل صحیح استخراج ویژگی ها و طبقه بندی بر اساس نوع داده ها استفاده نشده است. همین طور وجود ضعف هایی در BSS، استفاده از ICA را ملزم می کند تا بتوان ساختاری مبتنی بر روان شناسی صوتی را بر اساس تئوری گشتالت ارائه داد. عدم مناسب بودن معیارهای ارزیابی هم چون دقت در جداسازی سیگنال در زمان طبقه بندی نهایی، یکی دیگر از مشکلات اکثر روش های موجود و پیشین است.

روش جمع آوری اطلاعات

مطالعه کتاب ها، مقالات و پایان نامه های معتبر در حوزه این پروژه و در کل مطالعه از نوع کتابخانه ای است.

روش تجزیه و تحلیل

استفاده از نرم افزار MATLAB به عنوان محیط شبیه ساز و همین طور استفاده از یک سری معیارهای ارزیابی که مقالات پایه تحقیق استفاده کرده اند تا بتوان تضمین رویکرد پیشنهادی را در مقایسه با آن ها انجام داد. مسلما می توان به دقت، انتروپی، فراخوان، حساسیت و نرخ ویژگی ها و مواردی از این دست اشاره نمود.

مراجع

[1] A. Lerch. An Introduction to Audio Content Analysis. Wiley-IEEE Press, 2012.
[2] L. Lu, H. J. Zhang, and H. Jiang. Content analysis for audio classification and segmentation. IEEE Transactions on Speech and Audio Processing, Volume 10, No. 7, 2002, Pages 504–516.
[3] I. McLoughlin. Applied Speech and Audio Processing: With MATLAB Examples, Nanyang Technological University, Cambridge University Press, 2009.

www.nn٤e.com – گروه تلمیرا

[4] C. C. Huang, J. F. Wang, and D. J. Wu. Audio signal segmentation algorithm. U.S. Patent No. 7,774, 203, 2010.
[5] M. Yanai, F. Sasaki, O. Tanaka, and M. Yasuoka. Blind source separation and two-signal localization in time-frequency domain considering time lag information: application to the case where one signal includes a reflected signal. Acoustical Science and Technology, Volume 35, No. 1, 2014, Pages 55–58.
[6] Y. Liang, J. Harris, S. M. Naqvi, G. Chen, and J. A. Chambers. Independent vector analysis with a generalized multivariate Gaussian source prior for frequency domain blind source separation. Signal Processing, Volume 105, Pages175–184, 2014.
[7] J. Ye, H. Jin, and Q. Zhang. Adaptive weighted orthogonal constrained algorithm for blind source separation. Digital Signal Processing, Volume 23, No. 2, 2013, Pages 514–521.
[8] P. Comon, and C. Jutten. Eds., Handbook of Blind Source Separation: Independent Component Analysis and Applications, Academic Press, 2010.
[9] G. Lio, and P. Boulinguez. Greater robustness of second order statistics than higher order statistics algorithms to distortions of the mixing matrix in blind source separation of human EEG: implications for single-subject and group analyses. NeuroImage, Volume 67, 2013, Pages 137–152.
[10] J. F. Cardoso, and A. Souloumiac. Jacobi angles for simultaneous diagonalization. SIAM Journal on Matrix Analysis and Applications, Volume 17, No. 1, 1996, Pages 161–164.
[11] K. Abed-Meraim, and A. Belouchrani. Algorithms for Joint Block Diagonalization. 2004.
[12] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: current directions and future challenges," Proceedings of the IEEE, Volume 96, No. 4, 2008, Pages 668–696.
[13] Mert Sirakaya. Gestalt Theory in Education and Practice. Academic Radiology, Volume 26, Issue 4, April 2019, Page e18.
[14] Saadia Zahid, Fawad Hussain, Muhammad Rashid, Muhammad Haroon Yousaf, and Hafiz Adnan Habib. Optimized Audio Classification and Segmentation Algorithm by Using Ensemble Methods. Mathematical Problems in Engineering, Volume 2015, 2015.
[15] Xu-Kui Yang, Dan Qu, Wen-Lin Zhang, and Wei-Qiang Zhang. An adapted data selection for deep learning-based audio segmentation in multi-genre broadcast channel. Digital Signal Processing, Volume 81, October 2018, Pages 8-15.
[16] Sébastien Lefèvre, Nicole Vincent. A two level strategy for audio segmentation. Digital Signal Processing, Volume 21, Issue 2, March 2011, Pages 270-277.
[17] José M. Leiva-Murillo, Sancho Salcedo-Sanz, Ascensión Gallardo-Antolín, and Antonio Artés-Rodríguez. A simulated annealing approach to speaker segmentation in audio databases. Engineering Applications of Artificial Intelligence, Volume 21, Issue 4, June 2008, Pages 499-508.
[18] Bartosz Ziółko. Fuzzy precision and recall measures for audio signals segmentation. Fuzzy Sets and Systems, Volume 279, 15 November 2015, Pages 101-111.
[19] Gen-Fang Chen, and Ya-Dong Wu. Segmentation of Singing, Speech and Instruments in Kunqu Audio Based on Zero-Crossing Rate. 2019 12th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 14-15 Dec. 2019.
[20] Yi-Chen Chen, Sung-Feng Huang, Hung-yi Lee, Yu-Hsuan Wang, and Chia-Hao Shen. Audio Word2vec: Sequence-to-Sequence Autoencoding for Unsupervised Learning of Audio Segmentation and Representation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume 27, Issue 9, Sept. 2019, Pages 1481-1493.

[21] Jie Pu, Yannis Panagakis, Stavros Petridis, Jie Shen, and Maja Pantic. Blind Audio–Visual Localization and Separation via Low-Rank and Sparsity. IEEE Transactions on Cybernetics, Volume 50, Issue 5, May 2020, Pages 2288–2301.
[22] Sun Jingzhou, Wang Yongbin, and Chen Xiaosen. Audio Segmentation and Classification Approach Based on Adaptive CNN in Broadcast Domain. 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), Beijing, China, China, 17-19 June 2019.

www.nn٤e.com – گروه نامیرا