ارائه یک روش جدید برای جداسازی گفتار از پس زمینه موسیقی مبتنی بر ضرایب موجک گسسته و شبکه عصبی نگاشت خود سازمانده

مریم خاشعی ورنامخواستی [*,۱]، سیّد سعید آیت [۲]،

۱- کارشناسی ارشد مهندسی کامپیوتر، دانشگاه پیام نور

۲- دانشیار گروه علمی مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور

مسأله جداسازی گفتار از پس زمینه موسیقی، یک مسأله جذاب ولی مشکل و چالش برانگیز می باشد. بدلیل اینکه اطلاعات محدودی در سیگنال ترکیب شده از موسیقی و گفتار وجود دارد. در این مقاله، ما از تبدیل موجک گسسته برای نمایش اطلاعات زمان- فرکانس سیگنال ترکیبی استفاده کردیم. در روش پیشنهادی، بعد از بدست آوردن انرژی ضرایب موجک گسسته در سطوح تجزیه شده، با استفاده از شبکه عصبی SOM، به خوشه بندی آنها پرداختیم. توانایی روش پیشنهادی، با استفاده از نتایج MOS و SDR تأیید می شود.

کلمات کلیدی: جداسازی موسیقی و گفتار، موجک، شبکه عصبی

۱. مقدمه

با پیشرفت مطالعات علمی از مغز انسان، ابزارهای ریاضی توسعه یافته زیادی برای کمک به ساخت سیستم های هوش محاسباتی، پردازش اطلاعات مغز انسان را شبیه سازی کرده اند. مسئله جداسازی سیگنال صدا از یک سیگنال ترکیبی (Mixture)، مبتنی بر خصوصیاتی است که ما برای تشکیل سیگنال خاص خارج می کنیم. در واقع، جداسازی صدا، یک عملیات برای استخراج دو یا چند صدای متمایز از یک منبع صدای ترکیب شده است. همه صدا ها در طبیعت خصوصیات صوتی مختلفی دارند و هیچ دو صدایی یکسان نیستند. مغز انسان به راحتی می تواند دو صدای مختلف در زمان یکسان را تشخیص دهد. دو سیگنال موسیقی (Music) و گفتاری (Speech) که به صورت غیر خطی با هم ترکیب شده اند جداکردن آنها با روش های خطی غیر ممکن است.

اغلب سیگنال های واقعی به صورت فرمت ستونی از مقادیر عددی در حوزه ی زمان به نمایش در می آیند. در بسیاری از موارد، خصوصیات متمایز کننده در حوزه فرکانس سیگنال می باشد و ویژگی هایی که به سختی در حوزه زمان قابل

[*] **Corresponding author: توضیحات مربوط به نویسنده اوّل**
Email: Maryam.khasheie@yahoo.com